ELECTRONIC WORD OF MOUTH (EWOM) AND MARKETING IMPLICATIONS

Miriam Alzate Barricarte

SUPERVISORS

Javier Cebollada Calvo
Marta Arce Urriza

Acknowledgements

This thesis is the culmination of almost five years of both academic and personal development. Along this long, but rewarding, trip there have been many people who have played, somehow or other, a role in helping me to reach the goals. These few words in the thesis are addressed to them.

First, I would like to express my foremost gratitude to my thesis directors Javier Cebollada and Marta Arce for your exemplary and constructive guidance, constant support and encouragement, as well as for your trust in me. You have always been there for anything I needed, and you have provided me not only your academic support but also your personal one. Your expertise, advice and your belief in my capabilities allowed me to complete my thesis with higher standards. I have definitely been very lucky of having you both as my thesis directors.

I would like to thank also to Public University of Navarre for giving me the chance of starting my research career with a university grant. Without it, this journey might not have been possible.

I would also like to thank Professor Pradeep K. Chintagunta for inviting me to the University of Chicago Booth School of Business to do a short research stay. Thanks for giving me the opportunity to spend an enriching academic and personal experience at Chicago Booth, where I had the immense luck of learning from some of the best professors in the world. In particular, thanks for giving me the chance of assisting a lecture from the Nobel Prize in Economic Science Richard H. Thaler. Thanks also for being available to help me with some specific doubts about my thesis and to give me some recommendations for future research.

Thanks to Carlos Ortega for your constant support on R and data analysis, you have been always there no matter the day or time in the week. It was a pleasure to receive quick answers from the R users list whenever I had a question. But I have been even luckier of having your direct support from the beginning of my thesis, when I did not have any experience using R. You have been the most important source of knowledge for my R training over these years. I am definitely sure that without your help this journey would have been much harder.

Thanks to every person belonging to the Business Department of the Public University of Navarra, and especially to those in the Marketing Group, for your constant willingness to help me over the journey.

Thanks go to all my friends from Villafranca who have made this experience easier and more enjoyable.

Finally, I would like to dedicate this thesis to all my family for always being supportive and for trusting me. Special thanks to my parents, Antonio and Loli, to my sister, Elena, to my grandmother, Conchi, and to my partner, David, for providing me always with unconditional care and for always been there to listen. To my uncle, Miguel Angel, who is not longer with us but who has always been very important in my life and who would be so proud of my achievements. This is for you all.

Abstract

This doctoral thesis studies various aspects related to a specific type of electronic word of mouth, also referred to as eWOM, which are online consumer reviews. On the one hand, this thesis studies how certain non-textual and textual characteristics of online reviews, together with the visibility of those reviews, influence various types of consumer behaviour, specifically the adoption of information and the purchasing behaviour. On the other hand, this research studies the textual content of online reviews to study the brand positioning and brand segmentation, and to analyse the positioning and associations with products. Prior to the empirical analysis, the literature on information processing, decision-making literature, branding and text mining are mainly reviewed. The empirical analysis differs between chapters. Chapter 1 explores how the visibility of online reviews moderates the relationship between certain characteristics of the reviews and the number of helpful votes they receive from consumers (referred to as review helpfulness). To do this, a count regression method is used, specifically, the Zero Inflated Negative Binomial (ZINB) regression. The results of the study reveal that review visibility has a strong moderating role in explaining the excess of reviews with zero votes. However, when explaining the positive number of helpful votes that online reviews receive, both review visibility and certain characteristics of the reviews have a direct impact on helpfulness but review visibility as a weak moderating effect. In addition, it is also observed that the effect of review visibility on review helpfulness depends also on the reviews that are more visible, if they are the "most helpful" or the "most recent" reviews. Chapter 2 studies how certain non-textual and textual characteristics of reviews influence product sales in different review visibility cases, using a database of online **reviews of the category of "blush" products over a nine-week period**. For the analysis, a panel data model is used, the system GMM model, which allows correcting for endogeneity better than other panel data models. The results reveal that all the characteristics analysed of the reviews have an impact on sales. However, the effect of textual characteristics varies depending on the visibility case studied. In general, it is observed that the characteristics of those reviews that are more visible because they are "more helpful" are those that have the most impact on product sales. Chapter 3 studies the textual content of reviews to extract brand associations and uses the lexicon-based text mining method Linguistic Inquiry and Word Count (LIWC, Pennebaker et al., 2015). With the associations extracted, brand positioning and brand segmentation analyses are conducted. Finally, chapter 4 continues with the study of the textual content of the reviews, but in this case several unsupervised machine learning algorithms for text mining are reviewed and applied.

Resumen

Esta tesis doctoral estudia varios aspectos relacionados con un tipo concreto de boca a boca electrónico (en inglés electronic Word-of-Mouth o eWOM), que son las reseñas de consumidores. Por un lado, se estudia cómo ciertas características no textuales y textuales de esas reseñas, junto a su visibilidad, influye en varios tipos de comportamiento de consumidor, en concreto en la adopción de información y en el comportamiento de compra. Por otro lado, se estudia el contenido textual de las reseñas y cómo este se puede utilizar tanto para estudiar el posicionamiento y segmentación de marcas, como para estudiar el posicionamiento y asociaciones con productos. Previamente al análisis empírico, se revisa principalmente la literatura sobre procesamiento de información, toma de decisiones e imagen de marca y minería de textos. El análisis empírico difiere entre capítulos. En el capítulo 1, se estudia cómo la visibilidad de las reseñas modera el impacto de la relación entre determinadas características de las reseñas y el número de votos útiles que reciben por parte de los consumidores. Para ello, se utiliza un método de regresión de conteo, en concreto, se usa una regresión Binomial Negativa de Ceros Inflados (en inglés Zero Inflated Negative Binomial o ZINB). Los resultados del estudio revelan que la visibilidad tiene un papel moderador especialmente relevante al explicar el exceso de reseñas con cero votos. Sin embargo, al explicar el número positivo de votos útiles de las reseñas, tanto la visibilidad como ciertas características de las reseñas tienen un impacto directo pero el efecto moderador de la visibilidad es menor. Además, se observa también que el efecto de la visibilidad sobre la utilidad de las reseñas depende de qué reseñas son más visibles, si son las reseñas **“más útiles” o las “más recientes”**. El capítulo 2 estudia cómo determinadas características no textuales y textuales de las reseñas influyen en la venta de productos en diferentes escenarios de visibilidad con una base de datos de nueve semanas. Para el análisis, se utiliza un modelo de datos de panel, en particular el modelo system GMM, que permite corregir la endogeneidad mejor que otros modelos de datos de panel. Los resultados revelan que todas las características de las reseñas analizadas tienen un impacto sobre las ventas. Sin embargo, el efecto de las características textuales varía dependiendo del escenario de visibilidad estudiado. En general, se observa que las características de aquellas reseñas **más visibles porque son “más útiles” son las que más** impactan en la venta de productos. El capítulo 3 estudia el contenido textual de las reseñas para extraer asociaciones de marca y utiliza el método de minería de textos basado en diccionarios Linguistic Inquiry and Word Count (LIWC, Pennebaker et al., 2015). Con las asociaciones extraídas, se estudia el posicionamiento y la segmentación de las marcas de la categoría de productos estudiada en la investigación, que son

cosméticos, y en concreto, colorete. Por último, en el capítulo 4 se sigue con el estudio del contenido textual de las reseñas, pero en este caso se revisan y aplican de forma empírica varios algoritmos de aprendizaje automático no supervisado para la minería de textos.

TABLE OF CONTENTS

Introduction.....	1
Research Background	1
Research Gaps and Research Objectives	2
Methodology	7
References.....	9
Chapter 1. Review Characteristics and Review Helpfulness: The Role of Review Visibility.....	25
1. Introduction	25
2. Theoretical Background and Research Model.....	28
2.1. What Makes a Review Helpful?	28
2.2. Consumer Voting Behavior: Why is Review Visibility Relevant?	30
2.3. Conceptual Model and Hypotheses Development	33
3. Research Methodology	40
3.1. Data	40
3.2. Variables	42
4. Data Analysis	49
4.1. Descriptive Statistics.....	49
4.2. Model Estimation	51
5. Results.....	53
5.1. Model diagnostics and robustness check.....	59
6. Discussion	61
6.1. Theoretical Contribution	62
6.2. Managerial Implications	63
6.3. Limitations and Future Research.....	64
Appendix - Chapter 1	66
References.....	70
Chapter 2. Online Reviews and Product Sales: The Role of Review Visibility	90
1. Introduction	90
2. Theoretical Background and Conceptual Model	93
3. Methodology	102
3.1. Data	102
3.2. Research Variables	105
3.3. Empirical Model and Estimation	109
4. Results.....	112
4.1. Descriptive Statistics.....	112
4.2. Model Findings.....	113

4.3. Misspecification Tests and Alternative Panel Data Models	119
5. Discussion	121
5.1. Theoretical Contribution.....	122
5.2. Managerial Implications.....	123
5.3. Limitations and Future Research	124
Appendix – Chapter 2	126
References.....	130
Chapter 3. Mining the Text of Online Reviews to Uncover Brand Associations and Build Positioning Maps	146
1. Introduction	146
2. Literature Review.....	149
2.1. The Concept of Brand Image, Brand Associations and Brand Positioning	149
2.2. The Measurement of Brand Image, Brand Associations and Brand Positioning	151
2.3. Mining the Text of Online Reviews to Explore Brand Associations and Brand Image	154
3. Research Procedure	158
4. Empirical Analysis	160
4.1. Data Acquisition.....	160
4.2. Data Pre-processing	162
4.3. Text Mining: Linguistic Inquiry and Word Count (LIWC)	162
4.4. Data Aggregation.....	163
4.5. Brand Positioning: Principal Component Analysis (PCA)	167
4.6. Identifying Brand Segments: Hierarchical Clustering	174
5. Conclusions.....	181
5.1. Limitations and Future Research	183
References.....	185
Chapter 4. What are Consumers Saying Online about your Products? Mining the Text of Online Reviews to Uncover Hidden Features	201
1. Introduction	201
2. Theoretical Background.....	204
2.1. The Persuasive Power of Language.....	204
2.2. The Use of Online Reviews to Study Brand Image.....	205
2.3. Mining the Text of Online Reviews	205
3. Methodology	207
3.1. Data	207
3.2. Data Analysis	208
4. Results.....	210
4.1. Descriptive Statistics.....	210

4.2. Text Mining Analysis.....	210
5. Conclusions.....	219
5.1. Limitations and Future Research	222
References.....	224
General Conclusions	240
Summary of Findings.....	240
Contribution to Theory	245
Conclusiones generales	252
Resumen de resultados.....	252
Contribución teórica	257

Introduction

Research Background

Word of Mouth (WOM) has been defined as an informal, person-to-person communication between a perceived noncommercial communicator and a receiver regarding a brand, a product, an organization, or a service (Harrison-Walker, 2001). WOM has been recognized in academic literature as one of the most important sources of information for consumers, and it has stand out for being a kind of communication that in some cases can be more influential than traditional marketing communication tools, such as advertising (Trusov et al., 2009).

The rise of the Internet during the last decade has changed the way people communicate, shop, look for information and interact with each other and with companies. In this context, the concept of electronic Word of Mouth (hereafter eWOM) has emerged and attracted the interest of many academic researchers. eWOM is defined as any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet (Hennig-Thurau et al., 2004). The differences between traditional WOM and eWOM are related to the speed of the diffusion of the information in cyberspace, the **possibility of accessing large volumes of information, the “many-to-many” nature of** online communications, the lack of geographical limitations and the measurability of eWOM. On the Internet, customers can write their opinions on different platforms such as online discussion forums, newsgroups, blogs, reviews´ sites, electronic bulletin boards, social networking sites and companies´ websites. Therefore, eWOM can be found on all those digital platforms having different formats (e.g., Facebook posts, tweets on Twitter, blog entries, etc.). In this thesis, we are going to focus on the study of one specific type of eWOM, online customer reviews, which are evaluations posted online at different platforms by consumers who have used or experienced a product or service (Kamble et al., 2020).

Hennig-Thurau et al. (2004) claimed that the information provided by consumers on opinions´ sites exert more influence among consumers than traditional marketing messages provided by firms. This fact has been constantly supported either by academic or by industry research. For example, the consultancy firm Deloitte (2014) revealed that 60% of consumers pointed out online reviews as the most trusted source of information, **while the “product manufacturer/service provider” was the most trusted source of**

information for only 12% of consumers. Other consultancy companies have also conducted research on online reviews. The consultancy firm BrightLocal conducts an annual survey to explore the ways in which consumers use online reviews to evaluate businesses. The BrightLocal report (2020) revealed that 87% of consumers read online reviews to evaluate businesses. This implies an increase from 81% in 2019 and an increase from 67% back in 2010. Moreover, 79% of consumers admitted trusting online reviews as much as personal recommendations from friends or family. Moreover, 94% of consumers also claimed that positive reviews make them more likely to use a business, while 92% revealed that negative reviews make them less likely to use a business.

Academic literature has also highlighted the power of online reviews to predict different types of consumer behavior such as information adoption decisions (Cyr et al., 2018; Filieri, McLeay, et al., 2018), purchase intentions (Jiménez & Mendoza, 2013; Kostyk et al., 2017; D. H. Park & Lee, 2008) and product sales in product categories such as hardware, books, movies and hotels (J. Chevalier & Mayzlin, 2006; Chintagunta et al., 2010; Hofmann et al., 2017; Xiaolin Li et al., 2019; Marchand et al., 2017).

Even though the study of eWOM, and in particular of online reviews, has attracted a lot of attention from academics last years, it is still a relatively new area that needs further research in order to build stronger conclusions and to explore the gaps found in literature.

Research Gaps and Research Objectives

The study of online reviews is a relatively new phenomenon, which has started to gain attention with the increasing growth of the Internet. Moreover, since the digital environment and how consumers behave online changes rapidly over time, continuous research in this context is always needed to understand the new trends.

Two areas of research can be identified in literature: the study of non-textual variables of online reviews and the study of textual characteristics. Non-textual variables are those not directly related to the review text and that are easier to identify and measure (e.g., review rating, reviewer number of reviews, presence or not of summary words, etc.). On the contrary, textual variables are those directly extracted from the text of online reviews (e.g., text length, textual sentiment, subjectivity, level of confidence expressed by the reviewer, etc.). The main stream of online reviews literature has focused on studying how several non-textual variables of online reviews influence two types of consumer decisions: (1) the likelihood of consumers to vote online reviews as helpful (also known

as review helpfulness) and (2) the likelihood of consumers to buy products. In general, previous research has found that there are several non-textual characteristics of online reviews that have a significant impact on both the likelihood that consumers vote online reviews as helpful and the likelihood of consumers to buy products. However, textual characteristics have been less explored. One of the main reasons is that the textual content of online reviews is qualitative in nature, making it difficult to analyze and to extract meaningful insights (K. Chen et al., 2015). When incorporating factors of the review text, a lot of studies have analysed the length of the review as a proxy of review informativeness. Some of them have also studied the overall sentiment of online reviews (positive vs. negative). However, the lack of studies and clear conclusions makes this area interesting for future research. Thus, we incorporate review textual variables in Chapter 1 and Chapter 2 to study review helpfulness and product sales and we directly focus on analysing the textual content of online reviews in Chapter 3 and Chapter 4.

This thesis has two faces. Chapter 1 and Chapter 2 adopt a conclusive research design and focus mainly on an econometric-style journey in relating review characteristics and their visibility to helpfulness and product sales, while Chapter 3 and Chapter 4 have a more exploratory focus and are much more practice-oriented in nature, since they focus on the study of brand image through the textual content of online reviews. In these chapters, we use text mining techniques to uncover the “hidden” dimensions of brand/product image (e.g., sentiments, psychological processes, attributes, etc.) expressed in online reviews’ texts. Therefore, we first focus on studying how review characteristics (both non-textual and textual) impact two types of consumer behaviour, which are the likelihood of voting and review as helpful and the likelihood to buy a product. Then, we continue studying the textual content of online reviews in depth, in order to have a broader idea of how text mining techniques can allow academics and practitioners uncover different dimensions of brand image that might be expressed in the text.

When analysing the impact of online reviews on review helpfulness and product sales, previous literature has implicitly assumed that every review for a product has the same probability of being viewed by consumers, which means that every review is usually considered to be equally influential in consumer behaviour. Thus, when studying the impact of different review characteristics on review helpfulness or product sales, previous literature assumes the characteristic of each review to be equally influential on consumer decision making. However, online reviews are not displayed individually but, in a sequence, next to other reviews, so the relative visibility of online reviews in the sequence is likely to influence consumer decision-making. In this line, literature in

decision making and information processing has revealed that consumers usually face information overload situations in online environments due to the large amount of information available (Beach, 1993; Häubl & Trifts, 2000), as it might happen when dealing with high volume of online reviews. In these complex environments, consumers cannot evaluate every single online review available for each product and, instead, they are likely to adopt selective processing strategies in order to reduce the cognitive effort of managing big volume of information.

The accessibility-diagnostics theory by Feldman and Lynch (1988), claim that the likelihood of using a piece of information for making a choice depends on its accessibility and its diagnosticity. Therefore, this theory suggests that not only those review characteristics that make a review diagnostic (e.g., review rating, review length reviewer information disclosure, etc.), but also the accessibility of the review is likely to influence consumers' information adoption decisions. For instance, the annual study of online reviews developed by the consultancy firm BrightLocal (2020) reveals that, on average, consumers read a maximum of 10 online reviews before making a decision, which might indicate that consumers are likely to base their decision only on a subset of all available reviews. Since consumers are more likely to read most visible online reviews, characteristics of those online reviews are likely to have a greater influence on consumer voting and purchasing behaviour.

Therefore, a gap is identified in literature since more research is needed in order to understand if the characteristics of every online review are equally influential on consumer decision making or, by contrast, the characteristics of more visible online reviews have an even greater impact. To try to cover this gap, we incorporate the variable review visibility in Chapter 1 and Chapter 2. Review visibility is proxied in this research by the rank order of the review in the sequence of reviews for a given product when consumers sort reviews using different mechanisms. We study the effect of review visibility when using two review sorting mechanisms, which are two of the most popular in online **reviews'** sites: **sorting by "most helpful" and sorting by "most recent" reviews.**

Moreover, a gap is also found in literature regarding how the consumer voting behaviour is understood. We suggest in Chapter 1 that there might be many online reviews with zero helpfulness because due to two different reasons: (1) not viewing the review and (2) viewing the review and deciding not to vote it as helpful. Therefore, with our empirical estimation we want to make a distinction between the antecedents making a review to have a positive number of helpful votes, as literature usually does, and the factors that might explain the reviews having zero helpful votes.

We formulate the following hypotheses in Chapter 1

- H1a. *More visible online reviews when ranked by the “most helpful” mechanism are more likely to be voted as helpful*
- H1a. *More visible online reviews when ranked by the “most recent” mechanism are more likely to be voted as helpful*
- H2a. *Review visibility when online reviews are ranked by the “most helpful” mechanism moderates the impact of review characteristics on review helpfulness*
- H2b. *Review visibility when online reviews are ranked by the “most recent” mechanism moderates the impact of review characteristics on review helpfulness*

So far, studies have not found consensus when explaining how the characteristics of online reviews, mainly rating and volume, impact product sales. Some scholars, such as **Babić Rosario et al. (2016)**, have revealed that the different effects of review variables on product sales depend on factors such as the platform where the online text is published (e.g. social network, review site, etc.), the type of product (e.g. hedonic vs. utilitarian), the metric used to build the review variable of interest (e.g. average, cumulative, incremental, etc.) and the design of the study (e.g. how endogeneity is controlled, the type of control variables, etc.) (**Babić Rosario et al., 2016**). In this research, we posit that review visibility can be also a factor that might impact the relationship between review characteristics and product sales. In this sense, effects might be different if we assume that consumers read every online review available for a product or if we assume that most visible reviews are more likely to be read.

Moreover, previous literature has mainly analysed the effect of non-textual review characteristics on product sales but has paid less attention to study the effect of textual variables. Therefore, we include in our study not only traditional non-textual variables but also some variables related to the textual content of reviews that have been proven to influence consumer behaviour.

We define the following hypotheses in Chapter 2:

- H1a. *Review non-textual features influence product sales considering different cases of review visibility*
- H1b. *Review textual features influence product sales considering different cases of review visibility*
- H2a. *The influence of review non-textual features on product sales is different depending on the review visibility case considered*

- H2b. The influence of review textual features on product sales is different depending on the review visibility case considered

Concerning the study of online reviews' textual content, most papers until now that study brand image and brand associations have relied on survey data, using multidimensional scales to analyse brand image. Others have adopted qualitative techniques to understand brand elicitations and to build associative networks and conceptual maps. A very small stream of literature has focused on analysing textual features of online reviews to explore brand image and brand associations. When analysing the text of online reviews, the **concept of "text mining" arises. Text mining** refers to the process of extracting useful and meaningful information from unstructured text (Netzer et al., 2012). The study of online reviews to uncover brand associations has been conducted in different product settings (e.g. hotels, restaurants and pc components) and using a wide range of text mining techniques, such as lexicon-based techniques (e.g. Linguistic Inquiry and Word Count, developed by Pennebaker et al., 2007) and unsupervised machine learning techniques (e.g. Latent Dirichlet Allocation, LDA). The selection of one text mining technique or another must rely on the type of data and on the research objectives. However, extant literature is still scarce and is quite heterogenous in terms of research objectives, research processes and text mining techniques used. Because Chapter 3 and Chapter 4 have a more exploratory focus, we formulate research objectives instead of hypotheses.

- To present a unified and structured procedure to explore brand image and brand positioning using the information contained in the text of online reviews

By addressing the previous objective in Chapter 3, we want to guide prospective scholars and practitioners in their text mining analysis process. To do that, we chose an easy to apply text mining technique, the Linguistic Inquiry and Word Count (LIWC) developed by Pennebaker et al. (2007). However, several text mining techniques based on unsupervised machine learning algorithms can be also used to explore the text of online reviews. In Chapter 3 we focus on the process of uncovering brand associations from online reviews using a lexicon-based method, the LIWC, to explore brand image and positioning. However, in Chapter 4, rather than using a lexicon-based method for text mining, we adopt different unsupervised machine learning algorithms to answer some research questions regarding the study of product image. Thus, the main objective of Chapter 4 is the following one:

- To show and to illustrate some methods of text mining with the aim of exploring product image uncovering information contained in online reviews.

Methodology

To conduct the different empirical analyses in the thesis, we collected online consumer reviews from a popular US cosmetics retailer website using web crawling. The set of online reviews collected were all the **reviews posted in the "blush" cosmetics category**. The data were collected on different dates, depending on the purpose of each chapter. As claimed by the report published by BrightLocal (2020), the industry of **"Hair/Beauty"** is one of the top-10 industries in which consumers usually read online reviews to evaluate products. 85% of consumers claim reading online reviews in this industry. Therefore, since consumers are likely to read reviews to evaluate cosmetics, they might be also likely to be influenced by the characteristics of those online reviews.

In Chapter 1, the analysis of the influence of review visibility on review helpfulness is grounded on the accessibility-diagnostics theory (Feldman & Lynch, 1988), which suggests that the likelihood of consumers to adopt a piece of information is influenced by both the accessibility and by the diagnostics of that information. Being equal the diagnostics of the information, the higher the accessibility, the more likely that information is adopted by consumers. In our case, an online review is a piece of information and adopting the information contained in online reviews is done by voting it as helpful. Thus, not only the characteristics of online reviews that make them diagnostic are likely to influence their helpfulness, but also the accessibility of those reviews for consumers. We apply a Zero Inflated Negative Binomial (ZINB) regression, which allows us to model review helpfulness in two parts. First, a logit part that allows us to understand the drivers explaining that online reviews do not receive helpful votes because they have not been viewed. Second, a negative binomial part that explains the likelihood that online reviews receive helpful votes. Data were collected on two different days, 4 October 2016 (63,985 online reviews belonging to 145 products) and 12 April 2017 (64,803 online reviews belonging to 142 products). For the empirical analysis, a final database of online reviews available at the online retailer on both dates was used (58,770 online reviews belonging to 110 products).

In Chapter 2 the analysis of the impact of online reviews on product sales under different cases of review visibility was also drawn on the accessibility-diagnostics theory (Feldman & Lynch, 1988). Similarly to Chapter 1, not only the characteristics of online reviews (e.g., rating, length or volume) are likely to influence consumers in their

purchasing decision-making, but also the accessibility of those reviews. Online reviews more visible for consumers might exert a greater influence in their decision-making. We used a specific panel data methodology to correct for endogeneity, the system generalized method of moments (system GMM) estimator developed by Arellano and Bover (1995) and Blundell and Bond (1998). Data were collected over nine consecutive weeks between 21 December 2016 (65,503 reviews belonging to 142 products), and 17 February 2017 (62,496 reviews of 131 products).

In Chapter 3 the study of textual content of online reviews to analyse brand positioning and brand segmentation is rooted in the brand image and text mining literature. Literature suggests that language is the way in which people express their internal thoughts and emotions, which makes the study of consumers' texts a very interesting tool to know how consumers perceive products and brands. The chapter proposes an easy-to-follow procedure to uncover emotional and psychological brand associations from the text of online reviews. In this procedure, different types of analyses are used: text mining analysis using the software LIWC (Pennebaker et al., 2015), brand positioning analysis using Principal Component Analysis (PCA) and brand segmentation analysis using Hierarchical Clustering. Online reviews collected on 17 February 2017 were used as the data for this chapter. The dataset is composed by a total of 62,496 reviews of 131 products and 44 brands, all the products available at the blush category of cosmetics.

Chapter 4 can be seen as an extension of Chapter 3. In this chapter we also analyse the textual content of online reviews to explore associations but, unlike Chapter 3, we use unsupervised machine learning algorithms instead of a lexicon-based method to uncover hidden features in the text (e.g., main topics discussed, product elicited, etc.). Moreover, rather doing the analysis at brand level to uncover brand associations, we do the analysis at product level, studying consumer associations with specific products (a brand can have more than one product in our data set). Online reviews collected on 17 February 2017 were also used to do the analyses in this chapter. The dataset is composed by a total of 62,496 reviews of 131 products and 44 brands, all the products available at the blush category of cosmetics.

References

- Aaker, D. A. (1991). Managing Brand Equity. *Managing Brand Equity*.
- Aaker, D. A. (1996). Building strong brands - aaker,da. *Building Strong Brands*, 13(4), 382–386.
- Ahani, A., Nilashi, M., Yadegaridehkordi, E., Sanzogni, L., Tarik, A. R., Knox, K., Samad, S., & Ibrahim, O. (2019). Revealing customers' satisfaction and preferences through online review analysis: The case of Canary Islands hotels. *Journal of Retailing and Consumer Services*, 51(July), 331–343.
- Ahmad, S. N., & Laroche, M. (2015). How do expressed emotions affect the helpfulness of a product review? Evidence from reviews using latent semantic analysis. *International Journal of Electronic Commerce*, 20(1), 76–111.
- Aiken, L. S., & West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*.
- Al-Obeidat, F., Spencer, B., & Kafeza, E. (2018). The Opinion Management Framework: Identifying and addressing customer concerns extracted from online product reviews. *Electronic Commerce Research and Applications*, 27, 52–64.
- Al., H. et. (2010). *Multivariate Analysis 7 th Edition*.
- Aljukhadar, M., Senecal, S., & Daoust, C.-E. (2012). Using Recommendation Agents to Cope with Information Overload. *International Journal of Electronic Commerce*, 17(2), 41–70.
- Ansari, A., & Mela, C. F. (2003). *E-Customization*. 40(2), 131–145.
- Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*, 57(8), 1485–1509.
- Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2), 277.
- Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1), 29–51.
- Areni, C. S. (2003). The Effects of Structural and Grammatical Variables on Persuasion: An Elaboration Likelihood Model Perspective. *Psychology and Marketing*, 20(4), 349–375.
- Arun, R., Suresh, V., Madhavan, C. E. V., & M.N, N. M. (2010). On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In *Advances in Knowledge Discovery and Data Mining* (pp. 391–403).
- Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. A. (2016). The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors. *Journal of Marketing Research*, 53(3), 297–318.
- Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, 0, 2200–2204.
- Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of Online Consumer Reviews: Readers' Objectives and Review Cues. *International Journal of Electronic Commerce*,

- 17(May 2015), 99–126.
- Baker, W. E. (2001). The Diagnosticity of Advertising Generated Brand Attitudes in Brand Choice Contexts. *Journal of Consumer Psychology*, 11(2), 129–139.
- Baker, W. E., Lutz, R. J., Journal, S., Spring, N., Taylor, P., Baker, W. E., & Lutz, R. J. (2000). Effectiveness An Empirical Test of an Updated Relevance- Accessibility Model of Advertising Effectiveness. *Journal of Advertising*, 29(1), 1–14.
- Bakhshi, S., Kanuparth, P., & Shamma, D. A. (2015). Understanding Online Reviews: Funny, Cool or Useful ? *CSCW 2015: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1270–1276.
- Baksi, A. K., & Panda, T. K. (2018). Branding destinations with multisensory brand associations and evaluating its impact on behavioural pattern under the intervention of multiplex phenomenon of relationship-branding. *Management Science Letters*, 8(11), 1169–1182.
- Balducci, B., & Marinova, D. (2018a). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, 46(4), 557–590.
- Balducci, B., & Marinova, D. (2018b). Unstructured data in marketing. In *Journal of the Academy of Marketing Science* (Vol. 46, Issue 4, pp. 557–590). Springer New York LLC.
- Beach, L. R. (1993). Broadening the definition of decision making: The role of prechoice screening option. *Psychological Science*, 4(4), 215–220.
- Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding Satisfied and Dissatisfied Hotel Customers: Text Mining of Online Hotel Reviews. *Journal of Hospitality Marketing and Management*, 25(1), 1–24.
- Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the Tribes: Using Text for Marketing Insight. *Journal of Marketing*, 84(1), 1–25.
- Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive Consumer Choice Processes. *Journal of Consumer Research*, 25(3), 187–217.
- Biehal, G., & Chakravarti, D. (1983). Information Accessibility as a Moderator of Consumer Choice. *Journal of Consumer Research*, 10(1), 1.
- Biswas, D., Grewal, D., & Roggeveen, A. (2010). How the order of sampled experiential products affects choice. *Journal of Marketing Research*, 47(3), 508–519.
- Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), 115–143.
- Brandt, C., De Mortanges, C. P., Bluemelhuber, C., & Van Riel, A. C. R. (2011). Associative Networks: A New Approach to Market Segmentation. *International Journal of Market Research*, 53(2), 187–207.
- Breugelmans, E., Campo, K., & Gijsbrechts, E. (2007). Shelf sequence and proximity effects on online grocery choices. *Marketing Letters*, 18(1–2), 117–133.
- BrightLocal. (2020). *Local Consumer Review Survey*. Available at: <https://www.brightlocal.com/research/local-consumer-review-survey/>
- Brown, C. L., & Krishna, A. (2004). The skeptical shopper: A metacognitive account for the effects of default options on choice. *Journal of Consumer Research*, 31(3), 529–539.

- Bruine de Bruin, W., & Keren, G. (2003). Order effects in sequentially judged options due to the direction of comparison. *Organizational Behavior and Human Decision Processes*, 92(1–2), 91–101.
- Buda, R., & Zhang, Y. (2000). Consumer product evaluation: the interactive effect of message framing, presentation order, and source credibility. *Journal of Product & Brand Management*, 9(4), 229–242.
- Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics Using Stata*. Stata Press Book, StataCorp LP.
- Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data, Second edition. In *Regression Analysis of Count Data, Second Edition*.
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72(7–9), 1775–1781.
- Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, 50(2), 511–521.
- Chaiken, S. (1987). *The heuristic model of persuasion*.
- Chapman, C., & Feit, E. M. (2015). *R for marketing research and analytics*.
- Chen, C. C., & Tseng, Y. De. (2011). Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, 50(4), 755–768.
- Chen, K., Kou, G., Shang, J., & Chen, Y. (2015). Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches. *Electronic Commerce Research and Applications*, 14(1), 58–74.
- Cheng, Y. H., & Ho, H. Y. (2015). Social influence’s impact on reader perceptions of online reviews. *Journal of Business Research*, 68(4), 883–887.
- Cheung, C. M. K., Lee, M. K. O., & Rabjohn, N. (2008). The impact of electronic word-of-mouth. In *Internet Research* (Vol. 18, Issue 3).
- Chevalier, J. A., & Goolsbee, A. (2003). Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics*, 1, 203–222.
- Chevalier, J., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 43(3), 345–354.
- Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science*, 29(5), 944–957.
- Cho, E., Fiore, A. M., & W. Russell, D. (2015). Validation of a Fashion Brand Image Scale Capturing Cognitive, Sensory, and Affective Associations: Testing Its Role in an Extended Brand Equity Model. *Psychology & Marketing*, 32(1), 28–48.
- Chua, A. Y. K., & Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior*, 54, 547–554.
- Chung, C.K., & Pennebaker, J. W. (2007). The psychological function of function words. *Social Communication: Frontiers of Social Psychology*, 343–359.
- Chung, Cindy K., & Pennebaker, J. W. (2007). Revealing dimensions of thinking in open-

- ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality*, 42(1), 96–132.
- Cohn, M. a., Mehl, M. R., & Pennebaker, J. W. (2004). Markers of Linguistic Psychological Change. *Psychological Science*, 15(10), 687–693.
- Cohn, M. a., Mehl, M. R., & Pennebaker, J. W. (2014). Markers of Linguistic Psychological Change Surrounding September 11, 2001. *Psychological Science*, 15(10), 687–693.
- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2014). *Markers of Linguistic Psychological Change Surrounding September*. 15(10), 687–693.
- Cui, G., Lui, H.-K., & Guo, X. (2012). The Effect of Online Consumer Reviews on New Product Sales. *International Journal of Electronic Commerce*, 17(1), 39–58.
- Cyr, D., Head, M., Lim, E., & Stibe, A. (2018). Using the elaboration likelihood model to examine online persuasion through website design. *Information and Management*, April 2017, 1–15.
- Darley, W. K., & Smith, R. E. (1993). Advertising Claim Objectivity: Antecedents and Effects. *Journal of Marketing*, 57(4), 100.
- Davis, D. F., Golcic, S. L., & Marquardt, A. (2009). Measuring brand equity for logistics services. *The International Journal of Logistics Management*, 20(2), 201–212.
- Deloitte. (2014). Deloitte Consumer Review: The growing power of consumers. 1–16.
- De Vries, L., Gensler, S., & Leeftang, P. S. H. (2012). Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing. *Journal of Interactive Marketing*, 26(2), 83–91.
- Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, 27(4), 293–307.
- Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4), 23–45.
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective Latent Concept Modeling for ad hoc information retrieval. *Document Numerique*, 17(1), 61–84.
- Drèze, X., & Zufryden, F. (2004). Measurement of online visibility and its impact on Internet traffic. *Journal of Interactive Marketing*, 18(1), 20–37.
- Duan, W., Gu, B., & Whinston, A. B. (2008a). Do online reviews matter? - An empirical investigation of panel data. *Decision Support Systems*, 45(4), 1007–1016.
- Duan, W., Gu, B., & Whinston, A. B. (2008b). The dynamics of online word-of-mouth and product sales-An empirical investigation of the movie industry. *Journal of Retailing*, 84(2), 233–242.
- Elberse, A., & Eliashberg, J. (2003). Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures. *Marketing Science*, 22(3), 329–354.
- Ellsworth, P. C., & Smith, C. A. (1988). From appraisal to emotion: Differences among unpleasant feelings. *Motivation and Emotion*, 12(3), 271–302.
- Ert, E., & Fleischer, A. (2016). Mere Position Effect in Booking Hotels Online. *Journal of Travel Research*, 55(3), 311–321.

- Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management*, 55, 62–73.
- Eslami, S. P., & Ghasemaghaei, M. (2018). Effects of online review positiveness and review score inconsistency on sales: A comparison by product involvement. *Journal of Retailing and Consumer Services*, 45(June), 74–80.
- Faircloth, J. B., Capella, L. M., & Alford, B. L. (2001). The Effect of Brand Attitude and Brand Image on Brand Equity. *Journal of Marketing Theory and Practice*, 9(3), 61–75.
- Feldman, J. M., & Lynch, J. G. (1988). Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior. *Journal of Applied Psychology*, 73(3), 421–435.
- Felfernig, A., Friedrich, G., Gula, B., Hitz, M., Kruggel, T., Leitner, G., Melcher, R., Riepan, D., Strauss, S., Teppan, E., & Vitouch, O. (2007). Persuasive recommendation: Serial position effects in knowledge-based recommender systems. *Proceedings of the 2nd International Conference on Persuasive Technology (PERSUASIVE '07)*, 283–295.
- Filieri, R., Hofacker, C. F., & Alguezaui, S. (2018). What makes information in online consumer reviews diagnostic over time? The role of review relevancy, factuality, currency, source credibility and ranking score. *Computers in Human Behavior*, 80, 122–131.
- Filieri, R., & McLeay, F. (2014). E-WOM and Accommodation: An Analysis of the Factors That Influence Travelers' Adoption of Information from Online Reviews. *Journal of Travel Research*, 53(1), 44–57.
- Filieri, R., McLeay, F., Tsui, B., & Lin, Z. (2018). Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services. *Information and Management*, 55(8), 956–970.
- Filieri, R., Raguseo, E., & Vitari, C. (2018). When are extreme ratings more helpful? Empirical evidence on the moderating effects of review characteristics and product type. *Computers in Human Behavior*, 88, 134–142.
- Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., & Freling, T. (2014). How Online Product Reviews Affect Retail Sales: A Meta-analysis. *Journal of Retailing*, 90(2), 217–232.
- Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., & Treinen, M. (2001). What makes web sites credible? A report on a large quantitative study. *Conference on Human Factors in Computing Systems - Proceedings*, 61–68.
- Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 19(3), 291–313.
- Gao, B., Hu, N., & Bose, I. (2017). Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews. *Decision Support Systems*, 95, 1–11.
- Gensler, S., Völckner, F., Egger, M., Fischbach, K., & Schoder, D. (2015). Listen to your customers: Insights into brand image using online consumer-generated product reviews. *International Journal of Electronic Commerce*, 20(1), 112–141.
- Ghose, A., & Ipeirotis, P. G. (2010). Estimating the Helpfulness and Economic Impact of Product Reviews. *Ieee Transactions on Knowledge and Data Engineering*, 1–15.

- Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science*, 31(3), 493–520.
- Girard, T., & Dion, P. (2010). Validating the search, experience, and credence product classification framework. *Journal of Business Research*, 63(9–10), 1079–1087.
- Gobet, F., Lane, P., Croker, S., Cheng, P., Jones, G., Oliver, I., & Pine, J. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5(6), 236–243.
- Godes, D., & Silva, J. C. (2012). Sequential and Temporal Dynamics of Online Opinion. *Marketing Science*, 31(3), 448–473.
- Godes, David, & Mayzlin, D. (2004). Using Online Conversations to Study Word-of-Mouth Communication. *Marketing Science*, 23(4), 545–560.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(SUPPL. 1), 5228–5235.
- Gu, B., Gu, B., & Park, J. (2012). The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products. *Information Systems Research*, 23(1), 182–196.
- Guo, J., Wang, X., & Wu, Y. (2020). Positive emotion bias: Role of emotional content from online customer reviews in purchase decisions. *Journal of Retailing and Consumer Services*, 52(July 2019).
- Guo, Y., Barnes, S. J., & Jia, Q. (2017a). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.
- Guo, Y., Barnes, S. J., & Jia, Q. (2017b). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.
- Hamby, A., Daniloski, K., & Brinberg, D. (2015). How consumer reviews persuade through narratives. *Journal of Business Research*, 68(6), 1242–1250.
- Haridasan, A. C., & Fernando, A. G. (2018). Online or in-store: unravelling consumer's channel choice motives. *Journal of Research in Interactive Marketing*, 12(2), 215–230.
- Harrison-Walker, L. J. (2001). The Measurement of Word-of-Mouth Communication and an Investigation of Service Quality and Customer Commitment As Potential Antecedents. *Journal of Service Research*, 4(1), 60–75.
- Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, 36(1), 20–38.
- Häubl, G., & Trifts, V. (2000). Consumer Decision Making in Online Shopping Environments : The Effects of Interactive Decision Aids Consumer Decision Making in Online Shopping Environments : The Effects of Interactive Decision Aids. *Marketing Science*, 19(1), 4–21.
- Haugtvedt, C. P., & Wegener, D. T. (1994). Message Order Effects in Persuasion: An Attitude Strength Perspective. *Journal of Consumer Research*, 21(1), 205.
- Hauser, J. R., & Koppelman, F. S. (1979). Alternative Perceptual Mapping Techniques : Relative Accuracy and Usefulness. *Journal of Marketing Research*, 16(4), 495–506.

- Henderson, G. R., Iacobucci, D., & Calder, B. J. (1998). Brand diagnostics: Mapping branding effects using consumer associative networks. *European Journal of Operational Research*, 111(2), 306–327.
- Heng, Y., Gao, Z., Jiang, Y., & Chen, X. (2018). Exploring hidden factors behind online food shopping from Amazon reviews: A topic mining approach. *Journal of Retailing and Consumer Services*, 42(January), 161–168.
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, 18(1), 38–52.
- Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective. *Journal of Consumer Research*, 17(4), 454.
- Herrmann, A., Goldstein, D. G., Stadler, R., Landwehr, J. R., Heitmann, M., Hofstetter, R., & Huber, F. (2011). The effect of default options on choice-Evidence from online product configurators. *Journal of Retailing and Consumer Services*, 18(6), 483–491.
- Ho-Dac, N. N., Carson, S. J., & Moore, W. L. (2013). The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter? *Journal of Marketing*, 77(6), 37–53.
- Hofmann, J., Clement, M., Völckner, F., & Hennig-Thurau, T. (2017). Empirical generalizations on the impact of stars on the economic success of movies. *International Journal of Research in Marketing*, 34(2), 442–461.
- Holbrook, M. B. (1978). Beyond Attitude Structure: Toward the Informational Determinants of Attitude. *Journal of Marketing Research*, 15(4), 545.
- Holtgraves, T., & Lasky, B. (1999). Linguistic power and persuasion. *Journal of Language and Social Psychology*, 18(2), 196–205.
- Hornik, K., & Grün, B. (2011). topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software*, 40(13), 1–30.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD'04*, 168.
- Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? the mediating role of online review sentiments on product sales. *Decision Support Systems*, 57(1), 42–53.
- Hu, Y., & Li, X. (2011). Context-Dependent Product Evaluations: An Empirical Analysis of Internet Book Reviews. *Journal of Interactive Marketing*, 25(3), 123–133.
- Huang, A. H., Chen, K., Yen, D. C., & Tran, T. P. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behavior*, 48, 17–27.
- Hutto, C. J., & Gilbert, E. E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14).". *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*. <http://sentic.net/>
- International, E. (2017). *Colour cosmetics in the US* (Issue May).
- Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011a). Language Style Matching Predicts Relationship Initiation and

- Stability. *Psychological Science*, 22(1), 39–44.
- Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011b). Language Style Matching Predicts Relationship Initiation and Stability. *Psychological Science*, 22(1), 39–44.
- Jalilvand, M. R., & Samiei, N. (2012). The effect of electronic word of mouth on brand image and purchase intention: An empirical study in the automobile industry in Iran. *Marketing Intelligence and Planning*, 30(4), 460–476.
- Jiang, S., Cai, S., Olle, G. O., & Qin, Z. (2015). Durable product review mining for customer segmentation. *Kybernetes*, 44(1), 124–138.
- Jiménez, F. R., & Mendoza, N. A. (2013). Too popular to ignore: The influence of online reviews on purchase intentions of search and experience products. *Journal of Interactive Marketing*, 27(3), 226–235.
- John, D. R., Loken, B., Kim, K., & Monga, A. B. (2006). Brand concept maps: A methodology for identifying brand association networks. *Journal of Marketing Research*, 43(4), 549–563.
- Johnson, E. J., Bellman, S., & Lohse, G. L. (2002). Defaults , Framing and Privacy :Why Opting In-Opting Out. *Marketing Letters*, 13(1), 5–15.
- Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, 80(4), 557–571.
- Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*, 33(2), 125–143.
- Kamble, Z., Doan, Q. T., Nguyen, H. N., Zeng, Z., & Zihui, L. L. (2020). The Significance of Online Reviews for Hotel Entrepreneurial Success. In *Entrepreneurial Development and Innovation in Family Businesses and SMEs* (pp. 75–86).
- Keller, K. L. (1993). *Conceptualizing, Measuring, and Managing Customer-Based Brand Equity*. 57(1), 1–22.
- Kelly, G. A. (1991). *The Psychology of Personal Constructs*.
- Khwaja, M. G., & Zaman, U. (2020). Configuring the evolving role of ewom on the consumers information adoption. *Journal of Open Innovation: Technology, Market, and Complexity*, 6(4), 1–13.
- Kim, H. B., Kim, W. G., & An, J. A. (2003). The effect of consumer-based brand equity on firms' financial performance. *Journal of Consumer Marketing*, 20(4–5), 335–351.
- Kim, M., & Lennon, S. (2007). The Effects of Visual and Verbal Information on Attitudes and Purchase Intentions in Internet Shopping. *Psychology & Marketing*, 24(9), 763–785.
- Kim, S. G., & Kang, J. (2018). Analyzing the discriminative attributes of products using text mining focused on cosmetic reviews. *Information Processing and Management*, 54(6), 938–957.
- King, M. F., & Balasubramanian, S. k. (1994). The effects of Expertise, End Goal, and Product Type on Adoption of Preference Formation Strategy. *Journal of the Academy of Marketing Science*, 22.

- Kostyk, A., Niculescu, M., & Leonhardt, J. M. (2017). Less is more: Online consumer ratings' format affects purchase intentions and processing. *Journal of Consumer Behaviour*, 16(5), 434–441.
- Kotler, P., & Armstrong, G. (2020). *Principles of Marketing* (18th Global). Pearson Education Limited.
- Kozinets, R. V., De Valck, K., Wojnicki, A. C., & Wilner, S. J. S. (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of Marketing*, 74(2), 71–89.
- Kübler, R. V., Colicev, A., & Pauwels, K. H. (2019). Social Media's Impact on the Consumer Mindset: When to Use Which Sentiment Extraction Tool? *Journal of Interactive Marketing*, 50, 136–155.
- Kudeshia, C., & Kumar, A. (2017). Social eWOM: does it affect the brand attitude and purchase intention of brands? *Management Research Review*, 40(3), 310–330.
- Lee, P. J., Hu, Y. H., & Lu, K. T. (2018). Assessing the helpfulness of online hotel reviews: A classification-based approach. *Telematics and Informatics*, 35(2), 436–445.
- Lee, S., & Choeh, J. Y. (2018). The interactive impact of online word-of-mouth and review helpfulness on box office revenue. *Management Decision*, 56(4), 849–866.
- Lee, S., & Choeh, J. Y. (2020). Using the social influence of electronic word-of-mouth for predicting product sales: The moderating effect of review or reviewer helpfulness and product type. *Sustainability (Switzerland)*, 12(19).
- Lee, T. Y., & Bradlow, E. T. (2011). Automated Marketing Research Using Online Customer Reviews. *Journal of Marketing Research*, 48(5), 881–894.
- Levinson, P. (2013). *New New Media*. Pearson.
- Lewsey, J. D., & Thomson, W. M. (2004). The utility of the zero-inflated Poisson and zero-inflated negative binomial models: A case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. *Community Dentistry and Oral Epidemiology*, 32(3), 183–189.
- Li, M., Huang, L., Tan, C.-H., & Wei, K.-K. (2013). Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features. *International Journal of Electronic Commerce*, 17(4), 101–136.
- Li, S. T., Pham, T. T., & Chuang, H. C. (2019). Do reviewers' words affect predicting their helpfulness ratings? Locating helpful reviewers by linguistics styles. *Information and Management*, 56(1), 28–38.
- Li, Xiaolin, Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information and Management*, 56(2), 172–184.
- Li, Xinxin, & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, 19(4), 456–474.
- Liang, T. P., Li, X., Yang, C. T., & Wang, M. (2015). What in Consumer Reviews Affects the Sales of Mobile Apps: A Multifacet Sentiment Analysis Approach. *International Journal of Electronic Commerce*, 20(2), 236–260.
- Liu, J., Cao, Y., Lin, C., Huang, Y., ... M. Z. the 2007 J. C. on, & 2007, U. (2007). Low-quality product review detection in opinion summarization. *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, June, 334–342.
- Liu, Q., & Karahanna, E. (2015). An agent-based modeling analysis of helpful vote on

- online product reviews. *Proceedings of the Annual Hawaii International Conference on System Sciences, 2015-March*, 1585–1595.
- Liu, X., Burns, A. C., & Hou, Y. (2017). An Investigation of Brand-Related User-Generated Content on Twitter. *Journal of Advertising*, 46(2), 236–247.
- Liu, Y. (2006). Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*, 70(3), 74–89.
- Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, 47, 140–151.
- Low, G. S., & Lamb, C. W. (2000). The measurement and dimensionality of brand associations. *Journal of Product & Brand Management*, 9(6), 350–370.
- Lozano, M. B., Martínez, B., & Pindado, J. (2016). Corporate governance, ownership and firm value: Drivers of ownership as a good corporate governance mechanism. *International Business Review*, 25(6), 1333–1343.
- Luca, M. (2011). Revenue : The Case of Reviews , Reputation , and Revenue : The Case of Yelp . com. *Harvard Business Review*, 1–40.
- Ludwig, S., de Ruyter, K., Friedman, M., Brügger, E. C., Wetzels, M., & Pfann, G. (2013a). More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*, 77(1), 87–103.
- Ludwig, S., de Ruyter, K., Friedman, M., Brügger, E. C., Wetzels, M., & Pfann, G. (2013b). More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*, 77(1), 87–103.
- Lynch, Jr., J. G., Marmorstein, H., & Weigold, M. F. (1988). Choices from Sets Including Remembered Brands: Use of Recalled Attributes and Prior Overall Evaluations. *Journal of Consumer Research*, 15(2), 169.
- Magerman, T., Van Looy, B., Baesens, B., & Debackere, K. (2011). Assessment of Latent Semantic Analysis (LSA) text mining algorithms for large scale mapping of patent and scientific publication documents. *FBE Research Report MSI_1114*.
- Magoulas, R., & Swoyer, S. (2020). *AI Adoption in the Enterprise 2020*.
- Mahr, D., Stead, S., & Odekerken-Schröder, G. (2019). Making sense of customer service experiences: a text mining review. *Journal of Services Marketing*, 33(1), 88–103.
- Malhotra, N. K. (1981). A Scale to Measure Self-Concepts, Person Concepts, and Product Concepts. *Journal of Marketing Research*, 18(4), 456.
- Mantonakis, A., Roderio, P., Lesschaeve, I., & Hastie, R. (2009). Order in choice: Effects of serial position on preferences. *Psychological Science*, 20(11), 1309–1312.
- Marchand, A., Hennig-Thurau, T., & Wiertz, C. (2017). Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. *International Journal of Research in Marketing*, 34(2), 336–354.
- Martínez Salinas, E., & Pina Pérez, J. M. (2009). Modeling the brand extensions' influence on brand image. *Journal of Business Research*, 62(1), 50–60.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.
- Moe, W. W., & Trusov, M. (2011). The Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing Research*, 48(3), 444–456.

- Mohammad, S., & Turney, P. (2010). Emotions Evoked by Common Words and Phrases. *Proceedings of the NAACL-HLT, June*, 26–34.
- Moon, S., & Kamakura, W. A. (2017). A picture is worth a thousand words: Translating product reviews into a product positioning map. *International Journal of Research in Marketing*, 34(1), 265–285.
- Mudambi, S., & Schuff, D. (2010). What makes a Helpful online Review? *MIS Quarter*, 28(4), 695–704.
- Munch, J. M., & Swasy, J. L. (1988). Rhetorical question, summarization frequency, and argument strength effects on recall. *Journal of Consumer Research*, 15(1), 69–76.
- Murphy, J., Hofacker, C., & Mizerski, R. (2006). Primacy and recency effects on clicking behavior. *Journal of Computer-Mediated Communication*, 11(2), 173–192.
- Nabi, R. L. (2003). Exploring the framing effects of emotion: Do discrete emotions differentially influence information accessibility, information seeking, and policy preference? *Communication Research*, 30(2), 224–247.
- Nazlan, N. H., Tanford, S., & Montgomery, R. (2018). The effect of availability heuristics in online consumer reviews. *Journal of Consumer Behaviour*, 17(5), 449–460.
- Nelson, P. (1970). *Information and Consumer Behavior*. 78(2), 311–329.
- Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine Your Own Business: Market-Structure Surveillance Through Text Mining. *Marketing Science*, 31(3), 521–543.
- Netzer, Oded, Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521–543.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Personality and Social Psychology Bulletin Lying Words : Predicting Deception From Linguistic Styles. *Personality and Social Psychology Bulletin*, 29(5), 665–675.
- Ngo-Ye, T. L., & Sinha, A. P. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, 61(1), 47–58.
- Nielsen. (2016). *The sweet smell of seasonal success*. Available at: <https://www.nielsen.com/au/en/insights/news/2016/the-sweet-smell-of-seasonal-success.html>
- Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *CEUR Workshop Proceedings*, 718, 93–98.
- Olson, J. C., & Muderrisoglu, A. (1979). The stability of responses obtained by free elicitation: Implications for measuring attribute salience and memory structure. *Advances in Consumer Research*, 6, 269–275.
- Pan, B., Zhang, L., & Law, R. (2013). The Complex Matter of Online Hotel Choice. *Cornell Hospitality Quarterly*, 41(1), 74–83.
- Pan, Y., & Zhang, J. Q. (2011). Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews. *Journal of Retailing*, 87(4), 598–612.
- Pang, J., & Qiu, L. (2016). Effect of online review chunking on product attitude: The moderating role of motivation to think. *International Journal of Electronic Commerce*, 20(3), 355–383.

- Park, D. H., & Lee, J. (2008). eWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications*, 7(4), 386–398.
- Park, E. J., Kim, E. Y., Funches, V. M., & Foxx, W. (2012). Apparel product attributes, web browsing, and e-impulse buying on shopping websites. *Journal of Business Research*, 65(11), 1583–1589.
- Park, H.-J., & Rabolt J., N. (2009). Cultural Value, Consumption Value, and Global Brand Image: A Cross-National Study. *Psychology & Marketing*, 26(8), 714–735.
- Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research*, 50, 67–83.
- Payan, J. M., & McFarland, R. G. (2005). Decomposing Influence Strategies: Argument Structure and Dependence as Determinants of the Effectiveness of Influence Strategies in Gaining Channel Member Compliance. *Journal of Marketing*, 69(3), 66–79.
- Payne, J. W. (1982). Contingent Decision Behavior. *Psychological Bulletin*, 92(2), 382.
- Peladeau, N. (2016). *WordStat: Content Analysis Module for SIM- STAT*.
- Pennebaker, J. W., Boyd, R., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. *Austin, TX: University of Texas at Austin, SEPTEMBER 2015*, 1–22.
- Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS ONE*, 9(12), 1–10.
- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The Development and Psychometric Properties of LIWC2007. *Austin, Tx: LIWC.Net*, 1(2).
- Petty, R. E., Tormala, Z. L., Hawkins, C., & Wegener, D. T. (2001). Motivation to think and order effects in persuasion: The moderating role of chunking. *Personality and Social Psychology Bulletin*, 27(3), 332–344.
- Pindado, J., Requejo, I., & de la Torre, C. (2011). Family control and investment-cash flow sensitivity: Empirical evidence from the Euro zone. *Journal of Corporate Finance*, 17(5), 1389–1409.
- Price, P. C., & Stone, E. R. (2004). Intuitive Evaluation of Likelihood Judgment Producers: Evidence for a Confidence Heuristic. *Journal of Behavioral Decision Making*, 17(1), 39–57.
- Puranam, D., Narayan, V., & Kadiyali, V. (2017). The effect of calorie posting regulation on consumer opinion: A flexible latent dirichlet allocation model with informative priors. *Marketing Science*, 36(5), 726–746.
- Purnawirawan, N., De Pelsmacker, P., & Dens, N. (2012). Balance and Sequence in Online Reviews: How Perceived Usefulness Affects Attitudes and Intentions. *Journal of Interactive Marketing*, 26(4), 244–255.
- Purnawirawan, N., Dens, N., & De Pelsmacker, P. (2012). Balance and Sequence in Online Reviews: The Wrap Effect. *International Journal of Electronic Commerce*, 17(2), 71–98.
- Racherla, P., & Friske, W. (2012). Perceived “usefulness” of online consumer reviews: An exploratory investigation across three services categories. *Electronic Commerce*

- Research and Applications*, 11(6), 548–559.
- Reynolds, T. J., & Gutman, J. (1979). Laddering theory, method, analysis, and interpretation. *Journal of Advertising Research*, 28(1), 11–31.
- Rodríguez-Díaz, M., Rodríguez-Díaz, R., & Espino-Rodríguez, T. (2018). Analysis of the Online Reputation Based on Customer Ratings of Lodgings in Tourism Destinations. *Administrative Sciences*, 8(3), 51.
- Roedder, D. J., Loken, B., Kim, K., & Monga, A. B. (2006). Brand concept maps: A methodology for identifying brand association networks. *Journal of Marketing Research*, 43(4), 549–563.
- Roodman, D. (2009a). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, 9(1), 86–136.
- Roodman, D. (2009b). Practitioners' corner: A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71(1), 135–158.
- Salehan, M., & Kim, D. J. (2015). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, 81, 30–40.
- Saumya, S., Singh, J. P., Baabdullah, A. M., Rana, N. P., & Dwivedi, Y. K. (2018). Ranking online consumer reviews. *Electronic Commerce Research and Applications*, 29(March), 78–89.
- Schnittka, O., Sattler, H., & Zenker, S. (2012). Advanced brand concept maps: A new approach for evaluating the favorability of brand association networks. *International Journal of Research in Marketing*, 29(3), 265–274.
- Selkirk, E. (1996). The prosodic structure of function words. In J. L. (Brown U. Morgan & K. (Brown U. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 187–214). Lawrence Erlbaum Associates, Inc.
- Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80(2), 159–169.
- Sheldon, P., & Bryant, K. (2016). Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in Human Behavior*, 58, 89–97.
- Shugan, S. M. (1980). The Cost of Thinking. *Journal of Consumer Research*, 7(2), 99–111.
- Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Kumar Roy, P. (2017). Predicting the “helpfulness” of online consumer reviews. *Journal of Business Research*, 70, 346–355.
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, 105(1), 131–142.
- Slovic, P. (1972). From Shakespeare to Simon: Speculations and some evidence. *Oregon Research Institute Bulletin*, 1–19. <https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/22179/36>
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality & Social Psychology*, 48(4), 813–838.
- Snizek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, 84(2), 288–307.

- Social Blade*. (2017). *Instagram stats*. Available at: <https://socialblade.com/>
- Spears, N., & Singh, S. N. (2004). Measuring attitude toward the brand and purchase intentions. *Journal of Current Issues and Research in Advertising*, 26(2), 53–66.
- Spenner, P., & Freeman, K. (2012). To keep your customers, keep it simple. *Harvard Business Review*, 90(5).
- SproutSocial. (2018). *Instagram stats*. Available at: <https://sproutsocial.com/insights/instagram-stats/>
- Statista. (2020). Available at: <https://www.statista.com/forecasts/997051/sources-of-information-about-products-in-the-us>
- Sun, M. (2012). How Does the Variance of Product Ratings Matter? *Management Science*, 58(4), 696–707.
- Sun, Q., Niu, J., Yao, Z., & Yan, H. (2019). Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level. *Engineering Applications of Artificial Intelligence*, 81(December 2018), 68–78.
- Tang, T. (Ya), Fang, E. (Er), & Wang, F. (2014). Is neutral really neutral? The effects of neutral user-generated content on product sales. *Journal of Marketing*, 78(4), 41–58.
<http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc11&NEWS=N&AN=2014-27450-004>
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
- Tirunillai, S., & Tellis, G. J. (2014). Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation. *Journal of Marketing Research*, 51(4), 463–479.
- Toubia, O., Iyengar, G., Bunnell, R., & Lemaire, A. (2019). Extracting Features of Entertainment Products: A Guided Latent Dirichlet Allocation Approach Informed by the Psychology of Media Consumption. *Journal of Marketing Research*, 56(1), 18–36.
- Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing*, 73(5), 90–102.
- Tybout, A. M., Sternthal, B., Malaviya, P., Bakamitsos, G. A., & Park, S. E. B. (2005). Information accessibility as a moderator of judgments: The role of content versus retrieval ease. *Journal of Consumer Research*, 32(1), 76–85.
- Van Hove, G., & Lievens, F. (2007). Social influences on organizational attractiveness: Investigating if and when word of mouth matters. *Journal of Applied Social Psychology*, 37(9), 2024–2047.
- Veloutsou, C., & Delgado-Ballester, E. (2018). New challenges in brand management. *Spanish Journal of Marketing - ESIC*, 22(3), 255–272.
- Vermeer, S. A. M., Araujo, T., Bernritter, S. F., & van Noort, G. (2019). Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media. *International Journal of Research in Marketing*, 36(3), 492–508.
- Vriens, M., Chen, S., & Schomaker, J. (2019). The evaluation of a brand association

- density metric. *Journal of Product and Brand Management*, 28(1), 104–116.
- Walker, M. A., Mehl, M. R., Moore, R. K., & Mairesse, F. (2007). Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research*, 30, 457–500.
- Wang, F., Liu, X., & Fang, E. (Er). (2015). User Reviews Variance, Critic Reviews Variance, and Product Sales: An Exploration of Customer Breadth and Depth Effects. *Journal of Retailing*, 91(3), 372–389.
- Wang, W., Feng, Y., & Dai, W. (2018). Topic analysis of online reviews for two competitive products using latent Dirichlet allocation. *Electronic Commerce Research and Applications*, 29(January), 142–156.
- Wang, X., Tang, L. (Rebecca), & Kim, E. (2019). More than words: Do emotional content and linguistic style matching matter on restaurant review helpfulness? *International Journal of Hospitality Management*, 77(August 2018), 438–447.
- Weathers, D., Swain, S. D., & Grover, V. (2015). Can online product reviews be more helpful? Examining characteristics of information content by product type. *Decision Support Systems*.
- Westerman, D., Spence, P. R., & Van Der Heide, B. (2014). Social Media as Information Source: Recency of Updates and Credibility of Information. *Journal of Computer-Mediated Communication*, 19(2), 171–183.
- Wickham, H. (2016). *tidyr: Easily tidy data with spread and gather functions*.
- Wickham, H. (2017). ggplot2 - Elegant Graphics for Data Analysis (2nd Edition). *Journal of Statistical Software*, 77(Book Review 2), 2–5.
- Wickham, H., & Francois, R. (2016). *dplyr: A Grammar of Data Manipulation*.
- Winkelmann, P. D. R. (2008). Econometric Analysis of Count Data. In *Econometric Analysis of Count Data*.
- Wong, C. U. I., & Qi, S. (2017). Tracking the evolution of a destination's image by text-mining online reviews - the case of Macau. *Tourism Management Perspectives*, 23, 19–29.
- Wooldridge, J. M. (2011). *Econometric analysis of cross section and panel data* (Issue November).
- Wu, J. (2017). Review popularity and review helpfulness: A model for user review effectiveness. *Decision Support Systems*, 97, 92–103.
- Xu, P., & Liu, D. (2019). Product engagement and identity signaling: The role of likes in social commerce for fashion products. *Information and Management*, 56(2), 143–154.
- Yang, H., & Cho, S. (2015). *Understanding Brands with Visualization and Keywords from eWOM Using Distributed Representation*.
- Yau, K. K. W., Wang, K., & Lee, A. H. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal*, 45(4), 437–452.
- Yazdani, E., Gopinath, S., & Carson, S. (2018). Preaching to the Choir: The Chasm Between Top-Ranked Reviewers, Mainstream Customers, and Product Sales. *Marketing Science*, 37(5), 838–851.
- Yin, G., Wei, L., Xu, W., & Chen, M. (2014). Yin, G., Wei, L., Xu, W., & Chen, M. (2014).

Exploring Heuristic Cues for Consumer Perceptions of Online Reviews Helpfulness: The Case of Yelp.com. PACIS 2014 Proceedings, 1–12. Exploring Heuristic Cues for Consumer Perceptions of Online Reviews Helpfulness. *PACIS 2014 Proceedings*, 1–12.

Zaltman, G. (1997). Rethinking Market Research : Putting People Back In. *Journal of Marketing Research*, 34(4), 424–437.

Zhang, J. (2019). What's yours is mine: exploring customer voice on Airbnb using text-mining approaches. *Journal of Consumer Marketing*, 36(5), 655–665.

Zhang, Ying, & Lin, Z. (2018). Predicting the helpfulness of online product reviews: A multilingual approach. *Electronic Commerce Research and Applications*, 27, 1–10.

Zhang, Yuchi, Moe, W. W., & Schweidel, D. A. (2017). Modeling the role of message content and influencers in social media rebroadcasting. *International Journal of Research in Marketing*, 34(1), 100–119.

Zhou, S., & Guo, B. (2017). The order effect on online review helpfulness: A social influence perspective. *Decision Support Systems*, 93, 77–87.

Zhu, F., & Zhang, X. (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*, 74(March), 133–148.

1

Review Characteristics and Review Helpfulness: The Role of Review Visibility

1. Introduction

Online consumer reviews are a type of electronic word-of-mouth (eWOM) **communication that can be defined as “peer-generated product evaluations posted on company or third-party websites”** (Mudambi & Schuff, 2010). Both academics and practitioners have highlighted the importance of online reviews for both consumers and companies. A study by the consultancy firm BrighLocal (2019) reveals that 82% of consumers read online reviews when evaluating a business and 76% trust online reviews as much as personal recommendations. Besides, the same study reveals that including online reviews on the retailer website makes the searchers to see the business as more trustworthy.

Previous literature has also highlighted the power of online reviews to predict product sales in product categories such as hardware, books, movies and hotels (Chevalier & Mayzlin, 2006; Chintagunta et al., 2010; Li et al., 2019; Marchand et al., 2017). However, not every online review might have an equal **influence on consumers’ purchase decisions**. It has been stated that those online reviews voted as helpful by other consumers have a higher influence on consumer decision making (Li et al., 2013; Singh et al., 2017). In this line, literature has claimed that a review is helpful if it aids consumers to understand and evaluate the quality and performance of products at one or more stages of the consumer decision-making journey (Mudambi & Schuff, 2010). Spenner and Freeman (2012) stated that the easier the decision-making process is for a consumer, the higher the probability of purchasing. In this line, online retailers holding more helpful reviews offer greater potential value for consumers (Huang et al., 2015) and consumers are more likely to buy products on their site (Huang et al., 2015; Spenner & Freeman, 2012). Most online **retailers include the question “Was this review helpful?” at the end of each online review**. Answering this question implies a consumer evaluation of the helpfulness of a review during the consumer decision-making journey (Mudambi & Schuff, 2010). The number of helpful votes of online reviews is an information cue so important for consumers that,

when available, they usually sort online reviews by the most helpful order (Liu & Karahanna, 2015).

Previous literature has mainly focused on exploring direct relationships between review characteristics (e.g., review rating and review length) and review helpfulness. However, online reviews are not displayed individually but in a sequence next to other reviews, so the relative visibility of reviews in the sequence is likely to influence consumer decision-making. Grounded on the accessibility-diagnostics theory (Feldman & Lynch, 1988), we posit that the number of helpful votes a review receives, referred as in previous literature as review helpfulness, might not only depend on those review characteristics that determine the diagnosticity of online reviews, but also on review accessibility factors, which determine how visible are online reviews for consumers. Based on the theory, the two dimensions might interact in explaining consumer decision-making.

The diagnosticity of the review can be described as the perceived ability of the review to provide consumers with relevant product information that helps them to understand and evaluate the quality and performance of the product (Filieri, Hofacker, et al., 2018). Overall, studies have claimed that an input's diagnosticity depends on whether it enables a decision maker to discriminate among alternatives and it depends on the characteristics of the input of information (Payne, 1982). To approach the review diagnosticity dimension, which determine how diagnostic is the review for the consumer, we focus on two groups of review characteristics that have been claimed to influence review helpfulness: review non-textual characteristics (e.g., review rating) and review textual characteristics (e.g., argument structure).

The review accessibility dimension is approached in this research by the concept of review visibility, which reflects the probability of a review to be viewed by the consumer. Review visibility is proxied in this research by the rank order of the review in the sequence of reviews for a given product when consumers sort reviews using different mechanisms. We explore the effect of review visibility when using two review sorting mechanisms, which are very common in reviews' web pages: sorting by "most helpful" and sorting by "most recent" reviews. As stated by Bruine de Bruin and Keren (2003), whenever options appear in sequence, judgments may be vulnerable to potential order effects. The effects of information accessibility in consumer decision-making might be even more relevant in online contexts, where consumers usually face information overload situations, as it happens when dealing with high volume of online reviews. As stated by the decision-making theory (Beach, 1993), when consumers are in complex environments they are often unable to evaluate all available alternatives in depth before

making a choice. Instead, they adopt selection mechanisms, such as sorting and filtering, to reduce the cognitive effort of evaluating a big number of alternatives (Häubl & Trifts, 2000). In this line, Liu and Karahanna (2015) claim that consumers read on average 7 reviews before making a decision. A study by the consultancy firm BrightLocal (2020) revealed also that 80 percent of consumers read a maximum of 10 reviews before trusting a business. These findings support the idea that consumers do not read all available reviews and, instead, they tend to focus on a small number of reviews to evaluate the product, being more visible reviews more likely to be evaluated by consumers.

Two main contributions can be pointed out from this research.

First, we incorporate the role of review visibility in explaining the relationship between those review characteristics that determine how diagnostic a review is for consumers and review helpfulness. If review visibility factors are not considered, as happens in most previous literature, we might overlook the likely impact of those factors on the consumer voting decision, and all the changes in review helpfulness would be attributed to the impact of review characteristics. In this research, we take review visibility as a moderator¹, since we suggest both a direct effect of review visibility on review helpfulness and an interactive effect of review visibility with review characteristics in explaining review helpfulness. In this way, we would be able to know the impact on the consumer voting behaviour of those characteristics of online reviews ranked in top positions, by mechanisms **such as “most recent” or “most helpful”**.

Second, we propose a novel perspective for understanding consumer voting behaviour. We posit that to better explore review helpfulness, we should consider that there might be an excess of online reviews with zero helpful votes due to two different reasons: (1) not viewing the review and (2) viewing the review and deciding not to vote it as helpful. Considering this, we apply a Zero Inflated Negative Binomial (ZINB) regression, which allows us to model review helpfulness in two parts. First, a logit part that allows us to understand the drivers explaining that online reviews do not receive helpful votes because they have not been viewed. Second, a negative binomial part that explains the likelihood that online reviews receive helpful votes.

From a managerial perspective, more understanding about the factors influencing the consumer voting behaviour might help online retailers to improve the design of user

¹ In some papers, what we call moderating variable in this research is also referred to as “quasi-moderator”. Quasi-moderators affect the true relationship between two variables and, at the same time, are independently associated to the dependent variable.

rating systems to make the navigation experience more satisfactory for consumers. To carry out the empirical analysis, data on online reviews belonging to the category of **“blush”** were gathered at two points of time in a six-month period from a US cosmetics online retailer.

The rest of the paper is structured as follows: First, the theoretical background and research model are presented. Then, the data and research methodology are explained. Afterwards, the conclusion, followed by a general discussion about theoretical and managerial implications is developed. Finally, last section shows the main limitations and areas for future research.

2. Theoretical Background and Research Model

2.1. What Makes a Review Helpful?

Previous review helpfulness literature has mainly focused on studying how specific characteristics of online reviews influence review helpfulness on different product settings. In general, these characteristics can be organized into two different groups: review non-textual characteristics and review textual characteristics (Baek et al., 2012; Cao et al., 2011; Gao et al., 2017; Li et al., 2013).

Review Non-textual Characteristics

Review non-textual features are those structured characteristics of online reviews that can be easily identified without reading the text of the review. The rating of the review, which captures whether the review provides a positive or negative evaluation of the product, has been the most explored non-content feature in previous literature. Findings regarding the effect of the rating on review helpfulness are ambiguous, which means that there is room for more research in this area. Some scholars state that negative reviews capture more consumers' attention and are more likely to be voted as helpful (Wu, 2017). However, others claim a positive relationship between rating and review helpfulness (Pan & Zhang, 2011). A big stream of literature in impression formation has focused on exploring how extreme opinions are perceived by consumers. Extreme ratings refer to a review that contains either an extremely positive or extremely negative evaluation of a product or service based on the rating score. Many scholars have found that extreme information is usually seen as less ambiguous and more explanatory than moderate information (Filleri, Raguseo, et al., 2018; Skowronski & Carlston, 1989). Overall,

existing research has shown that extreme ratings are more likely to be voted as helpful by consumers (Filieri, Raguseo, et al., 2018; Liu & Park, 2015; Park & Nicolau, 2015).

Information processing and persuasion theories suggest that communicator credibility influences consumers to agree or disagree with a message, in the way that those statements made by communicators considered as experts can usually be considered veridical (Chaiken, 1987). In an online environment, where reviewers are usually anonymous, consumers tend to trust on several reviewer characteristics disclosed in the review, which are a sign of source credibility. In this line, review helpfulness literature has claimed that the information provided by high credible sources, is perceived as more helpful and reliable (Cheung et al., 2008). For example, online reviews written by reviewers who post many online reviews at a specific platform are more likely to receive **helpful votes because they are seen as “experienced” reviewers** (Cheng & Ho, 2015; Racherla & Friske, 2012). Reviewer personal information disclosure is also a sign of reviewer credibility. Some scholars state that reviews containing reviewers’ identity information are more likely to be voted as helpful (Baek et al., 2012; Forman et al., 2008). Those studies indicate that the **reviewer’s** real name and the location disclosure positively influence review helpfulness.

Review Textual Characteristics

Review textual cues have been less studied in previous literature. Some articles have explored review content by either using review length as a proxy of argument quality (Chevalier & Mayzlin, 2006; Huang et al., 2015) or review rating as a proxy of affect (Chua & Banerjee, 2016; Singh et al., 2017). In this line, previous literature has widely claimed that longer reviews usually provide more information to help consumers in the decision-making process and it has been found evidence to state that longer reviews are more likely to receive helpful votes (Baek et al., 2012; Cheng & Ho, 2015; Mudambi & Schuff, 2010; Racherla & Friske, 2012). A smaller stream of research has focused on other textual aspects of online reviews, such as emotions. Some scholars have claimed that different emotions (e.g. fear vs. anger) have different effects on the outcome variable, even if the emotions are of the same valence (Nabi, 2003). In this line, Ahmad and Laroche (2015) conducted a more exhaustive analysis of review emotions to explore how hope, happiness, anxiety and disgust influence review helpfulness. They revealed that emotions expressed with certainty, regardless of the valence, have a higher impact on review helpfulness. For example, both happiness and hope are positive emotions, and disgust and anxiety are negative emotions. However, happiness and disgust are associated with certainty and hope and anxiety with uncertainty, according to cognitive

appraisal theory (Ahmad & Laroche, 2015). Wang et al., (2019) explored review helpfulness not only including affective characteristics of online reviews, but also linguistic aspects. They found that linguistic aspects have a significant impact on review helpfulness in the hotel industry, especially the use of prepositions and auxiliary verbs. As described by Ludwig et al. (2013), the linguistic style is a combination of two different categories of words: lexical words, which include adjectives, nouns, verbs and most adverbs, and function words, which include prepositions, pronouns, auxiliary verbs, conjunctions, grammatical articles or particles (Selkirk, 1996). The linguistic style of the review may serve as identity-descriptive information that shapes consumers evaluations of the review and the product (Ludwig et al., 2013a). In fact, social psychology and communication theories show that the way or style in which a person communicates elicits relational perceptions in the communication partner and influences consumer judgments and behaviors (Ludwig et al., 2013a; Smith & Ellsworth, 1985). In this line, Li et al., (2019) found that the linguistic style of a person writing language has a significant impact on tracking helpful reviewers. In this context, some scholars have explored how the degree of objectivity of a review influences its helpfulness (Ghose & Ipeiroitis, 2011; Singh et al., 2017). However, there is not a consensus in the direction of the effects. Some of them have found that subjectivity positively influences review helpfulness (Chen & Tseng, 2011), others have claimed that those reviews containing a mixture of subjective and objective elements are more helpful (Archak et al., 2011; Ghose et al., 2012), while others did not find relationship between subjectivity and review helpfulness (Liu et al., 2007).

As claimed by Zhou and Guo (2017), most existing academic papers have explored review helpfulness from the perspective of a single review, but online reviews are rarely presented in isolation but as a part of a set of reviews. However, a very small part of literature has considered the role of review visibility when exploring review helpfulness.

2.2. Consumer Voting Behavior: Why is Review Visibility Relevant?

Review visibility is defined in this research as the probability of a review to be viewed by a consumer. This probability is based on the rank order when using different review sorting tools available on the website (Godes & Silva, 2012; Zhou & Guo, 2017). The concept of review visibility is based on the study of Godes and Silva (2012). The term "view" in our research does not necessarily imply reading the text of a review. Sometimes, the consumer may only scan it and look at features such as review rating or reviewer identity disclosure. Once consumers view a specific online review, they can decide if voting it as helpful. Then, review visibility becomes relevant when analysing review

helpfulness because, to be able to vote a review as helpful, consumers need to have previously viewed the specific review. If a review is not viewed by consumers, they cannot evaluate its characteristics (e.g., rating and length) and they cannot vote it as helpful. This decision making journey referred to in this research as "Consumer Voting Journey" is shown in Figure 1 and it is based on the traditional decision-making journey proposed by Payne (1982). This journey underlines the main steps that a consumer might follow in a website before facing the decision of voting a specific review as yes helpful, not helpful or no voting. Appendix A shows the sequence of screenshots corresponding to

Figure 1. Consumer decision-making journey when voting online reviews: Consumer Voting Journey (based on Payne, 1982)

each stage of the suggested “Consumer Voting Journey” on the cosmetics online retailer explored. This journey is explained to justify the relevancy of review visibility when exploring review helpfulness.

A typical “Consumer Voting Journey” starts when consumers are interested in a specific category and visit an online retailer website category page, where a list of products from the category is displayed (*product consideration* stage). When a specific product of the set is considered and clicked on (product consideration does not imply buying the product but only considering it to buy), the product page and its reviews are shown on the website. At this point, consumers are in a *review consideration* stage. Many online reviews (sometimes thousands) are usually available to evaluate the product, so consumers, to reduce the information overload, must decide which ones to read (when speaking about reading a review, it might imply either reading the review text or just scanning the visible features). Some studies have revealed that consumers, on average, read between 5 and 10 reviews of a product before making a purchase decision (Q. Liu & Karahanna, 2015). Websites usually offer some filtering and sorting tools to display the reviews. When consumers decide to read a specific online review, they reach the *voting decision* stage, where they evaluate the usefulness of the review and decide if voting it as helpful or not helpful, or not voting. The review helpfulness variable can be understood as the aggregation of every individual “Consumer Voting Journey”. Our research focus starts on the *review consideration* stage, so we assume consumers have reached the product page at the online retailer. There are different paths consumers can follow before reaching the product page, but it is not the focus of this research.

To explore the role of review visibility in explaining review helpfulness we assume in the empirical research that the product page containing the reviews has been presented to the consumer. In other words, the consumer has reached the *review consideration* stage. Although we do not know how each individual consumer sorts online reviews at the *review consideration* stage, we do know the exact rank order of each review according to the available website sorting mechanisms, so we are able to analyse how this rank order influences the number of helpful votes online reviews receive².

We believe that our approach to review visibility adds value to the existent review helpfulness literature, providing a greater understating of actual consumer voting behaviour. If only direct effects of review characteristics on helpfulness are considered, as it is traditionally done in the literature, we would leave out the likely moderating influence of visibility on the consumer voting behaviour. Note that online reviews may have zero or a low number of yes helpful votes not because the review is not helpful or diagnostic itself, but because it is less visible than other reviews for a given product when **sorting by mechanisms such as “most recent” or “most helpful”**. In fact, most online retailers provide a default sorting mechanism, which might bias the final consumer decision of voting a review as helpful.

Very few scholars have somehow incorporated the effect of review visibility when studying review helpfulness. Zhou and Guo (2017) explored how the order of online reviews when sorting by most recent date influences review helpfulness from the perspective of the social influence theory. They use some review and reviewer characteristics as moderating variables to characterize a reviewer's susceptibility or response to social influence in rating behaviour. They found a negative relationship between review order and review helpfulness, which means that all else being equal, the helpfulness of reviews in a sequence will decline over order. This finding supports the argument that social influence in online reviews might exert a detrimental effect on the helpfulness of subsequent reviews. However, the authors did not explore social influence **at other review order scenarios, such as when reviews are ordered by the “most helpful”** ranking. Besides, they only explored two review characteristics, review rating and review length, leaving out the effect of other features, such as textual characteristics.

Wu (2017), explored three elements of review effectiveness: review polarity, review helpfulness and review persuasiveness. In the study, the author incorporated the effect

² Even though previous theories of information processing explain consumer decisions at the individual level, they have been widely used in previous literature to explain review helpfulness at the aggregate level, as it is in our case (Baek et al., 2012; Cheng & Ho, 2015).

of the order of the review by most helpful but not the effect of other alternative ordering mechanisms, such as order by most recent. Besides, Wu (2017) only explored a direct effect of review order on review helpfulness and not a moderation effect. Overall, both studies claim that the rank order of online reviews are likely to influence review helpfulness. These studies focus only on some specific review and reviewer characteristics, leaving out relevant factors, such as review content features, which might be also important when exploring review helpfulness.

Therefore, due to the scarce literature linking review visibility and review helpfulness, we are interested in getting more insights of how review visibility influence review helpfulness. First, unlike previous literature we include two review sorting mechanisms in our study (most helpful and most recent). Since most helpful and most recent reviews provide different information, which might have different characteristics, the information contained in those reviews might also exert different influence on consumer behaviour. Second, we adopt a different perspective than previous scholars incorporating the role of review visibility. We suggest that there might be an excess of online reviews with zero helpfulness votes not because they are bad quality, but because they are unlikely to be viewed by consumers. Therefore, we study review helpfulness in two stages. First, we study the factors, including review visibility, which explains that online reviews do not receive helpful votes because they have not been viewed. Second, we study the factors, including also review visibility, that impact the likelihood that online reviews receive helpful votes.

2.3. Conceptual Model and Hypotheses Development

Figure 2 shows the conceptual model proposed in this research. The main objective is to examine how review visibility moderates the effect of review characteristics on review helpfulness. The accessibility-diagnostics theory (Feldman & Lynch, 1988) is taken as the theoretical foundation of our conceptual model. It is an information processing theory that explains the determinants of consumers' adoption of information and it has been widely used in the consumer psychology field to explore the relationship between information and consumer choices (Bettman et al., 1998) and between advertising and brand choice (Baker, 2001; Baker et al., 2000).

Figure 2. Conceptual Model

The accessibility-diagnostics theory (Feldman & Lynch, 1988) states that the probability that any piece of information will be adopted as an input for making a choice depends on the accessibility of that input, the accessibility of the alternative inputs and the diagnosticity or perceived relevance of the input (Herr et al., 1991; Van Hove & Lievens, 2007). Holding constant the diagnosticity of alternative inputs, any factor that influences the accessibility of an input affects its adoption (Lynch et al., 1988). The second premise of the theory, accessibility of the alternative inputs, states that given two inputs in memory, A and B, any factor that increases the accessibility of B will decrease the use of A. Even when many diagnostic inputs are available in memory, decreasing the accessibility of one increases the attempt to retrieve another (Lynch et al., 1988). In our research, online reviews are the inputs of information that consumers evaluate before making a choice about the products.

From the perspective of the accessibility-diagnostics theory, the consumer decision of voting a review as helpful might imply an adoption of the information contained in the review, and this decision might not only depend on the characteristics of the review, which might determine how diagnostic the review is for consumers, but also on the visibility of online reviews, which might determine how visible or accessible the review is for consumers.

Relationship between Review Visibility and Review Helpfulness

In this research, the accessibility dimension is approached by review visibility, which depends on the rank order of online reviews when using specific sorting tools. Sorting tools affect the order in which reviews are displayed at the product page in the online retailer website. Depending on this, some reviews are more visible than others, and this visibility might play a role in the final voting decision.

In line with the accessibility-diagnostics theory is the theory about serial position effects and sequential bias, which claims that consumers are more likely to adopt the information to which they are more exposed (Haugtvedt & Wegener, 1994; Jonas et al., 2001; Mantonakis et al., 2009). Scholars have focused on studying how order effects, particularly primacy and recency effects, influence consumer behaviour.

Primacy effect refers to “a cognitive bias that occurs when the first item of a sequence is chosen over all other items because the cognitive capacity of the short term memory at the beginning of the list is still unburdened, compared to its status in the middle of the list” (Haugtvedt & Wegener, 1994; Purnawirawan et al., 2012). In contrast, the recency effect refers to “the tendency to recall or take into account the last item in a list or sequence” (Haugtvedt & Wegener, 1994; Purnawirawan et al., 2012).

Most papers have explored the effect of primacy and recency in product choice (Biswas et al., 2010; Mantonakis et al., 2009) and information persuasion (Haugtvedt & Wegener, 1994; Petty et al., 2001). For example, Biswas et al. (2010) observed a recency effect when sampling experiential products (e.g. beverages, music) but primacy effects when sampling non experiential products (e.g. scissors). Regarding information persuasion, Haugtvedt and Wegener (1994) found that in situations that foster high levels of message elaboration consumers are influenced by primacy effects, whereas in situations that foster low levels of message elaboration, consumers experience recency effects.

Although previous literature has mainly focused on studying the effect of position effects in memory and intentions, there is also research that suggest that position effects go beyond memory and can influence actual behaviour (Murphy et al., 2006). It has been claimed that when information about a single entity (e.g. a product) is presented sequentially in online environments, there can be primacy effects (Drèze & Zufryden, 2004), recency effects (Buda & Zhang, 2000) or both effects simultaneously (Mantonakis et al., 2009; Purnawirawan et al., 2012). As claimed by Nazlan et al. (2018), sequential

effects are even more important in the online purchasing environment, because consumers have to scroll down to view additional information. Breugelmans et al. (2007) explored how position effects in online grocery stores influences consumer choices and they found that product choice probability increased when product was presented in the first screen or located near focal items. For instance, Felfernig et al. (2007) found that a product presented in first position was chosen 2.5 times more often than any other product, regardless of its quality or other characteristics (Purnawirawan et al., 2012). Ert and Fleischer (2016) revealed that hotel position on a web page influences the hotel choice, although the position has nothing to do with the hotel attributes. In this line, Pan et al. (2013) found that consumers' attention was drawn to hotels at the top positions of the web page. Ansari and Mela (2003) found also both primacy and recency effects in their study about serial positions effects in online clicking behaviour, although primacy effects were stronger.

Overall, literature studying position effects in online environments has found that consumers' shopping decisions are particularly influenced by primacy effects, so that items on the first screen of the online retailer are more likely to be selected (Breugelmans et al., 2007; Ert et al., 2016). Thus, the serial position effects theory can also be used to justify the relevancy of review visibility (Haugtvedt & Wegener, 1994; Purnawirawan et al., 2012). In this line, primacy effects might be particularly relevant in an online review setting, in the way that consumers might be more likely to evaluate those reviews appearing in first positions at the online retailer.

In the same line, consultancy reports studying online reviews have revealed that consumers read on average a maximum of ten online reviews before making a decision about a firm, product or service (BrightLocal, 2020). These findings might also indicate that online reviews appearing in top ranked positions at the online retailer are more likely to be evaluated.

In our research, review visibility is proxied using two of the most relevant review sorting mechanisms for consumers: sorting by most helpful and sorting by most recent (BrightLocal, 2020; Liu & Karahanna, 2015; Saumya et al., 2018; Wu, 2017). However, there are other factors that might also influence review visibility, such as other sorting mechanisms (e.g., sorting by highest or lowest rating) and filtering options (e.g., filtering by physical appearance of the reviewer). Depending on the mechanism we use to proxy review visibility, some reviews are going to be more visible than others, but every mechanism is going to increase the probability that a review will be visible for consumers. According to the accessibility-diagnostics theory (Feldman & Lynch, 1988) and as

claimed by Herr, Kardes and Kim (2002), any factor that increases the accessibility of an input of information, should also increase the likelihood which that input will be adopted. Moreover, we can find further evidence to justify the relevancy of both review sorting mechanisms, most helpful and most recent, in consumer decision-making and, thus, in explaining review helpfulness.

Influence of “most helpful” reviews

There is evidence that consumers experience the “wisdom of the crowd” effect when evaluating online reviews (Liu & Karahanna, 2015; Zhou & Guo, 2017). This effect refers to the belief that the aggregation of many people’s judgements is a better approximation to the truth than an individual judgement. In this line, Zhou and Guo (2017) also claimed that social informational influence occurs as a consequence of the tendency of prospective consumers to conform to previous consumers’ opinions. For instance, if prospective consumers know that many other consumers have already bought the product and or that other consumers have highly rated the product, they might be more likely to select it as a promising alternative and they might have a better attitude towards the product (Filiari & McLeay, 2014; Pang & Qiu, 2016). Liu and Karahanna (2015) found in an experiment that when sorting by most helpful and most recent options were available, 70 percent of consumers in their sample sorted online reviews in Amazon.com **by the “most helpful” criterion, while 30 percent sorted them by the “most recent” mechanism.** As stated by Singh et al. (2017), since most helpful reviews have higher exposure to consumers, they normally become even more helpful due to a social influence effect. Lee, Hu and Lu (2018) and Saumya et al. (2018) also state in their research that those reviews in top positions in the ranking by “most helpful” are more likely to be evaluated by consumers. Thus, we hypothesize:

*H1a: More visible online reviews when **ranked by the “most helpful”** mechanism are more likely to be voted as helpful*

Influence of “most recent” reviews

Previous literature has also highlighted the importance of information recency in consumer behaviour. Westerman et al (2014) pointed out the relevancy of recency in explaining source credibility in online environments. In the same line, Fogg et al. (2001) found that consumers associated websites that update information more frequently with higher credibility. Other scholars, such as Levinson (2013), claim that social networks hallmark is the immediacy of messages, which is one of the factors that make them more

credible for consumers. Nevertheless, the effect of information recency has been scarcely explored in the review helpfulness literature. Zhou and Guo (2017) found that the order of online reviews when ranked by recency was negatively related to review helpfulness. Since order is inversely related to visibility (lower-order reviews are those in top ranked positions and, thus, more visible), this result suggests that review visibility when ordering by most recent is positively related to review helpfulness.

The study of online reviews has also captured the attention of practitioners. In this line, the consultancy company BrightLocal (2020) revealed that recency was one of the most important factors of online reviews for consumers when judging a business, close to factors such as review rating and text sentiment. The study reveals that recency was a very important factor for 80 percent of consumers and 50 percent of them said that online reviews have greater impact on their decisions if they have been written within the last two weeks. One of the reasons might be that consumers want to know up-to-date information about those businesses, products and services they are interested in. Since they can be modified over time, consumers are interested in knowing how the business, the product or the service performs at present.

The relevancy of the most recent sorting mechanism for consumer decision-making might not only be because of the date itself but also to the fact that it is the default review sorting mechanism at the online retailer. As defined by Brown and Krishna (2004), a default can be interpreted as an option that the individual receives to the extent that s/he does not willingly decide on something else. Existing research supports the idea that consumers are biased by the default. For example, Johnson et al. (2016) claim that consumers consider defaults to reduce the cognitive effort required to make a decision.

In this line, information processing theories reveal that many consumers usually adopt the information that is readily available in order to reduce the cognitive effort associated with decision making (Häubl & Trifts, 2000; Nazlan et al., 2018). For example, Slovic (1972) suggested that consumers tend to use only the information that is explicitly displayed, and they will use it in the form it is displayed because that behaviour reduces the cognitive effort required to process information (Bettman et al., 1998). Herrmann et al. (2011) also claimed that defaults influence decision-making behaviour even when consumers do not actually select the default option. Thus, review visibility when sorting by most recent, which is also the default mechanism at the online retailer explored, is likely to influence consumer voting decisions. Following previous findings, we hypothesize:

H1b: More visible online reviews when ranked by the “most recent” mechanism are more likely to be voted as helpful

Moderating Role of Review Visibility in the Relationship between Review Characteristics and Review Helpfulness

The diagnosticity of the review can be described as the perceived ability of the review to provide consumers with relevant product information that helps them to understand and evaluate the quality and performance of the product (Filieri, Hofacker, et al., 2018). Overall, studies have claimed that an input’s diagnosticity depends on whether it enables a decision maker to discriminate among alternatives and it depends on the characteristics of the input of information (Payne, 1982). Therefore, review characteristics, such as review rating and reviewer information disclosure, might determine how consumers perceive the review to be diagnostic (Filieri, Hofacker, et al., 2018).

Most papers in the review helpfulness literature have focused on exploring the direct relationships between review characteristics and review helpfulness (Cao et al., 2011; Baek et al., 2012). However, on the basis of the accessibility-diagnosticsity theory (Feldman & Lynch, 1988), we cannot see accessibility and diagnosticsity as two factors independent one from each other. On the contrary, diagnostic information that is also more accessible, is more likely to be adopted. Thus, the consumer decision of voting a review as helpful might not only depend on the diagnosticity of the review itself, which is determined by review characteristics, but also on the accessibility of that review, which is determined by the visibility of the review.

In this line, previous literature has explored the moderating role of information accessibility in consumer persuasion and brand choice. For example, Tybout et al. (2005) revealed a moderating role of accessibility in consumer judgements for two different brands, comparing the role of information content versus retrieval ease. They claimed that, when relevant knowledge is highly accessible or not at all accessible, judgments are based on the content of information considered. However, between those extremes of information accessibility, judgments are based on perceived ease of information retrieval. In a similar line, Biehal and Chakravarti (1983) claimed that accessibility differences may influence how the information is used in subsequent brand choice situations, affecting final choice outcomes. For example, previously encountered brands for which information is difficult to retrieve might be ignored in subsequent choices, in favour of brands for which information is externally available, which might lead to the

choice of suboptimal brands. When exploring the effect of Word-of-Mouth (WOM) on persuasion, Herr et al. (1991) claimed that information accessibility mediates the effect of WOM content characteristics on persuasion. They found a strong impact of information accessibility on persuasion, which was reduced when more diagnostic information was available. Likewise, Pan et al. (2013) studied online hotel choice and they found that consumers' attention was drawn to hotels at top positions of the hotel booking website and that information located at those positions had a stronger influence on consumers' decisions.

Thus, based on previous findings and on the accessibility-diagnostics theory (Feldman & Lynch, 1988) we expect that the accessibility of online reviews, approached by review visibility variables, moderates the relationship between those characteristics that make a review diagnostic and review helpfulness. Thus, we hypothesize:

H2a: Review visibility when online reviews are ranked by the “most helpful” mechanism moderates the impact of review characteristics on review helpfulness

H2b: Review visibility when online reviews are ranked by the “most recent” mechanism moderates the impact of review characteristics on review helpfulness

3. Research Methodology

3.1. Data

Figure 3. Typical online consumer review at the online retailer

We collected online consumer reviews from a popular US cosmetics retailer website, which was placed in the top-50 shopping sites in the US in May 2017 according to

alexa.com. Figure 3 shows an example of a typical online consumer review on the cosmetics website. We collected all the online reviews posted in the online store for all the products in the "blush" category.

Data were collected on two different days, October the 4th 2016 (63,985 online reviews belonging to 145 products) and April the 12th 2017 (64,803 online reviews belonging to 142 products). Instead of taking the total number of helpful votes at a point of time as our dependent variable, as previous papers usually do, this research takes the change of the variable in an approximately six-month period. We chose this period because we observed that in a shorter time frame, online reviews received a very low number of helpful votes. This means that we would have a much higher percentage of zeros in our dependent variable so exploring it would have more limitations. Then, our empirical model focused on the analysis of those online reviews available at the online retailer on both days (58,770 online reviews belonging to 110 products). Of all the cosmetics categories, **"blush" was selected because it has low sales seasonality, as opposed to other categories such as perfumes or sun cream, avoiding, in this way, seasonality effect over the period.** Blush products are uniformly bought over the year, as they are not usually subject of special events' gifts and those women who use blush products, use it in a regular way. In contrast, as revealed by Nielsen (2016), the fragrances category is particularly seasonal. For example, the eight weeks leading Christmas accounts for around 35% of annual fragrances sales. Other events, such as Valentine's Day and Mother's Day, are also key dates **accounting for fragrances' sales growth. To check the consumer interest over the year in the "blush" category, we used the information provided by the Google Trends tool, as we did not have information about monthly sales of the different cosmetics categories.** Figure 4 shows the interest over the year of the **search terms "blush" and "perfume" in the US. In line with Nielsen (2016), the search term "perfume" is especially relevant in November and December, representing the weeks leading up Christmas. In contrast, the interest in the search term "blush" is more uniformly distributed over time. Therefore, fewer external factors might be affecting the**

voting behaviour of those consumers voting online reviews of blush products. The selection of this product category allows us to reduce the omitted variable bias.

3.2. Variables

The dependent variable in our empirical model is *ReviewHelpfulness*, measured by the incremental number of yes helpful votes of online reviews between October the 4th 2016 and April the 12th 2017. Our main objective is to explore the influence of review visibility on review helpfulness, and this visibility, which is based on the review rank order when sorting, varies over time. If we had taken the cumulative number of yes helpful votes at one specific date as the dependent variable, we would not have been able to know the corresponding rank order of reviews at the time consumers voted the review as helpful. Taking the change of yes helpful votes as our dependent variable allows us to better

Figure 4. Interest over a year period (October the 4th 2016-October the 4th 2017) of the search terms “blush” and “perfume” in the US (Google Trends, 2018)

The numbers reflect the search interest in relation to the maximum value of a graph in a given region and period. A value of 100 indicates the maximum popularity of a term, while 50 and 0 indicate that a term is half popular in relation to the maximum value of the term. Therefore, we cannot compare the relative interest of one term over another, but only the seasonality of the term.

control the effect of review visibility on those helpful votes received by online reviews over that period, as we know the rank order of online reviews at the beginning and at the end of it.

Table 1 shows the description and instrumentation of every variable introduced in the model.

Table 1

Description and instrumentation of research variables

Variable	Description
<i>Dependent variable</i>	
ReviewHelpfulness	Incremental number of yes helpful votes between October the 4 th 2016 and April the 12 th 2017
<i>Review visibility variables</i>	
MHelpfulV	Review visibility probability when sorting by “most helpful” on October the 4th 2016
MRecentV	Review visibility probability when sorting by “most recent” on October the 4th 2016
Review characteristics: Non-textual variables	
1Star	A dummy variable that takes the value 1 if the review is 1-star and 0 otherwise.
5Star	A dummy variable that takes the value 1 if the review is 5-star and 0 otherwise.
Summary	A dummy variable that takes the value 1 if the review has summary words and 0 otherwise.
ReviewerReviews	Number of reviews written by the reviewer at the online retailer website
ReviewerExpenditure	Category of expenditure of the reviewer depending on the annual expenditure at the retailer (4 categories: not specified, low, medium, high)
PhysicalInfo	Number of physical attributes revealed by the reviewer. Attributes are age, skin type, skin tone and eye colour
Length	Number of words of the online review
<i>Review characteristics: Textual variables</i>	
Confidence	A high number suggests that the author is speaking from the perspective of high expertise and is confident; low numbers suggest a more tentative and humble style
Subjectivity	Higher numbers are associated with a more honest, personal, and disclosing text; lower numbers suggest a more guarded, distanced form of discourse
ArgumentStructure	A high number reflects formal, logical, and hierarchical thinking; lower numbers reflect more informal, personal, here and now, and narrative thinking
<i>Control variables</i>	
PBestV	Product visibility probability when sorting by “bestselling” on October the 4th 2016

PNewV	Product visibility probability when sorting by “new” on October the 4 th 2016
PTopRatedV	Product visibility probability when sorting by “top rated” on October the 4 th 2016
PPriceV	Product visibility probability when sorting by “lowest price” on October the 4 th 2016
TopBrand	A dummy variable that takes the value 1 if the brand is one of USA top-10 bestselling brands in 2016 and 0 otherwise (Euromonitor International, 2017).
BrandInt	Search interest in Google of the brand on October the 4 th 2016
BrandIntChange	Search interest in Google of the brand on April the 12 th 2016- Search interest in Google of the brand on October the 4 th 2016

Review Visibility Variables

Review visibility variables, *MHelpfulV* and *MRecentV*, were built based on the rank order of reviews at the beginning of the period (October the 4th 2016) when using two sorting tools available on the online retailer, sorting by most helpful and sort by most recent³.

By default, online reviews were shown by the most recent criterion. These variables were built following the process explained below and illustrated in Figure 5 Product visibility variables, included as controls in the model, were built following the same process. Review visibility variables are used in this research to approach the accessibility dimension of the accessibility-diagnostics theory (Feldman and Lynch, 1988) and they are built based on the study of Godes and Silva (2012), which has also been used by other scholars such as Zhou and Guo (2017) and Hu and Li (2011).

The total number of reviews for a product (N_{pr}) was ordered on both collection dates according to each sorting criterion. Then, review order variables (one for each sorting criterion) were created. In those variables, value 1 represented the first review in the rank and N_{pr} represented the last review in the rank, respectively. To build the order variables, every product and every review available on the website on each of the two dates was

³ The cosmetics website has also the option of sorting online reviews by lowest and by highest rating. However, we decided not to include these two mechanisms in our study since the order of online reviews when sorting by highest rating is highly correlated to the order by most recent (Pearson correlation of 71%). Around 70% of online reviews are 5-star reviews on the website and, when several reviews share the same rating, the second ordination criterion is the most recent date. Therefore, the moderating effect of the visibility by highest rating might be quite similar to the one by most recent. The correlation between review visibility by lowest rating and by most recent review is not so high (31%), since a very few proportion of reviews on the website are 1-star reviews. However, to account for the effect of extreme review ratings we included two dummies in the model: if the review is 1-star and if the review is 5-stars. Future research could build a model introducing the moderating effect of visibility by lowest rating.

considered. When sorting reviews by “most helpful” and by “most recent date”, several online reviews shared the same number of helpful votes or the same date. In those cases, it was adopted the approach that Godes and Silva (2012) followed to capture the order of those online reviews sharing the same publication date. The following formula was applied to build the order by “most recent date” review. Let’s d' represent the publication date of review r . For each d' , it was formed $S_{d'} \equiv \{r: d_r = d'\}$, which represents the set of products for which $d_r = d'$. Then, the variable order was operationalized as $(d') \equiv \sum_{d < d'} N(S_d) + 1$, where $N(S_d)$ is the cardinality of set S_d . This method assigns the same order to every review with the same publication date. For the rest of reviews, the order is always 1 plus the number of reviews with most recent publication dates (Godes and Silva 2012). The same process was followed to order reviews when sorting by “most helpful”, since there are reviews with the same number of yes helpful votes. In this case, we considered that the most recent publication date was the second ordering criteria at the website when two reviews have the same number of helpful votes. The process was as follows:

Figure 5. Flow chart followed to build review visibility variables

1. Inverse of rank orders were computed to make the model interpretation smoother. If not inverted, a value of 1 is better than a value of 2, 2 is better than 3, and so on. When dealing with ranks in literature, many papers use the inverse of the rank instead of the rank itself (Cui, Lui and Guo, 2012). A visibility probability value was assigned to each review depending on its rank position in the sequence of reviews according to each sorting criterion. Then, each review within the review set for a product was assigned the visibility probability of $1/(\text{review order})$. For example, for a product with 100 reviews, if we sort reviews by “most helpful”, the 1st review in the sequence had a visibility probability of $(1/1) = 1$, and the last review in that sequence had a visibility probability of $(1/100) = 0.01$.

2. By default, each page of reviews shows a total of 5 reviews. Therefore, a review page order variable was built to control this effect. As revealed by several studies, consumers usually read between 5 and 10 reviews before making a decision (Liu and Karahanna, 2015; BrightLocal, 2020). Therefore, reviews in first pages are also more accessible and might have a greater influence on consumers. Wu (2017) also incorporates the notion of the review page in his study and online reviews in top 3 pages were more likely to be read and voted as helpful. In our research, if a review, for example, was shown in page 1 when **ordering by “most helpful”**, it was assigned a **page order value of 1** for that sorting criterion. If it was shown in page 2, it adopted value 2 and so on.

3. As in the case of the individual review visibility probabilities, the inverse of the page rank order was computed, assuming that reviews in the first page were more likely to be viewed. This page visibility probability was also obtained from the formula $1/(\text{page order})$.

4. A final review visibility variable for each sorting criterion was built as an average between the individual review visibility probability and the page visibility probability. In this way, there were not reviews in our database with visibility probability of zero. Even if they were placed as the last item in the rank list, the page visibility probability gave them a probability of been viewed. Because consumers could reach reviews without using the studied sorting aids, there would not be realistic to assign visibility probabilities of zero.

Review Characteristics: Non-textual Variables

Review non-textual characteristics are part of the information provided in online reviews and the value of the variables was directly collected from the online retailer. The variables *1Star*, *5Star*, *Summary*, *ReviewerReviews*, *ReviewerExpenditure* and *PhysicalInfo* are incorporated in this research as main characteristics of online reviews. Table 1 describes each variable. *1Star* and *5Star* dummy variables are used to capture the effect of extreme ratings on review helpfulness. *Summary* was introduced to capture the influence of writing summary words about the product in the review (e.g., luminous, good-quality, value for money). *ReviewerExpenditure* is a categorical variable that is based on the annual expenditure of the reviewer at the website. There are 3 different labels provided by the retailer: “low expenditure” if reviewer’s total expenditure is less than \$300 per year; “medium expenditure” if it is between \$300 and \$1000 per year; “high expenditure” if it is more than \$1000 per year. Those who are not registered at the online

retailer are referred to in this research as “not specified expenditure” reviewers. The variable *PhysicalInfo* is built considering the number of personal and physical aspects revealed by the reviewer when writing the online review. Reviewers can provide information regarding age, eye color, skin tone and skin type. When reviewers do not provide information about any of the aspects, the variable takes value of 0, when all the aspects are revealed, the variable takes value of 4.

Review Characteristics: Textual Variables

To build the textual variables, the text mining analysis software Linguistic Inquiry and Word Count (LIWC), developed by Chung and Pennebaker (2007) was used. The scores provided are based on the percentage of words belonging to different categories over the total number of words. Because it is a word counting system, there could entail a limitation of measure. However, the validity of LIWC has been proved in more than 100 studies carried out for different kinds of texts, such as online blogs (M. a. Cohn et al., 2004) and instant messaging (Ireland et al., 2011a; Ludwig et al., 2013a). Therefore, even though is a word counting system, the results have been proved to be highly precise. The complete set of online reviews was analysed using the LIWC software, which provided us scores for more than a hundred of content variables, based on the text of the reviews. For this research, we incorporated three summary variables provided directly by LIWC: *Clout* (Kacewicz et al., 2014), *Authentic* (Newman et al., 2003) and *Analytic* (Pennebaker et al., 2014). For a better comprehension of the nomenclature, we have used different names in this research, *Confidence*, *Subjectivity* and *ArgumentStructure*, respectively. However, we took the definition and interpretation of the variables directly from the original studies (Kacewicz et al., 2014; Newman et al., 2003; Pennebaker et al., 2015), which are shown in Table 1. The three summary variables provided by LIWC are built as a factor or other content variables. For example, the variable *Analytic* or *ArgumentStructure* reflects how well structured is the text and it was built by Pennebaker *et al.* (2014) as a factor of other 8 content variables provided by the LIWC: article, preposition, personal pronoun, impersonal pronoun, auxiliary verb, conjunction, adverb and negation. Appendix C shows examples of online reviews having high vs. low values in the LIWC output for each variable of interest. Moreover, we incorporated the variable *Length*, which reflects the number of words of the online review.

Control Variables

Brand and product related variables are introduced as controls in the model. By using these control variables, we try to capture the likely effect of market factors, both product

and brand related, on review helpfulness. *TopBrand* was introduced to control for the market share of each brand in the US market in 2016 and was built using the information **provided by the report “Colour Cosmetics in the US”** (Euromonitor International, 2017). Appendix B shows the list of brands and market shares. *BrandInt* and *BrandIntChange* were introduced to capture both the consumer interest on each brand at the beginning of the six-month period and the change in that interest over the period. These two variables were built using the Google Trends tool, which provides the consumer interest of a different search term over a specific period. As far as product information is concerned, product visibility variables are introduced to control the influence that the rank order of products at the website might have on review helpfulness. The variable *PBestV* is seen as a measure of product popularity since it captures the inverse of the rank order of products when sorting by bestselling. The variable *PTopRatedV* captures the inverse of rank order of products when they are sorted by best average rating and it could be a measure of both product quality and popularity. In fact, some scholars claim that when prospective consumers know that many other consumers have already bought the product and/or other consumers have highly rated the product, they might be more likely to select it as a promising alternative and they might have a better attitude towards that product (Filieri & McLeay, 2014; Pang & Qiu, 2016). The variable *PNewV* captures the inverse rank order of products when sorted by new. The online retailer provides this label to those products that are either new or have a new format available. Amongst all new products, the second ordering criterion used by the online retailer to show the list of products is the bestselling order. Park et al. (2012) suggest that an online retailer **incorporating new products might influence consumers’ attitudes towards those** products and might lead to a greater impulse buying consumer behaviour and therefore,. Cui, Lui and Guo (2012) found that online reviews affect the sales of new products. **Products labelled as “new” in the online retailer might capture more consumers’** attention, as they are more visible when consumers sort by the “new” aid. Finally, the variable *PPriceV* is the inverse of the rank order of products when sorting by lowest price. Instead of taking the price itself, we took the order by price to be in line with our research, which focuses on the role of visibility. Price has been considered as the top attraction for online shoppers as it is present in all shopping situations (Park et al., 2012). Some scholars, as Weathers et al. (2015), found that online reviews of products with higher prices are less likely to be voted as helpful. However, empirical knowledge about the role of price on review helpfulness is very scarce.

4. Data Analysis

4.1. Descriptive Statistics

Table 2 and Table 3 show the main descriptive statistics and the correlation matrix of the variables.

Table 2

Descriptive statistics

Variable	n	Mean	Sd	Min	Max
ReviewHelpfulness	58770	0,02	0,23	0,00	22,00
MHelpfulV	58770	0,11	0,07	0,10	1,00
MRecentV	58770	0,11	0,07	0,10	1,00
1Star	58770	0,02	0,15	0,00	1,00
5Star	58770	0,69	0,46	0,00	1,00
Summary	58770	0,68	0,47	0,00	1,00
ReviewerReviews	58770	23,05	161,99	0,00	5425,00
PhysicalInfo	58770	2,02	1,39	0,00	4,00
Length	58770	51,84	43,02	1,00	1032,00
Confidence	58770	29,94	24,50	1,00	99,00
Subjectivity	58770	51,09	33,11	1,00	99,00
ArgumentStructure	58770	45,19	28,35	1,00	99,00
PBestV	58770	0,51	0,12	0,17	0,63
PNewV	58770	0,53	0,01	0,53	0,63
PTopRatedV	58770	0,48	0,11	0,17	1,00
PPriceV	58770	0,36	0,14	0,17	1,00
TopBrand	58770	0,09	0,28	0,00	1,00
BrandInt	58770	37,72	13,00	0,00	58,00
BrandIntChange	58770	-0,87	11,50	-27,00	64,00

Table 3

Correlations among research variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
(1) ReviewHelpfulness	1,00																		
(2) MHelpfulV	0,29	1,00																	
(3) MRecentV	0,24	0,23	1,00																
(4) 1Star	0,01	0,03	0,02	1,00															
(5) 5Star	0,01	-0,01	-0,03	-0,23	1,00														
(6) Summary	0,06	0,06	-0,10	-0,11	0,13	1,00													
(7) ReviewerReviews	0,00	0,10	-0,01	0,00	0,00	0,00	1,00												
(8) PhysicalInfo	0,10	0,09	-0,06	-0,01	0,02	0,51	0,02	1,00											
(9) Length	0,10	0,19	0,02	0,01	-0,06	0,06	0,14	0,13	1,00										
(10) Confidence	-0,02	-0,01	-0,02	-0,09	0,20	0,03	-0,02	-0,05	-0,19	1,00									
(11) Subjectivity	0,00	-0,01	-0,01	0,05	-0,08	-0,01	0,00	0,03	0,14	-0,49	1,00								
(12) ArgumentStructure	0,01	0,02	0,01	-0,04	0,01	0,02	0,01	0,01	-0,02	0,23	-0,11	1,00							
(13) PBestV	-0,07	-0,18	-0,18	-0,07	0,11	0,03	-0,02	-0,02	-0,08	0,05	-0,02	0,01	1,00						
(14) PNewV	0,03	-0,01	0,00	0,02	-0,05	-0,11	-0,01	-0,06	0,10	-0,02	0,00	0,01	0,10	1,00					
(15) PTopRatedV	-0,03	-0,11	-0,11	-0,11	0,16	0,06	0,00	0,03	-0,09	0,05	-0,05	0,00	0,49	0,01	1,00				
(16) PPriceV	0,03	0,02	0,03	0,01	-0,03	-0,03	0,00	0,00	0,00	-0,03	0,01	0,00	-0,22	0,03	0,04	1,00			
(17) TopBrand	-0,01	0,01	0,00	-0,01	-0,01	-0,02	0,00	0,02	-0,02	-0,01	-0,01	0,01	0,02	0,05	-0,10	0,17	1,00		
(18) BrandInt	-0,05	-0,11	-0,11	-0,03	0,09	0,10	0,00	0,03	-0,06	0,02	0,00	0,00	0,42	0,13	0,33	-0,42	-0,18	1,00	
(19) BrandIntChange	0,05	0,04	0,05	-0,01	0,01	-0,05	0,00	0,03	0,04	0,01	-0,01	0,01	0,15	0,18	0,19	-0,13	-0,13	0,03	1,00

A total of 69 percent of the 58,770 online reviews were 5-star reviews, and only a 2 percent were 1-star reviews. This means that the average review average rating is high, in line with data analysed in previous studies (Baek et al., 2012; Chevalier & Mayzlin, 2006). Among the top-10 bestselling face make-up brands in the US in 2016, our online retailer only sells 5 brands in the blush category. Only 9 percent of the total reviews are online reviews of products belonging to those 5 brands. The rest 91 percent belongs to the remaining 36 brands available on the online retailer. A total of 6.6 percent of the total reviews are reviews written by reviewers labelled as “high expenditure”, 7.5 percent are reviews by “medium expenditure” reviewers and 60.6 percent are reviews written by “low expenditure” reviewers. The remaining 25.3 percent of reviews are written by reviewers who do not specify their level of expenditure.

4.2. Model Estimation

To test the hypotheses, a zero-inflated negative binomial regression model (ZINB) was built. The selection of this model is due to several reasons. First, our dependent variable is a count variable and has over dispersion (Mean=0.02, Sd=0.21). Besides, most online reviews in our dataset did not increment the number of helpful votes over the period analysed, which means that the dependent variable had a high percentage of zeros. A negative binomial model fits better than a Poisson because the latter assumes equality between the mean and the variance, and it was not the case in our data. Moreover, in line with the proposed “Consumer Voting Journey”, we assume that the zero outcome in our data can be due to two different processes: “not view the review” vs. “view the review and not give it a helpful vote”. If a consumer does not view the review, he cannot vote it as helpful, and the only possible outcome is zero, referred to in the model as false zero. If a consumer views the review, he can also make the decision of not giving it a helpful vote, referred to as true zero, or he can vote it as helpful. Thus, using a ZINB model allows us to explore the two processes that might be involved in the consumer voting behaviour. One process models the excess of zeros due to non-visibility of reviews (1a), and another for the count number of helpful votes once the review is viewed (1b):

$$y_{jk} \sim 0, \quad \text{with probability } q_{jk} \quad (1a)$$

$$y_{jk} \sim g(\lambda), \quad \text{with probability } 1 - q_{jk}$$

$$\lambda = 0, 1, 2, 3, 4 \dots \quad (1b)$$

Equation 1a accounts for the excess of zeros, and it is modelled as a logistic regression, which includes a set of regressor variables (the z 's). In the logistic regression, the response modelled is the probability of y_{jk} to be a false 0. The logistic regression

coefficients give the change in log that the response is a false zero (not viewing the review) for a unit change in the regressor variable, holding all other regressors constant. Equation 1b accounts for the number of helpful votes, and it is modelled as a negative binomial regression. The dependent variable is the log of the conditional mean, meaning that the negative binomial regression coefficients (x' s) give the change in the log mean of helpful votes for a unit increase in the regressor variable, holding all other regressors constant. In our analysis, the x' s and z' s are the same because we consider that the same variables can influence both equations. The logistic part of the ZINB model allows us to explore the factors that make a review not to be viewed (to be a false zero), whereas the negative binomial part allows us to explore the factors that make a review to receive a count number of helpful votes. Coefficients from the ZINB model can be interpreted as semi-elasticities (Cameron & Trivedi, 2009; Winkelmann, 2008), so they measure the proportional change in y as the regressors change. The variables *ReviewerReviews* and *Length* have been transformed to logarithms to comply with the normality requirement, as has been done traditionally in previous review helpfulness literature (Gao et al., 2017; Racherla & Friske, 2012). Besides, all the variables have been standardized in the empirical model to reduce the multicollinearity that might arise in a model with interaction terms (Aiken & West, 1991). ZINB models have previously been used in studies dealing with count data, especially in medical papers (Lewsey & Thomson, 2004; Yau et al., 2003). However, few scholars in the review helpfulness literature have used ZINB models (up to our knowledge only Yin et al., 2014 and Bakhshi et al., 2015). This might be because, up to now, little attention has been paid to the two different processes generating the zero outcome in online reviews context (false and true zeros).

A concern that might arise from our research is the presence of endogeneity, especially due to omitted variable bias, such as the expenditure of the different brands on advertising. This bias could come from any market-related variable that we are not considering in our empirical model. To try to minimize it, **we selected the “blush”** category because the interest on the category is quite constant over the year, which means that blush products are likely to be less susceptible to external market effects. Besides, we have introduced several product and brand variables into the model as controls. To explore if our empirical model is biased by endogeneity, we ran two instrumental variables (IV) regressions⁴, a Two Stage Least Square (2SLS) and an IV Poisson, and we

⁴ We ran two instrumental variable regressions, which allowed us to identify potential endogenous variables and to correct the potential endogeneity by using instrumental variables. We identified review visibility as the potential endogenous variable, since there might be factors influencing visibility not considered in our empirical model. We used the number of product reviews as instrument for visibility. The reason is that the number of reviews of a product influences the visibility of each individual review but not (at least directly) the helpful votes received by it. A Two Stage Least Squares (2SLS) regression and an instrumental variable (IV) Poisson regression were

concluded that results are very similar when correcting or not for endogeneity. Because of that, and because it is the model that better fits our theoretical and conceptual framework, the ZINB model was finally adopted.

To explore the moderation effect of each review visibility variable, five different specifications were estimated and compared (Models 1 to 5, Table 3). To compare between models, three widely used goodness of fit indices were computed: the Akaike's information criterion (AIC), the Schwarz Bayesian information criteria (BIC) and the Log-likelihood ratio test.

5. Results

Table 4 shows the output of the ZINB models. Model 1 is a model where all reviews are assumed to have the same probability of being viewed, which means that review visibility (based on sorting mechanisms) is not considered. This represents the model traditionally used in previous review helpfulness literature. In Models 2 and 3 we assume that consumers sort online reviews by the most helpful criterion, in the way that most helpful online reviews are more visible. In Models 4 and 5, we assume that consumers sort online reviews by the most recent criterion, so it incorporates the effect of review visibility when sorting my most recent reviews, which is also the default order in which reviews are shown displayed at the retailer. Models 2 and 4 show main effects and we add interaction effects of review visibility in Models 3 and 5, which are the ones we use to discuss the results. Although not shown in the study, we have also estimated the model using average visibility variables between both collection dates instead of the initial review visibility as robustness check. The objective was to verify that the way of approaching the visibility variables did not produce a difference in the results. Results in both cases are almost the same, so a model that incorporates the initial review visibility variables was finally chosen.

estimated. As proposed by Wooldridge (2011), the IV Poisson is the best option to use when endogeneity is present and we are dealing when a count dependent variable, even if we have excess of zeros. For the IV Poisson, following Cameron and Trivedi (2013), a Control Function (CF) estimation was applied. We compared the results of OLS vs. 2SLS; and of standard Poisson vs. Poisson CF, and we found very similar results in terms of coefficient significance and sign. As far as we know, endogeneity correction has not been developed for the ZINB model so, since the results of 2SLS and instrumental Poisson were not much different from the models without endogeneity correction, we decided to use the ZINB model in our empirical section, because it is the one that better fits our theoretical and conceptual framework.

Table 4
ZINB regression output

	Model 1	Model 2	Model 3	Model 4	Model 5
	No visibility considered	Most helpful visibility Without interaction	Most recent visibility With interaction	Most recent visibility Without interaction	Most recent visibility With interaction
a) Logit part					
Visibility (MHelpfulV)		-20.66***	-6.73***		
Visibility (MRecentV)				-0.17***	-25.41***
1Star	1.44***	2.55***	-10.99	-0.02	0.62***
5Star	-0.39***	-0.35**	-0.68***	0.05	0.21
Summary	-0.04	-9.79***	-1.18**	-3.19***	-3.66***
<i>ReviewerReviews</i>	-0.07	0.10	0.34**	0.08	-0.56***
PhysicalInfo	-0.72***	-0.94***	-1.55***	-1.48***	-1.17***
ReviewerExpenditure: Low	0.25	0.47	-1.39**	2.93***	0.96*
ReviewerExpenditure: Medium	0.11	0.20	-1.19*	1.80**	0.93
ReviewerExpenditure: High	-0.68	-21.43***	-1.49**	0.93	0.40
<i>Length</i>	-0.24***	-0.16	-0.46**	-0.24**	-0.09
Confidence	0.21*	-0.05	0.69**	0.12	0.10
Subjectivity	0.003	-0.12	1.08***	0.17	-0.04
ArgumentStructure	-0.20**	-0.30**	-1.04***	-0.18	-0.14
1Star * Visibility			3.92		5.42***
5Star * Visibility			-3.42***		0.05
Summary * Visibility			-5.87***		-22.37***
<i>ReviewerReviews</i> * Visibility			2.71***		-2.44**
PhysicalInfo * Visibility			-6.72***		-1.22
ReviewerExpenditure: Low * Visibility			-10.33***		-0.12
ReviewerExpenditure: Medium * Visibility			-5.97**		2.96
ReviewerExpenditure: High * Visibility			-4.82**		2.39
<i>Length</i> * Visibility			-2.04**		0.37
Confidence* Visibility			4.96***		-0.66
Subjectivity* Visibility			7.90***		0.24
ArgumentStructure* Visibility			-5.08***		0.34
PBestV	0.14	0.31*	0.49***	0.56***	0.52***
PNewV	-0.16**	-0.10	-0.12	-0.22***	-0.31***
PTopRatedV	-0.56***	-0.55***	-0.66***	-0.06	-0.45***
PPriceV	-19.13***	-12.71***	-8.90***	-0.08	-0.49***
TopBrand	0.20*	0.35**	0.36***	-0.36***	-0.10
BrandInt	-0.08	-0.26**	-0.15	0.70***	-0.10
BrandIntChange	0.36***	0.25**	0.20**	-0.16*	0.24***

Constant	-1.16*	4.22***	-1.45	1.82**	-1.08*
b) Negative binomial part					
Visibility (MHelpfulV)		0.14***	0.22***		
Visibility (MRecentV)				0.26***	2.07***
1Star	-0.15***	-0.10***	-0.12***	0.03	-0.01
5Star	0.55***	0.33***	0.35***	0.24***	0.32***
Summary	1.35***	0.86***	1.52***	-1.04***	0.33
ReviewerReviews	-0.07*	-0.07*	-0.12**	0.01	-0.11**
PhysicalInfo	0.28***	0.28***	0.30***	0.03	0.04
ReviewerExpenditure: Low	0.12	0.27	0.02	0.88***	0.60**
ReviewerExpenditure: Medium	0.50**	0.66***	0.45*	0.95***	0.67**
ReviewerExpenditure: High	0.28	0.09	0.37	0.99***	0.70***
Length	0.48***	0.30***	0.28***	0.45***	0.49***
Confidence	0.01	-0.12**	-0.09	-0.02	0.04
Subjectivity	-0.14***	-0.16***	-0.10	-0.09	-0.16**
ArgumentStructure	0.01	0.003	-0.03	0.01	0.02
1Star * Visibility			-0.002		-0.02
5Star * Visibility			0.002		-0.02
Summary * Visibility			-0.14***		-2.56***
ReviewerReviews * Visibility			0.01		0.01
PhysicalInfo * Visibility			-0.02		0.02
ReviewerExpenditure: Low * Visibility			0.03		-0.14***
ReviewerExpenditure: Medium * Visibility			0.03		-0.13**
ReviewerExpenditure: High * Visibility			-0.01		-0.16***
Length * Visibility			-0.01		0.01
Confidence* Visibility			0.02		0.0004
Subjectivity* Visibility			0.005		0.01
ArgumentStructure* Visibility			-0.01		-0.02
PBestV	0.09*	0.03	0.11**	-0.02	0.19***
PNewV	0.15***	0.18***	0.16***	0.13***	0.07**
PTopRatedV	-0.004	0.03	0.02	0.08**	0.04
PPriceV	0.32***	0.22***	0.25***	0.21***	0.16***
TopBrand	-0.10**	-0.08**	-0.07*	-0.19***	-0.15***
BrandInt	-0.14***	-0.14***	-0.11***	0.04	-0.11***
BrandIntChange	0.18***	0.14***	0.14***	0.03	0.12***
Constant	-3.79***	-3.82***	-3.97***	-2.90***	-3.01***
Observations	58,770	58,770	58,770	58,770	58,770
Log Likelihood	-4,028.54	-3,914.67	-3,855.08	-4,058.33	-3,709.06
AIC	8,139.09	7,915.33	7,844.15	8,202.67	7,552.12
BIC	8,507.33	8,301.53	8,445.90	8,588.87	8,153.87

Vuong	11.98***	12.45***	12.99***	5.37***	11.87***
-------	----------	----------	----------	---------	----------

Note: Significance levels are denoted in asterisk: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Variables in *italics* have been transformed to logarithms. All variables in the model have been standardized

First, it can be observed that Model 1, which does not consider review visibility, is one of the models with the worse goodness of fit measures, together with Model 4. It has the highest values of AIC and BIC and relatively low value for log-likelihood (Cameron & Trivedi, 2009). Therefore, this might suggest that incorporating review visibility variables entails an improvement over only using review characteristics as determinants of review helpfulness. When comparing models without interaction terms to models with interaction terms, Model 2 vs. Model 3 and Model 4 vs. Model 5, it can be observed a better fit of models including interactions, especially comparing Model 4 to. Model 5. Considering only models with interactions, Model 3 and Model 5, we observe that Model 5 has better goodness of fit measures than Model 3. By default, online reviews are displayed by the most recent mechanism at the studied online retailer, so it might be a reason why Model 5 has better fit than Model 3, in which consumers should re-order the initial set of online reviews by the most helpful mechanism. To further analyse the ZINB output, Model 3 and Model 5 are then used. When explaining the findings of the ZINB models, a distinction should be made between the results in the logit and in the negative binomial part. The logit part explains the likelihood that a review is not voted because it has not been viewed by the consumer. The negative binomial part explains the likelihood that the review receives a count number of helpful votes. Since this research focuses on analysing the role of review visibility in explaining the number of helpful votes online reviews receive, we will place particular emphasis in discussing the results of the negative binomial part of Model 3 and Model 5.

Hypotheses *H1a* and *H1b* propose that more visible online reviews when sorting by most helpful and by most recent are more likely to be voted as helpful. Both visibility variables, *MHelpfulV* and *MRecentV*, are significant in the two parts of the ZINB model, so *H1a* and *H1b* are **supported. In the logit part, they are significant and negative** ($\delta = -6.73$, $p < 0.005$; $\delta = -25.41$, $p < 0.05$, respectively). This means that the higher the visibility of online reviews when sorting either by most helpful or by most recent, the less likely the review does not receive helpful votes because it has not been viewed. In the negative binomial part, both *MHelpfulV* and *MRecentV* are also significant but in this case the **coefficient is positive** ($\delta = 0.22$, $p < 0.005$; $\delta = 2.07$, $p < 0.05$), meaning that the higher the visibility of online reviews the greater the likelihood of receiving helpful votes. Compared to the logit part, the two coefficients of review visibility have a lower magnitude in the

negative binomial part. Thus, findings might suggest that review visibility plays an important role in explaining the excess of online reviews with zero helpful votes or, in other words, the likelihood that online reviews do not receive votes because they have not been viewed. However, review visibility has a smaller effect in explaining the count **number of helpful votes that online reviews receive. The proposed “Customer Voting Journey” might be a reason** behind this finding since consumers might be influenced by the rank of online reviews when they are choosing which reviews to read but, once a specific review is read, the characteristics of the review themselves are likely to influence the consumer decision or voting the review as helpful.

Hypotheses *H2a* and *H2b* posit that review visibility when online reviews are ranked by the “most helpful” mechanism (*H2a*) and by the “most recent” mechanism (*H2b*) moderates the effect of review characteristics on review helpfulness. To analyse this effect, the interaction terms between review visibility and review characteristics in Model 3 and Model 5 should be assessed.

As regards to Model 3, where we analyse the interaction effect of *MHelpfulIV*, it can be observed that every interaction term is significant in the logit part, except from *1Star x Visibility*, while only one interaction term, *Summary x Visibility*, is significant in the negative binomial part. In Model 5, we only find three significant interaction terms in the logit part and two significant interactions in the negative binomial part. Thus, *H2a* and *H2b* are partially supported. Overall, it can be observed from the results that the moderating role of review visibility in explaining review helpfulness might depend on three factors: if online reviews are sorted by most helpful (Model 3) or by most recent (Model 5); if we are explaining the excess of zeros (logit part) or the number of helpful votes (negative binomial part); and the specific review characteristic that is being analysed. To further analyse *H2a* and *H2b* a distinction has been made between the logit and the negative binomial part, in which we explore interactions in more detail. Since we have adopted an empirical model that implies two processes, the moderating effect of review visibility might be different when explaining the excess of online reviews with zero helpful votes and the positive number of votes.

Looking at the logit part of Model 3, results suggest that review visibility has a strong moderating role in explaining the excess of online reviews with zero votes when consumers sort reviews by the most helpful mechanism. However, we observe in Model 5 that review visibility has a weaker moderating role in explaining the excess of zeros when consumers sort reviews by most recent.

As regards to the output of the negative binomial part, which allows us to analyse the likelihood that online reviews receive a count number of helpful votes, we observe in Model 3 that the moderating role of review visibility is much weaker than in the logit part and only one interaction term, *Summary x Visibility*, is significant. In Model 5, the moderating role of review visibility in the negative binomial part is also weak but it is similar than in the logit part. Two interaction terms are significant in this part, *Summary x Visibility* and *ReviewerExpenditure x Visibility*. Therefore, it can be observed that the relationship between review visibility and the number of helpful votes that online reviews receive is in general additive and not interactive. In the one hand, review visibility variables, *MHelpfulV* and *MRecentV*, are significant in the negative binomial part of Model 3 and Model 5. On the other hand, many review characteristics variables are also significant in the two models. However, only the interaction terms *Summary x Visibility* in Model 3, and *Summary x Visibility* and *ReviewerExpenditure x Visibility* in Model 5 are significant. Thus, review visibility by itself is a factor that triggers helpful votes, but the moderating role of visibility in explaining the number of helpful votes is weak.

If we look at the effect of those review characteristics that are not moderated by review visibility in the negative binomial part of Model 3 and Model 5, we observe that the sign of significant coefficients is the same in both models. Thus, regardless the assumption we make in terms of the consumer sorting choice, *5Star*, *ReviewerExpenditure: Medium* and *Length* positively influence the likelihood that online reviews receive helpful votes, while *ReviewerReviews* negatively influences the likelihood that online reviews receive helpful votes. However, we also notice that some review characteristics, *1Star*, *Summary* and *PhysicalInfo* have no impact in Model 5, while they have a significant impact in Model 3. On the contrary, *ReviewerExpenditure: Low*, *ReviewerExpenditure: High* and *Subjectivity* are significant in Model 5 but not in Model 3.

Overall, we find different results in the logit and in the negative binomial part of each model. What these findings might suggest is that, when explaining the helpfulness of online reviews, the process of voting a review as helpful might not be a one-step process but rather we should consider that consumers might face different sub-processes when **voting online reviews as helpful. Linking these findings to the proposed “Consumer Voting Journey”, one could think that when consumers are in the “review consideration” stage and they are deciding which reviews to read, they might be influenced by different factors than when they are in the “voting stage”, where they have already read a specific review and have to decide if voting it as helpful or not. As expected, review visibility has a more important role when consumers are deciding which reviews to read than when they have to decide if voting the review as helpful or not.**

Differences are also found between Model 3 and Model 5. A possible explanation is that, although the two sorting mechanisms enhance the review visibility likelihood, they capture different aspects of online reviews: the number of helpful votes and the publication date. When consumers evaluate one set of online reviews or another, they already know that they are either the reviews with the highest number of helpful votes or the reviews most recently published. This information might influence consumers subsequent decisions, such as which reviews to read and then, if voting the reviews as helpful or not. On the one hand, when consumers evaluate reviews in top-ranked positions by the most helpful mechanism, they draw from the fact that many other consumers have found the information contained in those reviews useful in their decision-making. **Thus, due to a probable “wisdom of the crowd” effect, consumers are likely to believe that the information contained in those most helpful reviews is a better approximation to the truth, and thus, more reliable (Liu & Karahanna, 2015).** On the contrary, being visible when sorting by most recent does not provide any additional information for prospective consumers about the usefulness of the review to evaluate the product. In those cases, consumers only know that a review is in top-ranked positions because it has been recently published, but the date by itself might not be diagnostic enough to play a role in the consumer decision of reading a review or not. In fact, since they are new reviews, they have not had time enough to receive helpful votes, so consumers might not know to what extent the information contained in those reviews is useful and reliable to make a good evaluation of the product. Therefore, consumer information processing might be different depending on the type of information they are evaluating. **In our case, consumers are not likely to process “most helpful” information in the same way as “most recent” information.**

5.1. Model diagnostics and robustness check

To ensure significance of our models, a likelihood ratio test was run to compare each of our ZINB models (Models 1-5, in Table 4) to a model where *ReviewHelpfulness* is regressed only on a constant term (Models 1.1-5.1). As shown in Table 4, for the 3 models $p\text{-value} < 2.2e^{-16}$, so the ZINB models are significant.

Table 4

Likelihood ratio test ZINB models

Model	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	41	-4,028.5			
1.1	3	-5,686.8	-38	3,316.5	<2.2e-16***
2	43	-3,914.7			
2.1	3	-5,686.8	-40	3,544.3	<2.2e-16***

3	67	-3,855.1			
3.1	3	-5,686.8	-64	3,663.5	<2.2e-16***
4	43	4,058.3			
4.1	3	5,686.8	-40	3,257	<2.2e-16***
5	67	3,709.1			
5.1	3	5,686.8	-64	3,955.5	<2.2e-16***

Note: Significance. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Furthermore, the Vuong test, which is a likelihood-ratio test that use the Kullback-Leiber information criterion, was used to see if a ZINB model implied an improvement over a standard NB model with the same coefficient but without the zero-inflation correction. The statistic tests the null hypothesis that the two models are equally close to the true data estimating process, against the alternative that one model is closer. The Vuong test statistic in the three models reported in Table 3 favoured the ZINB model over a NB. This means that using the ZINB model is a significant improvement over the standard NB.

To go further in confirming the ZINB significance, several goodness of fit measures are compared: the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the log-likelihood test to other possible count models: Poisson (P), Negative Binomial (NB), Zero-inflated Poisson (ZIP), Zero-altered Poisson (ZAP) and Zero-altered Negative Binomial (ZANB). Table 5 summarises the goodness of fit measures of each model. The ZINB model is the one that better fitted the data, as it had the lowest AIC index, the lowest BIC, and the highest log-likelihood value.

Table 5

Goodness of fit measures and comparisons between count models (NB, ZINB, P, ZIP, ZAP and ZANB)

	NB	ZINB	P	ZIP	ZAP	ZANB
Model 1						
AIC	8,768.4	8,139.1	9,442.5	8,339.4	8,799.5	8,638.9
BIC	8,957	8,507.3	9,622.1	8,698.7	9,158.8	9,007.2
Log-likelihood	-4,363.2	-4,028.5	-4,701.3	-4,129.7	-4,359.7	-4,278.5
Model 2						
AIC	8,427	7,915.3	8,811.5	8,285.2	8,358	8,284.8
BIC	8,624.6	8,301.5	9,000.1	8,662.4	8,735.3	8,671
Log-likelihood	-4,191.5	-3,914.7	-4,384.8	-4,100.6	-4,137	-4,099.4
Model 3						
AIC	8,363.5	7,844.2	8,696.7	7,654.7	8,292.5	8,225.3
BIC	8,668.9	8,445.9	8,993.1	8,247.5	8,885.7	8,827
Log-likelihood	-4,147.8	-3,855.1	-4,315.4	-3,761.4	-4,080.5	-4,045.6
Model 4						
AIC	8,545	8,202.7	9,146.4	8,369.1	8,549.6	8,403.2
BIC	8,742.6	8,588.9	9,335	8,746.4	8,926.8	8,789.4
Log-likelihood	-4,250.5	-4,058.3	-4,552.2	-4,142.6	-4,232.8	-4,158.6
Model 5						
AIC	8,485.2	7,552.1	8,973.3	8,048.5	8,430.9	8,335.6

BIC	8,790.5	8,153.9	9,269.7	8,641.3	9,023.6	8,937.4
Log-likelihood	-4,208.6	-3,709.1	-4,453.6	-3,958.2	-4,149.4	-4,100.8

6. Discussion

When studying the impact of review characteristics on review helpfulness previous literature assumes the characteristic of each review to be equally influential on consumer decision making. However, online reviews are shown in a sequence next to other reviews, and not individually, so the relative visibility of online reviews in the sequence is likely to influence consumer decision-making. In fact, there is evidence to believe that consumers do not evaluate all available information and, instead, they tend to be more influenced by more accessible information (BrightLocal, 2020; Feldman & Lynch, 1988; Q. Liu & Karahanna, 2015). In this line, this paper studies the role of review visibility in explaining the relationship between the characteristics of the review and the number of helpful votes that online reviews receive. We take review visibility as a moderator, sometimes rereferred to in literature as quasi-moderator, suggesting that the variable has a direct impact on review helpfulness but also an interactive effect with review characteristics.

We use a ZINB regression in the empirical analysis drawing from the premise that the excess of online reviews with zero helpful votes might be due to two different processes: first the process of viewing or not a specific online review, and second, the process of voting it as helpful or not. If a consumer does not view a specific online review, the online possible outcome is zero. However, when a consumer views a review, he can make the decision of either not voting it or voting it as helpful. In this research, review visibility is proxied based on the rank order of reviews when we assume consumers sort by two mechanisms: most helpful and most recent.

Overall, findings reveal that the role of review visibility is different depending on the sorting assumption made, if consumers sort online reviews by the most helpful or most recent mechanism. Regardless the sorting assumption made, the moderating role of review visibility is stronger when explaining the excess of zeros, or in other words, the likelihood that online reviews do not receive helpful votes because they have not been viewed. Once online reviews are viewed, the moderating role of review visibility in explaining the number of helpful votes is less relevant. Moreover, findings reveal that review visibility might play a different role depending the type of information consumers are evaluating (most helpful or most recent) so that consumer information processing and decision-making is not likely to be the same when evaluating either most helpful or most recent online reviews. Although both sorting mechanisms influence the likelihood

that online reviews are viewed by consumers, the set of online reviews shown when using one mechanism or another is different and reviews do not have the same characteristics. When evaluating most helpful online reviews, consumers draw from the premise that many other consumers have found the information contained in those reviews as helpful, so consumers usually perceive that information as more credible. When evaluating most recent online reviews, consumers just know that those reviews have been recently published, but they cannot access the reliability of that information, as they do in the case of most helpful reviews. Thus, our findings suggest that the information processing of the two sets of online reviews is different, as well as the triggers of helpful votes.

6.1. Theoretical Contribution

This study contributes to the review helpfulness literature in the following ways.

First of all, the accessibility-diagnosticsity theory (Feldman & Lynch, 1988) has been taken as the basis to study how these two dimensions of online reviews, accessibility and diagnosticsity, might influence review helpfulness, which is a measure of the adoption of the information. On the one hand, review accessibility is approached in this research by review visibility when sorting by two mechanisms: most helpful and most recent. On the other hand, review characteristics determine how diagnostic a review is considered by consumers. We posit that the accessibility of online reviews, approached by review visibility variables, have both a direct impact on review helpfulness and an interactive impact with review characteristics in explaining review helpfulness.

Most previous literature has mainly explored direct relationships between review characteristics and review helpfulness (Ahmad & Laroche, 2015; Cao et al., 2011; Ghose & Ipeiritos, 2010; Zhang & Lin, 2018). This literature implicitly assumes each review as independent one from each other and proposes the characteristics of online reviews as the main drivers of review helpfulness. However, individual online reviews are part of a set of reviews, so consumer voting decisions might not only be influenced by the characteristics of online reviews, which determine how diagnostic are them for consumers, but also by the visibility of reviews, which determines how accessible reviews are for consumers.

A small stream of research has included the notion of review visibility to study review helpfulness. Wu (2017), for instance, uses the review order when sorting by most helpful as an independent variable that directly impacts review helpfulness. Zhou and Guo (2017) explore how some review characteristics moderate the relationship between

review order by the most recent rank and review helpfulness. However, there might be other ordering mechanisms, such as the most helpful mechanism, which might also impact on review helpfulness. Therefore, we not only include the most recent sorting mechanism but also the most helpful one. Our findings suggest that both approaches to review visibility, when we assume that consumers sort online reviews either by “the most helpful” or “the most recent” mechanism, are relevant in explaining review helpfulness. Moreover, rank order mechanisms might have a different effect on explaining reviews with zero helpful votes and reviews with a positive number of helpful votes, which has not been tested in previous studies.

Therefore, this study posits that, to better explore review helpfulness, we should consider that the excess of online reviews with zero helpful votes might be due to two different reasons: (1) not viewing the review and (2) viewing the review and deciding not to vote it as helpful. By applying a Zero Inflated Negative Binomial (ZINB) regression, which models review helpfulness in two parts, we can understand first, the drivers explaining that online reviews do not receive helpful votes because they have not been viewed and, second, the drivers explaining the likelihood that online reviews receive helpful votes. Overall, previous literature has not made distinctions between the two processes behind the zero outcome and have overlooked the fact that we find an excess of online reviews with zero helpful votes (Ahmad & Laroche, 2015; Cao et al., 2011; Ghose & Ipeiritis, 2010).

Our findings suggest that the role of review visibility is different if we are explaining the excess of online reviews with zero helpful votes or if we are explaining the factors leading to helpful votes. This might indicate that, when exploring review helpfulness, it is important to bear in mind that consumers might face different sub-processes before voting a review as helpful.

6.2. Managerial Implications

Academics and practitioners have highlighted the power of online reviews to influence product sales, so many companies are hosting online reviews in their web sites to enable consumers to discuss about their products or services and to increase consumer engagement. However, it is not only important for companies to host online reviews, but also to understand the way consumers process this information.

A recommendation derived from our results is that marketers should provide more and better online tools to allow consumers to organize online reviews according to

consumers' preferences. For example, considering that consumers might see the information contained in most helpful reviews as more reliable, firms could highlight the information provided by those reviews to reduce the cognitive effort of consumers associated to process all the information contained in the reviews. These could be done by incorporating a subsection showing a summary of the characteristics of most helpful reviews, uncovering, for example, the profile of the reviewers who write most helpful reviews and some aspects of the text of those reviews. In fact, after data were collected, the online retailer has introduced some changes in line with our findings. For example, they do not longer show online reviews by the default mechanism of most recent rank, but by the most helpful order.

Our findings might also help companies to know the characteristics of online reviews that make them more likely to be voted as helpful. In this way, firms could identify those helpful reviews to analyse the information contained in those opinions with the objective of solving consumers' claims and of adapting their products to consumers' needs.

6.3. Limitations and Future Research

In this paper, the research was limited to one category of experience products, the **category of “blush”, but future research could be expanded to other more seasonal categories**, such as fragrances, and to the context of search products. The research may be also extended to services, such as hotels or restaurants. In terms of data, aggregate information was used, and proxy variables were built to approach review visibility probabilities. To build the variables, ranks generated by the retailer sorting tools were considered, but filtering aids could also be incorporated to the analysis. Future research could also use click stream data to know the actual ordering used by the consumers on the online retailer. To explore the approximately six-month change of review helpfulness, two points of time were chosen and the change of the dependent variable between two dates was explored. Future research could build a panel of reviews and collect weekly data to better approach the evolution over time and to better correct for endogeneity due to omitted variable bias. Review visibility is approached by the rank order of reviews based on the most helpful and most recent criteria, but future research could explore other sorting criteria, such as the lowest or highest rating orders, and the effect of filtering mechanisms on review visibility. Three content style variables were introduced into the model, as they are summary variables built as a factor of other content variables, but further research might consider the effect of other specific content features. Future research might also examine the effect of review visibility in other outcomes such as product sales. Lastly, this research explored the number of yes helpful votes of online

reviews, but the same research could be replicated taking the number of not helpful votes as the dependent variable. Further research could also compare both results and identify similarities and differences.

Appendix - Chapter 1

A. “Consumer Voting Journey” screen shots

Figure A.1. Product consideration stage screenshot (when sorting products by “bestselling”)

Figure A.2. Review consideration stage screenshot (review visibility when sorting reviews by “most helpful”)

Figure A.3. Voting decision stage screenshot (review characteristics can be evaluated)

Appendix B. Market Shares of Facial Make-up brands in the US: % Value 2013-2016

Table B.1

Brand Shares of Facial Make-up in the US: % Value 2013-2016 (Source: Euromonitor International, 2017)

% retail value rsp Brand (GBO)	Company (NBO)	2013	2014	2015	2016
Cover Girl	Coty Inc	-	-	-	9.4
Maybelline (L'Oréal Groupe)	L'Oréal USA Inc	7.5	7.3	7.2	7.3
Mac	Estée Lauder Cos Inc	7.3	7.4	7.3	6.5
Clinique	Estée Lauder Cos Inc	7.2	6.9	6.7	6.5
L'Oréal Paris (L'Oréal Groupe)	L'Oréal USA Inc	7.0	6.5	6.2	5.8
Lancôme (L'Oréal Groupe)	L'Oréal USA Inc	4.7	4.6	4.5	4.1
Urban Decay (L'Oréal Groupe)	Urban Decay Cosmetics	3.2	3.4	3.6	4.0
Mary Kay	Mary Kay Inc	4.3	4.4	4.2	3.9
Estée Lauder	Estée Lauder Cos Inc	3.9	3.8	4.0	3.7
Revlon	Revlon Inc	4.7	4.2	4.0	3.6
bareMinerals (Shiseido Co Ltd)	Bare Escentuals Inc	3.7	3.8	3.4	3.1
Neutrogena (Johnson & Johnson Inc)	Johnson & Johnson Consumer Products Inc	3.0	3.2	3.0	2.8
Smashbox	Estée Lauder Cos Inc	2.0	2.1	2.3	2.6
Nars (Shiseido Co Ltd)	Shiseido Americas Corp	1.5	1.5	1.7	1.8
Chanel (Chanel SA)	Chanel USA Inc	2.1	2.1	2.0	1.8
BeneFit (LVMH Moët Hennessy Louis Vuitton SA)	LVMH Perfums & Cosmetics USA Inc	1.4	1.6	1.6	1.7
Bobbi Brown	Estée Lauder Cos Inc	2.0	2.1	1.9	1.7
Rodan + Fields	Rodan & Fields LLC	0.1	0.1	0.1	1.4
Too Faced	Estée Lauder Cos Inc	-	-	-	1.4
Almay (Revlon Inc)	Revlon Products Corp	1.8	1.6	1.4	1.2
Laura Mercier	Shiseido Americas Corp	-	-	-	1.0

Appendix C. Examples of online reviews having high vs. low values for the content style variables

Table C.1

Examples of online reviews having high vs. low values for the content style variables and explanations

Variable	High numerical value (99/100)	Low numerical value (1/100)
Confidence Higher values suggest that the author is speaking from the perspective of high expertise and is confident	<p><i>"They are just so pretty. The colours work well on almost everyone. From very, very light to darker complexions. This is a staple and every MUA loves them (almost everyone). It gives your face a pretty glow - you look well rested even when you're not. Get it- you'll love it!"</i></p> <p>Why does it have high value? The reviewer demonstrates to be familiar about the product reviewed. She speaks about the different colours available of the product and she knows how the product fits to different skin types.</p>	<p><i>"I've never purchase "expensive" makeup before so I was skeptical it would be any better than any other makeup. Boy was I wrong. Laguna looks dark but it is an amazing bronze color. No orange in sight. I haven't tried the Orgasm blush yet but I'm sure I won't be disappointed and even if i don't love it on my cheks I'm sure it'll make a great light pink eyeshadow. Love NARS!"</i></p> <p>Why does it have low value? The reviewer admits she had not bought this kind of product before, so her opinion is seen as more "inexperienced" because is based on an unique experience with this type of product.</p>
Subjectivity Higher values are associated with a more honest, personal, and disclosing text	<p><i>"I love the O-Glow, I started using it when it first came out and haven't stopped. I don't wear much makeup but this I wear everyday. It doesn't look like too much makeup it just looks natural. I don't know what I would do without this!"</i></p> <p>Why does it have high value? The reviewer just reveals her feelings and emotions about the product. She says she likes the product but she does not provide more technical information.</p>	<p><i>"This product is super easy to apply. If I want a light sheen, I apply it with a brush and if I want it darker I apply it directly to the face and blend. It's the perfect shade for a romantic looking flush."</i></p> <p>Why does it have low value? The reviewer provides more technical information about the characteristics of the product and the way of using it.</p>
Argument structure Higher values reflect formal, logical, and hierarchical thinking	<p><i>"I was really excited when I purchased this duo. I've never had NARS product, and I heard so many good things about it I just had to buy it. It's very loose and shimmery. I do like the shimmer, it's not a cheap-looking shimmer, but it's not long-lasting, and I really didn't think it was worth it if it wasn't going to last me. It didn't really last me even with a primer! Also, the bronzer, with my skin tone, hardly even showed. I returned it, sadly"</i></p>	<p><i>"Really glad I bought this- didn't know I had a natural blush already until I put it on. However I've just found that Laguna just works so well with me ~Love it!"</i></p>

Why does it have high value?
The reviewer provides a connected **“story” to express the opinion**, using many connected sentences for it. She explains the complete **“buying experience”**. **First, she** admitted she had never bought that product nor that brand. Then, she said that she had good expectations about it so she decided to buy it. However, she explains later why the product did not meet her expectations. Finally, she admitted having returned the product.

Why does it have low value?
The reviewer does not provide the reasons why she likes the product; she just says that it works well on her. Only two sentences are used to express her opinion, with not much hierarchical thinking. She does not connect sentences to express the reason why likes the product. After reading the review, prospective consumers do not know about her previous experience with the product nor her expectations

References

- Aaker, D. A. (1991). Managing Brand Equity. *Managing Brand Equity*.
- Agnihotri, A., & Bhattacharya, S. (2016). Online Review Helpfulness: Role of Qualitative Factors. *Psychology & Marketing*, 33(11), 1006–1017.
- Ahani, A., Nilashi, M., Yadegaridehkordi, E., Sanzogni, L., Tarik, A. R., Knox, K., Samad, S., & Ibrahim, O. (2019). Revealing customers' satisfaction and preferences through online review analysis: The case of Canary Islands hotels. *Journal of Retailing and Consumer Services*, 51(July), 331–343.
- Ahmad, S. N., & Laroche, M. (2015). How do expressed emotions affect the helpfulness of a product review? Evidence from reviews using latent semantic analysis. *International Journal of Electronic Commerce*, 20(1), 76–111.
- Aiken, L. S., & West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*.
- Al-Obeidat, F., Spencer, B., & Kafeza, E. (2018). The Opinion Management Framework: Identifying and addressing customer concerns extracted from online product reviews. *Electronic Commerce Research and Applications*, 27, 52–64.
- Al., H. et. (2010). *Multivariate Analysis 7 th Edition*.
- Aljukhadar, M., Senecal, S., & Daoust, C.-E. (2012). Using Recommendation Agents to Cope with Information Overload. *International Journal of Electronic Commerce*, 17(2), 41–70.
- Ansari, A., & Mela, C. F. (2003). *E-Customization*. 40(2), 131–145.
- Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*, 57(8), 1485–1509.
- Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2), 277.
- Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1), 29–51.
- Areni, C. S. (2003). The Effects of Structural and Grammatical Variables on Persuasion: An Elaboration Likelihood Model Perspective. *Psychology and Marketing*, 20(4), 349–375.
- Arun, R., Suresh, V., Madhavan, C. E. V., & M.N, N. M. (2010). On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In *Advances in Knowledge Discovery and Data Mining* (pp. 391–403).

- Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, 0, 2200–2204.
- Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of Online Consumer Reviews: Readers' Objectives and Review Cues. *International Journal of Electronic Commerce*, 17(May 2015), 99–126.
- Baker, W. E. (2001). The Diagnosticity of Advertising Generated Brand Attitudes in Brand Choice Contexts. *Journal of Consumer Psychology*, 11(2), 129–139.
- Baker, W. E., Lutz, R. J., Journal, S., Spring, N., Taylor, P., Baker, W. E., & Lutz, R. J. (2000). Effectiveness An Empirical Test of an Updated Relevance- Accessibility Model of Advertising Effectiveness. *Journal of Advertising*, 29(1), 1–14.
- Bakhshi, S., Kanuparth, P., & Shamma, D. A. (2015). Understanding Online Reviews: Funny, Cool or Useful ? *CSCW 2015: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1270–1276.
- Balducci, B., & Marinova, D. (2018a). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, 46(4), 557–590.
- Balducci, B., & Marinova, D. (2018b). Unstructured data in marketing. In *Journal of the Academy of Marketing Science* (Vol. 46, Issue 4, pp. 557–590). Springer New York LLC.
- Beach, L. R. (1993). Broadening the definition of decision making: The role of prechoice screening option. *Psychological Science*, 4(4), 215–220.
- Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding Satisfied and Dissatisfied Hotel Customers: Text Mining of Online Hotel Reviews. *Journal of Hospitality Marketing and Management*, 25(1), 1–24.
- Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the Tribes: Using Text for Marketing Insight. *Journal of Marketing*, 84(1), 1–25.
- Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive Consumer Choice Processes. *Journal of Consumer Research*, 25(3), 187–217.
- Biehal, G., & Chakravarti, D. (1983). Information Accessibility as a Moderator of Consumer Choice. *Journal of Consumer Research*, 10(1), 1.
- Bilgicer, T., Jedidi, K., Lehmann, D. R., & Neslin, S. A. (2015). Social Contagion and Customer Adoption of New Sales Channels. *Journal of Retailing*, 91(2), 254–271.
- Biswas, D., Grewal, D., & Roggeveen, A. (2010). How the order of sampled experiential products affects choice. *Journal of Marketing Research*, 47(3), 508–519.

- Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), 115–143.
- Bosman, D. (2013). The review credibility of electronic word-of-mouth communication on e-commerce platforms. *Management Dynamics*, 22(3), 29–45.
- Brack, A. D., & Benkenstein, M. (2012). The effects of overall similarity regarding the customer-to-customer-relationship in a service context. *Journal of Retailing and Consumer Services*, 19(5), 501–509.
- Breugelmans, E., Campo, K., & Gijsbrechts, E. (2007). Shelf sequence and proximity effects on online grocery choices. *Marketing Letters*, 18(1–2), 117–133.
- BrightLocal. (2020). *Local Consumer Review Survey*. Available at: <https://www.brightlocal.com/research/local-consumer-review-survey/>
- Brocato, E. D., Voorhees, C. M., & Baker, J. (2012). Understanding the Influence of Cues from Other Customers in the Service Experience: A Scale Development and Validation. *Journal of Retailing*, 88(3), 384–398.
- Brown, C. L., & Krishna, A. (2004). The skeptical shopper: A metacognitive account for the effects of default options on choice. *Journal of Consumer Research*, 31(3), 529–539.
- Bruine de Bruin, W., & Keren, G. (2003). Order effects in sequentially judged options due to the direction of comparison. *Organizational Behavior and Human Decision Processes*, 92(1–2), 91–101.
- Buda, R., & Zhang, Y. (2000). Consumer product evaluation: the interactive effect of message framing, presentation order, and source credibility. *Journal of Product & Brand Management*, 9(4), 229–242.
- Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics Using Stata*. Stata Press Book, StataCorp LP.
- Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data, Second edition. In *Regression Analysis of Count Data, Second Edition*.
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72(7–9), 1775–1781.
- Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, 50(2), 511–521.
- Chaiken, S. (1980). Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion. *Journal of Personality and Social Psychology*, 39(5), 752–766.
- Chaiken, S. (1987). *The heuristic model of persuasion*.

- Chapman, C., & Feit, E. M. (2015). *R for marketing research and analytics*.
- Chen, C. C., & Tseng, Y. De. (2011). Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, 50(4), 755–768.
- Chen, K., Kou, G., Shang, J., & Chen, Y. (2015). Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches. *Electronic Commerce Research and Applications*, 14(1), 58–74.
- Cheng, Y. H., & Ho, H. Y. (2015). Social influence's impact on reader perceptions of online reviews. *Journal of Business Research*, 68(4), 883–887.
- Cheung, C. M. K., Lee, M. K. O., & Rabjohn, N. (2008). The impact of electronic word-of-mouth. In *Internet Research* (Vol. 18, Issue 3).
- Chevalier, J. A., & Goolsbee, A. (2003). Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics*, 1, 203–222.
- Chevalier, J. a, & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 43(3), 345–354.
- Chevalier, J., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 43(3), 345–354.
- Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science*, 29(5), 944–957.
- Cho, E., Fiore, A. M., & W.Russell, D. (2015). Validation of a Fashion Brand Image Scale Capturing Cognitive, Sensory, and Affective Associations: Testing Its Role in an Extended Brand Equity Model. *Psychology & Marketing*, 32(1), 28–48.
- Chua, A. Y. K., & Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior*, 54, 547–554.
- Chung, C.K., & Pennebaker, J. W. (2007). The psychological function of function words. *Social Communication: Frontiers of Social Psychology*, 343–359.
- Chung, Cindy K., & Pennebaker, J. W. (2007). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality*, 42(1), 96–132.
- Cohn, M. a., Mehl, M. R., & Pennebaker, J. W. (2004). Markers of Linguistic Psychological Change. *Psychological Science*, 15(10), 687–693.
- Cohn, M. a., Mehl, M. R., & Pennebaker, J. W. (2014). Markers of Linguistic Psychological Change Surrounding September 11, 2001. *Psychological Science*, 15(10), 687–693.

- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2014). *Markers of Linguistic Psychological Change Surrounding September. 15*(10), 687–693.
- Cui, G., Lui, H.-K., & Guo, X. (2012). The Effect of Online Consumer Reviews on New Product Sales. *International Journal of Electronic Commerce*, 17(1), 39–58.
- Cyr, D., Head, M., Lim, E., & Stibe, A. (2018). Using the elaboration likelihood model to examine online persuasion through website design. *Information and Management*, April 2017, 1–15.
- Darley, W. K., & Smith, R. E. (1993). Advertising Claim Objectivity: Antecedents and Effects. *Journal of Marketing*, 57(4), 100.
- Davis, D. F., Golicic, S. L., & Marquardt, A. (2009). Measuring brand equity for logistics services. *The International Journal of Logistics Management*, 20(2), 201–212.
- De Vries, L., Gensler, S., & Leeﬂang, P. S. H. (2012). Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing. *Journal of Interactive Marketing*, 26(2), 83–91.
- Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, 27(4), 293–307.
- Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4), 23–45.
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective Latent Concept Modeling for ad hoc information retrieval. *Document Numerique*, 17(1), 61–84.
- Drèze, X., & Zufryden, F. (2004). Measurement of online visibility and its impact on Internet traffic. *Journal of Interactive Marketing*, 18(1), 20–37.
- Duan, W., Gu, B., & Whinston, A. B. (2008a). Do online reviews matter? - An empirical investigation of panel data. *Decision Support Systems*, 45(4), 1007–1016.
- Duan, W., Gu, B., & Whinston, A. B. (2008b). The dynamics of online word-of-mouth and product sales-An empirical investigation of the movie industry. *Journal of Retailing*, 84(2), 233–242.
- Duan, W., Gu, B., & Whinston, A. B. (2008c). Do online reviews matter? — An empirical investigation of panel data. *Decision Support Systems*, 45(4), 1007–1016.
- Elberse, A., & Eliashberg, J. (2003). Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures. *Marketing Science*, 22(3), 329–354.
- Ellsworth, P. C., & Smith, C. A. (1988). From appraisal to emotion: Differences among unpleasant feelings. *Motivation and Emotion*, 12(3), 271–302.

- Ert, E., & Fleischer, A. (2016). Mere Position Effect in Booking Hotels Online. *Journal of Travel Research*, 55(3), 311–321.
- Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management*, 55, 62–73.
- Eslami, S. P., & Ghasemaghaei, M. (2018). Effects of online review positiveness and review score inconsistency on sales: A comparison by product involvement. *Journal of Retailing and Consumer Services*, 45(June), 74–80.
- Faircloth, J. B., Capella, L. M., & Alford, B. L. (2001). The Effect of Brand Attitude and Brand Image on Brand Equity. *Journal of Marketing Theory and Practice*, 9(3), 61–75.
- Feldman, J. M., & Lynch, J. G. (1988). Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior. *Journal of Applied Psychology*, 73(3), 421–435.
- Felfernig, A., Friedrich, G., Gula, B., Hitz, M., Kruggel, T., Leitner, G., Melcher, R., Riepan, D., Strauss, S., Teppan, E., & Vitouch, O. (2007). Persuasive recommendation: Serial position effects in knowledge-based recommender systems. *Proceedings of the 2nd International Conference on Persuasive Technology (PERSUASIVE '07)*, 283–295.
- Filieri, R., Hofacker, C. F., & Alguezaui, S. (2018). What makes information in online consumer reviews diagnostic over time? The role of review relevancy, factuality, currency, source credibility and ranking score. *Computers in Human Behavior*, 80, 122–131.
- Filieri, R., & McLeay, F. (2014). E-WOM and Accommodation: An Analysis of the Factors That Influence Travelers' Adoption of Information from Online Reviews. *Journal of Travel Research*, 53(1), 44–57.
- Filieri, R., McLeay, F., Tsui, B., & Lin, Z. (2018). Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services. *Information and Management*, 55(8), 956–970.
- Filieri, R., Raguseo, E., & Vitari, C. (2018). When are extreme ratings more helpful? Empirical evidence on the moderating effects of review characteristics and product type. *Computers in Human Behavior*, 88, 134–142.
- Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., & Freling, T. (2014). How Online Product Reviews Affect Retail Sales: A Meta-analysis. *Journal of Retailing*, 90(2), 217–232.
- Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., & Treinen, M. (2001). What makes web sites credible? A report on a large quantitative study. *Conference on Human Factors in Computing Systems - Proceedings*, 61–68.

- Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 19(3), 291–313.
- Gao, B., Hu, N., & Bose, I. (2017). Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews. *Decision Support Systems*, 95, 1–11.
- Gensler, S., Völckner, F., Egger, M., Fischbach, K., & Schoder, D. (2015). Listen to your customers: Insights into brand image using online consumer-generated product reviews. *International Journal of Electronic Commerce*, 20(1), 112–141.
- Ghose, A., & Ipeiroitis, P. G. (2010). Estimating the Helpfulness and Economic Impact of Product Reviews. *Ieee Transactions on Knowledge and Data Engineering*, 1–15.
- Ghose, A., & Ipeiroitis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10), 1498–1512.
- Ghose, A., Ipeiroitis, P. G., & Li, B. (2012). Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science*, 31(3), 493–520.
- Girard, T., & Dion, P. (2010). Validating the search, experience, and credence product classification framework. *Journal of Business Research*, 63(9–10), 1079–1087.
- Gobet, F., Lane, P., Croker, S., Cheng, P., Jones, G., Oliver, I., & Pine, J. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5(6), 236–243.
- Godes, D., & Silva, J. C. (2012). Sequential and Temporal Dynamics of Online Opinion. *Marketing Science*, 31(3), 448–473.
- Godes, David, & Mayzlin, D. (2004). Using Online Conversations to Study Word-of-Mouth Communication. *Marketing Science*, 23(4), 545–560.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(SUPPL. 1), 5228–5235.
- Gu, B., Gu, B., & Park, J. (2012). The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products. *Information Systems Research*, 23(1), 182–196.
- Guo, J., Wang, X., & Wu, Y. (2020). Positive emotion bias: Role of emotional content from online customer reviews in purchase decisions. *Journal of Retailing and Consumer Services*, 52(July 2019).
- Guo, Y., Barnes, S. J., & Jia, Q. (2017a). Mining meaning from online ratings and

- reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.
- Guo, Y., Barnes, S. J., & Jia, Q. (2017b). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.
- Hamby, A., Daniloski, K., & Brinberg, D. (2015). How consumer reviews persuade through narratives. *Journal of Business Research*, 68(6), 1242–1250.
- Harrison-Walker, L. J. (2001). The Measurement of Word-of-Mouth Communication and an Investigation of Service Quality and Customer Commitment As Potential Antecedents. *Journal of Service Research*, 4(1), 60–75.
- Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, 36(1), 20–38.
- Häubl, G., & Trifts, V. (2000). Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids. *Marketing Science*, 19(1), 4–21.
- Haugtvedt, C. P., & Wegener, D. T. (1994). Message Order Effects in Persuasion: An Attitude Strength Perspective. *Journal of Consumer Research*, 21(1), 205.
- Heng, Y., Gao, Z., Jiang, Y., & Chen, X. (2018). Exploring hidden factors behind online food shopping from Amazon reviews: A topic mining approach. *Journal of Retailing and Consumer Services*, 42(January), 161–168.
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, 18(1), 38–52.
- Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective. *Journal of Consumer Research*, 17(4), 454.
- Herrmann, A., Goldstein, D. G., Stadler, R., Landwehr, J. R., Heitmann, M., Hofstetter, R., & Huber, F. (2011). The effect of default options on choice-Evidence from online product configurators. *Journal of Retailing and Consumer Services*, 18(6), 483–491.
- Ho-Dac, N. N., Carson, S. J., & Moore, W. L. (2013). The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter? *Journal of Marketing*, 77(6), 37–53.
- Hofmann, J., Clement, M., Völckner, F., & Hennig-Thurau, T. (2017). Empirical generalizations on the impact of stars on the economic success of movies.

- International Journal of Research in Marketing*, 34(2), 442–461.
- Holbrook, M. B. (1978). Beyond Attitude Structure: Toward the Informational Determinants of Attitude. *Journal of Marketing Research*, 15(4), 545.
- Holtgraves, T., & Lasky, B. (1999). Linguistic power and persuasion. *Journal of Language and Social Psychology*, 18(2), 196–205.
- Hornik, K., & Grün, B. (2011). topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software*, 40(13), 1–30.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '04*, 168.
- Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? the mediating role of online review sentiments on product sales. *Decision Support Systems*, 57(1), 42–53.
- Hu, Y., & Li, X. (2011). Context-Dependent Product Evaluations: An Empirical Analysis of Internet Book Reviews. *Journal of Interactive Marketing*, 25(3), 123–133.
- Huang, A. H., Chen, K., Yen, D. C., & Tran, T. P. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behavior*, 48, 17–27.
- Hutto, C. J., & Gilbert, E. E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14).". *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*.
- International, E. (2017). *Colour cosmetics in the US* (Issue May).
- Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011a). Language Style Matching Predicts Relationship Initiation and Stability. *Psychological Science*, 22(1), 39–44.
- Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011b). Language Style Matching Predicts Relationship Initiation and Stability. *Psychological Science*, 22(1), 39–44.
- Jalilvand, M. R., & Samiei, N. (2012). The effect of electronic word of mouth on brand image and purchase intention: An empirical study in the automobile industry in Iran. *Marketing Intelligence and Planning*, 30(4), 460–476.
- Jiang, S., Cai, S., Olle, G. O., & Qin, Z. (2015). Durable product review mining for customer segmentation. *Kybernetes*, 44(1), 124–138.
- Jiménez, F. R., & Mendoza, N. A. (2013). Too popular to ignore: The influence of online reviews on purchase intentions of search and experience products. *Journal of Interactive Marketing*, 27(3), 226–235.

- John, D. R., Loken, B., Kim, K., & Monga, A. B. (2006). Brand concept maps: A methodology for identifying brand association networks. *Journal of Marketing Research*, 43(4), 549–563.
- Johnson, E. J., Bellman, S., & Lohse, G. L. (2002). Defaults , Framing and Privacy :Why Opting In-Opting Out. *Marketing Letters*, 13(1), 5–15.
- Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, 80(4), 557–571.
- Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*, 33(2), 125–143.
- Kamble, Z., Doan, Q. T., Nguyen, H. N., Zeng, Z., & Zihui, L. L. (2020). The Significance of Online Reviews for Hotel Entrepreneurial Success. In *Entrepreneurial Development and Innovation in Family Businesses and SMEs* (pp. 75–86).
- Keller, K. L. (1993). *Conceptualizing, Measuring, and Managing Customer-Based Brand Equity*. 57(1), 1–22.
- Khwaja, M. G., & Zaman, U. (2020). Configuring the evolving role of ewom on the consumers information adoption. *Journal of Open Innovation: Technology, Market, and Complexity*, 6(4), 1–13.
- Kim, H. B., Kim, W. G., & An, J. A. (2003). The effect of consumer-based brand equity on firms' financial performance. *Journal of Consumer Marketing*, 20(4–5), 335–351.
- Kim, M., & Lennon, S. (2007). The Effects of Visual and Verbal Information on Attitudes and Purchase Intentions in Internet Shopping. *Psychology & Marketing*, 24(9), 763–785.
- King, M. F., & Balasubramanian, S. k. (1994). The effects of Expertise, End Goal, and Product Type on Adoption of Preference Formation Strategy. *Journal of the Academy of Marketing Science*, 22.
- Kostyk, A., Niculescu, M., & Leonhardt, J. M. (2017). Less is more: Online consumer ratings' format affects purchase intentions and processing. *Journal of Consumer Behaviour*, 16(5), 434–441.
- Kotler, P., & Armstrong, G. (2020). *Principles of Marketing* (18th Globa). Pearson Education Limited.
- Kozinets, R. V., De Valck, K., Wojnicki, A. C., & Wilner, S. J. S. (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal*

- of Marketing*, 74(2), 71–89.
- Kübler, R. V., Colicev, A., & Pauwels, K. H. (2019). Social Media's Impact on the Consumer Mindset: When to Use Which Sentiment Extraction Tool? *Journal of Interactive Marketing*, 50, 136–155.
- Kudeshia, C., & Kumar, A. (2017). Social eWOM: does it affect the brand attitude and purchase intention of brands? *Management Research Review*, 40(3), 310–330.
- Lee, J., Park, D. H., & Han, I. (2008). The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications*, 7(3), 341–352.
- Lee, P. J., Hu, Y. H., & Lu, K. T. (2018). Assessing the helpfulness of online hotel reviews: A classification-based approach. *Telematics and Informatics*, 35(2), 436–445. <https://doi.org/10.1016/j.tele.2018.01.001>
- Lee, S., & Choeh, J. Y. (2018). The interactive impact of online word-of-mouth and review helpfulness on box office revenue. *Management Decision*, 56(4), 849–866.
- Lee, S., & Choeh, J. Y. (2020). Using the social influence of electronic word-of-mouth for predicting product sales: The moderating effect of review or reviewer helpfulness and product type. *Sustainability (Switzerland)*, 12(19).
- Lee, T. Y., & Bradlow, E. T. (2011). Automated Marketing Research Using Online Customer Reviews. *Journal of Marketing Research*, 48(5), 881–894.
- Levinson, P. (2013). *New Media*. Pearson.
- Lewsey, J. D., & Thomson, W. M. (2004). The utility of the zero-inflated Poisson and zero-inflated negative binomial models: A case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. *Community Dentistry and Oral Epidemiology*, 32(3), 183–189.
- Li, M., Huang, L., Tan, C.-H., & Wei, K.-K. (2013). Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features. *International Journal of Electronic Commerce*, 17(4), 101–136.
- Li, S. T., Pham, T. T., & Chuang, H. C. (2019). Do reviewers' words affect predicting their helpfulness ratings? Locating helpful reviewers by linguistics styles. *Information and Management*, 56(1), 28–38.
- Li, Xiaolin, Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information and Management*, 56(2), 172–184.
- Li, Xinxin, & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, 19(4), 456–474.
- Liang, T. P., Li, X., Yang, C. T., & Wang, M. (2015). What in Consumer Reviews Affects the Sales of Mobile Apps: A Multifacet Sentiment Analysis Approach. *International*

- Journal of Electronic Commerce*, 20(2), 236–260.
- Lin, C.-L., Lee, S.-H., & Horng, D.-J. (2011). The effects of online reviews on purchasing intention: The moderating role of need for cognition. *Social Behavior and Personality: An International Journal*, 39(1), 71–81.
- Liu, J., Cao, Y., Lin, C., Huang, Y., ... M. Z. the 2007 J. C. on, & 2007, U. (2007). Low-quality product review detection in opinion summarization. *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, June, 334–342.
- Liu, Q., & Karahanna, E. (2015). An agent-based modeling analysis of helpful vote on online product reviews. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2015-March, 1585–1595.
- Liu, Y. (2006). Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*, 70(3), 74–89.
- Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, 47, 140–151.
- Low, G. S., & Lamb, C. W. (2000). The measurement and dimensionality of brand associations. *Journal of Product & Brand Management*, 9(6), 350–370.
- Lozano, M. B., Martínez, B., & Pindado, J. (2016). Corporate governance, ownership and firm value: Drivers of ownership as a good corporate governance mechanism. *International Business Review*, 25(6), 1333–1343.
- Ludwig, S., de Ruyter, K., Friedman, M., Brügger, E. C., Wetzels, M., & Pfann, G. (2013a). More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*, 77(1), 87–103.
- Ludwig, S., de Ruyter, K., Friedman, M., Brügger, E. C., Wetzels, M., & Pfann, G. (2013b). More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*, 77(1), 87–103.
- Lynch, Jr., J. G., Marmorstein, H., & Weigold, M. F. (1988). Choices from Sets Including Remembered Brands: Use of Recalled Attributes and Prior Overall Evaluations. *Journal of Consumer Research*, 15(2), 169.
- Mackinnon, S. P., Jordan, C. H., & Wilson, A. E. (2011). Birds of a Feather Sit Together: Physical Similarity Predicts Seating Choice. *Personality and Social Psychology Bulletin*, 37(7), 879–892.
- Magerman, T., Van Looy, B., Baesens, B., & Debackere, K. (2011). Assessment of Latent Semantic Analysis (LSA) text mining algorithms for large scale mapping of patent and scientific publication documents. *FBE Research Report MSI_1114*.

- Magoulas, R., & Swoyer, S. (2020). *AI Adoption in the Enterprise 2020*.
- Mahr, D., Stead, S., & Odekerken-Schröder, G. (2019). Making sense of customer service experiences: a text mining review. *Journal of Services Marketing*, 33(1), 88–103.
- Malhotra, N. K. (1981). A Scale to Measure Self-Concepts, Person Concepts, and Product Concepts. *Journal of Marketing Research*, 18(4), 456.
- Mantonakis, A., Rodero, P., Lesschaeve, I., & Hastie, R. (2009). Order in choice: Effects of serial position on preferences. *Psychological Science*, 20(11), 1309–1312.
- Marchand, A., Hennig-Thurau, T., & Wiertz, C. (2017). Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. *International Journal of Research in Marketing*, 34(2), 336–354.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1), 415–444.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.
- Moe, W. W., & Trusov, M. (2011). The Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing Research*, 48(3), 444–456.
- Mohammad, S., & Turney, P. (2010). Emotions Evoked by Common Words and Phrases. *Proceedings of the NAACL-HLT, June*, 26–34.
- Moon, S., & Kamakura, W. A. (2017). A picture is worth a thousand words: Translating product reviews into a product positioning map. *International Journal of Research in Marketing*, 34(1), 265–285.
- Mudambi, S. ., & Schuff, D. (2010). What makes a Helpful online Review? *MIS Quarter*, 28(4), 695–704.
- Munch, J. M., & Swasy, J. L. (1988). Rhetorical question, summarization frequency, and argument strength effects on recall. *Journal of Consumer Research*, 15(1), 69–76.
- Murphy, J., Hofacker, C., & Mizerski, R. (2006). Primacy and recency effects on clicking behavior. *Journal of Computer-Mediated Communication*, 11(2), 173–192.
- Nabi, R. L. (2003). Exploring the framing effects of emotion: Do discrete emotions differentially influence information accessibility, information seeking, and policy preference? *Communication Research*, 30(2), 224–247.
- Nazlan, N. H., Tanford, S., & Montgomery, R. (2018). The effect of availability heuristics in online consumer reviews. *Journal of Consumer Behaviour*, 17(5), 449–460.
- Nelson, P. (1970). *Information and Consumer Behavior*. 78(2), 311–329.
- Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine Your Own Business:

- Market-Structure Surveillance Through Text Mining. *Marketing Science*, 31(3), 521–543.
- Netzer, Oded, Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521–543.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Personality and Social Psychology Bulletin Lying Words: Predicting Deception From Linguistic Styles. *Personality and Social Psychology Bulletin*, 29(5), 665–675.
- Ngo-Ye, T. L., & Sinha, A. P. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, 61(1), 47–58.
- Nielsen. (2016). *The sweet smell of seasonal success*. <https://www.nielsen.com/au/en/insights/news/2016/the-sweet-smell-of-seasonal-success.html>
- Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *CEUR Workshop Proceedings*, 718, 93–98.
- Pan, B., Zhang, L., & Law, R. (2013). The Complex Matter of Online Hotel Choice. *Cornell Hospitality Quarterly*, 41(1), 74–83.
- Pan, Y., & Zhang, J. Q. (2011). Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews. *Journal of Retailing*, 87(4), 598–612.
- Pang, J., & Qiu, L. (2016). Effect of online review chunking on product attitude: The moderating role of motivation to think. *International Journal of Electronic Commerce*, 20(3), 355–383.
- Park, D.-H., Lee, J., & Han, I. (2007). The Effect of On-Line Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement. *International Journal of Electronic Commerce*, 11(4), 125–148.
- Park, D. H., & Kim, S. (2008). The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews. *Electronic Commerce Research and Applications*, 7(4), 399–410.
- Park, D. H., & Lee, J. (2008). eWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications*, 7(4), 386–398.
- Park, E. J., Kim, E. Y., Funches, V. M., & Foxx, W. (2012). Apparel product attributes, web browsing, and e-impulse buying on shopping websites. *Journal of Business Research*, 65(11), 1583–1589.
- Park, H.-J., & Rabolt J., N. (2009). Cultural Value, Consumption Value, and Global

- Brand Image: A Cross-National Study. *Psychology & Marketing*, 26(8), 714–735.
- Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research*, 50, 67–83.
- Payan, J. M., & McFarland, R. G. (2005). Decomposing Influence Strategies: Argument Structure and Dependence as Determinants of the Effectiveness of Influence Strategies in Gaining Channel Member Compliance. *Journal of Marketing*, 69(3), 66–79.
- Payne, J. W. (1982). Contingent Decision Behavior. *Psychological Bulletin*, 92(2), 382.
- Peladeau, N. (2016). *WordStat: Content Analysis Module for SIM- STAT*.
- Pennebaker, J. W., Boyd, R., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. *Austin, TX: University of Texas at Austin, SEPTEMBER 2015*, 1–22.
- Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS ONE*, 9(12), 1–10.
- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The Development and Psychometric Properties of LIWC2007. *Austin, Tx: LIWC.Net*, 1(2).
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In *Advances in experimental social psychology* (Vol. 19, pp. 123–205).
- Petty, Richard E., Tormala, Z. L., Hawkins, C., & Wegener, D. T. (2001). Motivation to think and order effects in persuasion: The moderating role of chunking. *Personality and Social Psychology Bulletin*, 27(3), 332–344.
- Pindado, J., Requejo, I., & de la Torre, C. (2011). Family control and investment-cash flow sensitivity: Empirical evidence from the Euro zone. *Journal of Corporate Finance*, 17(5), 1389–1409.
- Price, P. C., & Stone, E. R. (2004). Intuitive Evaluation of Likelihood Judgment Producers: Evidence for a Confidence Heuristic. *Journal of Behavioral Decision Making*, 17(1), 39–57.
- Puranam, D., Narayan, V., & Kadiyali, V. (2017). The effect of calorie posting regulation on consumer opinion: A flexible latent dirichlet allocation model with informative priors. *Marketing Science*, 36(5), 726–746.
- Purnawirawan, N., De Pelsmacker, P., & Dens, N. (2012). Balance and Sequence in Online Reviews: How Perceived Usefulness Affects Attitudes and Intentions. *Journal of Interactive Marketing*, 26(4), 244–255.
- Purnawirawan, N., Dens, N., & De Pelsmacker, P. (2012). Balance and Sequence in

- Online Reviews: The Wrap Effect. *International Journal of Electronic Commerce*, 17(2), 71–98.
- Racherla, P., & Friske, W. (2012). Perceived “usefulness” of online consumer reviews: An exploratory investigation across three services categories. *Electronic Commerce Research and Applications*, 11(6), 548–559.
- Rodríguez-Díaz, M., Rodríguez-Díaz, R., & Espino-Rodríguez, T. (2018). Analysis of the Online Reputation Based on Customer Ratings of Lodgings in Tourism Destinations. *Administrative Sciences*, 8(3), 51.
- Roodman, D. (2009a). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, 9(1), 86–136.
- Roodman, D. (2009b). Practitioners’ corner: A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71(1), 135–158.
- Ruiz-Mafe, C., Chatzipanagiotou, K., & Curras-Perez, R. (2018). The role of emotions and conflicting online reviews on consumers’ purchase intentions. *Journal of Business Research*, 89(June 2017), 336–344.
- Salehan, M., & Kim, D. J. (2015). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, 81, 30–40.
- SanJosé-Cabezudo, R., Gutiérrez-Arranz, A. M., & Gutiérrez-Cillán, J. (2009). The Combined Influence of Central and Peripheral Routes in the Online Persuasion Process. *CyberPsychology & Behavior*, 12(3), 299–308.
- Saumya, S., Singh, J. P., Baabdullah, A. M., Rana, N. P., & Dwivedi, Y. K. (2018). Ranking online consumer reviews. *Electronic Commerce Research and Applications*, 29(March), 78–89.
- Schnittka, O., Sattler, H., & Zenker, S. (2012). Advanced brand concept maps: A new approach for evaluating the favorability of brand association networks. *International Journal of Research in Marketing*, 29(3), 265–274.
- Selkirk, E. (1996). The prosodic structure of function words. In J. L. (Brown U. Morgan & K. (Brown U. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 187–214). Lawrence Erlbaum Associates, Inc.
- Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers’ online choices. *Journal of Retailing*, 80(2), 159–169.
- Sheldon, P., & Bryant, K. (2016). Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in Human Behavior*, 58, 89–97.
- Shugan, S. M. (1980). The Cost of Thinking. *Journal of Consumer Research*, 7(2), 99–111.

- Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Kumar Roy, P. (2017). Predicting the “helpfulness” of online consumer reviews. *Journal of Business Research*, 70, 346–355.
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, 105(1), 131–142.
- Slovic, P. (1972). From Shakespeare to Simon: Speculations and some evidence. *Oregon Research Institute Bulletin*, 1–19.
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality & Social Psychology*, 48(4), 813–838.
- Snizek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, 84(2), 288–307.
- Social Blade*. (2017). <https://socialblade.com/>
- Spears, N., & Singh, S. N. (2004). Measuring attitude toward the brand and purchase intentions. *Journal of Current Issues and Research in Advertising*, 26(2), 53–66.
- Spenner, P., & Freeman, K. (2012). To keep your customers, keep it simple. *Harvard Business Review*, 90(5).
- SproutSocial. (2018). *Instagram stats*. Available at: <https://sproutsocial.com/insights/instagram-stats/>
- Statista*. (2020). Available at: <https://www.statista.com/forecasts/997051/sources-of-information-about-products-in-the-us>
- Sun, M. (2012). How Does the Variance of Product Ratings Matter? *Management Science*, 58(4), 696–707.
- Sun, Q., Niu, J., Yao, Z., & Yan, H. (2019). Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level. *Engineering Applications of Artificial Intelligence*, 81(December 2018), 68–78.
- Sussman, S. W., & Siegal, W. S. (2003). Informational influence in organizations: An integrated approach to knowledge adoption. *Information Systems Research*, 14(1), 47–65.
- Tang, T. (Ya), Fang, E. (Er), & Wang, F. (2014). Is neutral really neutral? The effects of neutral user-generated content on product sales. *Journal of Marketing*, 78(4), 41–58.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
- Tirunillai, S., & Tellis, G. J. (2014). Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation. *Journal of*

- Marketing Research*, 51(4), 463–479.
- Toubia, O., Iyengar, G., Bunnell, R., & Lemaire, A. (2019). Extracting Features of Entertainment Products: A Guided Latent Dirichlet Allocation Approach Informed by the Psychology of Media Consumption. *Journal of Marketing Research*, 56(1), 18–36.
- Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing*, 73(5), 90–102.
- Tybout, A. M., Sternthal, B., Malaviya, P., Bakamitsos, G. A., & Park, S. E. B. (2005). Information accessibility as a moderator of judgments: The role of content versus retrieval ease. *Journal of Consumer Research*, 32(1), 76–85.
- Van Hove, G., & Lievens, F. (2007). Social influences on organizational attractiveness: Investigating if and when word of mouth matters. *Journal of Applied Social Psychology*, 37(9), 2024–2047.
- Veloutsou, C., & Delgado-Ballester, E. (2018). New challenges in brand management. *Spanish Journal of Marketing - ESIC*, 22(3), 255–272.
- Vermeer, S. A. M., Araujo, T., Bernritter, S. F., & van Noort, G. (2019). Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media. *International Journal of Research in Marketing*, 36(3), 492–508.
- Vriens, M., Chen, S., & Schomaker, J. (2019). The evaluation of a brand association density metric. *Journal of Product and Brand Management*, 28(1), 104–116.
- Walker, M. A., Mehl, M. R., Moore, R. K., & Mairesse, F. (2007). Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research*, 30, 457–500.
- Wang, F., Liu, X., & Fang, E. (Er). (2015). User Reviews Variance, Critic Reviews Variance, and Product Sales: An Exploration of Customer Breadth and Depth Effects. *Journal of Retailing*, 91(3), 372–389.
- Wang, W., Feng, Y., & Dai, W. (2018). Topic analysis of online reviews for two competitive products using latent Dirichlet allocation. *Electronic Commerce Research and Applications*, 29(January), 142–156.
- Wang, X., Tang, L. (Rebecca), & Kim, E. (2019). More than words: Do emotional content and linguistic style matching matter on restaurant review helpfulness? *International Journal of Hospitality Management*, 77(August 2018), 438–447.
- Weathers, D., Swain, S. D., & Grover, V. (2015). Can online product reviews be more helpful? Examining characteristics of information content by product type. *Decision Support Systems*.

- Westerman, D., Spence, P. R., & Van Der Heide, B. (2014). Social Media as Information Source: Recency of Updates and Credibility of Information. *Journal of Computer-Mediated Communication*, 19(2), 171–183.
- Wickham, H. (2016). *tidyr: Easily tidy data with spread and gather functions*.
- Wickham, H. (2017). ggplot2 - Elegant Graphics for Data Analysis (2nd Edition). *Journal of Statistical Software*, 77(Book Review 2), 2–5.
- Wickham, H., & Francois, R. (2016). *dplyr: A Grammar of Data Manipulation*.
- Winkelmann, P. D. R. (2008). Econometric Analysis of Count Data. In *Econometric Analysis of Count Data*.
- Wong, C. U. I., & Qi, S. (2017). Tracking the evolution of a destination's image by text-mining online reviews - the case of Macau. *Tourism Management Perspectives*, 23, 19–29.
- Wooldridge, J. M. (2011). *Econometric analysis of cross section and panel data* (Issue November).
- Wright, K. (2000). Perceptions of on-line support providers: An examination of perceived homophily, source credibility, communication and social support within on-line support groups. *Communication Quarterly*, 48(1), 44–59.
- Wu, J. (2017). Review popularity and review helpfulness: A model for user review effectiveness. *Decision Support Systems*, 97, 92–103.
- Xu, P., & Liu, D. (2019). Product engagement and identity signaling: The role of likes in social commerce for fashion products. *Information and Management*, 56(2), 143–154.
- Yang, H., & Cho, S. (2015). *Understanding Brands with Visualization and Keywords from eWOM using Distributed Representation*.
- Yau, K. K. W., Wang, K., & Lee, A. H. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal*, 45(4), 437–452.
- Yazdani, E., Gopinath, S., & Carson, S. (2018). Preaching to the Choir: The Chasm Between Top-Ranked Reviewers, Mainstream Customers, and Product Sales. *Marketing Science*, 37(5), 838–851.
- Yin, G., Wei, L., Xu, W., & Chen, M. (2014). Yin, G., Wei, L., Xu, W., & Chen, M. (2014). Exploring Heuristic Cues for Consumer Perceptions of Online Reviews Helpfulness: The Case of Yelp.com. PACIS 2014 Proceedings, 1–12. Exploring Heuristic Cues for Consumer Perceptions of Online Reviews Helpfulness. *PACIS 2014 Proceedings*, 1–12.
- Zhang, J. (2019). What's yours is mine: exploring customer voice on Airbnb using text-

- mining approaches. *Journal of Consumer Marketing*, 36(5), 655–665.
- Zhang, K. Z. K., Zhao, S. J., Cheung, C. M. K., & Lee, M. K. O. (2014). Examining the influence of online reviews on consumers' decision-making: A heuristic-systematic model. *Decision Support Systems*, 67, 78–89.
- Zhang, Ying, & Lin, Z. (2018). Predicting the helpfulness of online product reviews: A multilingual approach. *Electronic Commerce Research and Applications*, 27, 1–10.
- Zhang, Yuchi, Moe, W. W., & Schweidel, D. A. (2017). Modeling the role of message content and influencers in social media rebroadcasting. *International Journal of Research in Marketing*, 34(1), 100–119.
- Zhou, S., & Guo, B. (2017). The order effect on online review helpfulness: A social influence perspective. *Decision Support Systems*, 93, 77–87.
- Zhu, F., & Zhang, X. (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*, 74(March), 133–148.

2

Online Reviews and Product Sales: The Role of Review Visibility

1. Introduction

Online consumer reviews are a type of electronic word-of-mouth (eWOM) **communication that can be defined as “peer-generated product evaluations posted on the company’s or a third party’s websites”** (Mudambi & Schuff, 2010). Academics and practitioners have highlighted the importance of online reviews for both consumers and companies.

A study by the consultancy firm BrighLocal (BrightLocal, 2020) reveals that 82% of consumers read online reviews when evaluating a business, and 76% rely on online reviews as much as personal recommendations. Besides, the same study reveals that including online reviews on the retailer website makes the searchers see the business as more trustworthy.

Academic literature has also highlighted the power of online reviews to predict different types of consumer behavior such as information adoption decisions (Cyr et al., 2018; Filieri, McLeay, et al., 2018; Khwaja & Zaman, 2020), purchase intentions (Jiménez & Mendoza, 2013; Kostyk et al., 2017; D. H. Park & Lee, 2008), and product sales in product categories such as hardware, books, movies, and hotels (J. Chevalier & Mayzlin, 2006; Chintagunta et al., 2010; Hofmann et al., 2017; S. Lee & Choeh, 2020; Xiaolin Li et al., 2019; Marchand et al., 2017). Some studies have also focused on exploring online reputation and image by analyzing product features revealed at online reviews (Rodríguez-Díaz et al., 2018), and others have studied review texts to uncover product features and sentiments (Q. Sun et al., 2019).

When exploring the role of online reviews to predict product sales, previous literature has implicitly assumed that every review for a product has the same probability of being viewed by consumers, so every review has been considered as equally influential in the consumer purchase decision. However, literature in decision-making has revealed that consumers usually face information overload situations in online environments, due to a large amount of information available (Beach, 1993; Häubl & Trifts, 2000), as it might

happen when dealing with a high volume of online reviews. In these complex environments, consumers cannot evaluate every single online review available for each product, and instead, they are likely to adopt selective processing strategies to reduce the cognitive effort of managing a big volume of information. For instance, the report published by BrightLocal (2020) reveals that, on average, consumers read a maximum of 10 online reviews before making a decision, which means that consumers are likely to base their decision only on a subset of all reviews. In the same line, the accessibility-diagnostics theory by Feldman and Lynch (Feldman & Lynch, 1988), claim that the likelihood of using a piece of information for making a choice depends both on its accessibility and its diagnosticity. Therefore, this theory might suggest that more accessible or visible online reviews are likely to be more used by consumers to make a choice.

In this research, we explore the relationship between online reviews and product sales by incorporating the notion of review visibility, which approaches the accessibility dimension of the accessibility-diagnostics theory by Feldman and Lynch (Feldman & Lynch, 1988). In line with other scholars (Archak et al., 2011; J. Chevalier & Mayzlin, 2006; Cui et al., 2012; M. Sun, 2012), product sales information is proxied in this study by the sales rank of products at the online retailer, which is obtained by web-scraping the web store.

Review visibility captures the rank order of online reviews for a product when using two important sorting mechanisms: most helpful and most recent. Online retailers show in their product web pages online reviews sorted by a specific default mechanism, which might vary between online retailers. Thus, even if consumers do not sort online reviews by themselves, reviews are already shown in a default sorting rank, many times in chronological order, in the way that some of them are more visible than others. Therefore, depending on the way consumers organize online reviews, some of them might have a greater impact on consumer decisions than others.

The diagnosticity of online reviews is described as the ability of the information to provide consumers with relevant product information that helps them to understand and evaluate the quality and performance of the product (Filleri, Hofacker, et al., 2018). As claimed by Payne (1982), an input's diagnosticity depends on whether it enables a decision maker to discriminate among alternatives and it depends on the characteristics of the input of information. Therefore, the diagnosticity of online reviews depends on the characteristics of the information contained in online reviews. To approach the diagnosticity dimension of online reviews, we use two sets of review variables. The first

set contains those review non-textual variables most used in previous literature: *volume*, *rating* and *rating_inconsistency*, which reflects the difference between each individual review rating and the product average rating. The second set includes three variables that summarize the textual content of online reviews for each product: *analytic*, *authentic*, and *clout*. These three summary variables are extracted from the last version of the text mining program Linguistic Inquiry and Word Count (LIWC) developed by Pennebaker et al. (Pennebaker et al., 2015). Previous literature has mainly analyzed the effect of non-textual review characteristics on product sales but has paid less attention to study the effect of textual variables. Therefore, we include in our study not only traditional non-textual variables but also some variables related to the textual content of reviews that have been proven to influence consumer behavior.

The main objective of this research is to study the impact of review non-textual and textual features on product sales in three cases of review visibility: (1) When every online review for a product is assumed to have the same probability of being viewed; (2) when we assume that consumers sort online reviews for a product by the most helpful mechanism, so most helpful reviews are more likely to be viewed and (3) when we assume that consumers sort online reviews for a product by the most recent mechanism, in the way that most recent online reviews are more likely to be viewed. So far, studies have revealed different effects of review variables on product sales depending of factors such as the platform, the type of product, the metric used to build the variable and the design of the study (Babić Rosario et al., 2016). In this research, we incorporate review visibility as a factor that might also impact the relationship between review characteristics and product sales. In this sense, effects might be different if we assume that consumers read every online review available for a product or if we assume that most visible reviews are more likely to been read.

To carry out the empirical analysis, we collected data from a US cosmetics online retailer. **Data belongs to the category of “blush” products, and it was collected on a weekly basis** over nine weeks between the 21 December 2016 and 17 February 2017. A panel data methodology was adopted to carry out the estimations.

The rest of the paper is structured as follows: First, the theoretical background and conceptual model are presented. Then, the data and research methodology are explained. Afterward, the discussion of results, followed by a general discussion about theoretical and managerial implications, is developed. Finally, the last section shows the main limitations and areas for future research.

2. Theoretical Background and Conceptual Model

2.1. Influence of Online Reviews on Product Sales

A great deal of literature has explored how the two main factors of online reviews, volume (number of online comments to the product or service) and valence or product average rating (average rating star given to a product), influence product sales (Chevalier & Mayzlin, 2006; Chintagunta et al., 2010; Li et al., 2019; Marchand et al., 2017; Moe & Trusov, 2011). However, the effect of these features on product sales are not unanimously clear. In fact, **Babić Rosario et al. (2016)** in their meta-analysis study, which analyses the effect of eWOM on product sales, reveal that the link between eWOM and product sales differs across platforms (social media platforms, review platforms and e-commerce platforms), products (tangible goods, services, digital products, utilitarian vs. hedonic products, high financial risk vs. low financial risk and mature vs. new products), eWOM metrics (how the variable is built, for example average, incremental and cumulative) and other study characteristics (e.g. how endogeneity is controlled, the type of control variables, etc.) (**Babić Rosario et al., 2016**). In our research, we suggest that differences might be also due to the role of review visibility. Results might be different if we assume that consumers read every online review available for a product, as in general previous literature has done, or, instead, we assume that most visible reviews are more likely to be read.

Most previous studies have revealed that a greater volume of online reviews of a product leads to an increase in product sales (Chevalier & Mayzlin, 2006; Duan et al., 2008a; Godes & Mayzlin, 2004; Y. Liu, 2006; Marchand et al., 2017), since a higher number of online reviews generates more product awareness and increases the perception of the product quality (Duan et al., 2008a). As stated by Marchand et al. (2017) the volume might communicate how many people find the product interesting. However, there are also studies that reveal that the volume of online reviews is not only a precursor of product sales, but also an outcome (Duan et al., 2008a), so this dual relationship might be considered to get to good estimations.

The effect of product rating on product sales remains less clear in the literature, which shows mixed results. Consumers usually associate positive online reviews with a better expected quality of the product, which leads to a positive attitude towards the product; while negative ratings are seen as complaints or denigration of the product, which leads to an unfavorable attitude towards the product (Floyd et al., 2014; Y. Liu, 2006). Some studies find that higher product ratings are positively related to product sales (J.

Chevalier & Mayzlin, 2006; Chintagunta et al., 2010; Dellarocas et al., 2007; Xiaolin Li et al., 2019); others reveal that the effect of rating on product sales is not significant (Duan et al., 2008a; Y. Liu, 2006). The effect of product rating on sales also has been explored using other rating measures, such as rating variance, rating inconsistency, and proportion of positive and negative reviews. However, no consensus has been reached. Chevalier and Mayzlin (2006) found that the greater the fraction of five-star reviews for books at Amazon.com, the better the sales rank of the book, whereas the higher the fraction of one-star reviews, the worse the sales rank of the book. Liu (2006) analyzed the impact of online reviews on box office revenues and found that, both the fraction of one-star and five-star reviews were not significant in explaining box office revenues.

In terms of review variance, Wang, Liu, and Fang (2015) found a negative impact of review variance on box office revenues. Moreover, in a movie context, Chintagunta et al. (2010) did not find a significant impact of variance on box office revenues. Furthermore, Sun (2012) stated that a higher standard deviation of Amazon.com ratings for books leads to higher relative sales only when the average product rating is low. Therefore, more research is needed to get a clearer understating of the effect of both product average rating and volume of online reviews on product sales.

A smaller stream of literature has explored the impact of review textual features on product sales. The study of narrative and its persuasion power has been widely explored in several research domains, such as communication, psychology, and marketing. Tausczik and Pennebaker (2010) claim that the words we use in daily life reflect who we are and the social relationships we are in, so language is the way in which people express their internal thoughts and emotions. Overall, studies on narrative persuasion have concluded that message characteristics have a strong power over different types of consumer behavior, such as purchase intention (M. Kim & Lennon, 2007), conversion rates (Ludwig et al., 2013a), liking and commenting brand posts (De Vries et al., 2012), and social media rebroadcasting (Zhang et al., 2017). However, so far, a relatively small number of scholars have explored the **influence of review textual features on consumers'** purchasing decisions. Although historically, the analysis of the text was complex, slow, and costly, the development of high-speed computers and new statistical methods has helped companies and researchers to go one step further in the study of texts and language (C.K. Chung & Pennebaker, 2007; Tausczik & Pennebaker, 2010). Therefore, scholars are increasingly paying attention to the study of the text of online reviews.

Most scholars exploring review text have focused on studying text sentiment (positive vs. negative). For example, Hu, Koh, and Reddy (2014), Liang, Li, Yang, and Wang (2015),

and Li et al. (2019) claim that more positive comments on the product lead to higher sales. Tang, Fang, and Wang (2014) reveal that neutral comments in terms of sentiment also impact product sales, and this impact depends on the amount of mixed and indifferent neutral comments and their relative strength. However, the effect of other review textual factors on product sales remains quite underexplored. For example, Yazdani et al. (2018) incorporates the role of three textual dimensions, adopted from Pennebaker et al. (2015), to explore the effect of text on product sales: affective content, social content, and informal language content. They find that product sales are positively influenced by reviews with higher affective and social content and by those that use more informal language. Overall, we observe that studies exploring review textual features usually conclude that the text of reviews influences consumer behavior and product sales. Therefore, not only non-textual aspects of online reviews (e.g., product average rating and volume) should be considered, but also review textual features. In our research, both non-textual and textual review features are incorporated to explore the effect of online reviews on product sales in three different cases of review visibility.

When exploring the effect of online reviews on product sales, different scholars have used different metrics to capture product sales. Some of them have used information on actual sales, for example, monthly sales (Zhu & Zhang, 2010), quarterly revenues (Luca, 2011), designated market areas (DMA) level sales (Yazdani et al., 2018) and daily revenues (Duan et al., 2008b). However, many papers have relied on sales-rankings as a proxy of actual product sales (Archak et al., 2011; Chevalier & Mayzlin, 2006; Cui et al., 2012; Sun, 2012). The main reason is that Chevalier and Goolsbee (2003) found that for Amazon.com, the relationship between $\ln(\text{sales})$ and $\ln(\text{sales ranks})$ is approximately linear. Our study also uses sales-rankings as a proxy of product sales.

2.2. Decision-Making and Information Processing in the Online Environment

Due to the growth of the internet, not only do more consumers articulate themselves online, but also search costs are lower than in offline situations (Aljukhadar et al., 2012). Therefore, when consumers evaluate online reviews to make purchase decisions, they might be confronted by too much information, which results in information overload situations (Park & Lee, 2008). In this scenario, consumers should select which online reviews to evaluate, since it is very difficult for them to evaluate all available reviews. It is well-known in the decision-making and information processing literature that in complex environments, individuals are often unable to evaluate all available alternatives because humans have limited information processing capacity (Beach, 1993). For

example, previous research in psychology has revealed that the span of information processing for humans is between five and nine chunks (Miller, 1956). In cognitive science, a familiar unit or chunk is defined as “a collection of elements having strong associations with one another, but weak associations with elements within other chunks” (Gobet et al., 2001).

In an online reviews context, Liu and Karahanna (2015) revealed that consumers read on average, seven reviews before making a decision. The evidence to believe that the accessibility or visibility of online reviews plays an important role in consumer decision-making is also grounded in the accessibility-diagnostics theory (Feldman & Lynch, 1988), which states that the probability that any piece of information will be adopted as an input for making a choice depends on the accessibility of that input, the accessibility of the alternative inputs, and the diagnosticity or perceived relevance of the input (Herr et al., 1991; Van Hove & Lievens, 2007). This theory conveys that the use of information to make choices varies positively with the accessibility of the information. Holding constant the accessibility and diagnosticity of alternative inputs, any factor that influences the accessibility of input affects its adoption (Lynch et al., 1988). The accessibility dimension of the accessibility-diagnostics theory helps consumers to reduce the cognitive effort needed when evaluating information in the online environment. In this line, Slovic (Slovic, 1972) suggests that consumers tend to use only the information that is explicitly displayed, and they will use it in the form it is displayed because that behavior reduces the cognitive effort required to process information (Bettman et al., 1998).

In line with the cognitive effort associated to information processing Aljukhadar et al. (2012) claimed that in complex choice situations, consumers are selective in acquiring and processing product information. According to Payne (1982), humans adapt their decision-making to specific situations and environments. For instance, Shugan (1980) described them as “cognitive misers”, who strive to reduce the amount of cognitive effort associated with decision-making. One way of dealing with complex decision environments, when alternatives are numerous and difficult to compare, is to use decision support systems, which are computer’ based technologies designed to assist individuals in making a decision. Decision support systems include decision aids that perform information processing tasks, such as search in a database or sort objects by some criterion. Individuals are usually good at selecting variables that are relevant in their decision-making process, but weak at retaining large amounts of information (Häubl & Trifts, 2000). Therefore, to help consumers deal with information overload situations, online retailers usually provide decision aids, such as sorting, in their online

review system. These aids allow consumers to reduce their review processing load by choosing those online reviews they want to read, and the order of review presentation they prefer (Pang & Qiu, 2016). As claimed by Häubl and Trifts (2000), decision aids have strong favorable effects on both the quality and the efficiency of purchase decisions, since they have the potential to change the way consumers search for product information and make purchase decisions.

2.3. Conceptual Model and Hypotheses Development

Figure 1 shows the conceptual model proposed in this research. Grounded on the accessibility-diagnostics theory (Feldman & Lynch, 1988), the main objective of this research is to explore the effect of review non-textual and textual variables on product sales in three independent cases of review visibility. First, when every online review for a product is assumed to have the same probability of being viewed by consumers (traditional approach in the literature); second, when we assume that consumers sort online reviews for a product by the most helpful mechanism, and third, when we assume that consumers sort online reviews for a product by the most recent mechanism, which is the default order in which online reviews are displayed on the online retailer.

Figure 1. Conceptual Model

The Role of Review Visibility in Explaining Purchasing Behavior

The selection of these two sorting mechanisms to approach review visibility in this study is due to several reasons.

On the one hand, we explore the visibility when sorting by most helpful online reviews because literature has pointed out the influential effect of review helpfulness on consumer decision-making, and it has been considered as a sign of review quality and diagnosticity (Archak et al., 2011; Mudambi & Schuff, 2010; Racherla & Friske, 2012).

There is evidence that consumers experience the “wisdom of the crowd” effect when evaluating online reviews (Liu & Karahanna, 2015; Zhou & Guo, 2017). This effect refers **to the belief that the aggregation of many people’s judgments is a better approximation to the truth than an individual judgment.** Thus, if consumers see that many other consumers have voted a review as helpful, they might be more likely to adopt that information since they consider it as more diagnostic and reliable. In this line, Zhou and Guo (2017) and Singh et al. (2017) revealed that consumers tend to experience a social influence due to the tendency of prospective consumers to conform to previous **consumers’ opinions.** Thus, future consumers are more likely to have a better attitude to the product and to choose it if they know that other consumers have previously bought it or have rated it positively (Filiari & McLeay, 2014; Pang & Qiu, 2016). In the same line, an experiment conducted by Liu and Karahanna (2015) claim that 70 percent of **consumers in their sample sorted online reviews in Amazon.com by the “most helpful” mechanism, while 30 percent sorted them by “most recent”.** Lee, Hu, and Lu (2018) and Saumya et al. (2018) also claim that most helpful reviews are more likely to be evaluated by consumers.

On the other hand, review visibility when sorting by most recent is also explored since it is one of the most relevant factors consumers pay attention to when they evaluate reviews. Previous literature has pointed out the importance of information recency in consumer behavior. For example, Westerman et al. (2014) highlighted the relevancy of recency in explaining source credibility in online environments. In the same line, Fogg et al. (2001) found that consumers associate websites that update information more frequently with higher credibility. Other scholars, such as Levinson (2013), claim that **social networks’ hallmark is the immediacy of messages, which is one of the factors that make them more credible for consumers.** In an online reviews context, the consultancy company BrightLocal (2020) revealed that recency was the most important factor of online reviews for consumers when judging a business, ahead of factors, such as review

rating and text sentiment. The study reveals that recency was the most important factor for 58 percent of consumers and 40 percent of them said that they online evaluate those reviews that are two weeks old or less. A possible explanation is that consumers want to know up-to-date information about those businesses, products, and services they are interested in. Since they can be modified over time, consumers are interested in knowing how the business, the product, or the service performs at present.

However, the most recent sorting mechanism might not only be relevant due to the role of the date itself, but also because it is the default review sorting mechanism at the online retailer. As defined by Brown and Krishna (2004), a default can be interpreted as an option that the individual receives to the extent that he does not willingly decide on something else. Existing literature supports the idea that consumers are biased by default. For example, Johnson et al. (2002) claim that consumers consider defaults to reduce the cognitive effort required to make a decision. In this line, information processing theories reveal that many consumers usually adopt the information that is readily available to reduce the cognitive effort associated with decision-making (Häubl & Trifts, 2000; Nazlan et al., 2018). Slovic (1972) suggested that consumers tend to use only the information that is explicitly displayed, and they will use it in the form it is displayed because that behavior reduces the cognitive effort required to process information (Bettman et al., 1998). Herrmann et al. (2011) also claimed that defaults influence decision-making behavior even when consumers do not actually select the default option. Thus, review visibility when sorting by most recent, which is also the default mechanism at the online retailer explored, is likely to be an important factor in influencing consumer voting decisions.

Evaluating the Diagnosticity of Information: Online Reviews' characteristics

The diagnosticity of the information provided by online reviews can be described as the perceived ability of the information to provide consumers with relevant product information that helps them to understand and evaluate the quality and performance of the product (Lynch et al., 1988). Overall, studies have claimed that an input's diagnosticity depends on whether it enables a decision-maker to discriminate among alternatives, and this depends on the characteristics of the input of information, which is represented by online reviews in our research (Payne, 1982).

We incorporate in our research two sets of review variables to approach the diagnosticity dimension of the theory. Firstly, we include those non-textual variables that previous

literature has claimed to influence product sales: *volume*, *rating*, and *rating_inconsistency* (Chevalier & Mayzlin, 2006; Duan et al., 2008a; Gu et al., 2012; Ho-Dac et al., 2013; Lee & Choeh, 2018). The second set of variables incorporated into our study are some directly related to the review text: *analytic*, *authentic*, and *clout*, which have been quite underexplored **in the online reviews' literature**. Literature analyzing textual content of online reviews is still scarce. However, those papers studying the persuasion of online texts have concluded that messages' characteristics have a strong power over different types of consumer behavior, such as purchase intention (M. Kim & Lennon, 2007), conversion rates (Ludwig et al., 2013a), liking and commenting brand posts (De Vries et al., 2012) and social media rebroadcasting (Zhang et al., 2017).

Textual variables were extracted from online reviews using the text mining tool Linguistic Inquiry and Word Count (Pennebaker et al., 2007). Although other review textual features could have been analyzed, we decided to include the so-called summary variables by Pennebaker et al. (2015) because they represent a broader picture of what is expressed in the text. Summary variables represent a factor of other textual variables, such as the number of personal pronouns, number of adverbs, prepositions, and negations. As suggested by Ludwig et al., (2013), review text communicates specific linguistic styles that allow reviewers to express their thoughts, experiences, and opinions. This linguistic style is then a combination of two different categories of words: Lexical words, which include adjectives, nouns, verbs, and most adverbs and function words, which include prepositions, pronouns, auxiliary verbs, conjunctions, grammatical articles, or particles (Selkirk, 1996). The review style may serve as identity-descriptive **information that shapes consumers' evaluations of the review and the product** (Ludwig et al., 2013a). Social psychology and communication theories show that the way or style in which a person communicates elicits relational perceptions in the communication partner and influences consumer judgments and behaviors (Ludwig et al., 2013a; Smith & Ellsworth, 1985).

The variable *analytic* represents how well the message is organized and structured in the review. As claimed by Areni (2003), constructing compelling arguments have to do with providing statements to support a given set of claims. Structural elements in verbal arguments are joined with connectives, words, or short phrases that link the propositions comprising an argument (Areni, 2003). These connectives might enhance the comprehension of arguments because they imply the conceptual relationship between the data and claim (Munch & Swasy, 1988). In the consumer behavior field, it has been shown that those messages with a more thorough argument structure have a stronger

positive impact on consumer beliefs and message acceptance (Areni, 2003; Payan & McFarland, 2005).

The variable *authentic* represents the level of subjectivity shown in the text. Earlier scholars in the marketing field have studied how objectivity influences the attitude towards advertising or other promotional communication (Darley & Smith, 1993; Holbrook, 1978). For example, Holbrook (1978) revealed that objective claims are perceived as more credible than subjective claims, and therefore, the message acceptance is higher and also the attitude towards the brand and the buying intentions. Darley and Smith (1993) also stated that objective claims are more effective than subjective claims in both print and radio media. In the context of online reviews, some scholars have explored how subjectivity influences the helpfulness of online reviews, which is represented by the number of helpful votes received by online reviews. However, there is not a consensus in the direction of the effects. Some of them have found that subjectivity positively influences review helpfulness (Chen & Tseng, 2011), others have claimed that those reviews containing a mixture of subjective and objective elements are more helpful (Archak et al., 2011; Ghose et al., 2012), while others did not find a relationship between subjectivity and review helpfulness (Liu et al., 2007).

The last textual variable incorporated in our research is *clout*, which represents the level of self-confidence shown by the reviewer in the review text (Pennebaker et al., 2015). In the psychology literature, the level of confidence of the advisor has been found to be important in reducing consumer uncertainty, especially in online settings (Snizek & Van Swol, 2001). **Confidence is defined as “the strength with which a person believes that a specific statement, opinion, or decision is the best possible”** (Snizek & Van Swol, 2001). The Judge-Advisor System paradigm (Snizek & Van Swol, 2001) reveals that high advisor confidence can act as a cue to expertise and can influence the judge to accept the advice. For example, Price and Stone (2004) revealed that when financial advisors expressed high confidence about stock forecasts, they were perceived as more knowledgeable and were more frequently chosen.

As far as we know, there are no studies exploring the effect of these textual variables, *analytic*, *authentic*, and *clout*, on product sales, but they have been proved to predict different types of outcomes in other fields, such as academic success and deception (Kacewicz et al., 2014; Newman et al., 2003; Pennebaker et al., 2014). Thus, we are interested in exploring how the selected textual variables influence product sales in each case of review visibility.

Hypotheses

Overall, previous literature has considered every review to have the same influence on consumer purchase decisions, but based on the accessibility-diagnostics theory (Feldman & Lynch, 1988), we posit that those reviews more accessible or visible for consumers are likely to be even more influential in consumer decision-making. Therefore, we expect that first, non-textual and textual features of online reviews influence product sales when considering different review visibility cases (and not only when every review is considered to have the same influence on consumer purchase decisions), and second, that the impact of non-textual and textual features of online reviews on product sales might be different depending on the review visibility case considered, since, for example, the characteristics of most helpful online reviews might be different to those of most recent online reviews. Thus, different sets of online reviews might have different effects on consumer purchase decisions.

Hence, we hypothesize as follows:

H1a. Review non-textual features influence product sales considering different cases of review visibility.

H1b. Review textual features influence product sales considering different cases of review visibility.

H2a. The influence of review non-textual features on product sales is different depending on the review visibility case considered.

H2b. The influence of review textual features on product sales is different depending on the review visibility case considered.

3. Methodology ---

3.1. Data

To carry out our research, we collected online consumer reviews from the product category of blush from a popular US cosmetics retailer website, which was placed in the top-50 shopping sites in the US in March 2017 according to Alexa.com. The data was obtained using web-scraping, so a robot was designed to collect the data of interest from the online retailer website. Using web-scraping, the data was stored in a structured format in Excel spreadsheets. Then, the databases were imported to R to conduct the

empirical analysis. Figure 2 shows an example of the review information collected from the online retailer for each product. In addition to review-related information, other product information was gathered from the online retailer: Brand name, product price, product size, product bestselling ranking, if the product was labeled as “new” and if the product was labeled as “exclusive”. Brand-related information was gathered from external sources: Brand number of followers on Instagram (*Social Blade*, 2017), brand market share (Euromonitor International, 2017), and if the brand is in the category of premium brands (Euromonitor International, 2017). Variables are described in Table 1.

Figure 2. Example of review information collected from the online retailer

To carry out the empirical analysis, we gathered data in a weekly basis over nine weeks between 21 December 2016 and 17 February 2017. First, we decided to select a relatively small period to ensure that environmental and market factors did not change too much, which allowed us to control for endogeneity, as well as possible in our empirical models. Second, the online retailer made some slight changes in the website design from the 17 February 2017 onwards, so our web-scraping robot was able to collect complete data until those changes were made. To make sure that a nine-week period was adequate for our empirical research, we consulted some econometric professors, who found the period

appropriate for our analyses. Only those products available at the online retailer each of the nine weeks were used to build the panel, resulting in a balanced panel of 119 products and 1,071 observations. On each date, we collected between 63,000 and 66,000 online reviews for the whole blush category (cumulative number of reviews of each product at each date), and we had two levels of information: Review-level information and product-level information. Since we were working with a panel of products, review information had to be aggregated to product-level variables. To aggregate review information, we considered the three review visibility cases, shown in the conceptual model in Figure 1.

In the cosmetics industry, sales usually show a seasonality pattern. As revealed by Nielsen (Nielsen, 2016), some categories of cosmetics, such as perfumes and sun cream, are very seasonal. However, blush can be considered a low seasonal cosmetics category, since these products are usually bought for personal and regular use over the year (Nielsen, 2016). Therefore, products in our panel are less likely to be influenced by seasonality patterns not recorded in our database.

Table 1

Research variables.

	Definition
<i>Dependent variable</i>	
<i>Ln_sales_rank_inverse_{it}</i>	The natural Log of the multiplicative inverse of the sales rank of product <i>i</i> at time <i>t</i> ($1/\text{sales rank}_{it}$)
<i>Independent variables</i>	
<i>Review non-textual variables</i>	
<i>Ln_volume_{it}</i>	The natural Log of the cumulative number of online consumer reviews for product <i>i</i> at time <i>t</i>
<i>Ln_rating_{ivt}</i>	The natural Log of the average of ratings for product <i>i</i> at time <i>t</i> considering review visibility case <i>v</i>
<i>Ln_rating_inconsistency_{ivt}</i>	The natural Log of the average difference in absolute value between review rating and product average rating for product <i>i</i> at time <i>t</i> considering review visibility case <i>v</i>
<i>Review textual variables</i>	
<i>Ln_analytic</i>	The natural Log of the average of analytical thinking shown in online reviews for product <i>i</i> at time <i>t</i> considering review visibility case <i>v</i> The variable captures the degree to which consumers use words that suggest formal, logical and hierarchical thinking patterns (Pennebaker et al., 2014). It is extracted using the text mining tool LIWC (Pennebaker et al., 2015).
<i>Ln_authentic</i>	The natural Log of the average of authenticity shown in online reviews for product <i>i</i> at time <i>t</i> considering review visibility case <i>v</i> The variable captures the degree to which consumers reveal themselves in an authentic or honest way, so their discourse is more personal and humble (Newman et al., 2003). It is extracted using the text mining tool LIWC (Pennebaker et al., 2015).
<i>Ln_clout</i>	The natural Log of the average of clout shown in online reviews for product <i>i</i> at time <i>t</i> considering review visibility case <i>v</i> The variable captures the relative social status, confidence or leadership displayed by consumers through their writing style

(Kacewicz et al., 2014). It is extracted using the text mining tool LIWC (Pennebaker et al., 2015).

Control variables

<i>Christmas_t</i>	Binary variable: 1 if it is between 21 st December and 5 th January; 0 otherwise.
<i>New_{it}</i>	Binary variable: 1 if product <i>i</i> at time <i>t</i> had the label of “new” ; 0 otherwise
<i>Exclusive_{it}</i>	Binary variable: 1 if product <i>i</i> had the exclusive label at time <i>t</i> ; 0 otherwise
<i>Ln_price_{it}</i>	The natural Log of the of the price per gram of product <i>i</i> at time <i>t</i>
<i>Ln_size_{it}</i>	The natural Log of the of the size in gr of product <i>i</i> at time <i>t</i>
<i>Brand_retailer_i</i>	Binary variable: 1 if product <i>i</i> 's brand belongs to the retailer private brand; 0 otherwise
<i>Ln_brand_followers_{it}</i>	The natural Log of the cumulative number of brand Instagram followers for product <i>i</i> at time <i>t</i> . Data collected from www.socialblade.com (Social Blade, 2017)
<i>Brand_top_i</i>	Binary variable: 1 if product <i>i</i> 's brand was in the top-10 bestselling brands in the US in 2016; 0 otherwise. Data from Euromonitor International (2017)
<i>Band_premium_i</i>	Binary variable: 1 if the product <i>i</i> 's brand is was categorized as premium brand in 2016; 0 otherwise. Data from Euromonitor International (2017)

3.2. Research Variables

Table 1 shows a description of the research variables considered in this study. We have two sets of explanatory variables. First, independent variables, which are directly related to the diagnosticity of online reviews, such as *ln_rating* and *ln_analytic*. Second, control variables, including product features, such as *ln_price* and *ln_size*, and brand features, such as *ln_brand_followers*. Following extant research (Elberse & Eliashberg, 2003; Y. Liu, 2006), we log-transformed every non-binary variable to smooth the distribution of the variables in the regression and to avoid distorting the estimation by outliers. In this way, estimated coefficients directly represent the elasticity of the variables. In the case of those variables that have zero values, we log-transformed the variables after adding the value of one.

Dependent Variable

The dependent variable in our study is the multiplicative inverse of the sales rank of the product, which is a proxy of product sales. We do not have information about actual product sales, but the online retailer shows a sales rank for each product category, which represents a snapshot of sales (units of product sold) for up to a week. The product sales rank is inversely related to its sales, which means that the first product in the sales rank in a specific product category is the one with the highest sales (in units) during the previous week. On the other hand, high sales ranks values represent lower sales.

According to Chevalier and Goolsbee (2003), the relationship between the actual volume of sales and the sales rank in Amazon.com is $\ln(\text{Sales}) = \beta_0 - \beta_1 * \ln(\text{sales_rank})$, which makes the relationship between $\ln(\text{sales})$ and $\ln(\text{sales_rank})$ approximately linear. Since sales rank is a log-linear function of sales with a negative slope, we adopt $\ln_sales_rank_inverse$ as our dependent variable.

Independent Variables

Every review non-textual and textual variable $\ln_rating$, $\ln_rating_inconsistency$, $\ln_analytic$, $\ln_authentic$, and $\ln_clout$ was aggregated to a product level (i) on each specific date (t) depending on the review visibility case (v) to estimate the model in Equation (2). The formula followed to aggregate review variables for each product is as follows:

$$X_{ivt} = \frac{X_{rt} * w_{rvt}}{\sum w_{rvt}} \quad (1)$$

In Equation (1), X_{ivt} is the product-aggregate variable, X_{rt} is the review-level variable to be aggregated, $r=1, \dots, R$ are the reviews of the product i and w is the review visibility weight based on the review visibility case v : (v_1) we assume all reviews have the same probability of being viewed, (v_2) we assume consumers sort online reviews by most helpful, so most helpful reviews are more likely to be viewed, (v_3) we assume consumers sort online reviews by most recent, so more recent reviews are more likely to be viewed. $Volume_{it}$ is already an aggregate variable that is not influenced by the review visibility case, since it captures the cumulative number of online reviews for each product at each date. The variable $\ln_volume$ is not affected by the review visibility case because it represents the cumulative number of reviews for each product (i) at each date (t).

In case 1, variables were aggregated in the same way as previous literature does, by giving each online review the same probability of being viewed and therefore, the same relative weight when aggregating them at a product level. For example, the product average rating resulting from case 1 is the same as the one provided by the online retailer, since it is the average of every individual review rating for each product. In case 2, review information was aggregated considering the rank order of each individual online review when sorting reviews for each product by the criterion of most helpful. Finally, in case 3, review information was aggregated considering the rank order of each individual online review for each product according to the most recent criterion, which was the predetermined sorting criterion used by the online retailer when data was collected.

In cases 2 and 3, we incorporate the effect of review visibility, which captures the rank order of online reviews when sorting by most helpful and by most recent, respectively. To compute the rank order of online reviews at each case, the approach proposed by Godes and Silva (D. Godes & Silva, 2012) was followed. For example, the following formula was applied to build the rank order of online reviews when sorting by the most **recent criterion**. Let's d' represent the publication date of review r . For each d' , it was formed $S_{d'} \equiv \{r: d_r = d'\}$, which represents the set of reviews for which $d_r = d'$. Then, the variable order was operationalized as $Order(d') \equiv \sum_{d < d'} N(S_d) + 1$, where $N(S_d)$ is the cardinality of set S_d . This method assigns the same order to every review with the same publication date. For the rest of the reviews, the order is always 1 plus the number of reviews with more recent publication dates (D. Godes & Silva, 2012). The same process was followed to order reviews when sorting by most helpful. In this case, for those reviews of the same product sharing the same number of helpful votes, the most recent publication date was the second ordering mechanism at the website, so it was the second ordering criterion used to build the variable *Order*.

In review visibility cases 2 (v_2) and 3 (v_3), we considered two approaches to build the aggregated review variables. In the first approach ($v_{2.1}$ and $v_{3.1}$), we assumed that online reviews have a decreasing probability of being viewed by consumers based on each rank order, order by most helpful in case 2, and order by most recent in case 3. In this first approach, the review visibility weight was operationalized as: $w = (1/Order) * 100$. In the second approach ($v_{2.2}$ and $v_{3.2}$), we assumed that consumers just read the first five online reviews when sorting by each criterion, because five is the number of online reviews displayed on the first page of the studied cosmetics online retailer when sorting by each criterion. In this case, a weight (w) of 1 was assigned to each of the five first reviews, while the rest of the reviews were given a weight (w) of 0.

Table 2 reports an example to illustrate how we built the *ln_rating* variable for a product that has a total of 10 online reviews when considering review visibility case 1 (v_1), where all reviews have the same probability of being viewed, and review visibility case 2 (v_2), where most helpful visibility is considered. The example shows the two different weighting approaches used (all the reviews have a decreasing probability of being viewed; only the top five most helpful reviews are viewed). The same process was followed to aggregate review variables in review visibility case v_3 (most recent visibility considered). As shown in Table 2, the final product-aggregated variable *ln_rating* is slightly different at each review visibility case.

Table 2

Example of aggregation process of review variables for a specific product I at a specific time t .

		Case 1		Case 2	
		Approach 1 (v_{1j})		Approach 1 ($v_{2,1}$)	Approach 2 ($v_{2,2}$)
Product (i)	Review Rating	All Reviews the Same Probability of Being Viewed	Review Visibility When Sorting by Most Helpful (Rank Order)	Review Visibility Weight (w) All Reviews Have a Decreasing Probability of Being Viewed When Sorting by Most Helpful (1/Review Rank Order)	Review Visibility Weight (w) Only Reviews in the First Page (top 5) are Viewed When Sorting by Most Helpful are Viewed
1	5	1	1	1	1
	4	1	2	0.5	1
	3	1	3	0.33	1
	5	1	4	0.25	1
	4	1	5	0.2	1
	3	1	6	0.16	0
	4	1	7	0.14	0
	5	1	8	0.12	0
	1	1	9	0.11	0
	2	1	10	0.1	0
Sum probabilities	of 10			2.93	5
Rating		Rating $_{v1}$ = 3.6 $5 * 1 + 4 * 1 + \dots +$		Rating $_{v2,1}$ = 4.12 $5 * 1 + 4 * 0.5 + \dots + 2$	Rating $_{v2,2}$ = 4.2 $5 * 1 + 4 * 1 + \dots + 1$
		10		2.93	5
ln_rating		Ln 3.6 = 1.28		Ln 4.12 = 1.42	Ln 4.2 = 1.44

Control Variables

Following previous literature, we incorporate to our empirical analysis some variables not related to online reviews to control for the possible effect of products, brands, and time features on product sales (Chintagunta et al., 2010; Duan et al., 2008b; Hofmann et al., 2017; Zhu & Zhang, 2010). To control for some time features, we include the variable Christmas, which is a dummy variable that captures if the collection date is within Christmas. In this way, we want to capture any effect that Christmas might have on specific products of the category.

In terms of product characteristics, we use information about every product feature provided by the online retailer: Price, size, if the product was labeled as “new” and if it was labeled as “exclusive”. The incorporation of product attributes as controls in our model is based on Decker and Trusov (2010), who studied consumer preferences from online reviews incorporating the effect of product attributes. In our research, product attributes are adapted to the cosmetics category. The “new” label at the online retailer

means either that the product has recently been added to the shopping category or that **a new color has been launched for the product**. The “exclusive” label means that the product is exclusively sold at the online retailer and at the brand website itself, but consumers cannot buy it at other retailers. We also include the variable *ln_price*, since the price has been considered as the top attraction for online shoppers (Park et al., 2012). The variable *ln_size* records the size in grams of the product, and it might be an important product feature in our cosmetics category.

Brands also play an important role in a cosmetics scenario, so we incorporate several variables **that control for brands’ characteristics** (Ho-Dac et al., 2013). The variable *brand_retailer* is a dummy that captures if the product is from the retailer’s private brand. The variable *ln_brand_followers* records the cumulative number of followers of the brand at Instagram on each date. The number of followers was collected from Socialblade.com. The variable *brand_top* is a dummy that records if the brand was in the top-10 bestselling brands in the facial make-up category in the US in 2016 (Euromonitor International, 2017). This variable was built to differentiate between strong and weak brands in the whole US market in terms of annual sales in the US in 2016 (Euromonitor International, 2017). The variable *brand_premium* is a dummy that **records if the brand was in the “premium” segment of color cosmetics in the US in 2016** (Euromonitor International, 2017). We included this variable because the annual report from Euromonitor International (Euromonitor International, 2017) revealed that the sales of premium color cosmetics brands increased a lot in 2016 with respect to those of mass brands.

3.3. Empirical Model and Estimation

We model the sales equation as follows:

$$\begin{aligned}
 \ln_sales_rank_inverse_{it} &= \alpha_0 + \alpha_1 \ln_sales_rank_inverse_{i,t-1} + \alpha_2 \ln_volume_{it} + \alpha_3 \ln_rating_{ivt} \\
 &+ \alpha_4 \ln_rating_inconsistency_{ivt} + \alpha_5 \ln_rating_{ivt} * \ln_rating_inconsistency_{ivt} \\
 &+ \alpha_6 \ln_analytic_{ivt} + \alpha_7 \ln_authentic_{ivt} + \alpha_8 \ln_clout_{ivt} + \alpha_9 \text{Christmas}_t \\
 &+ \alpha_{10} \text{new}_{it} + \alpha_{11} \text{exclusive}_{it} + \alpha_{12} \ln_price_{it} + \alpha_{13} \ln_size_{it} + \alpha_{14} \text{brand_retailer}_i \\
 &+ \alpha_{15} \ln_brand_followers_{it} + \alpha_{16} \text{top_brand}_i + \alpha_{17} \text{brand_premium}_i + \varepsilon_{it}
 \end{aligned} \quad (2)$$

In Equation (2), *i* represents the product, *t* represents the time, and *v* is the review visibility case. The model is independently estimated for each of the review visibility cases. Review variables (*ln_rating*, *ln_rating_inconsistency*, *ln_analytic*, *ln_authentic*,

and *In_clout*) differ among review visibility cases, since their aggregation to product-level variables depend on the weighting given to each review based on its visibility. However, neither *In_volume* nor control variables (*new*, *exclusive*, *In_price*, *In_size*, *brand_retailer*, *In_brand_followers*, *top_brand*, and *brand_premium*) depend on review visibility, since these variables are already at a product-level.

As revealed in previous literature, endogeneity should be considered when exploring the influence of online reviews on product sales because not accounting for it could bias the results (Chintagunta et al., 2010; Xu & Liu, 2019; Yazdani et al., 2018). As in previous papers, endogeneity is an issue in our study, due to several reasons. First, there might be reverse causality between the volume of online reviews and the sales rank of a product, which is our dependent variable. *Volume* is a variable that represents the interest generated by a product, and it has usually been proved to impact product sales. However, product sales might also impact the number of reviews that products receive, since, as claimed by Hennig-Thurau et al. (2017), **“success breeds success”**. Another important source of endogeneity in our model is the presence of unobserved variables associated with the product and the environment, that can make the regressor to be correlated with the error structure. For example, product promotion strategies are not contemplated in our data and could influence the sales of the product on specific dates. Although we include some control variables trying to account for some product and environmental factors, there might be other unobserved ones that could bias our estimations. Another important issue in our model is the dynamic component of the dependent variable since past sales **ranks of the product might influence the current sales rank**. Again, **“success breeds success” and being in top positions in the bestsellers list might lead to continuing** in those top positions, due to a social influence effect (Cheng & Ho, 2015; Liu & Karahanna, 2015; Marchand et al., 2017).

Considering the panel structure of our data, to account for the dynamic effect of the dependent variable and to be able to correct for endogeneity, we estimate the model using panel data methodology, specifically the system generalized method of moments (system GMM) estimator, pioneered by Arellano and Bover (1995) and Blundell and Bond (1998).

The system GMM estimator has some advantages over other estimators, such as the ordinary least square (OLS) estimator (Lozano et al., 2016; Pindado et al., 2011). First, it allows us to control for the individual effect or unobserved heterogeneity, such as the product quality, which might influence the sales of products. By modeling it as individual effects, η_i , we can control this heterogeneity in products to avoid biased results. In this

line, the error term in our model, ϵ_{it} , is divided into three components: The individual effect, η_i ; the time dummies, d_t , which allow us to control for the effect of macroeconomic variables on product sales; and the random disturbance, v_i . Besides, the system GMM estimator aids to reduce the endogeneity problem. Endogeneity implies that the error term is correlated with some of the explanatory variables, and this correlation violates one of the main assumptions of OLS estimator. This correlation usually occurs, due to two reasons: (1) When important variables are omitted from the model, also called **“omitted variable bias”** and (2) when the dependent variable is a predictor of the explanatory variable and not only a response to it, referred to as **“simultaneity bias”** or **“reverse causality”**.

As happens in many studies, many of the explanatory variables may suffer from the endogeneity problem. To deal with this problem, Instrumental Variables (IV) models, such as the Two Stage Least Squares (2SLS) and Three Stage Least Squares (3SLS), have been widely used in previous literature (Duan et al., 2008a; Elberse & Eliashberg, 2003; Forman et al., 2008). However, finding instrumental variables that meet the two conditions required for instruments is very difficult, since they should be correlated with the endogenous explanatory variable, but uncorrelated with the error term of the model. To solve the issue, the GMM estimator provides the solution of using the lagged values of the explanatory variables as instruments for the endogenous variables, since these lags are highly correlated with the regressors that they instrument.

Two different GMM estimators can be used, the difference GMM (Arellano & Bond, 1991) and the system GMM (Arellano & Bover, 1995; Blundell & Bond, 1998). However, the difference GMM suffers the problem of weak instruments, so we use in this research the system GMM, which overcomes that problem. To employ the system GMM procedure, we should indicate those explanatory variables that are likely to be endogenous in our model. We have considered that every review variable, *ln_volume*, *ln_rating*, *ln_rating_inconsistency*, *ln_analytic*, *ln_authentic*, *ln_clout*, and the variable *ln_price*, might be endogenous in our model, because they might suffer either from **“omitted variable bias”** or from **“simultaneity bias”**. The rest of the variables are treated as exogenous, some of them are specific characteristics of the product and the brand collected from the online retailer website (*new*, *exclusive*, *ln_size*, and *brand_retailer*), and others are brand-specific features collected from external sources (*brand_top* and *brand_premium*). In the system GMM model, we estimate two equations: Equation in differences, in which the instruments are the right-hand-side variables in levels, and equation in levels, where the instruments are the right-hand-side variables in differences. To estimate the system GMM model, we used the package *xtabond2* in Stata,

following Roodman (2009a). We transformed to logarithms all the non-binary variables to avoid distorting the estimation by outliers (Marchand et al., 2017). Besides, all the non-binary variables were standardized to reduce the multicollinearity that might arise in a model with interaction terms (Aiken & West, 1991).

4. Results

4.1. Descriptive Statistics

Table 3 reports the descriptive statistics of the variables used in the research. For a better interpretation, we use the original variables instead of the log-transformed variables. As far as review aggregated variables are concerned, we show in the table the descriptive statistics for each of the review visibility cases. We can observe that descriptive statistics change depending on the review visibility case.

Table 3

Descriptive statistics

Variable	N	Mean	SD	Min	Max
Sales_rank_inverse	1,062	68.13	39.01	1	146
Volume	1,062	523.18	1604.49	1	16404
Rating _{v1}	1,062	4.29	0.38	2.9	5
Rating _{v2.1}	1,062	4.29	0.49	1.91	5
Rating _{v2.2}	1,062	4.39	0.66	1.6	5
Rating _{v3.1}	1,062	4.17	0.49	2.4	5
Rating _{v3.2}	1,062	4.18	0.66	2	5
Rating_inconsistency _{v1}	1,062	0.02	0.02	0	0.34
Rating_inconsistency _{v2.1}	1,062	0.23	0.21	0	1.06
Rating_inconsistency _{v2.2}	1,062	0.41	0.37	0	2.5
Rating_inconsistency _{v3.1}	1,062	0.22	0.2	0	1.06
Rating_inconsistency _{v3.2}	1,062	0.41	0.36	0	2
Analytic _{v1}	1,062	46.16	6.02	11	70.86
Analytic _{v2.1}	1,062	47.06	8.43	11	70.54
Analytic _{v2.2}	1,062	49.52	10.94	11	72.72
Analytic _{v3.1}	1,062	44.44	8.86	11	75.47
Analytic _{v3.2}	1,062	44.02	12.10	10.02	75.33
Authentic _{v1}	1,062	49.96	7.71	27.39	73.34
Authentic _{v2.1}	1,062	49.54	10.48	19.22	76.78
Authentic _{v2.2}	1,062	49.18	15.38	2.24	80.39
Authentic _{v3.1}	1,062	48.69	11.43	11.94	79.90
Authentic _{v3.2}	1,062	49.52	15.15	7.40	92.01
Clout _{v1}	1,062	27.14	5.66	8.65	64.45
Clout _{v2.1}	1,062	26.50	6.56	5.90	52.76
Clout _{v2.2}	1,062	27.29	9.77	6.64	64.45
Clout _{v3.1}	1,062	26.64	7.56	7.13	72.27
Clout _{v3.2}	1,062	26.59	10.92	2.33	64.45
Christmas	1,062	0.33	0.47	0	1
New	1,062	0.08	0.27	0	1
Exclusive	1,062	0.27	0.44	0	1
Price	1,062	5.63	3.87	0.35	26.25
Size	1,062	8.8	8.22	0.8	57

Brand_retailer	1,062	0.07	0.25	0	1
Brand_followers	1,062	3,512,608	3,649,264	2,837	14,000,000
Brand_top	1,062	0.09	0.29	0	1
Brand_premium	1,062	0.32	0.47	0	1

4.2. Model Findings

Table 4 shows the output of the system GMM regression. Five models are presented depending on the case of review visibility assumed and on the weighting approach followed.

Table 4

Output of system GMM regression.

	Case 1 (v1) No Visibility Considered	Case 2 (v2) Most Helpful Visibility	Case 3 (v3) Most Recent Visibility		
	Model 1	Model 2	Model 3	Model 4	Model 5
	Case v1	Case v2.1	Case v2.2	Case v3.1	Case v3.2
	All Reviews Same Probability of Being Viewed	All Reviews Decreasing Probability	Five Most Helpful Reviews	All Reviews Decreasing Probability	Five Most Recent Reviews
L1_In_sales_rank_inverse	0.919 *** (0.00)	0.889 *** (0.00)	0.767 *** (0.00)	0.937 *** (0.00)	0.942 *** (0.00)
Ln_volume	0.072 *** (0.01)	0.064 *** (0.01)	0.093 *** (0.01)	0.134 *** (0.01)	0.075 *** (0.01)
Ln_ratingmi	0.116 *** (0.00)	0.256 *** (0.01)	0.578 *** (0.01)	0.108 *** (0.01)	0.274 *** (0.01)
Ln_rating_inconsistency	0.506 *** (0.02)	1.012 *** (0.03)	1.949 *** (0.01)	0.369 *** (0.03)	0.552 *** (0.04)
Ln_rating x Ln_rating_inconsistency	-0.476 *** (0.02)	-0.953 *** (0.03)	-1.872 *** (0.01)	-0.312 *** (0.03)	-0.423 *** (0.03)
Ln_Analytic	-0.016 *** (0.00)	0.031 *** (0.00)	0.008 * (0.00)	-0.048 *** (0.00)	-0.012 ** (0.00)
Ln_Authentic	0.035 *** (0.00)	-0.070 *** (0.01)	-0.077 *** (0.01)	-0.042 *** (0.01)	-0.020 ** (0.01)
Ln_Clout	0.030 *** (0.01)	-0.077 *** (0.00)	-0.105 *** (0.00)	-0.013 ** (0.00)	-0.034 *** (0.01)
Christmas	-0.003 (0.00)	0.003 (0.00)	0.015 *** (0.00)	0.008 (0.00)	0.005 (0.00)
New	0.100 *** (0.01)	0.105 *** (0.01)	0.352 *** (0.01)	0.017 * (0.01)	0.092 *** (0.01)
Exclusive	0.163 *** (0.01)	0.218 *** (0.01)	0.280 *** (0.01)	0.145 *** (0.01)	0.186 *** (0.02)
Ln_price	0.321 *** (0.01)	0.294 *** (0.01)	0.060 *** (0.01)	0.133 *** (0.01)	0.033 * (0.01)
Ln_size	0.273 *** (0.01)	0.256 *** (0.02)	0.061 *** (0.01)	0.108 *** (0.01)	0.045 *** (0.01)
Brand_retailer	0.309 *** (0.02)	0.293 *** (0.03)	0.255 *** (0.04)	0.159 *** (0.03)	0.072 * (0.03)
Brand_followers	-0.053 *** (0.00)	-0.027 *** (0.00)	-0.053 *** (0.01)	-0.068 *** (0.01)	-0.057 *** (0.01)

Top_brand	0.057 ** (0.02)	0.059 (0.04)	-0.011 (0.03)	0.048 (0.02)	0.046 * (0.02)
Brand_premium	0.013 (0.01)	0.036 ** (0.01)	0.147 *** (0.02)	0.007 (0.01)	0.058 *** (0.02)
Constant	-0.078 *** (0.01)	-0.353 *** (0.02)	-0.200 *** (0.02)	-0.070 *** (0.01)	-0.107 *** (0.01)
Observations	944	944	944	944	944
z1	4.1e + 06 (17)	4.1e + 06 (17)	5.6e + 07 (17)	1.7e + 06 (17)	7.6e + 05 (17)
z2	63.30 (6)	47.90 (6)	109.89 (6)	133.99 (6)	47.25 (6)
Hansen	70.19, $p = 0.666$	79.70, $p = 0.364$	82.28, $p = 0.291$	86.86, $p = 0.185$	84.60, $p = 0.234$
AR (2)	0.66, $p = 0.512$	0.65, $p = 0.518$	0.40, $p = 0.687$	0.64, $p = 0.522$	0.60, $p = 0.548$

Notes: All time dummies are included, but not reported in the table to save space. The system GMM estimator is the two-step estimator. Robust standard errors are shown in parenthesis. Following Roodman (Roodman, 2009a), the instrument matrix is collapsed, and we use two lags of the explanatory variables as instruments in the equation in differences and one lag in the equation in levels. Windmeijer correction is not applied to standard errors, so they could be downwards biased. It is not applied because we have a relatively small sample of products, although they represent the complete set of products available at the online retailer in the blush category. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

We observe that review non-textual and textual variables are significant in every model: *ln_volume*, *ln_rating*, *ln_rating_inconsistency*, *ln_analytic*, *ln_authentic*, and *ln_clout*. The interaction term *ln_rating*ln_rating_inconsistency* is also significant in every model, so *H1a* and *H1b* are supported. Therefore, both review non-textual and textual features influence product sales not only in the traditional case of review visibility (case 1), where every review is assumed to have the same probability of being viewed, but also in the rest of cases (cases 2 and 3), where we assume that consumers sort online reviews either by the most helpful mechanism (case 2) or by the most recent mechanism (case 3).

To test *H2a* and *H2b*, we should look at the possible differences between review variables coefficients among models. To graphically show the results from models 1 to 5, we represent in Figures 3 and 4 the review non-textual and textual variables coefficients, respectively, in the different models. As far as *H2a* is concerned, we observe that the coefficient sign of review non-textual variables is the same in every model, while the magnitude changes among models. Therefore, we can support *H2a* because we observe differences among review visibility cases. *H2b* is also supported, since the coefficient magnitude, and even the sign, of review textual variables differs among review visibility cases. In fact, we find bigger differences between review visibility cases when dealing with review textual variables.

Figure 3. Non-textual product variables coefficients in system GMM models

Figure 4. Textual product variables coefficients in system GMM models

Differences between review visibility cases have been further explored.

First, we observe that when we compare between the two approaches of review visibility within the same case (Model 2 vs. Model 3, and Model 4 vs. Model 5), the approach where we assume that consumers view either the top five most helpful or the top five most recent online reviews (Model 3 and Model 5, respectively) has greater review variable coefficients than those in the approach where we assume consumers view every online review in a decreasing order when sorting either by most helpful or by most recent (Model 2 and Model 4, respectively). These findings might suggest that review features of the top five ranked online reviews (either top five most helpful or top five most recent) have a stronger influence on consumer purchase decisions.

Second, we also notice that review variables coefficients are higher in case 2 than in case 3. Therefore, this might indicate that information contained in most helpful online reviews is likely to have a greater impact on consumer purchasing behavior than the information in most recent online reviews. A possible explanation is that consumers **might experience a “wisdom of the crowd” effect when they evaluate the most helpful** online reviews (Liu & Karahanna, 2015). This effect refers to the fact that consumers might believe that, since many other consumers have voted the information contained in those reviews as helpful, that information about the product might be a better approximation to the truth, so consumers are more likely to rely on it when making a purchase. Moreover, if we look at case 2, we observe that coefficients are bigger in Model 3 than in Model 2, which might indicate that those top five online reviews have a strong **influence on consumers’ purchase behavior. This influence is greater than if** we consider every individual online review with its corresponding visibility probability, represented by Model 2. Therefore, these findings might not only suggest that most helpful reviews are more influential than most recent reviews, but also that those online reviews placed on the first page of online reviews of each product have even a greater impact in **consumers’ purchase behavior.**

Overall, we could say that if we just considered Model 1, in which we assume all reviews have the same probability of being viewed (approach traditionally used in previous literature), we could get to misleading conclusions because the strength, and even the sign, of some effects, is not the same as it is in the other review visibility cases. For example, the coefficient of *ln_rating* is $\delta = 0.116$ in Model 1, while it is $\delta = 0.578$ in Model 3. Therefore, we observe that the product average rating has a higher impact when we assume that consumers read the top five most helpful reviews of each product. In other words, this might indicate that the impact of the average rating of the five most helpful

online reviews is greater than the impact of the overall product average rating of the product. In this line, another pattern we observe is that the effect of the review non-textual variables, *ln_rating* and *ln_rating_inconsistency*, and also the interaction term *ln_rating x ln_rating_in* is greater (they have a higher coefficient) when we assume that consumers evaluate either the top five most helpful reviews ($v_{2.2}$) or the top five most recent reviews ($v_{3.2}$), than when we assume that consumers evaluate every online review following either the most helpful ($v_{2.1}$) or the most recent rank order ($v_{3.1}$). However, we cannot observe this pattern in the case of review textual variables.

Concerning the variable *L1_ln_sales_rank_inverse*, it is positive and significant in every model, which means that the bestselling rank of the previous week positively impacts the bestselling rank of the current week. This confirms the dynamic behavior of the dependent variable in our model. Besides, this finding is even more relevant in our context, where consumers are likely to be influenced by a social influence effect when they are choosing between products within a category. Since consumers believe that many people have bought those products in top positions in the bestselling list, they are likely to continue buying those products, due to the social influence effect. The variable *ln_volume* is always positive and significant, so the higher number of online reviews of a product, the more likely the product is in top positions of the bestselling rank. Coefficients for *L1_ln_sales_rank_inverse* and *ln_volume* are quite steady amongst models, so it might indicate that the effect of those variables on the sales rank does not depend much on the different cases of review visibility.

Ln_rating is also positive and significant in each model. Therefore, the better the product average rating, the better the bestselling position of the product. This means that regardless of the case of review visibility, the average rating always has a positive impact on the bestselling rank. However, we observe bigger differences in terms of coefficients magnitude. *Ln_rating* has a stronger impact when it is built considering the most helpful visibility of online reviews (case 2). Therefore, the higher the average rating of most helpful online reviews, the stronger the positive effect of *ln_rating* on the bestselling rank. It means that when the average rating of those reviews in top positions when sorting by the most helpful criterion is high, it has a greater positive impact on the bestselling ranking. This finding makes sense because it implies that those online reviews in top positions by the most helpful ranking are not only positive (high stars rating), but also, they have been voted as helpful by other consumers, which means that many other consumers have found the information provided by the review useful or diagnostic. On the other hand, the effect of *ln_rating* when considering review visibility by most recent ($v_{3.1}$ and $v_{3.2}$) is also significant, but it is smaller than in case 2. Thus, the product average

rating of the most recent online reviews also has a positive effect on the bestselling rank, but the effect is smaller than the one of the most helpful reviews. A possible explanation is that the date itself does not provide any extra information for consumers about the usefulness or diagnosticity of online reviews—it just means that the review has been recently published. However, the number of helpful votes is, by itself, rich information provided by online reviews.

The effect of *ln_rating_inconsistency* is positive and significant in every model. It means that the higher the difference between each individual review rating and the product average rating, the better the impact on the bestselling rank. Thus, it might be good for products to have online reviews whose ratings are different from the product average rating. **This might indicate that those products that have more “extreme” online reviews,** are more likely to be in better bestselling positions. A possible reason is that, since most online reviews at the online retailer are very positive (5-star online reviews), it is good for the product to have also negative online reviews. In this way, consumers can know both the positive and negative features of the product. Being aware of both the positive and negative information makes consumers have a better attitude towards the product because they might believe they have more real information than if they have only positive or only negative information. If we compare among models, there are also differences in the magnitude of coefficients. Again, the effect of *ln_rating_inconsistency* is stronger when we assume that online reviews are sorted by the most helpful criterion (case 2) rather than the most recent criterion (case 3). This might indicate that the presence of both the positive and negative online reviews in top positions of the most helpful ranking has a greater positive impact on the product bestselling ranking. As in the case of the *ln_rating*, being in top positions in the most helpful rank means that many other consumers have found the information of that online reviews useful or diagnostic. So, both, positive and negative reviews in top positions of that ranking have been useful for consumers, and therefore, prospective consumers find that information more trustworthy and closer to reality. If we had just looked at case 1, we would think that the effect is much stronger than it is when we consider review visibility.

We have also incorporated to the model an interaction term between *ln_rating* and *ln_rating_inconsistency*. We observe that in every model the interaction term is negative and significant. This indicates that the effect of *ln_rating* on *ln_sales_rank_inverse* is mitigated by *ln_rating_inconsistency*. In other words, when there is a high difference between individual review ratings, and the product average rating, the effect of the product average rating on the product bestselling rank is reduced. **Thus, the presence of “extreme” online reviews makes the *ln_rating* itself to be less**

influential on the product bestselling rank. As in previous cases, this relationship is **stronger in case 2 than in case 3. Therefore, when the presence of “extreme” reviews in the top most helpful ranking is high, the effect of the average rating of those most helpful online reviews on the product bestselling rank is smaller.**

Finally, we observe that the effect of the review textual variables *ln_analytic*, *ln_authentic*, and *ln_clout* is significant in every model, but there are some differences in both sign and magnitude. *Ln_analytic* has a negative impact in case 1 and case 3, while it is positive in case 2. Having more organized, logical, and hierarchical written online reviews is positive when we are in case 2, where consumers evaluate online reviews based on the most helpful criterion. However, this feature of online reviews has a negative impact on sales when we are in case 1, when we assume that all reviews have the same visibility, and in case 3, when we assume that consumers evaluate online reviews based on the most recent criterion. Thus, we might think that consumer decision-making changes depending on the set of online reviews they view and evaluate. *Ln_authentic*, and *ln_clout* positively influence product sales in case 1, but they both have a significant and negative coefficient in the rest of the models. Therefore, if only Model 1 was evaluated, which is the one traditionally used, we might think that first, products with more personal and humble online reviews (high values in *ln_authentic*) and second, products with online reviews showing high reviewer confidence and leadership (high values in *ln_clout*) are more likely to be sold. However, we observe the opposite effect if we consider the other review visibility cases. When we assume that all reviews do not have the same probability of being viewed and consumers evaluate reviews based on either the most helpful or more recent criterion, we observe that both *ln_authentic* and *ln_clout* negatively influence product sales.

Overall, we observe that just considering one review visibility case (case 1) might lead to **biased conclusions, since Model 1's output differs from the rest of the models. To get a broader picture of the effect of online reviews on product sales, several cases of review visibility should be explored.**

4.3. Misspecification Tests and Alternative Panel Data Models

Four misspecifications tests are conducted to check the validity of the models and are reported in Table 5. First, two Wald tests of the joint significance of the reported coefficients (*z1*) and time dummy variables (*z2*) are reported, with degrees of freedom in parentheses. The null hypothesis for *z1* claims no relationship between the explanatory variables, and the null hypothesis for *z2* posit no relationship between time dummy

variables. The two Wald tests indicate that there is joint significance of explanatory variables and time dummy variables. Second, the Hansen test verifies the validity of the instruments or, in other words, the lack of correlation between the instruments and the random disturbance of the error term. The null hypothesis is that the instruments are not valid so failure to reject the null hypothesis means that the instruments are valid. We do not reject the null hypothesis, so our instruments are valid. Finally, the AR(2) test (Arellano & Bond, 1991) was conducted to test the lack of second order serial correlation of the first differenced residuals. The null hypothesis is that the residuals are serially uncorrelated. Therefore, if the null hypothesis is not rejected, it provides evidence that there is no second-order serial correlation and the GMM estimator is consistent. The AR (2) tests in our models indicate that we cannot reject the null hypothesis, so there is no second-order serial correlation and the GMM is consistent. Overall, the four tests indicate that the models are well specified.

Table 5

Most helpful visibility—all reviews, decreasing in probability of being viewed (case $v_{2,1}$)

	(1)	(2)	(3)	(4)
	OLS	FE	RE	System GMM
L1.ln_sales_rank_inverse	0.915 *** (0.01)	0.643 *** (0.02)	0.915 *** (0.01)	0.889 *** (0.00)
Ln_volume	0.063 *** (0.01)	-0.730 ** (0.26)	0.063 *** (0.01)	0.064 *** (0.01)
Ln_rating	0.054 ** (0.02)	0.030 (0.08)	0.054 ** (0.02)	0.256 *** (0.01)
Ln_rating_inconsistency	0.182 (0.09)	0.397 * (0.18)	0.182 (0.09)	1.012 *** (0.03)
Ln_rating x Ln_rating_inconsistency	-0.173 * (0.09)	-0.322 (0.18)	-0.173* (0.09)	-0.953 *** (0.03)
Ln_analytic	-0.004 (0.01)	-0.071 (0.06)	-0.004 (0.01)	0.031 *** (0.00)
Ln_authentic	-0.003 (0.01)	-0.066 (0.05)	-0.003 (0.01)	-0.070 *** (0.01)
Ln_clout	-0.002 (0.01)	-0.049 (0.06)	-0.002 (0.01)	-0.077 *** (0.00)
Christmas	0.014 (0.04)	-0.001 (0.04)	0.014 (0.04)	0.003 (0.00)
New_label	0.112 * (0.05)	0.205 * (0.08)	0.112 * (0.05)	0.105 *** (0.01)
Exclusive_label	0.144 *** (0.03)	0.945 *** (0.21)	0.144 *** (0.03)	0.218 *** (0.01)
Ln_price	0.043 (0.03)	0.000 (.)	0.043 (0.03)	0.294 *** (0.01)
Ln_size	0.027 (0.02)	-0.302 (1.58)	0.027 (0.02)	0.256 *** (0.02)
Brand_retailer	0.004 (0.06)	0.000 (.)	0.004 (0.06)	0.293 *** (0.03)
Ln_brand_followers	-0.020	0.014	-0.020	-0.027 ***

	(0.01)	(0.07)	(0.01)	(0.00)
Brand_top	−0.008	0.000	−0.008	0.059
	(0.04)	(.)	(0.04)	(0.04)
Brand_premium	0.038	0.000	0.038	0.036**
	(0.03)	(.)	(0.03)	(0.01)
Constant	−0.087	−0.462 *	−0.087	−0.353 ***
	(0.06)	(0.19)	(0.06)	(0.02)
Observations	944	944	944	944

Notes: All time dummies are included, but not reported in the table to save space. All non-binary variables are standardized. Robust standard errors are shown in parenthesis. $P < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

There is theoretical and empirical evidence that the system GMM is the panel data model that better controls for unobserved heterogeneity and endogeneity of explanatory variables, so it is the one with less estimation bias (Lozano et al., 2016; Pindado et al., 2011; Roodman, 2009b).

To explore the output of other commonly used panel data models, which do not control for endogeneity, and to compare it to the system GMM results, we have estimated those models for each case of review visibility. Table 5 reports the output of the different panel data models when we consider case $v_{2,1}$, where we assume consumers sort online reviews by the most helpful order and all reviews have a decreasing probability of being viewed. In column 1, we show the results of the Ordinal Least Squares (OLS) estimator, and columns 2 and 3 report the results of the Fixed Effects (FE) and Random Effects (RE) estimators. Finally, column 4 shows the output of the adopted system GMM estimator. Focusing on review variables, we observe some differences concerning review numeric variables, but not a clear pattern. For example, *In_rating* is significant in every model, but not in the FE model, and *In_rating_inconsistency* is significant in the FE and system GMM models, but not in the OLS and RE models. We observe a clearer pattern in terms of review text variables, which are only significant in the system GMM model. Thus, we can conclude that not dealing with endogeneity in our analysis might bias the results. We have estimated every model (OLS, FE, RE, and system GMM) for the rest of the review visibility cases (v_1 , $v_{2,2}$, $v_{3,1}$, and $v_{3,2}$), and overall, results follow the same pattern as in the discussed case $v_{2,1}$, shown in Table 5. Comparison tables for each review visibility case are shown in Appendix A (Table A1–A4).

5. Discussion

In this paper, we propose a conceptual framework to explore the impact of online reviews on product sales. The framework incorporates the role of review visibility when exploring the relationship between online reviews and product sales. We develop the model in a cosmetics context by gathering information about products and online reviews from the

whole category of “blush” products on a cosmetic’s online retailer over nine weeks. By comparing four models that incorporate different cases of review visibility to a baseline model, which does not consider the effect of review visibility, we demonstrate that the incorporation of review visibility is important because the magnitude of the results is different from one assumption to another.

Our findings reveal that non-textual variables: the average rating of the product, the volume of reviews of the product and the rating inconsistency of the product, have a significant and positive impact on product sales at every visibility scenario, even though the magnitude of the effect varies over them. We incorporate three textual variables extracted from the lexicon-based method for text mining Linguistic Inquiry and Word Count: *analytic*, which represents if the review has been written in an organized, logical, and hierarchical way; *authentic*, which measures if the review is written in a personal and humble way; and *clout*, which records the confidence expressed by the reviewer in the text. The effect of textual variables *analytic*, *authentic* and *clout* is always significant explaining product sales but varies in terms of sign and magnitude over the scenarios.

5.1. Theoretical Contribution

This study makes a major theoretical contribution. The extant literature on online reviews and product sales assumes that when consumers evaluate online reviews to make a purchase decision, every available online review for each product has the same probability of being viewed by consumers. However, decision-making theories (Häubl & Trifts, 2000; Payne, 1982) claim that consumers usually suffer from information overload in complex situations, and they are unable to evaluate all available alternatives. Instead, they usually adopt selective processing strategies to reduce the cognitive effort of managing a big volume of information. In this line, the accessibility-diagnostics theory (Feldman & Lynch, 1988) has highlighted that not only the quality and relevancy of the information (diagnostics), but also its accessibility influences consumer information adoption decisions. Therefore, we add to previous literature the notion of review visibility, which approaches the concept of accessibility in theory. In line with the decision-making theory (Häubl & Trifts, 2000), we explore review visibility under three main assumptions: when every review of a product has the same probability of being viewed; when consumers sort online reviews by the most helpful mechanism, and most helpful online reviews are more likely to be viewed; finally, when consumers sort reviews by the most recent mechanism (predetermined at the online retailer), in the way that most recent online reviews are more likely to be viewed. Our findings are in line with both theories and reveal that the effect of online reviews on product sales varies

depending on what reviews consumers view and evaluate. Different sets of online reviews, such as the most helpful reviews and the most recent reviews, might lead to different consumer decisions, since they provide different types of information. Thus, review visibility should be considered somehow when explaining the relationship between online reviews and product sales. A major finding is that the effect of review variable on product sales is higher when consumers evaluate most helpful reviews. This might suggest that information contained in most helpful online reviews is likely to have a greater influence on consumer purchasing behavior than the information in most recent online reviews. This might be due to the social influence effect that makes that, since many other consumers have considered the information contained in those reviews as helpful, future consumers evaluating most helpful reviews might think that the information contained in those reviews is a better approximation to the truth.

Another important contribution lies in integrating into the study review non-textual variables, which have been widely studied in previous literature, and review textual variables. Although the literature on review textual content is scarce, we corroborate previous findings showing that narrative characteristics influence consumer behavior (Areni, 2003; C. C. Chen & Tseng, 2011). Our findings reveal that the effect of review non-textual variables has the same sign across visibility scenarios, while different magnitudes. However, the sign and magnitude of textual variables vary across visibility scenarios. On the one hand, *authentic* and *clout* have a significant and negative impact on product sales in every review visibility scenario, either when consumers are evaluating most helpful or most recent online reviews. On the other hand, *analytic* has a significant and positive effect on product sales when consumers evaluate most helpful reviews, but a significant and negative impact in the case of most recent reviews.

5.2. Managerial Implications

Our findings have some managerial implications. Firstly, this study corroborates previous literature and industry reports that highlight the power of online reviews to influence product sales. We observe that every review variable incorporated into our analysis has an impact on the product bestselling ranking. This impact is significant independent from the review visibility case considered. Moreover, because review variables also influence product sales in cases where we assume consumers sort online reviews either by the most helpful or by the most recent criterion, managers should pay special attention to those online reviews appearing in top positions. The information contained in those online reviews is going to be influential in prospective consumers, so companies could analyze it to improve the current products or to launch new ones.

Considering that consumers use sorting tools to reduce the cognitive effort of managing big amounts of information, managers could incorporate new sorting mechanisms to help consumers in their decision-making. If more sorting tools available, consumers could select online reviews based on the most preferred criterion. For example, sorting tools based on text and reviewer features could be added. In fact, we observe that when we assume that consumers sort online reviews by the most helpful order ($v_{2.2}$), those online reviews in top positions (those with more helpful votes) have a greater influence on product sales, since those review variables in our empirical model have the biggest coefficient magnitudes. In line with these findings, the online retailer introduced some changes after we collected the data for the research. For example, they do not longer show online reviews by the predetermined criterion of the most recent order, but by the most helpful criterion. This corroborates our finding revealing that most helpful online reviews are likely to be more influential on consumer shopping behavior.

5.3. Limitations and Future Research

This paper explores the relationship between review non-textual and textual variables and product sales in three different review visibility cases. However, future work could expand the research to other review visibility cases, such as when we assume that consumers sort online reviews by the highest rating or by the lowest rating mechanisms.

In this work, we focus on three review textual variables, which are obtained from the dictionary-based tool LIWC (Pennebaker et al., 2015), to analyze the effect of review text on product sales. Future research could deepen on the study of review textual features and could incorporate other LIWC variables, such as the use of informal language and the specific motivations (e.g., social status and power) evoked by consumers in the text. However, other (more sophisticated) text mining methods based on machine learning algorithms could be used to uncover other relevant textual aspects of online reviews, such as consumer perceptions and brand image. For example, following the line of Ngo-Ye and Sinha (Ngo-Ye & Sinha, 2014), we could analyze review texts to study the influence of reviewer engagement characteristics on product sales.

Other reviewer non-textual characteristics, such as reviewer expertise, reviewer reputation, and reviewer identity, could also be assessed. We could also use supervised machine learning methods, as proposed by Vermeer et al. (Vermeer et al., 2019), to detect satisfied and dissatisfied consumers from online reviews, with the objective of exploring, for example, if online reviews written by satisfied consumers lead to higher sales and those written by dissatisfied consumers drive lower sales. Another interesting stream of

research could also study how online reviews, both non-textual and textual features, influence reviewer perceptions of products or brands over time.

We have carried out the analysis with information about one cosmetics category, blush. More product categories could be added to the analysis to compare between them and analyze differences. It would be also interesting to add more weeks to the analysis to have more time information. Moreover, online reviews from other industries, different from cosmetics, could be analyzed to see if the results could be generalized or if they are industry dependent. In terms of dates, we are dealing with a period that includes the Christmas holidays, in which consumers tend to increase their purchases. Even though blush products are not as stationary as other cosmetics, such as perfume, other periods of time could be analyzed and compared.

Appendix – Chapter 2

Table A1

All reviews—same probability of being viewed (case v_1).

	(1)	(2)	(3)	(4)
	OLS	FE	RE	System GMM
L1.ln_sales_rank_inverse	0.918 *** (0.01)	0.644 *** (0.03)	0.918 *** (0.01)	0.919 *** (0.00)
Ln_volume	0.058 *** (0.01)	-0.889 ** (0.28)	0.058 *** (0.01)	0.072 *** (0.01)
Ln_rating	0.035 (0.02)	0.013 (0.08)	0.035 (0.02)	0.116 *** (0.00)
Ln_rating_inconsistency	0.081 (0.15)	0.614 * (0.26)	0.081 (0.15)	0.506 *** (0.02)
Ln_rating x Ln_rating_inconsistency	-0.064 (0.15)	-0.594 * (0.26)	-0.064 (0.15)	-0.476 *** (0.02)
Ln_analytic	0.004 (0.01)	-0.019 (0.06)	0.004 (0.01)	-0.016 *** (0.00)
Ln_authentic	0.001 (0.01)	0.018 (0.05)	0.001 (0.01)	0.035 *** (0.00)
Ln_clout	-0.001 (0.01)	-0.041 (0.08)	-0.001 (0.01)	0.030 *** (0.01)
Christmas	0.011 (0.04)	-0.005 (0.04)	0.011 (0.04)	-0.003 (0.00)
New_label	0.118 * (0.05)	0.217 ** (0.08)	0.118 ** (0.05)	0.100 *** (0.01)
Exclusive_label	0.141 *** (0.03)	0.959 *** (0.21)	0.141 *** (0.03)	0.163 *** (0.01)
Ln_price	0.039 (0.03)	0.000 (.)	0.039 (0.03)	0.321 *** (0.01)
Ln_size	0.024 (0.02)	-0.497 (1.58)	0.024 (0.02)	0.273 *** (0.01)
Brand_retailer	0.001 (0.05)	0.000 (.)	0.001 (0.05)	0.309 *** (0.02)
Ln_brand_followers	-0.018 (0.01)	0.008 (0.07)	-0.018 (0.01)	-0.053 *** (0.00)
Brand_top	-0.004 (0.04)	0.000 (.)	-0.004 (0.04)	0.057 ** (0.02)
Brand_premium	0.038 (0.03)	0.000 (.)	0.038 (0.03)	0.013 (0.01)
Constant	-0.076 (0.05)	-0.312 *** (0.07)	-0.076 (0.05)	-0.078 *** (0.01)
Observations	944	944	944	944

Notes: All time dummies are included, but not reported in the table to save space. All non-binary variables are standardized. Robust standard errors are shown in parenthesis. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A2

Most helpful visibility—five most helpful reviews are viewed (case $v_{2.2}$).

	(1)	(2)	(3)	(4)
	OLS	FE	RE	System GMM
L1.ln_sales_rank_inverse	0.913 *** (0.01)	0.581 *** (0.02)	0.913 *** (0.01)	0.767 *** (0.00)
Ln_volume	0.068 *** (0.01)	−0.037 (0.27)	0.068 *** (0.01)	0.093 *** (0.01)
Ln_rating	0.064 ** (0.02)	0.138 (0.09)	0.064 ** (0.02)	0.578 *** (0.01)
Ln_rating_inconsistency	0.132 ** (0.05)	1.799 *** (0.24)	0.132 ** (0.05)	1.949 *** (0.01)
Ln_rating x Ln_rating_inconsistency	−0.136 ** (0.05)	−1.618 *** (0.23)	−0.136 ** (0.05)	−1.872 *** (0.01)
Ln_analytic	−0.009 (0.01)	−0.126 * (0.06)	−0.009 (0.01)	0.008 * (0.00)
Ln_authentic	0.019 (0.01)	−0.053 (0.08)	0.019 (0.01)	−0.077 *** (0.01)
Ln_clout	0.008 (0.01)	0.093 (0.05)	0.008 (0.01)	−0.105 *** (0.00)
Christmas	0.015 (0.04)	0.016 (0.03)	0.015 (0.04)	0.015 *** (0.00)
New_label	0.128 ** (0.05)	0.315 *** (0.08)	0.128 ** (0.05)	0.352 *** (0.01)
Exclusive_label	0.146 *** (0.03)	0.803 *** (0.20)	0.146 *** (0.03)	0.280 *** (0.01)
Ln_price	0.059 * (0.03)	0.000 (.)	0.059 * (0.03)	0.060 *** (0.01)
Ln_size	0.037 (0.02)	−0.648 (1.49)	0.037 (0.02)	0.061 *** (0.01)
Brand_retailer	0.024 (0.05)	0.000 (.)	0.024 (0.05)	0.255 *** (0.04)
Ln_brand_followers	−0.021 (0.01)	0.011 (0.06)	−0.021 (0.01)	−0.053 *** (0.01)
Brand_top	0.001 (0.04)	0.000 (.)	0.001 (0.04)	−0.011 (0.03)
Brand_premium	0.038 (0.03)	0.000 (.)	0.038 (0.03)	0.147 *** (0.02)
Constant	−0.085 (0.05)	−0.305 *** (0.07)	−0.085 (0.05)	−0.200 *** (0.02)
Observations	944	944	944	944

Notes: All time dummies are included, but not reported in the table to save space. All non-binary variables are standardized. Robust standard errors are shown in parenthesis. $P < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A3

Most recent visibility—all reviews, decreasing probability of being viewed (case v_{3,1}).

	(1)	(2)	(3)	(4)
	OLS	FE	RE	System GMM
L1.ln_sales_rank_inverse	0.920 *** (0.01)	0.654 *** (0.02)	0.920 *** (0.01)	0.937 *** (0.00)
Ln_volume	0.064 *** (0.01)	-0.827 ** (0.25)	0.064 *** (0.01)	0.134 *** (0.01)
Ln_rating	0.046 * (0.02)	0.006 (0.04)	0.046 * (0.02)	0.108 *** (0.01)
Ln_rating_inconsistency	0.102 (0.10)	0.076 (0.17)	0.102 (0.10)	0.369 *** (0.03)
Ln_rating x Ln_rating_inconsistency	-0.083 (0.10)	-0.057 (0.16)	-0.083 (0.10)	-0.312 *** (0.03)
Ln_analytic	0.003 (0.01)	-0.012 (0.02)	0.003 (0.01)	-0.048 *** (0.00)
Ln_authentic	-0.022 (0.01)	-0.025 (0.02)	-0.022 (0.01)	-0.042 *** (0.01)
Ln_clout	-0.021 (0.01)	-0.019 (0.02)	-0.021 (0.01)	-0.013 ** (0.00)
Christmas	0.014 (0.04)	-0.000 (0.04)	0.014 (0.04)	0.008 (0.00)
New_label	0.099 * (0.04)	0.203 * (0.08)	0.099 * (0.04)	0.017 * (0.01)
Exclusive_label	0.142 *** (0.03)	0.983 *** (0.21)	0.142 *** (0.03)	0.145 *** (0.01)
Ln_price	0.028 (0.03)	0.000 (.)	0.028 (0.03)	0.133 *** (0.01)
Ln_size	0.014 (0.02)	-0.154 (1.59)	0.014 (0.02)	0.108 *** (0.01)
Brand_retailer	-0.006 (0.06)	0.000 (.)	-0.006 (0.06)	0.159 *** (0.03)
Ln_brand_followers	-0.014 (0.01)	0.015 (0.07)	-0.014 (0.01)	-0.068 *** (0.01)
Brand_top	-0.006 (0.04)	0.000 (.)	-0.006 (0.04)	0.048 (0.02)
Brand_premium	0.032 (0.03)	0.000 (.)	0.032 (0.03)	0.007 (0.01)
Constant	-0.076 (0.05)	-0.324 *** (0.07)	-0.076 (0.05)	-0.070 *** (0.01)
Observations	944	944	944	944

Notes: All time dummies are included, but not reported in the table to save space. All non-binary variables are standardized. Robust standard errors are shown in parenthesis. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A4

Most helpful visibility—five most recent reviews are viewed (case $v_{3,2}$).

	(1)	(2)	(3)	(4)
	OLS	FE	RE	System GMM
L1.ln_sales_rank_inverse	0.920 *** (0.01)	0.660 *** (0.03)	0.920 *** (0.01)	0.942 *** (0.00)
Ln_volume	0.059 *** (0.01)	-0.823 *** (0.24)	0.059 *** (0.01)	0.075 *** (0.01)
Ln_rating	0.087 *** (0.02)	0.069 (0.04)	0.087 *** (0.02)	0.274 *** (0.01)
Ln_rating_inconsistency	0.179 ** (0.07)	0.144 (0.11)	0.179 ** (0.07)	0.552 *** (0.04)
Ln_rating x Ln_rating_inconsistency	-0.152 ** (0.06)	-0.116 (0.09)	-0.152 ** (0.06)	-0.423 *** (0.03)
Ln_analytic	0.016 (0.01)	0.010 (0.02)	0.016 (0.01)	-0.012 ** (0.00)
Ln_authentic	-0.003 (0.01)	0.007 (0.02)	-0.003 (0.01)	-0.020 ** (0.01)
Ln_clout	-0.016 (0.01)	0.004 (0.02)	-0.016 (0.01)	-0.034 *** (0.01)
Christmas	0.014 (0.04)	0.000 (0.04)	0.014 (0.04)	0.005 (0.00)
New_label	0.112 * (0.04)	0.200 * (0.08)	0.112 * (0.04)	0.092 *** (0.01)
Exclusive_label	0.148 *** (0.03)	0.929 *** (0.21)	0.148 *** (0.03)	0.186 *** (0.02)
Ln_price	0.033 (0.02)	0.000 (.)	0.033 (0.02)	0.033 * (0.01)
Ln_size	0.020 (0.02)	-0.424 (1.59)	0.020 (0.02)	0.045 *** (0.01)
Brand_retailer	0.004 (0.05)	0.000 (.)	0.004 (0.05)	0.072 * (0.03)
Ln_brand_followers	-0.016 (0.01)	0.014 (0.07)	-0.016 (0.01)	-0.057 *** (0.01)
Brand_top	0.005 (0.04)	0.000 (.)	0.005 (0.04)	0.046 * (0.02)
Brand_premium	0.040 (0.03)	0.000 (.)	0.040 (0.03)	0.058 *** (0.02)
Constant	-0.080 (0.05)	-0.307 *** (0.07)	-0.080 (0.05)	-0.107 *** (0.01)
Observations	944	944	944	944

Notes: All time dummies are included, but not reported in the table to save space. All non-binary variables are standardized. Robust standard errors are shown in parenthesis. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

References

- Aaker, D. A. (1991). Managing Brand Equity. *Managing Brand Equity*.
- Aaker, D. A. (1996). Building strong brands - aaker,da. *Building Strong Brands*, 13(4), 382–386.
- Ahani, A., Nilashi, M., Yadegaridehkordi, E., Sanzogni, L., Tarik, A. R., Knox, K., Samad, S., & Ibrahim, O. (2019). Revealing customers' satisfaction and preferences through online review analysis: The case of Canary Islands hotels. *Journal of Retailing and Consumer Services*, 51(July), 331–343.
- Ahmad, S. N., & Laroche, M. (2015). How do expressed emotions affect the helpfulness of a product review? Evidence from reviews using latent semantic analysis. *International Journal of Electronic Commerce*, 20(1), 76–111.
- Aiken, L. S., & West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*.
- Al-Obeidat, F., Spencer, B., & Kafeza, E. (2018). The Opinion Management Framework: Identifying and addressing customer concerns extracted from online product reviews. *Electronic Commerce Research and Applications*, 27, 52–64.
- Al., H. et. (2010). *Multivariate Analysis 7 th Edition*.
- Aljukhadar, M., Senecal, S., & Daoust, C.-E. (2012). Using Recommendation Agents to Cope with Information Overload. *International Journal of Electronic Commerce*, 17(2), 41–70.
- Ansari, A., & Mela, C. F. (2003). *E-Customization*. 40(2), 131–145.
- Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*, 57(8), 1485–1509.
- Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2), 277.
- Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1), 29–51.
- Areni, C. S. (2003). The Effects of Structural and Grammatical Variables on Persuasion: An Elaboration Likelihood Model Perspective. *Psychology and Marketing*, 20(4), 349–375.
- Arun, R., Suresh, V., Madhavan, C. E. V., & M.N, N. M. (2010). On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In *Advances in Knowledge Discovery and Data Mining* (pp. 391–403). <http://hdl.handle.net/10722/26088>
- Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. A. (2016). The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors. *Journal of Marketing Research*, 53(3), 297–318.
- Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, 0, 2200–2204.
- Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of Online Consumer Reviews: Readers'

- Objectives and Review Cues. *International Journal of Electronic Commerce*, 17(May 2015), 99–126.
- Baker, W. E. (2001). The Diagnosticity of Advertising Generated Brand Attitudes in Brand Choice Contexts. *Journal of Consumer Psychology*, 11(2), 129–139.
- Baker, W. E., Lutz, R. J., Journal, S., Spring, N., Taylor, P., Baker, W. E., & Lutz, R. J. (2000). Effectiveness An Empirical Test of an Updated Relevance- Accessibility Model of Advertising Effectiveness. *Journal of Advertising*, 29(1), 1–14.
- Bakhshi, S., Kanuparth, P., & Shamma, D. A. (2015). Understanding Online Reviews: Funny, Cool or Useful? *CSCW 2015: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1270–1276.
- Baksi, A. K., & Panda, T. K. (2018). Branding destinations with multisensory brand associations and evaluating its impact on behavioural pattern under the intervention of multiplex phenomenon of relationship-branding. *Management Science Letters*, 8(11), 1169–1182.
- Balducci, B., & Marinova, D. (2018a). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, 46(4), 557–590.
- Balducci, B., & Marinova, D. (2018b). Unstructured data in marketing. In *Journal of the Academy of Marketing Science* (Vol. 46, Issue 4, pp. 557–590). Springer New York LLC.
- Beach, L. R. (1993). Broadening the definition of decision making: The role of prechoice screening option. *Psychological Science*, 4(4), 215–220.
- Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding Satisfied and Dissatisfied Hotel Customers: Text Mining of Online Hotel Reviews. *Journal of Hospitality Marketing and Management*, 25(1), 1–24.
- Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the Tribes: Using Text for Marketing Insight. *Journal of Marketing*, 84(1), 1–25.
- Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive Consumer Choice Processes. *Journal of Consumer Research*, 25(3), 187–217.
- Biehal, G., & Chakravarti, D. (1983). Information Accessibility as a Moderator of Consumer Choice. *Journal of Consumer Research*, 10(1), 1.
- Biswas, D., Grewal, D., & Roggeveen, A. (2010). How the order of sampled experiential products affects choice. *Journal of Marketing Research*, 47(3), 508–519.
- Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), 115–143.
- Brandt, C., De Mortanges, C. P., Bluemelhuber, C., & Van Riel, A. C. R. (2011). Associative Networks: A New Approach to Market Segmentation. *International Journal of Market Research*, 53(2), 187–207.
- Breugelmans, E., Campo, K., & Gijsbrechts, E. (2007). Shelf sequence and proximity effects on online grocery choices. *Marketing Letters*, 18(1–2), 117–133.
- BrightLocal. (2020). *Local Consumer Review Survey*. Available at: <https://www.brightlocal.com/research/local-consumer-review-survey/>
- Brown, C. L., & Krishna, A. (2004). The skeptical shopper: A metacognitive account for the effects of default options on choice. *Journal of Consumer Research*, 31(3), 529–539.

- Bruine de Bruin, W., & Keren, G. (2003). Order effects in sequentially judged options due to the direction of comparison. *Organizational Behavior and Human Decision Processes*, 92(1–2), 91–101.
- Buda, R., & Zhang, Y. (2000). Consumer product evaluation: the interactive effect of message framing, presentation order, and source credibility. *Journal of Product & Brand Management*, 9(4), 229–242.
- Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics Using Stata*. Stata Press Book, StataCorp LP.
- Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data, Second edition. In *Regression Analysis of Count Data, Second Edition*.
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72(7–9), 1775–1781.
- Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, 50(2), 511–521.
- Chaiken, S. (1987). *The heuristic model of persuasion*.
- Chapman, C., & Feit, E. M. (2015). *R for marketing research and analytics*.
- Chen, C. C., & Tseng, Y. De. (2011). Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, 50(4), 755–768.
- Chen, K., Kou, G., Shang, J., & Chen, Y. (2015). Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches. *Electronic Commerce Research and Applications*, 14(1), 58–74.
- Cheng, Y. H., & Ho, H. Y. (2015). Social influence’s impact on reader perceptions of online reviews. *Journal of Business Research*, 68(4), 883–887.
- Cheung, C. M. K., Lee, M. K. O., & Rabjohn, N. (2008). The impact of electronic word-of-mouth. In *Internet Research* (Vol. 18, Issue 3).
- Chevalier, J. A., & Goolsbee, A. (2003). Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics*, 1, 203–222.
- Chevalier, J., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 43(3), 345–354.
- Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science*, 29(5), 944–957.
- Cho, E., Fiore, A. M., & W. Russell, D. (2015). Validation of a Fashion Brand Image Scale Capturing Cognitive, Sensory, and Affective Associations: Testing Its Role in an Extended Brand Equity Model. *Psychology & Marketing*, 32(1), 28–48.
- Chua, A. Y. K., & Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior*, 54, 547–554.
- Chung, C.K., & Pennebaker, J. W. (2007). The psychological function of function words. *Social Communication: Frontiers of Social Psychology*, 343–359.
- Chung, Cindy K., & Pennebaker, J. W. (2007). Revealing dimensions of thinking in open-

- ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality*, 42(1), 96–132.
- Cohn, M. a., Mehl, M. R., & Pennebaker, J. W. (2004). Markers of Linguistic Psychological Change. *Psychological Science*, 15(10), 687–693.
- Cohn, M. a., Mehl, M. R., & Pennebaker, J. W. (2014). Markers of Linguistic Psychological Change Surrounding September 11, 2001. *Psychological Science*, 15(10), 687–693.
- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2014). *Markers of Linguistic Psychological Change Surrounding September*. 15(10), 687–693.
- Cui, G., Lui, H.-K., & Guo, X. (2012). The Effect of Online Consumer Reviews on New Product Sales. *International Journal of Electronic Commerce*, 17(1), 39–58.
- Cyr, D., Head, M., Lim, E., & Stibe, A. (2018). Using the elaboration likelihood model to examine online persuasion through website design. *Information and Management*, April 2017, 1–15.
- Darley, W. K., & Smith, R. E. (1993). Advertising Claim Objectivity: Antecedents and Effects. *Journal of Marketing*, 57(4), 100.
- Davis, D. F., Golcic, S. L., & Marquardt, A. (2009). Measuring brand equity for logistics services. *The International Journal of Logistics Management*, 20(2), 201–212.
- De Vries, L., Gensler, S., & Leeftang, P. S. H. (2012). Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing. *Journal of Interactive Marketing*, 26(2), 83–91.
- Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, 27(4), 293–307.
- Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4), 23–45.
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective Latent Concept Modeling for ad hoc information retrieval. *Document Numerique*, 17(1), 61–84.
- Drèze, X., & Zufryden, F. (2004). Measurement of online visibility and its impact on Internet traffic. *Journal of Interactive Marketing*, 18(1), 20–37.
- Duan, W., Gu, B., & Whinston, A. B. (2008a). Do online reviews matter? - An empirical investigation of panel data. *Decision Support Systems*, 45(4), 1007–1016.
- Duan, W., Gu, B., & Whinston, A. B. (2008b). The dynamics of online word-of-mouth and product sales-An empirical investigation of the movie industry. *Journal of Retailing*, 84(2), 233–242.
- Elberse, A., & Eliashberg, J. (2003). Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures. *Marketing Science*, 22(3), 329–354.
- Ellsworth, P. C., & Smith, C. A. (1988). From appraisal to emotion: Differences among unpleasant feelings. *Motivation and Emotion*, 12(3), 271–302.
- Ert, E., & Fleischer, A. (2016). Mere Position Effect in Booking Hotels Online. *Journal of Travel Research*, 55(3), 311–321.
- Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management*, 55, 62–73.

- Eslami, S. P., & Ghasemaghaei, M. (2018). Effects of online review positiveness and review score inconsistency on sales: A comparison by product involvement. *Journal of Retailing and Consumer Services*, 45(June), 74–80.
- Faircloth, J. B., Capella, L. M., & Alford, B. L. (2001). The Effect of Brand Attitude and Brand Image on Brand Equity. *Journal of Marketing Theory and Practice*, 9(3), 61–75.
- Feldman, J. M., & Lynch, J. G. (1988). Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior. *Journal of Applied Psychology*, 73(3), 421–435.
- Felfernig, A., Friedrich, G., Gula, B., Hitz, M., Kruggel, T., Leitner, G., Melcher, R., Riepan, D., Strauss, S., Teppan, E., & Vitouch, O. (2007). Persuasive recommendation: Serial position effects in knowledge-based recommender systems. *Proceedings of the 2nd International Conference on Persuasive Technology (PERSUASIVE '07)*, 283–295.
- Filieri, R., Hofacker, C. F., & Alguezaui, S. (2018). What makes information in online consumer reviews diagnostic over time? The role of review relevancy, factuality, currency, source credibility and ranking score. *Computers in Human Behavior*, 80, 122–131.
- Filieri, R., & McLeay, F. (2014). E-WOM and Accommodation: An Analysis of the Factors That Influence Travelers' Adoption of Information from Online Reviews. *Journal of Travel Research*, 53(1), 44–57.
- Filieri, R., McLeay, F., Tsui, B., & Lin, Z. (2018). Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services. *Information and Management*, 55(8), 956–970.
- Filieri, R., Raguseo, E., & Vitari, C. (2018). When are extreme ratings more helpful? Empirical evidence on the moderating effects of review characteristics and product type. *Computers in Human Behavior*, 88, 134–142.
- Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., & Freling, T. (2014). How Online Product Reviews Affect Retail Sales: A Meta-analysis. *Journal of Retailing*, 90(2), 217–232.
- Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., & Treinen, M. (2001). What makes web sites credible? A report on a large quantitative study. *Conference on Human Factors in Computing Systems - Proceedings*, 61–68.
- Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 19(3), 291–313.
- Gao, B., Hu, N., & Bose, I. (2017). Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews. *Decision Support Systems*, 95, 1–11.
- Gensler, S., Völckner, F., Egger, M., Fischbach, K., & Schoder, D. (2015). Listen to your customers: Insights into brand image using online consumer-generated product reviews. *International Journal of Electronic Commerce*, 20(1), 112–141.
- Ghose, A., & Ipeirotis, P. G. (2010). Estimating the Helpfulness and Economic Impact of Product Reviews. *Ieee Transactions on Knowledge and Data Engineering*, 1–15.
- Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science*, 31(3), 493–520.

- Girard, T., & Dion, P. (2010). Validating the search, experience, and credence product classification framework. *Journal of Business Research*, 63(9–10), 1079–1087.
- Gobet, F., Lane, P., Croker, S., Cheng, P., Jones, G., Oliver, I., & Pine, J. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5(6), 236–243.
- Godes, D., & Silva, J. C. (2012). Sequential and Temporal Dynamics of Online Opinion. *Marketing Science*, 31(3), 448–473.
- Godes, David, & Mayzlin, D. (2004). Using Online Conversations to Study Word-of-Mouth Communication. *Marketing Science*, 23(4), 545–560.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(SUPPL. 1), 5228–5235.
- Gu, B., Gu, B., & Park, J. (2012). The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products. *Information Systems Research*, 23(1), 182–196.
- Guo, J., Wang, X., & Wu, Y. (2020). Positive emotion bias: Role of emotional content from online customer reviews in purchase decisions. *Journal of Retailing and Consumer Services*, 52(July 2019).
- Guo, Y., Barnes, S. J., & Jia, Q. (2017a). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.
- Guo, Y., Barnes, S. J., & Jia, Q. (2017b). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.
- Hamby, A., Daniloski, K., & Brinberg, D. (2015). How consumer reviews persuade through narratives. *Journal of Business Research*, 68(6), 1242–1250.
- Haridasan, A. C., & Fernando, A. G. (2018). Online or in-store: unravelling consumer's channel choice motives. *Journal of Research in Interactive Marketing*, 12(2), 215–230.
- Harrison-Walker, L. J. (2001). The Measurement of Word-of-Mouth Communication and an Investigation of Service Quality and Customer Commitment As Potential Antecedents. *Journal of Service Research*, 4(1), 60–75.
- Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, 36(1), 20–38.
- Häubl, G., & Trifts, V. (2000). Consumer Decision Making in Online Shopping Environments : The Effects of Interactive Decision Aids Consumer Decision Making in Online Shopping Environments : The Effects of Interactive Decision Aids. *Marketing Science*, 19(1), 4–21.
- Haugtvedt, C. P., & Wegener, D. T. (1994). Message Order Effects in Persuasion: An Attitude Strength Perspective. *Journal of Consumer Research*, 21(1), 205.
- Hauser, J. R., & Koppelman, F. S. (1979). Alternative Perceptual Mapping Techniques : Relative Accuracy and Usefulness. *Journal of Marketing Research*, 16(4), 495–506.
- Henderson, G. R., Iacobucci, D., & Calder, B. J. (1998). Brand diagnostics: Mapping branding effects using consumer associative networks. *European Journal of Operational Research*, 111(2), 306–327.

- Heng, Y., Gao, Z., Jiang, Y., & Chen, X. (2018). Exploring hidden factors behind online food shopping from Amazon reviews: A topic mining approach. *Journal of Retailing and Consumer Services*, 42(January), 161–168.
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, 18(1), 38–52.
- Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective. *Journal of Consumer Research*, 17(4), 454.
- Herrmann, A., Goldstein, D. G., Stadler, R., Landwehr, J. R., Heitmann, M., Hofstetter, R., & Huber, F. (2011). The effect of default options on choice-Evidence from online product configurators. *Journal of Retailing and Consumer Services*, 18(6), 483–491.
- Ho-Dac, N. N., Carson, S. J., & Moore, W. L. (2013). The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter? *Journal of Marketing*, 77(6), 37–53.
- Hofmann, J., Clement, M., Völckner, F., & Hennig-Thurau, T. (2017). Empirical generalizations on the impact of stars on the economic success of movies. *International Journal of Research in Marketing*, 34(2), 442–461.
- Holbrook, M. B. (1978). Beyond Attitude Structure: Toward the Informational Determinants of Attitude. *Journal of Marketing Research*, 15(4), 545.
- Holtgraves, T., & Lasky, B. (1999). Linguistic power and persuasion. *Journal of Language and Social Psychology*, 18(2), 196–205.
- Hornik, K., & Grün, B. (2011). topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software*, 40(13), 1–30.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD'04*, 168.
- Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? the mediating role of online review sentiments on product sales. *Decision Support Systems*, 57(1), 42–53.
- Hu, Y., & Li, X. (2011). Context-Dependent Product Evaluations: An Empirical Analysis of Internet Book Reviews. *Journal of Interactive Marketing*, 25(3), 123–133.
- Huang, A. H., Chen, K., Yen, D. C., & Tran, T. P. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behavior*, 48, 17–27.
- Hutto, C. J., & Gilbert, E. E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14).". *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*. <http://sentic.net/>
- International, E. (2017). *Colour cosmetics in the US* (Issue May).
- Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011a). Language Style Matching Predicts Relationship Initiation and Stability. *Psychological Science*, 22(1), 39–44.
- Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011b). Language Style Matching Predicts Relationship Initiation and

- Stability. *Psychological Science*, 22(1), 39–44.
- Jalilvand, M. R., & Samiei, N. (2012). The effect of electronic word of mouth on brand image and purchase intention: An empirical study in the automobile industry in Iran. *Marketing Intelligence and Planning*, 30(4), 460–476.
- Jiang, S., Cai, S., Olle, G. O., & Qin, Z. (2015). Durable product review mining for customer segmentation. *Kybernetes*, 44(1), 124–138.
- Jiménez, F. R., & Mendoza, N. A. (2013). Too popular to ignore: The influence of online reviews on purchase intentions of search and experience products. *Journal of Interactive Marketing*, 27(3), 226–235.
- John, D. R., Loken, B., Kim, K., & Monga, A. B. (2006). Brand concept maps: A methodology for identifying brand association networks. *Journal of Marketing Research*, 43(4), 549–563.
- Johnson, E. J., Bellman, S., & Lohse, G. L. (2002). Defaults , Framing and Privacy :Why Opting In-Opting Out. *Marketing Letters*, 13(1), 5–15.
- Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, 80(4), 557–571.
- Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*, 33(2), 125–143.
- Kamble, Z., Doan, Q. T., Nguyen, H. N., Zeng, Z., & Zihui, L. L. (2020). The Significance of Online Reviews for Hotel Entrepreneurial Success. In *Entrepreneurial Development and Innovation in Family Businesses and SMEs* (pp. 75–86).
- Keller, K. L. (1993). *Conceptualizing, Measuring, and Managing Customer-Based Brand Equity*. 57(1), 1–22.
- Kelly, G. A. (1991). *The Psychology of Personal Constructs*.
- Khwaja, M. G., & Zaman, U. (2020). Configuring the evolving role of ewom on the consumers information adoption. *Journal of Open Innovation: Technology, Market, and Complexity*, 6(4), 1–13.
- Kim, H. B., Kim, W. G., & An, J. A. (2003). The effect of consumer-based brand equity on firms' financial performance. *Journal of Consumer Marketing*, 20(4–5), 335–351.
- Kim, M., & Lennon, S. (2007). The Effects of Visual and Verbal Information on Attitudes and Purchase Intentions in Internet Shopping. *Psychology & Marketing*, 24(9), 763–785.
- Kim, S. G., & Kang, J. (2018). Analyzing the discriminative attributes of products using text mining focused on cosmetic reviews. *Information Processing and Management*, 54(6), 938–957.
- King, M. F., & Balasubramanian, S. k. (1994). The effects of Expertise, End Goal, and Product Type on Adoption of Preference Formation Strategy. *Journal of the Academy of Marketing Science*, 22.
- Kostyk, A., Niculescu, M., & Leonhardt, J. M. (2017). Less is more: Online consumer ratings' format affects purchase intentions and processing. *Journal of Consumer Behaviour*, 16(5), 434–441.

- Kotler, P., & Armstrong, G. (2020). *Principles of Marketing* (18th Global). Pearson Education Limited.
- Kozinets, R. V., De Valck, K., Wojnicki, A. C., & Wilner, S. J. S. (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of Marketing*, 74(2), 71–89.
- Kübler, R. V., Colicev, A., & Pauwels, K. H. (2019). Social Media's Impact on the Consumer Mindset: When to Use Which Sentiment Extraction Tool? *Journal of Interactive Marketing*, 50, 136–155.
- Kudeshia, C., & Kumar, A. (2017). Social eWOM: does it affect the brand attitude and purchase intention of brands? *Management Research Review*, 40(3), 310–330.
- Lee, P. J., Hu, Y. H., & Lu, K. T. (2018). Assessing the helpfulness of online hotel reviews: A classification-based approach. *Telematics and Informatics*, 35(2), 436–445.
- Lee, S., & Choeh, J. Y. (2018). The interactive impact of online word-of-mouth and review helpfulness on box office revenue. *Management Decision*, 56(4), 849–866.
- Lee, S., & Choeh, J. Y. (2020). Using the social influence of electronic word-of-mouth for predicting product sales: The moderating effect of review or reviewer helpfulness and product type. *Sustainability (Switzerland)*, 12(19).
- Lee, T. Y., & Bradlow, E. T. (2011). Automated Marketing Research Using Online Customer Reviews. *Journal of Marketing Research*, 48(5), 881–894.
- Levinson, P. (2013). *New New Media*. 223. <https://books.google.com/books?id=G009uwAACAAJ&pgis=1>
- Lewsey, J. D., & Thomson, W. M. (2004). The utility of the zero-inflated Poisson and zero-inflated negative binomial models: A case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. *Community Dentistry and Oral Epidemiology*, 32(3), 183–189.
- Li, M., Huang, L., Tan, C.-H., & Wei, K.-K. (2013). Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features. *International Journal of Electronic Commerce*, 17(4), 101–136.
- Li, S. T., Pham, T. T., & Chuang, H. C. (2019). Do reviewers' words affect predicting their helpfulness ratings? Locating helpful reviewers by linguistics styles. *Information and Management*, 56(1), 28–38.
- Li, Xiaolin, Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information and Management*, 56(2), 172–184.
- Li, Xinxin, & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, 19(4), 456–474.
- Liang, T. P., Li, X., Yang, C. T., & Wang, M. (2015). What in Consumer Reviews Affects the Sales of Mobile Apps: A Multifacet Sentiment Analysis Approach. *International Journal of Electronic Commerce*, 20(2), 236–260.
- Liu, J., Cao, Y., Lin, C., Huang, Y., ... M. Z. the 2007 J. C. on, & 2007, U. (2007). Low-quality product review detection in opinion summarization. *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, June, 334–342. <http://www.aclweb.org/anthology/D07-1035>
- Liu, Q., & Karahanna, E. (2015). An agent-based modeling analysis of helpful vote on online product reviews. *Proceedings of the Annual Hawaii International*

- Conference on System Sciences, 2015-March*, 1585–1595.
- Liu, X., Burns, A. C., & Hou, Y. (2017). An Investigation of Brand-Related User-Generated Content on Twitter. *Journal of Advertising*, 46(2), 236–247.
- Liu, Y. (2006). Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*, 70(3), 74–89.
- Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, 47, 140–151.
- Low, G. S., & Lamb, C. W. (2000). The measurement and dimensionality of brand associations. *Journal of Product & Brand Management*, 9(6), 350–370.
- Lozano, M. B., Martínez, B., & Pindado, J. (2016). Corporate governance, ownership and firm value: Drivers of ownership as a good corporate governance mechanism. *International Business Review*, 25(6), 1333–1343.
- Luca, M. (2011). Revenue : The Case of Reviews , Reputation , and Revenue : The Case of Yelp . com. *Harvard Business Review*, 1–40.
- Ludwig, S., de Ruyter, K., Friedman, M., Brügger, E. C., Wetzels, M., & Pfann, G. (2013a). More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*, 77(1), 87–103.
- Ludwig, S., de Ruyter, K., Friedman, M., Brügger, E. C., Wetzels, M., & Pfann, G. (2013b). More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*, 77(1), 87–103.
- Lynch, Jr., J. G., Marmorstein, H., & Weigold, M. F. (1988). Choices from Sets Including Remembered Brands: Use of Recalled Attributes and Prior Overall Evaluations. *Journal of Consumer Research*, 15(2), 169.
- Magerman, T., Van Looy, B., Baesens, B., & Debackere, K. (2011). Assessment of Latent Semantic Analysis (LSA) text mining algorithms for large scale mapping of patent and scientific publication documents. *FBE Research Report MSI_1114*.
- Magoulas, R., & Swoyer, S. (2020). *AI Adoption in the Enterprise 2020*.
- Mahr, D., Stead, S., & Odekerken-Schröder, G. (2019). Making sense of customer service experiences: a text mining review. *Journal of Services Marketing*, 33(1), 88–103.
- Malhotra, N. K. (1981). A Scale to Measure Self-Concepts, Person Concepts, and Product Concepts. *Journal of Marketing Research*, 18(4), 456.
- Mantonakis, A., Roderio, P., Lesschaeve, I., & Hastie, R. (2009). Order in choice: Effects of serial position on preferences. *Psychological Science*, 20(11), 1309–1312.
- Marchand, A., Hennig-Thurau, T., & Wiertz, C. (2017). Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. *International Journal of Research in Marketing*, 34(2), 336–354.
- Martínez Salinas, E., & Pina Pérez, J. M. (2009). Modeling the brand extensions' influence on brand image. *Journal of Business Research*, 62(1), 50–60.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.
- Moe, W. W., & Trusov, M. (2011). The Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing Research*, 48(3), 444–456.

- Mohammad, S., & Turney, P. (2010). Emotions Evoked by Common Words and Phrases. *Proceedings of the NAACL-HLT, June*, 26–34.
- Moon, S., & Kamakura, W. A. (2017). A picture is worth a thousand words: Translating product reviews into a product positioning map. *International Journal of Research in Marketing*, 34(1), 265–285.
- Mudambi, S., & Schuff, D. (2010). What makes a Helpful online Review? *MIS Quarter*, 28(4), 695–704.
- Munch, J. M., & Swasy, J. L. (1988). Rhetorical question, summarization frequency, and argument strength effects on recall. *Journal of Consumer Research*, 15(1), 69–76.
- Murphy, J., Hofacker, C., & Mizerski, R. (2006). Primacy and recency effects on clicking behavior. *Journal of Computer-Mediated Communication*, 11(2), 173–192.
- Nabi, R. L. (2003). Exploring the framing effects of emotion: Do discrete emotions differentially influence information accessibility, information seeking, and policy preference? *Communication Research*, 30(2), 224–247.
- Nazlan, N. H., Tanford, S., & Montgomery, R. (2018). The effect of availability heuristics in online consumer reviews. *Journal of Consumer Behaviour*, 17(5), 449–460.
- Nelson, P. (1970). *Information and Consumer Behavior*. 78(2), 311–329.
- Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine Your Own Business: Market-Structure Surveillance Through Text Mining. *Marketing Science*, 31(3), 521–543.
- Netzer, Oded, Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521–543.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Personality and Social Psychology Bulletin Lying Words : Predicting Deception From Linguistic Styles. *Personality and Social Psychology Bulletin*, 29(5), 665–675.
- Ngo-Ye, T. L., & Sinha, A. P. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, 61(1), 47–58.
- Nielsen. (2016). *The sweet smell of seasonal success*. Available at: <https://www.nielsen.com/au/en/insights/news/2016/the-sweet-smell-of-seasonal-success.html>
- Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *CEUR Workshop Proceedings*, 718, 93–98.
- Olson, J. C., & Muderrisoglu, A. (1979). The stability of responses obtained by free elicitation: Implications for measuring attribute salience and memory structure. *Advances in Consumer Research*, 6, 269–275.
- Pan, B., Zhang, L., & Law, R. (2013). The Complex Matter of Online Hotel Choice. *Cornell Hospitality Quarterly*, 41(1), 74–83.
- Pan, Y., & Zhang, J. Q. (2011). Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews. *Journal of Retailing*, 87(4), 598–612.
- Pang, J., & Qiu, L. (2016). Effect of online review chunking on product attitude: The moderating role of motivation to think. *International Journal of Electronic Commerce*, 20(3), 355–383.

- Park, D. H., & Lee, J. (2008). eWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications*, 7(4), 386–398.
- Park, E. J., Kim, E. Y., Funches, V. M., & Foxx, W. (2012). Apparel product attributes, web browsing, and e-impulse buying on shopping websites. *Journal of Business Research*, 65(11), 1583–1589.
- Park, H.-J., & Rabolt J., N. (2009). Cultural Value, Consumption Value, and Global Brand Image: A Cross-National Study. *Psychology & Marketing*, 26(8), 714–735.
- Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research*, 50, 67–83.
- Payan, J. M., & McFarland, R. G. (2005). Decomposing Influence Strategies: Argument Structure and Dependence as Determinants of the Effectiveness of Influence Strategies in Gaining Channel Member Compliance. *Journal of Marketing*, 69(3), 66–79.
- Payne, J. W. (1982). Contingent Decision Behavior. *Psychological Bulletin*, 92(2), 382.
- Peladeau, N. (2016). *WordStat: Content Analysis Module for SIM- STAT*.
- Pennebaker, J. W., Boyd, R., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. *Austin, TX: University of Texas at Austin, SEPTEMBER 2015*, 1–22.
- Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS ONE*, 9(12), 1–10.
- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The Development and Psychometric Properties of LIWC2007. *Austin, Tx: LIWC.Net*, 1(2).
- Petty, R. E., Tormala, Z. L., Hawkins, C., & Wegener, D. T. (2001). Motivation to think and order effects in persuasion: The moderating role of chunking. *Personality and Social Psychology Bulletin*, 27(3), 332–344.
- Pindado, J., Requejo, I., & de la Torre, C. (2011). Family control and investment-cash flow sensitivity: Empirical evidence from the Euro zone. *Journal of Corporate Finance*, 17(5), 1389–1409.
- Price, P. C., & Stone, E. R. (2004). Intuitive Evaluation of Likelihood Judgment Producers: Evidence for a Confidence Heuristic. *Journal of Behavioral Decision Making*, 17(1), 39–57.
- Puranam, D., Narayan, V., & Kadiyali, V. (2017). The effect of calorie posting regulation on consumer opinion: A flexible latent dirichlet allocation model with informative priors. *Marketing Science*, 36(5), 726–746.
- Purnawirawan, N., De Pelsmacker, P., & Dens, N. (2012). Balance and Sequence in Online Reviews: How Perceived Usefulness Affects Attitudes and Intentions. *Journal of Interactive Marketing*, 26(4), 244–255.
- Purnawirawan, N., Dens, N., & De Pelsmacker, P. (2012). Balance and Sequence in Online Reviews: The Wrap Effect. *International Journal of Electronic Commerce*, 17(2), 71–98.
- Racherla, P., & Friske, W. (2012). Perceived “usefulness” of online consumer reviews: An exploratory investigation across three services categories. *Electronic Commerce*

- Research and Applications*, 11(6), 548–559.
- Reynolds, T. J., & Gutman, J. (1979). Laddering theory, method, analysis, and interpretation. *Journal of Advertising Research*, 28(1), 11–31.
- Rodríguez-Díaz, M., Rodríguez-Díaz, R., & Espino-Rodríguez, T. (2018). Analysis of the Online Reputation Based on Customer Ratings of Lodgings in Tourism Destinations. *Administrative Sciences*, 8(3), 51.
- Roedder, D. J., Loken, B., Kim, K., & Monga, A. B. (2006). Brand concept maps: A methodology for identifying brand association networks. *Journal of Marketing Research*, 43(4), 549–563.
- Roodman, D. (2009a). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, 9(1), 86–136.
- Roodman, D. (2009b). Practitioners' corner: A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71(1), 135–158.
- Salehan, M., & Kim, D. J. (2015). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, 81, 30–40.
- Saumya, S., Singh, J. P., Baabdullah, A. M., Rana, N. P., & Dwivedi, Y. K. (2018). Ranking online consumer reviews. *Electronic Commerce Research and Applications*, 29(March), 78–89.
- Schnittka, O., Sattler, H., & Zenker, S. (2012). Advanced brand concept maps: A new approach for evaluating the favorability of brand association networks. *International Journal of Research in Marketing*, 29(3), 265–274.
- Selkirk, E. (1996). The prosodic structure of function words. In J. L. (Brown U. Morgan & K. (Brown U. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 187–214). Lawrence Erlbaum Associates, Inc.
- Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80(2), 159–169.
- Sheldon, P., & Bryant, K. (2016). Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in Human Behavior*, 58, 89–97.
- Shugan, S. M. (1980). The Cost of Thinking. *Journal of Consumer Research*, 7(2), 99–111.
- Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Kumar Roy, P. (2017). Predicting the “helpfulness” of online consumer reviews. *Journal of Business Research*, 70, 346–355.
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, 105(1), 131–142.
- Slovic, P. (1972). From Shakespeare to Simon: Speculations and some evidence. *Oregon Research Institute Bulletin*, 1–19.
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality & Social Psychology*, 48(4), 813–838.
- Snizek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, 84(2), 288–307.
- Social Blade*. (2017). <https://socialblade.com/>

- Spears, N., & Singh, S. N. (2004). Measuring attitude toward the brand and purchase intentions. *Journal of Current Issues and Research in Advertising*, 26(2), 53–66.
- Spenner, P., & Freeman, K. (2012). To keep your customers, keep it simple. *Harvard Business Review*, 90(5).
- SproutSocial. (2018). *Instagram stats*. <https://sproutsocial.com/insights/instagram-stats/>
- Statista. (2020). <https://www.statista.com/forecasts/997051/sources-of-information-about-products-in-the-us>
- Sun, M. (2012). How Does the Variance of Product Ratings Matter? *Management Science*, 58(4), 696–707.
- Sun, Q., Niu, J., Yao, Z., & Yan, H. (2019). Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level. *Engineering Applications of Artificial Intelligence*, 81(December 2018), 68–78.
- Tang, T. (Ya), Fang, E. (Er), & Wang, F. (2014). Is neutral really neutral? The effects of neutral user-generated content on product sales. *Journal of Marketing*, 78(4), 41–58.
<http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc11&NEWS=N&AN=2014-27450-004>
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
- Tirunillai, S., & Tellis, G. J. (2014). Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation. *Journal of Marketing Research*, 51(4), 463–479.
- Toubia, O., Iyengar, G., Bunnell, R., & Lemaire, A. (2019). Extracting Features of Entertainment Products: A Guided Latent Dirichlet Allocation Approach Informed by the Psychology of Media Consumption. *Journal of Marketing Research*, 56(1), 18–36.
- Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing*, 73(5), 90–102.
- Tybout, A. M., Sternthal, B., Malaviya, P., Bakamitsos, G. A., & Park, S. E. B. (2005). Information accessibility as a moderator of judgments: The role of content versus retrieval ease. *Journal of Consumer Research*, 32(1), 76–85.
- Van Hove, G., & Lievens, F. (2007). Social influences on organizational attractiveness: Investigating if and when word of mouth matters. *Journal of Applied Social Psychology*, 37(9), 2024–2047.
- Veloutsou, C., & Delgado-Ballester, E. (2018). New challenges in brand management. *Spanish Journal of Marketing - ESIC*, 22(3), 255–272.
- Vermeer, S. A. M., Araujo, T., Bernritter, S. F., & van Noort, G. (2019). Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media. *International Journal of Research in Marketing*, 36(3), 492–508.
- Vriens, M., Chen, S., & Schomaker, J. (2019). The evaluation of a brand association density metric. *Journal of Product and Brand Management*, 28(1), 104–116.

- Walker, M. A., Mehl, M. R., Moore, R. K., & Mairesse, F. (2007). Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research*, 30, 457–500.
- Wang, F., Liu, X., & Fang, E. (Er). (2015). User Reviews Variance, Critic Reviews Variance, and Product Sales: An Exploration of Customer Breadth and Depth Effects. *Journal of Retailing*, 91(3), 372–389.
- Wang, W., Feng, Y., & Dai, W. (2018). Topic analysis of online reviews for two competitive products using latent Dirichlet allocation. *Electronic Commerce Research and Applications*, 29(January), 142–156.
- Wang, X., Tang, L. (Rebecca), & Kim, E. (2019). More than words: Do emotional content and linguistic style matching matter on restaurant review helpfulness? *International Journal of Hospitality Management*, 77(August 2018), 438–447.
- Weathers, D., Swain, S. D., & Grover, V. (2015). Can online product reviews be more helpful? Examining characteristics of information content by product type. *Decision Support Systems*.
- Westerman, D., Spence, P. R., & Van Der Heide, B. (2014). Social Media as Information Source: Recency of Updates and Credibility of Information. *Journal of Computer-Mediated Communication*, 19(2), 171–183.
- Wickham, H. (2016). *tidyr: Easily tidy data with spread and gather functions*.
- Wickham, H. (2017). ggplot2 - Elegant Graphics for Data Analysis (2nd Edition). *Journal of Statistical Software*, 77(Book Review 2), 2–5.
- Wickham, H., & Francois, R. (2016). *dplyr: A Grammar of Data Manipulation*.
- Winkelmann, P. D. R. (2008). Econometric Analysis of Count Data. In *Econometric Analysis of Count Data*.
- Wong, C. U. I., & Qi, S. (2017). Tracking the evolution of a destination's image by text-mining online reviews - the case of Macau. *Tourism Management Perspectives*, 23, 19–29.
- Wooldridge, J. M. (2011). *Econometric analysis of cross section and panel data* (Issue November).
- Wu, J. (2017). Review popularity and review helpfulness: A model for user review effectiveness. *Decision Support Systems*, 97, 92–103.
- Xu, P., & Liu, D. (2019). Product engagement and identity signaling: The role of likes in social commerce for fashion products. *Information and Management*, 56(2), 143–154.
- Yang, H., & Cho, S. (2015). *Understanding Brands with Visualization and Keywords from eWOM Using Distributed Representation*.
- Yau, K. K. W., Wang, K., & Lee, A. H. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal*, 45(4), 437–452.
- Yazdani, E., Gopinath, S., & Carson, S. (2018). Preaching to the Choir: The Chasm Between Top-Ranked Reviewers, Mainstream Customers, and Product Sales. *Marketing Science*, 37(5), 838–851.
- Yin, G., Wei, L., Xu, W., & Chen, M. (2014). Yin, G., Wei, L., Xu, W., & Chen, M. (2014). Exploring Heuristic Cues for Consumer Perceptions of Online Reviews Helpfulness: The Case of Yelp.com. PACIS 2014 Proceedings, 1–12. Exploring Heuristic Cues for

- Consumer Perceptions of Online Reviews Helpfulness. *PACIS 2014 Proceedings*, 1–12.
- Zaltman, G. (1997). Rethinking Market Research : Putting People Back In. *Journal of Marketing Research*, 34(4), 424–437.
- Zhang, J. (2019). What's yours is mine: exploring customer voice on Airbnb using text-mining approaches. *Journal of Consumer Marketing*, 36(5), 655–665.
- Zhang, Ying, & Lin, Z. (2018). Predicting the helpfulness of online product reviews: A multilingual approach. *Electronic Commerce Research and Applications*, 27, 1–10.
- Zhang, Yuchi, Moe, W. W., & Schweidel, D. A. (2017). Modeling the role of message content and influencers in social media rebroadcasting. *International Journal of Research in Marketing*, 34(1), 100–119.
- Zhou, S., & Guo, B. (2017). The order effect on online review helpfulness: A social influence perspective. *Decision Support Systems*, 93, 77–87.
- Zhu, F., & Zhang, X. (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*, 74(March), 133–148.

3

Mining the Text of Online Reviews to Discover Brand Associations and Build Positioning Maps

1. Introduction

Brand positioning is a crucial step in marketing strategy. Markets are overcrowded with products, and, to simplify the buying process, consumers organize products into categories and position them in their minds (Kotler & Armstrong, 2020). The position of a brand is formed by a complex set of perceptions, images, and emotions that consumers associate with it, and in comparison, with competing products. To position a product, companies need to understand how consumers perceive products in the category. Traditionally, companies gathered information from consumers surveying them. With the growth of the Internet and social media, there is an overwhelming amount of online information for both consumers and firms. In particular, consumers share information through electronic word-of-mouth (eWOM). Whereas there is a large amount of literature about how consumers generate and use this information for purchasing decisions (J. Chevalier & Mayzlin, 2006; Hofmann et al., 2017), less attention has been paid to study how firms can gain insights from it (Verma et al. 2021), as for example for brand positioning.

One of the main types of eWOM are online consumer reviews. Other types of eWOM include comments in social networks (e.g., posts on Facebook and tweets in Twitter), comments in forums, blog entries, etc. Online consumer reviews contain mainly text (Berger et al., 2020), **and can be defined as any “positive, neutral, or negative evaluation of a product, a service, a person, or a brand presumably posted by former customers on websites that host consumer reviews”** (Filieri, Hofacker, et al., 2018). Exploring the text of online reviews offers companies the opportunity to expand and deepen their knowledge in aspects such as consumer preferences, brand image, brand associations, and brand positioning (Balducci & Marinova, 2018b; Hartmann et al., 2019; Kübler et al., 2019). However, the text content of online reviews is qualitative in nature, which has

made it difficult to analyze and to extract meaningful insights in the past (K. Chen et al., 2015). However, the development of high-speed computers and new statistical methods has helped companies and researchers to go one step further in the study of texts and language (C.K. Chung & Pennebaker, 2007; Tausczik & Pennebaker, 2010).

Several text mining techniques are available to explore online reviews, which are usually classified into machine learning and lexicon-based methods (Hartmann et al., 2019; Kübler et al., 2019). The selection of one method over another relies mainly on two aspects: the specific research objectives of the firm (some of them shown in Table 1), and the skills required to conduct the analysis. Machine learning algorithms require higher expertise and greater levels of computational skills, making it difficult for small and medium firms to implement them. In this line, a report developed by Magoulas and Swoyer (2020) reveals that one of the main reasons why companies do not adopt further Artificial Intelligence (AI) algorithms, such as machine **learning, is the “lack of skilled people/ difficulty hiring the required roles”**. On the other hand, lexicon-based methods, which rely on established dictionaries of words, offer the opportunity to analyze the text of online reviews in an easier and more intuitive way, which is more appropriate for small and medium companies.

This research contributes to existing literature by presenting a structured and unified procedure to use the information contained in the text of online consumer reviews to explore brand image and brand positioning. Moreover, we posit that textual variables uncovered from online reviews might be the base for brand segmentation. So far, there are few papers in literature studying brand image and more precisely, brand positioning, through the text of online opinions. By “structured **procedure**” we mean a procedure that defines clear stages to conduct the brand positioning analysis using online reviews from the beginning to the end, focusing on the text mining process to extract brand associations. Moreover, “**unified procedure**” means that we have reviewed the literature in text mining, brand positioning and brand segmentation to put ideas together into one unique study, which finally presents an easy-to-follow guide or procedure combining different techniques for text mining, brand positioning and brand segmentation.

Most papers in literature have adopted multidimensional scales to analyze brand image (Cho et al., 2015; Davis et al., 2009; John et al., 2006; Kim et al., 2003; Low & Lamb, 2000; Malhotra, 1981; Park & Rabolt J., 2009). Others have studied brand image and brand positioning relying on techniques to elicit brand associations. The goal of these studies is to build “associative networks” that can be mapped into a conceptual map. These techniques ranges from most qualitative, such as free associations and free

response questions, to more structured techniques, such as repertory-grid techniques and laddering techniques (Haridasan & Fernando, 2018; Henderson et al., 1998; Olson & Muderrisoglu, 1979; Reynolds & Gutman, 1979). However, the growth of the Internet and the huge availability of online consumer opinions presents an opportunity to study the textual content of those opinions to understand the image and positioning of brands.

In terms of managerial implications, the main contribution of this research is to provide a comprehensive and easy-to-follow procedure, that small and medium companies can follow to understand brand associations and to build positioning maps using data from online consumer reviews. Moreover, we show how those associations can be a base for a brand segmentation strategy, in the way that we show how the market can be segmented to take advantage of the existing perceptions of consumers with respect to one brand relative to competitors' brands. We **say that it is an “easy-to-follow procedure” because** the text mining analysis is based on a lexicon-based method, the Linguistic Inquiry and Word Count (LIWC) developed by Pennebaker et al. (2007). This method is accessible in terms of price and intuitive to use, since the researcher does not need knowledge on machine learning. When using the LIWC, the textual files to analyze are uploaded to the software and it automatically provides the information related to more than 90 textual variables, including psychological associations. For the rest of the key tasks in the research procedure after the text mining analysis (brand positioning and brand segmentation), we also suggest to apply some of the most used approaches, Principal Component Analysis for brand positioning and Hierarchical Clustering for brand segmentation. In this case, these tasks are conducted in the software R, which is a free software for both individuals and companies. Companies can conduct brand positioning and brand segmentation analyses with existing packages created in R for those purposes.

To illustrate the proposed procedure with an empirical application, we use the whole set of online consumer reviews belonging to a category of cosmetics products (blush) available at a popular US cosmetics online retailer on February the 17th 2017. A total set of 62,496 online reviews belonging to 44 different cosmetics brands was analysed. We found that words related to positive emotions are the most common. Moreover, words representing perceptual processes, such as *see* and *feel*, words associated to *body* and those reflecting *time* and *space* issues are also quite related to the product category. Based on these associations, we build a positioning map of the brands in category and identified four main segments of brands.

The remainder of this paper is structured as follows. In section 2, we overview the related literature of brand associations, brand image and brand positioning, and how these

concepts have been measured up to now; we also review the literature of text mining approaches and tools used to uncover emotions and topics from text. In section 3 we present the research procedure proposal to explore brand positioning using online consumer reviews. In section 4, we conduct the empirical analysis to illustrate the research procedure proposal. In section 5, we discuss the findings, by analysing the main managerial implications and highlighting the main limitations and areas for future research.

2. Literature Review

We review literature in three main areas: the concept and relevance of brand associations, and product positioning for marketing strategy; how brand associations and positioning maps have been built until now; and the incipient literature on the use of text mining techniques to identify brand associations, and product attributes from text generated by consumers, from which to build positioning maps.

2.1. The Concept of Brand Image, Brand Associations and Brand Positioning

To understand and contextualize the relevancy of brand associations and product positioning, we have to go back to the brand equity literature, pioneered by Aaker (1991) and Keller (1993). As defined by Aaker (1991), brand associations can be defined as **“brands assets and liabilities that include anything linked in memory to a brand”**. John et al., (2006) consider that consumers might associate a brand with attributes, features, **usage situations, etc.** Keller (1998) define brand associations as **“informational nodes linked to the brand in memory that contain the meaning of the brand for consumers”**. Brand associations are important to companies and to consumers (Low & Lamb, 2000): companies use brand associations to create positive attitudes and to suggest benefits and to position the brand. Consumers use brand associations to process, organize and retrieve information in memory for purchase decision making (Aaker, 1991).

The set of brand associations forms the brand image, which can be defined as the consumer´s perception of the brand´s tangible and intangible associations (Faircloth et al., 2001). Both brand image and brand associations, represent the meaning of the brand to people. This meaning is usually formed from consumers´ own experiences with the brand, from the firm´s marketing mix activities (Aaker, 1991) and from the opinion of other consumers.

Veloutsou and Delgado-Ballester (2018) review what remains unchanged and what has changed in the evolving literature on brand management, and they point out that one of the challenging areas of research in branding is the co-creation of brands with agents that do not work in the company. With no doubt, one of the agents with an increasing voice in last years are consumers, who generate eWOM and participate actively in the communication of brand emotions and associations to other consumers.

Overall, the study of brand associations has captured the attention from both academia and the industry because they represent the base for purchase decisions and brand loyalty (Aaker, 1991; Keller, 1993). From a managerial point of view, creating favorable brand associations is the basis of a successful brand strategy. Several scholars have found that favorable brand image and brand attitudes have a positive impact on purchase intentions (Jalilvand & Samiei, 2012; Kudeshia & Kumar, 2017; Spears & Singh, 2004). As claimed by Henderson et al. (1998), brand associations that evoke positive affect, as well as cognitive considerations of benefits, provide consumers reasons to buy a brand or a product.

Another important way in which associations create value to the firm is by providing a basis for brand/product positioning (Aaker, 1991), **where positioning is defined as “the way a brand/product is defined by consumers on important attributes”, or “the place it occupies in consumer’s minds relative to competing brands/products”** (Kotler & Armstrong, 2020). The positioning of the product, or of the brand, can be done based on the important associations (e.g., attributes, and benefits) consumers use to evaluate the different alternatives available in the market.

We also introduce in this research the concept of brand segmentation. According to Kotler and Armstrong (2020), “market segmentation is the sub-dividing of market into homogeneous sub-sections of customers, where any sub-section may conceivably be selected as a market target to be reached with a distinct marketing mix.” Although market segmentation refers to the process of grouping consumers based on different characteristics, segmentation can also be done for brands or products. In this sense, brands or products can be grouped into groups based on different characteristics such as price, quality, market share, etc. By studying brand segmentation, one can identify the different groups of brands there are in a specific industry based on different characteristics of interest. If we study brand positioning together with brand segmentation one can, for example, identify saturated segments in the industry, where many brands are available, and interesting segments in the industry to enter which are underexploited. As claimed by Henderson et al. (1998), brand segmentation helps

managers to know how can the market be segmented to take advantage of the existing perceptions of consumers with respect to my brand relative to other brands.

2.2. The Measurement of Brand Image, Brand Associations and Brand Positioning

Most studies in literature have approached brand image as a multidimensional concept, which has been measured adopting some multidimensional scales available in literature (Baksi & Panda, 2018; Cho et al., 2015; Davis et al., 2009; John et al., 2006; Kim et al., 2003; Low & Lamb, 2000; Malhotra, 1981; Park & Rabolt J., 2009). We can find plenty of studies using scales to measure brand image as a multidimensional concept.

Some studies use general brand image scales, not adapted to the product category. For example, Martinez and Pina (2009) analyze the effect of brand extensions on brand image, using a brand image scale adapted from previous studies. In the scale, the authors evaluate functional image (e.g., “the products have a high quality”), affective image (e.g., “the brand is nice”) and reputation (e.g., “It is one of the best brands in the sector”). There is also a stream of literature that considers brand image as being related to the product category within which the brand is marketed. For example, Low and Lamb (2000) propose an empirical procedure to test the conceptualization of brand associations on three dimensions: brand image, brand attitude and perceived quality, and find support for it. The authors use a multidimensional scale adapted for each of the nine product categories analyzed (shampoo, soft drinks, mustard, watches, cereals, washing machines, raisin bread, golf clubs and computer games). They use dimensions, such as “I think that the shampoo in this advertisement is: Unfriendly/Friendly, Outdated/Modern and Not Useful/Useful”. In a similar line, Baksi and Panda (2018) analyze destination image of several cities in India using a survey that included 43 dimensions adapted from scales in previous studies (e.g. “safe and secure environment”, “entertainment in festivals” and “physical landscape of the destination”). Vriens et al. (2019) perform also a study of brand associations on two categories, smartphones and beer. They used and compared two different methods to identify brand associations, based on the type of questions raised in surveys (open free association questions or predefined attributes association questions). When using open free association questions, they raised the following question: “Please think about brand X. What pops into your mind? It can be images, feelings, anything at all that you like or dislike, positive or negative thoughts” (Vriens et al., 2019). When using predefined attributes, participants were asked to indicate whether the associated brands to a set of predefined brand attributes, some of them functional and others emotional.

Overall, literature on brand image considers that when measuring brand associations not only physical attributes should be considered, but also functional, emotional and self-expressive benefits c). We show in Table 1 two examples of multidimensional scales used in previous literature to measure brand image, one adopting amore general scale and another developing a scale based on the product category.

As far as brand positioning is concerned, John et al. (2006) review the different techniques available to build brand concept maps. They distinguish between qualitative (consumer mapping) and quantitative (analytical mapping) techniques to build maps of brand associations. In the case of consumer mapping, qualitative techniques are used to obtain the information and brand associations are elicited from consumers, who are then asked to construct networks of these associations as links to the brand and to one another. Analytical mapping involves getting the information from quantitative techniques, such as surveys.

As claimed by Brandt et al. (2011), Brand Concept Maps (BCM) have been used in the marketing literature since 1990s, but the authors pointed out that there is a clear lack of empirical research on quantitative concept mapping techniques and their applications in branding. BCM are based on the assumption that the structure of a concept map reveals the inherent content (concepts and their associations) and relationships (links between concepts and associations) represented in a persons mind (Brandt et al., 2011).

Henderson et al. (1998) pioneered quantitative mapping techniques using mental models called “**associative networks**”. The nodes in this associative network can include aspects such as brand name, product name, features of the product, people and occasions (Aaker, 1996; Henderson et al., 1998). Consumers´ associative networks can contain one or more than one firm or brand. The process of using associative network models to study brand image and positioning has three stages: data elicitation, representation of data as graph-theoretical or spatial structures, and network analytic techniques (Henderson et al., 1998). Regarding data collection, several techniques are used for brand associations´ elicitation, ranging from most qualitative technique, such as free association and free response (Olson & Muderrisoglu, 1979) to more structured techniques, such as repertory-grid (Kelly, 1991), laddering (Reynolds & Gutman, 1979) and pairwise similarities (Hauser & Koppelman, 1979).

One of the best-known techniques designed to understand mental models or brand associative networks is the ZMET developed by Zaltman (1997), which combines visual and narrative aspects and consists of three stages: elicitation, mapping and aggregation.

In the elicitation stage, several participants are recruited for in-depth interviews. During the interview, the repertory grid method and laddering process are used to make participants to elicit associations. At the mapping stage, participants are asked to create a map illustrating the connection among important associations. Finally, at the aggregation stage data are codified and associations are chosen based on how frequently they are mentioned. The ZMET as a qualitative nature and it is very labor intensive, since it is based on interviews and requires interviews to be thoroughly trained in cognitive psychology. Besides, respondents should be willing to participate in interviews. Henderson et al. (1998) and Roedder et al. (2006) were pioneered in developing quantitative tools to capture brand image. Their methods were based on capturing data from larger sample sizes using surveys. As Stated by Brandt et al. (2011) one of the disadvantages of quantitative methods (compared with ZMET or other qualitative techniques) is that they put emphasis on the conscious parts of brand evaluation, while qualitative techniques based enable the researcher to elicit also “hidden” or unconscious information. They advice to combine the strengths of both, quantitative and qualitative methods.

In general, most literature exploring brand image focuses on one or a few products or brands in a category to conduct the research (Baksi & Panda, 2018; Low & Lamb, 2000). Thus, their brand image analyses are not followed by a brand positioning study. In the online reviews literature, we find a few number of papers that explore brand positioning through the text of online reviews. Guo et al. (2017a) studied the positioning of hotels using Correspondence Analysis (CE) and based on the attributes extracted from online reviews using the machine learning method Latent Dirichlet Allocation (LDA). Liu et al. (2017) used also LDA to extract the main attributes associated to brands in different categories (fast food, department store, footwear, telecommunications and electronics). They analyze and compare between four brands in each category. To conduct the analysis, they computed the frequency in which attributes were related to each brand. However, they do not draw findings into perceptual maps. Wang et al. (2018) apply also a LDA analysis to compare attributes associated to two competing products (two wireless mice). Along our empirical analysis, we pretend to provide a broader picture of how brand associations elicited in online reviews’ texts can not only be used to understand brand image but also to understand brand positioning and serve as a basis for segmentation.

2.3. Mining the Text of Online Reviews to Explore Brand Associations and Brand Image

Although the study of brand image has traditionally been conducted adopting either multidimensional scaling throughout surveys or using qualitative techniques to elicit brand associations and build conceptual maps, the wide availability of consumers' opinions in the digital environment together with the increasing number of available tools to study unstructured data, as it is the case of online texts, has shifted the attention to the study of brand image using the language and narrative used in the text of eWOM. However, literature is still scarce.

The use of eWOM data over survey data to analyze brand associations and brand positioning has some advantages. One of the most relevant characteristics of eWOM is that it is spontaneous (Marchand et al., 2017; Yang & Cho, 2015), so consumers are more likely to express their true brand perceptions and to inform about their behaviors. It occurs without direct prompting or influence by marketers and it is usually motivated by a desire to help others, warn others or to communicate status (Kozinets et al., 2010). Moreover, in contrast to surveys that are a primary source of information, eWOM is a secondary source of information that is widely available online, so it is quite easy to collect with techniques such as web scrapping and social networks APIs. Nevertheless, we can still find advantages for the use of survey data. First, questionnaires are developed by the researcher so they can be carefully designed to explore specific research objectives. Second, since survey data has been used for a long time, companies and scholars have more expertise in the required techniques to design and analyze questionnaires. Besides, because eWOM is not sought, it might not address the objectives of the research. Finally, and probably the main two disadvantages of eWOM are that, first, due to the huge quantity of texts, firms and scholars might suffer from information overload situations; and second, the analysis of eWOM texts is a relatively new domain, so companies and scholars do not usually have the knowledge and expertise required.

The use of eWOM over qualitative techniques, such as the ZMET, to study brand image has also advantages. The main benefit is that the researcher can collect huge amounts of online consumer opinions in an automatic way. Therefore, the sample size to study brand associations is much larger than in qualitative techniques, which rely mainly on in-depth interviews. Moreover, as eWOM is spontaneous, consumers are likely to express real feelings and perceptions. One of the main disadvantages of eWOM over qualitative techniques might be based on the elicitation of unconscious associations. Consumers are not likely to express those most unconscious associations with brands in online opinions.

Studying the textual content of eWOM offers scholars and companies a huge opportunity to deepen in the study of brand image. The study of narrative and language persuasion has been widely explored in several research domains, such as communication, psychology and marketing (Areni, 2003; Hamby et al., 2015; Holtgraves & Lasky, 1999; Li et al., 2019). Tausczik and Pennebaker (2010) claim that language is the way in which people's express their internal thoughts and emotions. In this line, psychologists have found that people's personality can be recognized by analyzing linguistic cues, which include aspects such as topics discussed, style, syntax, lexicon and type of speech (Walker et al., 2007). Although the study of narrative and persuasion was rooted outside the digital environment, **the concept of "text mining"** has arise to refer to the process of extracting useful and meaningful information from large amounts of text that can be found online (Netzer et al., 2012). Text mining is composed by the set of techniques and technologies that are used to explore large amounts of text, automatically or semi-automatically, and discover repetitive patterns, trends or rules that explain the behavior of the text.

Scholars who have studied the textual content of online reviews have conducted the study in different product settings (e.g. hotels, restaurants and pc components) and using a wide range of text mining techniques, such as lexicon-based techniques (e.g. Linguistic Inquiry and Word Count, developed by Pennebaker et al., 2007) and unsupervised machine learning techniques (e.g. Latent Dirichlet Allocation, LDA). The text mining techniques selected for each study depend on factors such as the type of data and the research objectives. Those papers whose objective is to analyze brand image using online opinions have used different processes and techniques for text mining. For example, Gensler et al. (2015) study the brand image of McDonald's using a machine learning method of natural language processing (NLP), which is called tokenization. Ahani et al. (2019) use a machine learning method, the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) adapted for text mining to study the satisfaction of tourists in the Canary Islands. Wong and Qi (2017) study the destination image in the tourism sector, applied to the case of Macau. The authors use two specific software, *NVivo10* and *IBM_ManyEyes* for text mining. Wang et al. (2018) use the natural language processing algorithm Latent Dirichlet Allocation (LDA) to analyze the topics or associations made by consumers to two competing products. Kim and Kang (2018) use two machine learning algorithms, the Latent Semantic Analysis (LSA), which is a supervised algorithm, and the LDA, which is an unsupervised algorithm, to study the attributes associated to Korean cosmetics products and other non-Korean products. Overall, it can be observed that extant literature using text mining to study brand image

usually focuses on exploring brand associations with one or two unique products or brands. However, text mining can be also used to compare between products and brands and to build brand positioning maps. Moreover, literature is quite heterogeneous in terms of processes and techniques used for text mining, so it is not clear enough to guide prospective scholars and practitioners in their text mining analysis process.

Table 1 shows a summary of some of the most common text mining tools used in previous literature based on the research objectives, adapted from Berger et al. (2019). In this line, a few number of papers have conducted research to unify disjoint literature on unstructured data analysis and text mining in social media (Balducci & Marinova, 2018b; Berger et al., 2020; Hartmann et al., 2019; Kübler et al., 2019). Balducci and Marinova (2018) contextualize the different types of unstructured data, such as images, text, video and voice, providing a synthesis of the characteristics of each type of data and its use in different marketing areas. Other scholars, such as Berger et al. (2019), focus on a specific type of unstructured data, which is text, and provide a review and discussion on the different methodologies used in text mining analysis. Hartmann et al. (2019) and Kübler et al. (2019) compare the performance of several text classification methods, including lexicon-based approaches (e.g. LIWC by Pennebaker et al., 2007) and machine learning algorithms (e.g. random forest and naïve Bayes).

Overall, a common trend that can be identified in literature is that papers applying text mining techniques can be classified into those using lexicon-based approaches and those using machine learning algorithms. Hartmann et al. (2019) conclude that, normally, machine learning algorithms perform a bit better than lexicon-based approaches, but the accuracy of the performance is, in many cases, only slightly better. Thus, the selection of one type of approach or another must rely on the research objectives and resources available, considering that lexicon-based methods are quicker and easier to apply. Managers, in many cases, should make a tradeoff between cost and a deeper understanding of content in social media (Kübler et al., 2019).

Considering that Hartmann et al. (2019) claim that machine learning methods' performance is only slightly better than lexicon-based methods, we incorporate in our research procedure a lexicon-based approach, the Linguistic Inquiry and Word Count (LIWC) developed by Pennebaker et al. (2007). Since we are not looking for uncovering very specific features with our text mining analysis, general brand associations can be conveniently extracted using a lexicon-based method, which does not require machine learning knowledge.

Although several lexicon-based tools are available, we propose using the LIWC due to several reasons. First, it is an easy to implement tool. It is a software whose license is accessible for everyone. Moreover, the handling of the texts in the software is very intuitive. The researcher should import the texts from any type of filetype (e.g., word, pdf, excel, etc.) and LIWC automatically analyse the texts and provides scores for each text and for each output variable. Thus, LIWC does not require any knowledge on machine learning methods. Second, it provides information of around 90 categories of textual variables, including several variables uncovering psychological associations. Thus, it is a powerful tool to extract a huge quantity of insights into the brand positioning analysis. Other lexicon-based available tools, such as *SentiWordNet* (Baccianella et al., 2010) and *AFINN* (Nielsen, 2011) focus on conducting sentiment analysis, with the aim of uncovering if the content of the text is positive, negative or neutral. However, they do not provide other type of information, such as psychological associations expressed by consumers in their texts. Moreover, rather than focusing on uncovering product attributes, as it has been widely done in previous literature using machine learning algorithms (Y. Guo et al., 2017b; Moon & Kamakura, 2017; W. Wang et al., 2018), the use of LIWC allow us to extract other type of brand associations from online reviews. As far as we know, there are no papers using the output of LIWC to analyse brand positioning and segmentation. Finally, LIWC has already been validated in a lot of studies in different domains, such as Psychology and Marketing (M. A. Cohn et al., 2014; Ireland et al., 2011b; Ludwig et al., 2013b).

Table 1

Most common text mining approaches and tools used in literature based on the research objectives (Berger et al., 2019)

Main Objective	Specific research objectives	Main Text mining tools	Papers
Entity (Word) extraction Extracting and identifying single words	Sentiment analysis Extract psychological associations Consumer and market trends	Dictionaries and lexicons (e.g., LIWC, NRC Emotion Lexicon, SentiWordNet) Machine learning classification tools (e.g., deep learning)	Lee and Bradlow (2011b) Kübler et al. (2019) Mahr et al. (2019) Zhang (2019) Ludwig et al. (2013)
Topic extraction Extracting the main topics discussed in the text	Summarizing the discussion Uncovering perceived product features Identifying consumer needs and market trends	Machine learning algorithms: LDA (Latent Dirichlet Allocation) LSA (Latent Semantic Analysis)	Tirunillai and Tellis (2014) Chen et al. (2015b) Guo et al. (2017) Puranam et al. (2017) Heng et al. (2018)

			Wang et al. (2018)
Relation extraction	Identifying problems mentioned with specific product features	Supervised machine learning	Schnittka et al. (2012)
Extracting and identifying relationships between words	Identifying product attributes mentioned positively/negatively	Deep learning	Gensler et al. (2015)
	Identifying events and consequences.		Oded Netzer et al. (2012)

3. Research Procedure

The main objective of the research is to show and illustrate a unified and structured procedure to explore brand image from the text of online reviews, with the goal of building a brand positioning map and identifying possible brand segments. Several tools and techniques are available for text mining purposes and the selection of the technique must be based on the research objective and expertise. Our procedure relies on a lexicon-based method, in this case LIWC (Pennebaker et al., 2015), because these type of methods are easier to implement than machine learning tools (Hartmann et al., 2019; Kübler et al., 2019). Thus, our proposed procedure can serve as an easy-to-follow guide for small and medium firms, which might not have the required knowledge and expertise to carry out a more sophisticated machine learning tool for text mining but want to benefit from the hidden content of their available eWOM data.

Table 2 briefly explains the issues involved at each stage of the proposed research procedure and illustrates how our research has addressed them.

Table 2

Proposed research procedure

Stage	Issues involved at each stage	Proposed approach (used in this research)
Data acquisition	<ul style="list-style-type: none"> • Downloading online reviews and other relevant information from the website of interest. Several techniques are available, for example, web scrapping. • Recording other relevant information from other websites (e.g., industry reports, google trends and analytics websites) 	<ul style="list-style-type: none"> • We downloaded online reviews and product information from a US cosmetics retailer website. • The website www.socialblade.com was used to get information about brand followers at Instagram. • A database of 62,496 online reviews of 131 products and 44 brands was built (all belonging to the “blush” category).
Data pre-processing	<ul style="list-style-type: none"> • Importing data to a statistical software (e.g., Excel, R and Stata). • To apply some machine learning text mining techniques, more pre-processing should be conducted: 	<ul style="list-style-type: none"> • Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2015) automatically conducts the necessary pre-processing steps, such as tokenization.

	<ul style="list-style-type: none"> ○ Tokenization. Deciding the unit of analysis (e.g., word or sentence). ○ Removing stop words. Depending on the research objective, some general words can be removed. 	
Text Mining	<ul style="list-style-type: none"> ● Text mining methods: <ul style="list-style-type: none"> ○ Machine learning algorithms (e.g., Latent Dirichlet Allocation). ○ Lexicon-based methods, such as LIWC (Pennebaker et al., 2015), WordStat (Peladeau, 2016) and NRC Emotion Lexicon (Mohammad & Turney, 2010) 	<ul style="list-style-type: none"> ● We used the lexicon-based software LIWC (Pennebaker et al., 2015). ● From around 90 textual variables available at the LIWC, we selected the 26 variables that belong to the “Psychological Processes” group.
Data aggregation	<ul style="list-style-type: none"> ● When interested in conducting either product-level or brand-level analyses, review variables should be aggregated. <ul style="list-style-type: none"> ○ <i>Issue to consider</i>: Deciding which products/brands should be included in the analysis. Some products/brands might have such a small number of online reviews, that the average might not be meaningful. 	<ul style="list-style-type: none"> ● We used R to conduct analyses done from this stage onwards. ● Review textual variables and other non-textual product variables were aggregated to a brand-level. <ul style="list-style-type: none"> ○ In this research, we left out brands having less than 9 online reviews (those in the quantile 5% for the variable <i>Brand_reviews</i>).
Brand Positioning	<ul style="list-style-type: none"> ● Dimensionality reduction techniques are the most used, such as correspondence analysis (CA), multidimensional scaling (MS) and principal component analysis (PCA). 	<ul style="list-style-type: none"> ● We conducted a PCA using the 26 brand aggregated textual variables. Several issues are addressed here: <ul style="list-style-type: none"> ○ Exploring correlations between textual variables. ○ Analysing the explained variance by each principal component (PC). ○ Computing the textual variable loadings for each PC and the brand loadings on each PC. ○ Draw a two-dimensional (2D) and a three-dimensional (3D) perceptual map.
Brand Segmentation	<ul style="list-style-type: none"> ● Different types of clustering methods are available (e.g., hierarchical clustering and k-means clustering). ● Clustering results can be graphically represented using dendograms and within perceptual or positioning maps. ● Clusters can be described by analysing variable means. 	<ul style="list-style-type: none"> ● Hierarchical clustering is used in this research. Some issues are addressed here: <ul style="list-style-type: none"> ○ Identifying the optimal number of clusters (k). Several methods can be used, we took the Elbow method. Our optimal number of clusters is k=4. ○ Analysing the clustering output in a dendogram. ○ Plotting the clustering output in the 2D perceptual map resulting from PCA to see the positioning of the brands in the map. <ul style="list-style-type: none"> ○ Describing the clusters using textual and non-textual variables means at each cluster.

The next section describes the data used to conduct the research and illustrates how the main steps of the proposed research procedure have been addressed: data acquisition, text mining, brand positioning and brand segmentation.

4. Empirical Analysis

4.1. Data Acquisition

To carry out our research, we collected a total of 62,496 online consumer reviews from a US cosmetics retailer website, which was placed in the top-50 shopping sites in the US in March 2017 according to *alexa.com*. The collected database represents the entire set of online reviews available at the website under the category of “blush” products on February the 17th 2017. These reviews belong to a total of 131 products of 44 different cosmetics brands.

Figure 1 shows a typical online consumer review at the cosmetic’s online retailer website, where we can observe both textual and non-textual cues. Review text is marked in blue in Figure 1 since it is the part of the review in which we are focusing in this research.

Figure 1. Typical online review at the online retailer website

Besides collecting the text of individual online reviews, we collected other product and brand non-textual information: brand name, product price, product bestselling ranking at the website, product average rating and product number of reviews. The bestselling rank represents a snapshot of sales at the online retailer for up to a week. The product sales rank is inversely related to its sales, which means that the first product in the sales rank in a specific product category is the one with highest sales during the previous week. To account for brand popularity, we also gathered information about the number of each brand at Instagram (*Social Blade*, 2017). Instagram was chosen since it is the fastest growing social network site globally and one of the most used by both companies and users to share photographs and videos in the beauty industry (Sheldon & Bryant, 2016; SproutSocial, 2018).

Table 3 shows the brands available under the category of blush at the online retailer together with some descriptive statistics of brand non-textual variables.

Table 3

Cosmetics brands included in the research and main descriptive statistics of non-textual variables

Brand	Products	Price (gr)	Rating	Bestselling_rank	Reviews	Followers
bareMinerals	2	23.7	4.6	25.5	1633.3	398,280
BECCA	3	5.1	4.5	43.5	358.4	1,511,811
Benefit Cosmetics	11	3.2	4.4	51.4	1613.4	5,865,273
Bite Beauty	1	5.1	4.5	3.0	1548.0	288,840
Black Up	1	7.8	4.7	127.0	21.0	56,769
Bobbi Brown	5	6.8	4.5	32.4	467.9	2,077,698
BURBERRY	3	7.2	4.6	62.2	46.9	8,698,164
Chosungah 22	3	2.2	3.3	116.7	12.6	11,333
Ciaté London	1	3.2	4.3	118.0	14.0	206,371
CLINIQUE	3	5.9	4.5	26.9	607.3	1,214,278
Dior	4	6.0	4.6	44.0	442.8	13,957,736
Estée Lauder	1	4.4	3.6	115.0	18.0	1,642,483
Giorgio Armani Beauty	1	2.4	3.7	33.0	89.0	110,379
Givenchy	2	6.1	4.6	75.6	116.2	7,647,395
Guerlain	4	6.0	4.1	92.4	12.8	529,655
Hourglass	4	9.9	4.4	12.2	1032.0	760,788
ILIA	1	6.8	4.6	56.0	18.0	45,251
KEVYN AUCOIN	3	4.7	4.1	61.9	39.7	196,188
Lancôme	3	6.8	4.6	64.0	179.5	1,461,331
Laura Mercier	4	6.9	4.5	61.9	571.9	1,134,996
MAKE UP FOR EVER	5	9.0	4.4	21.9	573.6	3,382,912
Marc Jacobs Beauty	2	5.6	4.6	25.7	1165.7	3,368
MILK MAKEUP	2	2.4	4.2	56.1	56.6	152,732
NARS	8	5.7	4.6	4.9	12765.1	4,083,316
NUDESTIX	6	11.9	4.3	87.6	93.6	72,033
Perricone MD	1	4.1	4.5	135.0	238.0	35,715
Retailer brand	9	2.8	4.3	28.2	909.5	10,474,573
rms beauty	1	5.2	4.8	25.0	4.0	100,986
Shiseido	1	5.3	4.1	132.0	8.0	198,447
Smashbox	4	4.5	4.1	66.1	760.8	2,777,377
stila	3	5.7	4.4	67.3	1025.8	1,960,925
Supergoop!	1	4.9	4.5	102.0	10.0	17,486
surratt beauty	1	8.0	4.0	41.0	51.0	23,465

tarte	5	4.5	4.6	18.5	3578.0	5,873,000
Tata Harper	1	8.4	4.1	57.0	41.0	83,349
The Estée Edit	1	4.7	4.1	114.0	33.0	1,642,483
TOM FORD	2	12.2	4.2	91.9	20.5	8,468
Too Cool For School	3	1.9	3.8	118.9	17.7	10,836
Too Faced	4	4.8	4.2	17.2	517.1	8,210,500
trèStiQue	2	5.6	4.7	117.0	1.7	15,935
Urban Decay	3	2.7	4.3	25.0	1202.0	7,690,863
Visearth	1	3.3	5.0	70.0	9.0	72,634
Wander Beauty	2	7.1	4.1	54.8	45.4	33,929
Yves Saint Laurent	2	4.2	4.4	35.3	199.7	2,246,873

(1) Number of Instagram followers on February the 17th 2017

4.2. Data Pre-processing

The software LIWC automatically conducts the pre-processing steps required for the text mining analysis, such as removing punctuation marks and tokenization. Tokenization is a way of separating a piece of text into smaller units called tokens. Tokens can be either words, characters, or subwords (n-gram characters). In the case of LIWC, it uses word tokenization, since the unit analysed is the word. Each output variable provided by LIWC is based on a count of words belonging to the predefined dictionary.

4.3. Text Mining: Linguistic Inquiry and Word Count (LIWC)

Of the approximately 90 variables of information provided by LIWC, some are straightforward, such as the category of articles and the category of personal pronouns. However, other emotional and psychological dimensions are more subjective⁵. The validity of the software has been confirmed in more than 100 studies that have applied this methodology, analysing online content such as instant messaging (Ireland et al., 2011b; Ludwig et al., 2013b) and online blogs (M. a. Cohn et al., 2014). Using the word count strategy, the linguistic indicators for each LIWC variable are calculated by the percentage of words that match the pre-defined dictionary. For example, to measure the degree of positive emotions in an online review, LIWC calculates the total number of times the words belonging to positive emotions in the pre-defined dictionary (e.g., “love”,

⁵ In those cases, human judges were required to evaluate the words suitable for each category. For subjective categories, an initial set of word candidates for each category was built from dictionaries, thesauruses, questionnaires and lists made by research assistants (Tausczik & Pennebaker, 2010). Then, groups of three judges independently rated if each candidate word was appropriate to each category. Finally, a word remained in the category if two out of the three judges agreed it should be included. A word was deleted from the category if at least two of three judges agreed it should be included. The final agreement of judges was 100%.

“nice” and “beautiful”) appear in the review and it is divided by the total number of words in the online review.

Selection of LIWC Variables

LIWC provides a total of approximately 90 output variables. Some of them are general descriptors (e.g., words per sentence), others are standard linguistic dimensions (e.g., articles and auxiliary verbs) and 53 variables belong to the so called “Psychological Processes” group. Within this group of 53 variables, there are 10 general variables or categories and 43 more specific variables or sub-categories. In this research, we use a set of 26 variables, which belong to 7 general variables or categories under the “Psychological Processes” group: *affect*, *social processes*, *perceptual processes*, *biological processes*, *drives*, *relativity* and *personal concerns*. Besides general descriptors and standard linguistic dimensions, some LIWC variables in the “Psychological Processes” group have been excluded from this analysis because they reflect aspects related to the narrative style of the text, which is the focus of this research. It is the case of *informal language* and *time orientation*, which have been excluded from our analysis because they represent stylistic features of the text. In the same line, the subcategory of *death*, which belongs to the category of *personal concerns*, has not been incorporated into our research because it is not relevant in our cosmetics context. Under the overall category of *affect*, LIWC provides two categories: *PosEmotions* and *NegEmotions*. The category of *NegEmotions* is in turn divided in three groups: *anger*, *anxiety* and *sadness*. However, the category of *PosEmotions* is not broken-down in more precise subcategories. To ensure a balance set of sentiments, we only use the general categories of *PosEmotions* and *NegEmotions* as variables in our empirical analysis. However, the excluded categories could be incorporated in other studies if they are in line with the specific research objectives.

4.4. Data Aggregation

The LIWC was used to analyse the text of every online review in our database, which is composed by a total of 62,496 online reviews belonging to 44 brands. However, to explore brand positioning and brand segmentation, we worked with brand average variables, therefore the LIWC output, which represents scores for each review, was aggregated into a brand level, following the equation:

$$T_b = \Sigma T_n / N_b \quad \text{Equation 1}$$

Tb represents the average textual score of brand b ; Tn represents the textual score of review n of brand b and N_b records the total number of online reviews of brand b .

We use brand average values to study patterns. In this line, some scholars, such as Li and Hitt (2008), establish a minimum number of online reviews needed to determine a trend from online reviews. If a brand has a very few number of online reviews, the average might not be meaningful, and a trend cannot be properly determined. Therefore, we decided to exclude from the analysis those brands having less than nine reviews ($N < 9$). This threshold corresponds to brands that belong to the quantile 5% in terms of brand average number of reviews. In our case, three brands are excluded from the analysis: trèStiQue, rms beauty and Shiseido. Therefore, instead of using 44 cosmetics brands for our analysis, the empirical analysis is conducted using 41 brands.

Table 4 reports the descriptive statistics of the 26 textual variables used in this research, together with examples of representative words of each category and descriptive statistics, once LIWC review scores are aggregated to brand averages. N represents the number of brands analysed. As they represent brand averages, the minimum statistic belongs to the brand having the minimum average score at each textual variable and the maximum statistic belongs to that brand having the highest average score at that each feature. For example, the variable *PosEmotions* measures the degree of positive emotions associated to each specific brand. A high value in this category means that consumers associate the brand with positive experiences.

Table 4

Textual variables extracted from LIWC and brand average descriptive statistics

Variable	Example of words	N	Mean	Sd	Min	Max
<i>Affect</i>						
PosEmotions	Amazing, benefit, excellent, fair	41	7.324	1.232	3.677	9.720
NegEmotions	Afraid, anxious, cruel, despair	41	0.659	0.238	0.129	1.244
<i>Social processes</i>						
Family	Cousin, honeymoon, marry, husband	41	0.051	0.061	0.000	0.309
Friend	Beloved, best friend, bud, classmate	41	0.052	0.056	0.000	0.251
Female references	Bride, daughter, ex-wife, girl	41	0.110	0.071	0.000	0.292
Male references	Boy, brother, fellow, man	41	0.020	0.020	0.000	0.089
<i>Perceptual processes</i>						
See	Appear, beauty, colour, shine	41	5.313	1.293	2.905	10.408
Hear	Listen, noise, quiet, speak	41	0.120	0.064	0.000	0.285
Feel	Cold, dry, hard, hot	41	2.080	0.446	1.458	4.203

<i>Biological processes</i>						
Body	Cheek, eye, face, facial	41	2.568	1.039	1.168	5.620
Health	Acne, allergy, pain, fitness	41	0.143	0.099	0.000	0.514
Sexual	Lover, nude, sexy	41	0.106	0.171	0.000	0.788
Ingestion	Diet, eat, fat, food	41	0.248	0.202	0.000	0.838
<i>Drives</i>						
Affiliation	Belong, colleague, reunion, party	41	1.455	0.475	0.301	2.400
Achievement	Able, ambition, confident, proud	41	1.317	0.233	0.864	1.863
Power	Beat, celebrity, comply, win	41	1.246	0.315	0.513	2.246
Reward	Achieve, advantage, benefit, earn	41	2.232	0.498	1.239	3.597
Risk	Alarm, avoid, dangerous, doubt	41	0.210	0.127	0.000	0.758
<i>Relativity</i>						
Motion	Approach, arrive, attend, carry	41	1.400	0.322	0.521	2.069
Space	Anywhere, back, big, broad	41	5.701	0.666	4.244	8.202
Time	After, age, always, never	41	3.829	0.765	2.002	6.235
<i>Personal Concerns</i>						
Work	Business, class, company, student	41	1.589	0.414	0.921	2.632
Leisure	Art, band, café, party	41	0.254	0.135	0.000	0.644
Home	Family, home, house, neighbour	41	0.048	0.048	0.000	0.192
Money	Affordable, bargain, buy, cheap	41	1.236	0.359	0.637	2.372
Religion	God, bless, demonic, karma	41	0.039	0.036	0.000	0.128

Figure 2 shows the mean values of the textual variables to visualize the relevancy of each type of consumer association in the whole category of blush online reviews. We observe that online reviews in this category are highly associated to words representing *posEmotions*, which might indicate that those consumers writing online reviews are quite satisfied with the brand. Besides, associations to *space* and *time* issues are also quite common. This might suggest that consumers make references to product usage experiences (where, when and process). Some perceptual (*see* and *feel*) and *body* associations are also quite relevant, which make sense in the context of cosmetics consumption, where perceptual and body-related experiences are likely to be important in the product usage. Associations related to *affiliation*, *achievement*, *power* and *reward* are also quite relevant in our setting, so one might think that consumers experience feelings such as fulfilment or social recognition when using blush products. In terms of personal concerns, consumers usually associate brands to *work* and *money* experiences.

Figure 2. Mean values of textual variables (brand averages)

Table 5 shows descriptive statistics of the pooled textual variables without brand aggregation, so they reflect statistics at a review-level. In this case, *N* represents the total number of online reviews in our database, the minimum statistic represents the review with the lowest score at each textual variable, while the maximum statistic belongs to the review with the highest score at each textual variable. For example, in the case of *NegEmotions*, online reviews score, on average, 0.590 out of 100, so the presence of negative emotions in our database is very small.

Table 5

Descriptive statistics of textual variables at a review-level (no brand averages)

	N	Mean	St. Dev.	Min	Max
PosEmotions	62,496	8.05	6.06	0.00	100.00
NegEmotions	62,496	0.59	1.35	0.00	33.00
Family	62,496	0.06	0.45	0.00	25.00
Friend	62,496	0.06	0.42	0.00	17.00
Female	62,496	0.15	0.81	0.00	25.00
Male	62,496	0.02	0.27	0.00	12.00
See	62,496	5.57	4.90	0.00	75.00
Hear	62,496	0.15	0.66	0.00	20.00
Feel	62,496	1.85	2.58	0.00	42.86
Body	62,496	2.66	3.02	0.00	50.00
Health	62,496	0.14	0.70	0.00	25.00
Sexual	62,496	0.33	1.15	0.00	33.00
Ingest	62,496	0.16	0.73	0.00	25.00
Affiliation	62,496	1.83	3.31	0.00	100.00
Achievement	62,496	1.35	2.30	0.00	50.00
Power	62,496	1.17	2.08	0.00	40.00
Reward	62,496	2.65	3.55	0.00	100.00
Risk	62,496	0.18	0.69	0.00	20.00
Motion	62,496	1.41	2.29	0.00	44.00

Space	62,496	5.51	4.35	0.00	60.00
Time	62,496	4.25	4.28	0.00	50.00
Work	62,496	1.41	2.43	0.00	50.00
Leisure	62,496	0.23	0.88	0.00	25.00
Home	62,496	0.05	0.37	0.00	17.00
Money	62,496	1.13	2.21	0.00	50.00
Religion	62,496	0.05	0.45	0.00	25.00

To better understand the LIWC output, Table 6 shows some examples of online reviews with high scores at some LIWC variables.

Table 6

Examples of online reviews with high scores at some LIWC categories

Variable	Review text	Review score (in %)	Variable Mean (in %)
NegEmotions	<i>I HATE Nars Orgasm! Not color payoff...its the worst.</i>	20.00	0.59
Family	<i>I bought several of these last Valentine's Day for me and my daughter and daughter-in-law and LOVE it. It is kind of a blush, a bit of a bronzer, but mainly just adds this luminescent glow of shimmer that is absolutely flattering. I liked it so much that I just bought them as birthday gifts for my niece and sister-in-law</i>	9.84	0.06
Feel	<i>Silky smooth and beautifully pigmented on my brown skin.</i>	33.33	1.85
Power	<i>Subtle shimmer, rich color, luxurious feel. Top notch!</i>	25.00	1.17
Leisure	<i>I recently wore this to a pool party and it stayed put! WINNER</i>	15.38	0.23
Money	<i>Wonderful product, worth every penny.</i>	40.00	1.13

4.5. Brand Positioning: Principal Component Analysis (PCA)

In order to create a brand positioning map, a Principal Component Analysis (PCA) involving 41 brands on 26 textual categories was carried out. PCA is a statistical method used to analyse interrelationships among a large number of variables, which are usually correlated, by reducing the dimensionality of the multivariate data and not losing important information (Al., 2010). To do so, the PCA method creates new variables, called principal components, which are a linear combination of the original variables. The first component records as much of the variance as possible from all variables as a single linear function. The second component captures as much variance as possible that remains after the first component, and so on (Chapman & Feit, 2015). PCA results in a perceptual map that shows the relative positioning of all brands.

Exploring Correlations

PCA is usually conducted when variables are correlated with each other. Correlations between textual variables are shown in Table 7. We observe some high positive correlations between variables such as *Space/Work*, *Feel/Power* and *Reward/PosEmotions/Affiliation*, *NegEmotions/Risk* and *Health/Home*. Therefore, the use of PCA is justified to avoid the problem of multicollinearity.

Table 7

Correlation matrix between textual variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1.PosEmotions	1.00																									
2.NegEmotions	-0.31*	1.00																								
3.Family	0.31*	0.38*	1.00																							
4.Friend	0.16	-0.13	-0.21	1.00																						
5.Female	0.57*	-0.12	0.48*	0.00	1.00																					
6.Male	-0.01	0.00	-0.07	0.03	-0.16	1.00																				
7.See	0.42*	-0.28*	-0.06	-0.03	0.41*	-0.17	1.00																			
8.Hear	0.07	0.01	0.02	0.15	0.07	0.16	0.15	1.00																		
9.Feel	-0.06	-0.07	-0.14	-0.10	-0.07	0.21	0.27*	0.21	1.00																	
10.Body	0.15	-0.04	0.30*	-0.12	0.04	-0.14	-0.27*	-0.04	-0.22	1.00																
11.Health	0.12	-0.02	0.27*	-0.11	0.07	0.26*	-0.35*	-0.06	-0.03	0.45*	1.00															
12.Sexual	0.12	0.05	-0.09	0.20	-0.07	0.04	-0.06	0.40*	-0.15	0.28*	-0.07	1.00														
13.Ingest	0.08	0.16	0.30*	-0.35*	-0.04	-0.01	-0.25	-0.02	0.14	0.61*	0.57*	-0.09	1.00													
14.Affiliation	0.84*	-0.22	0.22	0.16	0.51*	0.05	0.23	0.05	-0.38*	0.14	0.13	0.19	-0.05	1.00												
15.Achievement	-0.31*	-0.23	-0.34*	0.14	-0.42*	-0.04	-0.36*	0.03	-0.22	0.11	0.00	0.01	-0.10	-0.16	1.00											
16.Power	-0.37*	0.19	-0.12	0.05	-0.30*	0.24	0.05	-0.12	0.42*	-0.42*	-0.21	-0.14	-0.22	-0.55*	-0.07	1.00										
17.Reward	0.64*	-0.23	0.25	0.21	0.53*	-0.11	0.02	-0.20	-0.21	0.26*	0.21	0.20	-0.01	0.61*	-0.14	-0.41*	1.00									
18.Risk	-0.25	0.53*	0.39*	-0.21	0.11	-0.16	-0.43*	-0.11	-0.35*	0.24	0.42*	-0.14	0.34*	-0.08*	0.11	-0.26*	0.02	1.00								
19.Motion	-0.14	-0.26*	-0.29*	0.27*	-0.13	-0.14	-0.48*	-0.09	-0.32*	0.09	0.06	0.09	-0.15	0.13	0.53*	-0.37*	0.27*	0.13	1.00							
20.Space	-0.66*	0.05	-0.35	-0.01	-0.45*	-0.03	-0.38*	-0.20	-0.12	0.01	0.06	0.00	0.03	-0.53*	0.27*	0.34*	-0.23	0.00	0.30*	1.00						
21.Time	0.18	-0.48*	-0.30*	0.44*	0.17	-0.17	0.17	-0.05	-0.17	-0.26*	-0.21	0.03	-0.45*	0.22	0.25	-0.26	0.37*	-0.33*	0.42*	0.07	1.00					
22.Work	-0.30*	0.13	-0.11	-0.15	-0.40*	-0.03	-0.32*	-0.07	0.08	0.33*	0.12	0.06	0.29*	-0.16	0.12	-0.02	-0.13	0.04	0.19	0.43*	-0.19	1.00				
23.Leisure	-0.20	-0.11	-0.06	0.09	-0.21	0.09	-0.57*	-0.02	-0.01	0.08	0.17	0.02	0.20	-0.13	0.16	-0.14	-0.04	0.16	0.42*	0.08	-0.02	-0.02	1.00			
24.Home	0.33*	0.04	0.31*	-0.01	0.24	0.27*	-0.06	-0.08	0.11	0.29*	0.58*	-0.07	0.23	0.39*	-0.15	-0.19	0.45*	0.28*	0.13	-0.23	-0.21	0.02	0.05	1.00		
25.Money	-0.20	0.06	-0.07	-0.02	-0.13	-0.01	0.13	0.32*	0.10	-0.34*	-0.34*	0.01	-0.28*	-0.32*	-0.02	0.36*	-0.38*	-0.17	-0.29*	0.02	-0.22	-0.43*	0.03	-0.36*	1.00	
26.Religion	-0.06	-0.05	-0.23	0.06	-0.20	-0.04	-0.10	0.22	-0.27*	-0.16	-0.21	0.44*	-0.35*	0.11	0.18	-0.04	-0.12	-0.20	0.14	0.08	0.09	-0.17	0.08	-0.25	0.20	1.00

Note: Significance. Codes: 0.1 '*'

Relevancy of Principal Components

Once PCA is conducted, we should look at the proportion of variance explained by each component. Table 8 shows the relative importance of the first 10 principal components obtained from the PCA, although the analysis reports a total of 26 components. Thanks to PCA, we can reduce the initial dimensionality from 26 to 10 while conserving the 83% of variance in our data. Besides, we can explain 46% of variance with just the first three components. In general, papers applying PCA methods adopt the first two principal components to build a two-dimensional positioning map. In our case, PC1 and PC2 account for a third of the variance of the data (33%) and if we add PC3, the explained variance raises to 46%, almost half of the variance. Because of that, instead of just showing a perceptual map on PC1 and PC2 (Figure 4), we also report a 3-dimensional perceptual map incorporating PC3 (Figure 3).

Table 8

Relevancy of the first 10 PCs

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Eigenvalue	2.17	2.00	1.84	1.85	1.35	1.25	1.13	1.00	0.95	0.91
Proportion of variance	0.18	0.15	0.13	0.08	0.07	0.06	0.05	0.04	0.04	0.03
Cumulative proportion of variance	0.18	0.33	0.46	0.54	0.61	0.67	0.72	0.76	0.80	0.83

Loadings and Positioning Map

Table 9 records the loadings of the textual variables in the three dimensions: PC1, PC2 and PC3. In PCA, variable loadings are interpreted as the coefficients in linear combination of the initial variables from which the principal components are constructed. Loadings help us to understand the relevancy and impact of each variable in each principal component.

Table 9

Textual variables loadings in the 3-dimensions (PC1, PC2 and PC3)

Textual variable	PC1	PC2	PC3
PosEmotions	-0.39	0.17	-0.02
NegEmotions	0.08	-0.19	-0.23
Family	-0.24	-0.14	-0.24
Friend	-0.02	0.16	0.24
Female	-0.33	0.13	-0.11
Male	0.03	-0.03	-0.08
See	-0.11	0.35	-0.21
Hear	0.01	0.11	-0.04

Feel	0.10	0.07	-0.26
Body	-0.18	-0.30	0.06
Health	-0.17	-0.32	-0.02
Sexual	-0.04	0.05	0.15
Ingest	-0.10	-0.35	-0.15
Affiliation	-0.38	0.11	0.12
Achievement	0.15	-0.08	0.33
Power	0.28	0.08	-0.22
Reward	-0.35	0.02	0.17
Risk	-0.07	-0.34	-0.04
Motion	-0.005	-0.11	0.46
Space	0.28	-0.16	0.16
Time	-0.06	0.24	0.35
Work	0.10	-0.26	0.07
Leisure	0.06	-0.17	0.18
Home	-0.26	-0.18	-0.05
Money	0.20	0.18	-0.15
Religion	0.09	0.13	0.19

We observe that PCA loadings have either positive or negative signs, depending on the direction of the contribution to the principal component dimension. For example, PC1 is highly positively associated to *Power*, *Space*, *Money* and *Achievement*, while it is highly and negatively associated to *PosEmotions*, *Affiliation*, *Reward*, *Home* and *Family*. PC2 is highly and positively related to *See*, *Money*, *Time* and *PosEmotions*, whereas it is quite negatively related to *Ingest*, *Risk* and *Health*. PC3 is highly positively related to *Motion*, *Time*, *Achievement* and *Friend*, while it is highly negatively associated to *Feel*, *Family*, *NegEmotions* and *Power*.

Table 10 shows the loading of each brand on the 3 components. The loadings obtained from PCA result in a three-dimensional (3D) map, displayed in Figure 3. Since there are several brands located in close positions in the 3D map, we should zoom in on the interactive map to better visualize and analyse the position of each individual brand. From the static version of the map shown in Figure 3 we can see that the brand Black Up has high and positive loadings in PC2 and PC3 and is almost in the middle of PC1. Therefore, this would tell us that the brand Black Up is highly associated to the associations related to PC2 (e.g., *See* and *Time*) and PC3 (e.g., *Time* and *Friend*). For a clearer visualization of brand associations and to make interpretations easier, we show and discuss the two-dimensional (2D) map on the two principal components, represented in Figure 4.

Table 10

Loading of brands on PC1, PC2 and PC3

Brand	PC1	PC2	PC3
bareMinerals	-2.95	0.67	0.14
BECCA	0.02	1.40	0.97
Benefit Cosmetics	-1.87	0.28	0.43
Bite Beauty	-0.44	-1.57	2.46
Black Up	-2.75	4.97	-3.20
Bobbi Brown	-0.45	1.07	-0.40
BURBERRY	0.81	0.98	-0.39
Chosungah 22	4.29	-3.05	-0.54
Ciaté London	1.37	-0.04	4.25
CLINIQUE	-0.52	1.12	0.07
Dior	-1.24	1.42	-0.87
Estée Lauder	1.46	1.09	-5.59
Giorgio Armani Beauty	4.23	-2.87	0.43
Givenchy	-0.54	1.95	-1.29
Guerlain	1.97	-1.02	-1.31
Hourglass	0.50	1.36	-0.75
ILIA	-1.82	-1.57	2.86
KEVYN AUCOIN	2.41	-0.31	-1.14
Lancôme	-1.19	1.27	-0.72
Laura Mercier	-1.23	0.92	-0.34
MAKE UP FOR EVER	0.23	0.41	0.17
Marc Jacobs Beauty	-1.11	0.97	1.08
MILK MAKEUP	-0.12	-2.13	0.32
NARS	-1.92	1.81	1.66
NUDESTIX	-0.90	-0.83	2.47
Perricone MD	0.99	0.15	2.38
Retailer brand	-1.15	1.35	0.17
Smashbox	-0.86	-0.19	0.68
stila	-0.55	-0.80	0.06
Supergoop!	-7.13	-6.82	-2.11
surratt beauty	3.55	-0.76	-2.37
tarte	-0.74	1.32	1.51
Tata Harper	0.17	-2.61	0.42
The Estée Edit	2.19	3.36	0.91
TOM FORD	1.82	-1.29	-4.07
Too Cool For School	1.47	-2.88	-1.17
Too Faced	-2.16	0.50	-0.94
Urban Decay	-1.04	1.01	0.86
Visart	4.39	1.40	2.21
Wander Beauty	0.64	-1.28	0.04
Yves Saint Laurent	0.17	-0.75	0.62

Figure 3. Three-dimensional perceptual map on PC1, PC2 and PC3

The two-dimensional positioning brand in Figure 4 results from the combination of two inputs. First, the direction and weight of the original textual variables considering both PC1 (X-axis) and PC2 (Y-axis); and second, the positioning of brands on the two dimensions (PC1 and PC2). The direction and weight of the original textual variables in the map is based on the loadings recorded in Table 9, while the position of brands in the map is based on loadings shown in Table 10. Looking at the position of brands in the map, we can analyse the similarity of brands according to the textual variables. For example, we observe that consumers strongly associate the brand bareMinerals with *PosEmotions*, *Affiliation* and *Female*. The brand Too Cool For School is highly associated with *Work* and the brand Viséart is the one most related to *Power*. We also notice that the brand Supergoop! is positioned far from the rest of brands in the map and it has high associations with *Family*, *Home*, *Health* and *Body*. Its position indicates that the brand

is highly differentiated in terms of those brand associations that highly contribute to PC1 and PC2. In order to better study similarities between brands and to identify possible brand clusters, we carry out a Hierarchical Clustering analysis.

Figure 4. Two-dimensional brand positioning map of PC1 & PC2

4.6. Identifying Brand Segments: Hierarchical Clustering

Although PCA helps us to see similarities between brands based on the different principal components, it is necessary to carry out a clustering analysis in order to clearly identify brand segments. Hierarchical clustering is one of the most common types of unsupervised learning, which is a type of machine learning algorithm, used to identify clusters from data. The objective of clustering is to create clusters in the way that brands within a cluster should be as similar as possible and brands in one cluster should be as dissimilar as possible from brands in another. The variables used as inputs in the

hierarchical clustering algorithm are the ones previously used in PCA (26 textual variables). Since our textual variables are continuous, we adopt the Euclidean distance to calculate the similarity between the clusters. To run the hierarchical clustering algorithm, all textual variables are previously scaled.

Optimum Number of Clusters (k)

In order to determine the optimum number of clusters (k) in our data set, we adopt the Elbow method. This method looks at the total within-cluster sum of squares (WSS) as a function of the number of clusters, in the way that one should choose a number of clusters so that adding another cluster doesn't improve much better the total WSS. Figure 5 shows the output of the Elbow method, indicating that the optimum number of clusters is $k=4$.

Figure 5. Optimum number of clusters (k)

Visualizing Clustering Results

The best way of visualizing the results of hierarchical clustering is to categorize the different objects, in our case brands, into a dendrogram, which is a type of tree diagram. Figure 6 shows the dendrogram obtained from the clustering analysis. The height of the dendrogram indicates the order in which the clusters were joined. We observe that one **cluster captures only a brand, “Black Up”, another cluster is composed by 21 different brands**, another compiles 9 brands and another one is formed by 10 brands. The higher the height of the linkage between brands, the greater the difference between those brands. For example, inside the second cluster (composed by 21 brands), the brands

“Bobbi Brown” and “Laura Mercier” are very similar, since the height of the linkage between them is very low, while they are more dissimilar to the brand “CLINIQUE”.

Figure 6. Hierarchical clustering dendrogram

Clusters can also be represented in a perceptual map, shown in Figure 7, by linking the results of both, PCA and hierarchical clustering. We observe in Figure 7 that cluster 3 is **only composed by the brand “Black Up”, which might indicate the brand holds a more differentiated position in the market in terms of brand associations relevant in PC1 and PC2.** Then, we have other three segments composed by several brands. In each segment, we observe some brands that are also more differentiated than others in the market. For **example, the brand “Supergoop!” in cluster 2 is characterized by very negative values at PC1 and PC2.**

Figure 7. Perceptual map with clusters

Describing Clusters Characteristics

An important stage in our proposed procedure is to describe the characteristics of each cluster or segment, in order to identify patterns and to identify similarities and differences between them. From a practitioner point of view, it is important to understand the differences between brand segments to be able to adapt marketing strategies to satisfy needs and behaviours at each cluster. To analyse the characteristics of the four segments in terms of brand associations, we should analyse the means of the textual variables, which are the inputs used to identify the clusters. Besides, some non-textual brand features are used to describe the four clusters. Although they have not been used as inputs for clustering because we want to identify segments based on textual brand associations, they might be relevant to know the specific characteristics of each clusters in terms of non-textual dimensions. Table 11 shows the mean of the textual and non-textual variables at each cluster. Figure 8, 9, 10 and 11 graphically represent the composition of each cluster in terms of textual associations and non-textual brand features. Moreover, we can see in Figure 12 a comparison between brand textual and non-textual features across segments.

Table 11

Cluster means for each brand variable (textual and brand features)

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Textual variables				
PosEmotions	0.389	0.223	1.945	-1.212

NegEmotions	-0.090	-0.065	-1.859	0.433
Family	0.069	0.121	-0.835	-0.170
Friend	0.179	-0.024	0.137	-0.368
Female	0.388	-0.163	1.705	-0.840
Male	-0.020	-0.194	-0.970	0.314
See	0.383	-0.689	3.940	-0.578
Hear	0.084	-0.429	0.859	0.124
Feel	-0.217	-0.437	1.462	0.703
Body	-0.409	1.510	-0.166	-0.482
Health	-0.287	0.610	-0.071	0.061
Sexual	0.066	0.326	-0.620	-0.370
Ingest	-0.450	1.015	-0.255	0.057
Affiliation	0.496	0.297	0.590	-1.368
Achievement	-0.067	0.172	-1.808	0.167
Power	-0.140	-0.752	-0.847	1.055
Reward	0.283	0.391	0.272	-0.973
Risk	-0.155	0.287	-1.646	0.232
Motion	0.030	0.494	-2.723	-0.234
Space	-0.312	0.170	-1.963	0.700
Time	0.448	-0.406	0.409	-0.616
Work	-0.384	0.596	-1.172	0.387
Leisure	-0.308	0.365	-1.885	0.507
Home	-0.054	0.527	-0.387	-0.323
Money	-0.097	-0.525	1.071	0.568
Religion	0.274	-0.257	-1.070	-0.237
Non-textual brand variables				
Price	0.112	0.056	0.466	-0.332
Rating	0.380	0.050	1.241	-0.967
Reviews	0.288	-0.219	-0.374	-0.370
Followers	0.548	-0.530	-0.672	-0.606
Bestselling_rank_inv*	0.070	0.295	-0.472	-0.365

*The variable *Bestselling_rank* has been transformed to its inverse, *Bestselling_rank_inv*, for interpretation purposes. The product sales rank is inversely related to its sales, which means that the first product in the sales rank in a specific product category is the one with highest sales. By doing the transformation, we can interpret higher values of *Bestselling_rank_inv* as better brand sales.

Figure 8. Cluster 1 composition

Figure 9. Cluster 2 composition

Figure 10. Cluster 3 composition

Figure 11. Cluster 4 composition

Figure 12. Comparisons of brand variables between clusters

By analysing Figure 12 to compare characteristics between clusters, we observe that Cluster 1 is characterized by medium-high prices, good average rating, relatively good sales and it clearly outweighs the rest of clusters in terms of number of online reviews received and number of followers at Instagram. Therefore, we could think that brands within this cluster are those quite popular amongst consumers and those with greater brand awareness, although the price is a bit higher than in other clusters. In terms of associations, brands within Cluster 1 do not have extreme scores at any textual variable, so we might think that these brands are not very differentiated in the market in terms of associations, and they might be brands targeted to the mass market. Cluster 2 is

represented by medium-low prices, neutral consumer ratings, the highest sales at the online retailer, a relatively high number of online reviews and it is the second cluster in terms of Instagram followers, although far behind cluster 1. Therefore, we can see that cluster 2 is composed by bestselling brands on the website that have a relatively high popularity. Cluster 2 stands out in several brand associations, since it is quite positively associated to *Body*, *Ingest*, *Affiliation*, *Reward*, *Motion* and *Home* and negatively associated to *Feel*, *Power* and *Money*. Clusters 3 and 4 are the ones with most extreme average scores at many brand textual variables. They are very similar in terms of brand textual dimensions, but they differ at brand non-textual features. They both are quite negatively related to *NegEmotions*, *Achievement*, *Risk*, *Motion*, *Space* and *Leisure* and strongly and positively related to *PosEmotions*, *Female*, *See*, *Feel*, *Money*. Cluster 3, which is composed only by the brand Black Up, is the one having the highest brand average rating and the highest brand average price but with the worst bestselling rank, the lowest number of online reviews and the lowest of total followers in Instagram. We can think that consumers have more favorable opinions on the brand, but it might not be targeted to the mass consumption, since it is more expensive and less popular in terms of sales, eWOM generation and social network followers. Finally, brands in cluster 4 are those with the lowest price, the lowest average rating, average sales and relatively low number of online reviews and Instagram followers. Thus, brands within this cluster are not only the cheapest and those receiving the least favorable opinions, but also some of the least popular amongst consumers.

5. Conclusions

So far, most papers in literature have adopted multidimensional scales to analyze brand image (Cho et al., 2015; Davis et al., 2009; John et al., 2006; Kim et al., 2003; Low & Lamb, 2000; Malhotra, 1981; Park & Rabolt J., 2009). Others have studied brand image and brand positioning using qualitative techniques to elicit brand associations. These techniques ranges from most qualitative, such as free associations and free response questions, to more structured techniques, such as repertory-grid techniques and laddering techniques (Haridasan & Fernando, 2018; Henderson et al., 1998; Olson & Muderrisoglu, 1979; Reynolds & Gutman, 1979). However, the availability of **consumers'** online opinions and the increasing availability of text mining tools, is starting to capture the attention of academics and practitioners when they want to study brand image. The stream of literature exploring the text of online reviews is quite diverse in terms of research objectives and methodological techniques used. Because of that, this research intends to suggest, explain and illustrate a procedure to extract brand associations from

the text of online reviews, with the objective of analyzing brand positioning and brand segmentation. Thus, this paper contributes to academic literature by providing more insights of how text mining can be a powerful tool to study brand image and brand associations and how findings from text mining can be the used to study brand positioning and segmentation.

First, this paper evaluates the different approaches of text mining that can be used in order to study brand positioning and brand segmentation. Then, and having in mind the problems that businesses might have to adopt sophisticated machine learning algorithms, a lexicon-based text mining method, the Linguistic Inquiry and Word Count (Pennebaker et al., 2007), is suggested in the procedure.

The main stage in the research procedure involves the text mining analysis, whose objective is to extract hidden brand associations from the text of online consumer reviews. Text mining techniques can be classified into machine learning algorithms and lexicon-based methods (Hartmann et al., 2019; Kübler et al., 2019). The selection of one tool over another must rely not only on the specific research objectives but also on the available resources (e.g., skilled people, technical infrastructure and data). As revealed by Magoulas and Swoyer (2020), one of the main barriers for Artificial Intelligence (AI) adoption in companies is the lack of skilled people or the difficulty to hire the required roles. Machine learning is a specific application of AI, and therefore, many companies find it difficult to carry out text mining analysis based on machine learning algorithms. In this scenario, we propose to follow a text mining procedure based on a lexicon-based text mining method, the LIWC, which is an easier to implement tool than machine learning, especially for small and medium companies. LIWC has been validated in many previous studies on several fields (M. A. Cohn et al., 2014; Ireland et al., 2011b; Ludwig et al., 2013b) and it allows us to uncover hidden brand associations from the text of online reviews.

Once brand associations are extracted from online reviews using LIWC, we suggest using them as inputs to explore brand positioning and brand segmentation. To do that, we apply first a Principal Component Analysis (PCA), which reduces the dimensionality of the data to build a brand positioning map and, second, we use Hierarchical Clustering to identify brand segments in that positioning map. Following the research procedure, we propose, companies can obtain several insights, such as knowing how brands are **positioned in consumers' minds, understanding** how similar or different are brands perceived with respect to its competitors and identifying where should be brands positioned to find a competitive advantage. In case companies need more inputs about

text mining techniques, especially based on machine learning algorithms, they could check the paper written by Berger et al. (2020), “Uniting the Tribes: Using Text for **Marketing Insights**”. **In this study, the authors** overview methodologies and metrics used in text analysis, providing a set of guidelines. Moreover, they point out some common metrics and challenges.

5.1. Limitations and Future Research

To illustrate the process, we use online reviews from one category of cosmetic products, **“blush”**. **Findings from** our empirical study are quite context-dependent, since brand associations are very different when dealing with a type of product or another. Therefore, although the research procedure we suggest can be used for any type of product or service category, the findings obtained must be evaluated without losing sight of the type of product or service we are dealing with. Moreover, we analyze in this research the broader picture in terms of **brand positioning in the “blush” category, without focusing on any** specific brand case, but individual brands should make greater efforts to study their individual cases and come to more specific conclusions.

The proposed procedure could be applied to analyze the text of online reviews of any type of product, service or brand. Online reviews available at any platform, either online retailers’ websites or reviews’ websites, have a similar structure and all appear in the product website. Therefore, it is easy to relate online reviews to the product or service of interest. The text mining stage of the procedure might be a bit different if we are analyzing other type of eWOM, such as comments in social networks or blog entries. In this case, we might need to identify the brand each comment, or blog entry is mentioning before uncovering brand associations and relating them to the product, service or brand. It is likely that we need to use machine learning algorithms for text mining in those cases.

Although the use of LIWC has been widespread in previous text mining literature, future research could use other lexicon-based methods available. Machine learning methods for text mining could be also used by companies if they need to extract more specific aspects from online reviews texts and if they have the required resources. In this research, we were especially interested in exploring other brand associations apart from the positive or negative sentiment expressed in online reviews, which is usually the focus in literature. However, other aspects hidden on texts could be also explored, such as features related to the reviewer writing style (e.g., informal writing, cognitive writing and time focus), which are also provided by LIWC.

In terms of research data, this research is carried out using online consumer reviews, but the same research procedure could be used using any type of eWOM, such as social networks and blogs. Brand associations are evaluated on a specific online retailer, it would be interesting to compare associations of brands in different online retailers, to see if brand associations are different depending on the retailer or they are not platform dependent. Moreover, we could carry out a survey to see if online brand associations are the same than overall brand associations, explored by questionnaires.

Moreover, we must consider that the profile of consumers who write online reviews might not be representative of all the population, and there might be a group of consumers who do not use the Internet to express their opinions. These consumers might still need be approached by traditional techniques, such as surveys. Ideally, we could combine both techniques and we would be able to compare the results in terms of brand image using online opinions and traditional methods, such as multidimensional scaling.

References

- Aaker, D. A. (1991). Managing Brand Equity. *Managing Brand Equity*.
- Aaker, D. A. (1996). Building strong brands - aaker,da. *Building Strong Brands*, 13(4), 382–386.
- Ahani, A., Nilashi, M., Yadegaridehkordi, E., Sanzogni, L., Tarik, A. R., Knox, K., Samad, S., & Ibrahim, O. (2019). Revealing customers' satisfaction and preferences through online review analysis: The case of Canary Islands hotels. *Journal of Retailing and Consumer Services*, 51(July), 331–343.
- Ahmad, S. N., & Laroche, M. (2015). How do expressed emotions affect the helpfulness of a product review? Evidence from reviews using latent semantic analysis. *International Journal of Electronic Commerce*, 20(1), 76–111.
- Aiken, L. S., & West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*.
- Al-Obeidat, F., Spencer, B., & Kafeza, E. (2018). The Opinion Management Framework: Identifying and addressing customer concerns extracted from online product reviews. *Electronic Commerce Research and Applications*, 27, 52–64.
- Al., H. et. (2010). *Multivariate Analysis 7 th Edition*.
- Aljukhadar, M., Senecal, S., & Daoust, C.-E. (2012). Using Recommendation Agents to Cope with Information Overload. *International Journal of Electronic Commerce*, 17(2), 41–70.
- Ansari, A., & Mela, C. F. (2003). *E-Customization*. 40(2), 131–145.
- Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*, 57(8), 1485–1509.
- Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2), 277.
- Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1), 29–51.
- Areni, C. S. (2003). The Effects of Structural and Grammatical Variables on Persuasion: An Elaboration Likelihood Model Perspective. *Psychology and Marketing*, 20(4), 349–375.
- Arun, R., Suresh, V., Madhavan, C. E. V., & M.N, N. M. (2010). On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In *Advances in Knowledge Discovery and Data Mining* (pp. 391–403). <http://hdl.handle.net/10722/26088>
- Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. A. (2016). The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors. *Journal of Marketing Research*, 53(3), 297–318.
- Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, 0, 2200–2204.
- Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of Online Consumer Reviews: Readers'

- Objectives and Review Cues. *International Journal of Electronic Commerce*, 17(May 2015), 99–126.
- Baker, W. E. (2001). The Diagnosticity of Advertising Generated Brand Attitudes in Brand Choice Contexts. *Journal of Consumer Psychology*, 11(2), 129–139.
- Baker, W. E., Lutz, R. J., Journal, S., Spring, N., Taylor, P., Baker, W. E., & Lutz, R. J. (2000). Effectiveness An Empirical Test of an Updated Relevance- Accessibility Model of Advertising Effectiveness. *Journal of Advertising*, 29(1), 1–14.
- Bakhshi, S., Kanuparth, P., & Shamma, D. A. (2015). Understanding Online Reviews: Funny, Cool or Useful? *CSCW 2015: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1270–1276.
- Baksi, A. K., & Panda, T. K. (2018). Branding destinations with multisensory brand associations and evaluating its impact on behavioural pattern under the intervention of multiplex phenomenon of relationship-branding. *Management Science Letters*, 8(11), 1169–1182.
- Balducci, B., & Marinova, D. (2018a). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, 46(4), 557–590.
- Balducci, B., & Marinova, D. (2018b). Unstructured data in marketing. In *Journal of the Academy of Marketing Science* (Vol. 46, Issue 4, pp. 557–590). Springer New York LLC.
- Beach, L. R. (1993). Broadening the definition of decision making: The role of prechoice screening option. *Psychological Science*, 4(4), 215–220.
- Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding Satisfied and Dissatisfied Hotel Customers: Text Mining of Online Hotel Reviews. *Journal of Hospitality Marketing and Management*, 25(1), 1–24.
- Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the Tribes: Using Text for Marketing Insight. *Journal of Marketing*, 84(1), 1–25.
- Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive Consumer Choice Processes. *Journal of Consumer Research*, 25(3), 187–217.
- Biehal, G., & Chakravarti, D. (1983). Information Accessibility as a Moderator of Consumer Choice. *Journal of Consumer Research*, 10(1), 1.
- Biswas, D., Grewal, D., & Roggeveen, A. (2010). How the order of sampled experiential products affects choice. *Journal of Marketing Research*, 47(3), 508–519.
- Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), 115–143.
- Brandt, C., De Mortanges, C. P., Bluemelhuber, C., & Van Riel, A. C. R. (2011). Associative Networks: A New Approach to Market Segmentation. *International Journal of Market Research*, 53(2), 187–207.
- Breugelmans, E., Campo, K., & Gijsbrechts, E. (2007). Shelf sequence and proximity effects on online grocery choices. *Marketing Letters*, 18(1–2), 117–133.
- BrightLocal. (2020). *Local Consumer Review Survey*. Available at: <https://www.brightlocal.com/research/local-consumer-review-survey/>
- Brown, C. L., & Krishna, A. (2004). The skeptical shopper: A metacognitive account for the effects of default options on choice. *Journal of Consumer Research*, 31(3), 529–539.

- Bruine de Bruin, W., & Keren, G. (2003). Order effects in sequentially judged options due to the direction of comparison. *Organizational Behavior and Human Decision Processes*, 92(1–2), 91–101.
- Buda, R., & Zhang, Y. (2000). Consumer product evaluation: the interactive effect of message framing, presentation order, and source credibility. *Journal of Product & Brand Management*, 9(4), 229–242.
- Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics Using Stata*. Stata Press Book, StataCorp LP.
- Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data, Second edition. In *Regression Analysis of Count Data, Second Edition*.
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72(7–9), 1775–1781.
- Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, 50(2), 511–521.
- Chaiken, S. (1987). *The heuristic model of persuasion*.
- Chapman, C., & Feit, E. M. (2015). *R for marketing research and analytics*.
- Chen, C. C., & Tseng, Y. De. (2011). Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, 50(4), 755–768.
- Chen, K., Kou, G., Shang, J., & Chen, Y. (2015). Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches. *Electronic Commerce Research and Applications*, 14(1), 58–74.
- Cheng, Y. H., & Ho, H. Y. (2015). Social influence’s impact on reader perceptions of online reviews. *Journal of Business Research*, 68(4), 883–887.
- Cheung, C. M. K., Lee, M. K. O., & Rabjohn, N. (2008). The impact of electronic word-of-mouth. In *Internet Research* (Vol. 18, Issue 3).
- Chevalier, J. A., & Goolsbee, A. (2003). Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics*, 1, 203–222.
- Chevalier, J., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 43(3), 345–354.
- Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science*, 29(5), 944–957.
- Cho, E., Fiore, A. M., & W. Russell, D. (2015). Validation of a Fashion Brand Image Scale Capturing Cognitive, Sensory, and Affective Associations: Testing Its Role in an Extended Brand Equity Model. *Psychology & Marketing*, 32(1), 28–48.
- Chua, A. Y. K., & Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior*, 54, 547–554.
- Chung, C.K., & Pennebaker, J. W. (2007). The psychological function of function words. *Social Communication: Frontiers of Social Psychology*, 343–359.
- Chung, Cindy K., & Pennebaker, J. W. (2007). Revealing dimensions of thinking in open-

- ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality*, 42(1), 96–132.
- Cohn, M. a., Mehl, M. R., & Pennebaker, J. W. (2004). Markers of Linguistic Psychological Change. *Psychological Science*, 15(10), 687–693.
- Cohn, M. a., Mehl, M. R., & Pennebaker, J. W. (2014). Markers of Linguistic Psychological Change Surrounding September 11, 2001. *Psychological Science*, 15(10), 687–693.
- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2014). *Markers of Linguistic Psychological Change Surrounding September*. 15(10), 687–693.
- Cui, G., Lui, H.-K., & Guo, X. (2012). The Effect of Online Consumer Reviews on New Product Sales. *International Journal of Electronic Commerce*, 17(1), 39–58.
- Cyr, D., Head, M., Lim, E., & Stibe, A. (2018). Using the elaboration likelihood model to examine online persuasion through website design. *Information and Management*, April 2017, 1–15.
- Darley, W. K., & Smith, R. E. (1993). Advertising Claim Objectivity: Antecedents and Effects. *Journal of Marketing*, 57(4), 100.
- Davis, D. F., Golcic, S. L., & Marquardt, A. (2009). Measuring brand equity for logistics services. *The International Journal of Logistics Management*, 20(2), 201–212.
- De Vries, L., Gensler, S., & Leeftang, P. S. H. (2012). Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing. *Journal of Interactive Marketing*, 26(2), 83–91.
- Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, 27(4), 293–307.
- Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4), 23–45.
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective Latent Concept Modeling for ad hoc information retrieval. *Document Numerique*, 17(1), 61–84.
- Drèze, X., & Zufryden, F. (2004). Measurement of online visibility and its impact on Internet traffic. *Journal of Interactive Marketing*, 18(1), 20–37.
- Duan, W., Gu, B., & Whinston, A. B. (2008a). Do online reviews matter? - An empirical investigation of panel data. *Decision Support Systems*, 45(4), 1007–1016.
- Duan, W., Gu, B., & Whinston, A. B. (2008b). The dynamics of online word-of-mouth and product sales-An empirical investigation of the movie industry. *Journal of Retailing*, 84(2), 233–242.
- Elberse, A., & Eliashberg, J. (2003). Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures. *Marketing Science*, 22(3), 329–354.
- Ellsworth, P. C., & Smith, C. A. (1988). From appraisal to emotion: Differences among unpleasant feelings. *Motivation and Emotion*, 12(3), 271–302.
- Ert, E., & Fleischer, A. (2016). Mere Position Effect in Booking Hotels Online. *Journal of Travel Research*, 55(3), 311–321.
- Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management*, 55, 62–73.

- Eslami, S. P., & Ghasemaghaei, M. (2018). Effects of online review positiveness and review score inconsistency on sales: A comparison by product involvement. *Journal of Retailing and Consumer Services*, 45(June), 74–80.
- Faircloth, J. B., Capella, L. M., & Alford, B. L. (2001). The Effect of Brand Attitude and Brand Image on Brand Equity. *Journal of Marketing Theory and Practice*, 9(3), 61–75.
- Feldman, J. M., & Lynch, J. G. (1988). Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior. *Journal of Applied Psychology*, 73(3), 421–435.
- Felfernig, A., Friedrich, G., Gula, B., Hitz, M., Kruggel, T., Leitner, G., Melcher, R., Riepan, D., Strauss, S., Teppan, E., & Vitouch, O. (2007). Persuasive recommendation: Serial position effects in knowledge-based recommender systems. *Proceedings of the 2nd International Conference on Persuasive Technology (PERSUASIVE '07)*, 283–295.
- Filieri, R., Hofacker, C. F., & Alguezaui, S. (2018). What makes information in online consumer reviews diagnostic over time? The role of review relevancy, factuality, currency, source credibility and ranking score. *Computers in Human Behavior*, 80, 122–131.
- Filieri, R., & McLeay, F. (2014). E-WOM and Accommodation: An Analysis of the Factors That Influence Travelers' Adoption of Information from Online Reviews. *Journal of Travel Research*, 53(1), 44–57.
- Filieri, R., McLeay, F., Tsui, B., & Lin, Z. (2018). Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services. *Information and Management*, 55(8), 956–970.
- Filieri, R., Raguseo, E., & Vitari, C. (2018). When are extreme ratings more helpful? Empirical evidence on the moderating effects of review characteristics and product type. *Computers in Human Behavior*, 88, 134–142.
- Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., & Freling, T. (2014). How Online Product Reviews Affect Retail Sales: A Meta-analysis. *Journal of Retailing*, 90(2), 217–232.
- Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., & Treinen, M. (2001). What makes web sites credible? A report on a large quantitative study. *Conference on Human Factors in Computing Systems - Proceedings*, 61–68.
- Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 19(3), 291–313.
- Gao, B., Hu, N., & Bose, I. (2017). Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews. *Decision Support Systems*, 95, 1–11.
- Gensler, S., Völckner, F., Egger, M., Fischbach, K., & Schoder, D. (2015). Listen to your customers: Insights into brand image using online consumer-generated product reviews. *International Journal of Electronic Commerce*, 20(1), 112–141.
- Ghose, A., & Ipeirotis, P. G. (2010). Estimating the Helpfulness and Economic Impact of Product Reviews. *Ieee Transactions on Knowledge and Data Engineering*, 1–15.
- Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science*, 31(3), 493–520.

- Girard, T., & Dion, P. (2010). Validating the search, experience, and credence product classification framework. *Journal of Business Research*, 63(9–10), 1079–1087.
- Gobet, F., Lane, P., Croker, S., Cheng, P., Jones, G., Oliver, I., & Pine, J. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5(6), 236–243.
- Godes, D., & Silva, J. C. (2012). Sequential and Temporal Dynamics of Online Opinion. *Marketing Science*, 31(3), 448–473.
- Godes, David, & Mayzlin, D. (2004). Using Online Conversations to Study Word-of-Mouth Communication. *Marketing Science*, 23(4), 545–560.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(SUPPL. 1), 5228–5235.
- Gu, B., Gu, B., & Park, J. (2012). The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products. *Information Systems Research*, 23(1), 182–196.
- Guo, J., Wang, X., & Wu, Y. (2020). Positive emotion bias: Role of emotional content from online customer reviews in purchase decisions. *Journal of Retailing and Consumer Services*, 52(July 2019).
- Guo, Y., Barnes, S. J., & Jia, Q. (2017a). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.
- Guo, Y., Barnes, S. J., & Jia, Q. (2017b). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.
- Hamby, A., Daniloski, K., & Brinberg, D. (2015). How consumer reviews persuade through narratives. *Journal of Business Research*, 68(6), 1242–1250.
- Haridasan, A. C., & Fernando, A. G. (2018). Online or in-store: unravelling consumer's channel choice motives. *Journal of Research in Interactive Marketing*, 12(2), 215–230.
- Harrison-Walker, L. J. (2001). The Measurement of Word-of-Mouth Communication and an Investigation of Service Quality and Customer Commitment As Potential Antecedents. *Journal of Service Research*, 4(1), 60–75.
- Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, 36(1), 20–38.
- Häubl, G., & Trifts, V. (2000). Consumer Decision Making in Online Shopping Environments : The Effects of Interactive Decision Aids Consumer Decision Making in Online Shopping Environments : The Effects of Interactive Decision Aids. *Marketing Science*, 19(1), 4–21.
- Haugtvedt, C. P., & Wegener, D. T. (1994). Message Order Effects in Persuasion: An Attitude Strength Perspective. *Journal of Consumer Research*, 21(1), 205.
- Hauser, J. R., & Koppelman, F. S. (1979). Alternative Perceptual Mapping Techniques : Relative Accuracy and Usefulness. *Journal of Marketing Research*, 16(4), 495–506.
- Henderson, G. R., Iacobucci, D., & Calder, B. J. (1998). Brand diagnostics: Mapping branding effects using consumer associative networks. *European Journal of Operational Research*, 111(2), 306–327.

- Heng, Y., Gao, Z., Jiang, Y., & Chen, X. (2018). Exploring hidden factors behind online food shopping from Amazon reviews: A topic mining approach. *Journal of Retailing and Consumer Services*, 42(January), 161–168.
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, 18(1), 38–52.
- Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective. *Journal of Consumer Research*, 17(4), 454.
- Herrmann, A., Goldstein, D. G., Stadler, R., Landwehr, J. R., Heitmann, M., Hofstetter, R., & Huber, F. (2011). The effect of default options on choice-Evidence from online product configurators. *Journal of Retailing and Consumer Services*, 18(6), 483–491.
- Ho-Dac, N. N., Carson, S. J., & Moore, W. L. (2013). The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter? *Journal of Marketing*, 77(6), 37–53.
- Hofmann, J., Clement, M., Völckner, F., & Hennig-Thurau, T. (2017). Empirical generalizations on the impact of stars on the economic success of movies. *International Journal of Research in Marketing*, 34(2), 442–461.
- Holbrook, M. B. (1978). Beyond Attitude Structure: Toward the Informational Determinants of Attitude. *Journal of Marketing Research*, 15(4), 545.
- Holtgraves, T., & Lasky, B. (1999). Linguistic power and persuasion. *Journal of Language and Social Psychology*, 18(2), 196–205.
- Hornik, K., & Grün, B. (2011). topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software*, 40(13), 1–30.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD'04*, 168.
- Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? the mediating role of online review sentiments on product sales. *Decision Support Systems*, 57(1), 42–53.
- Hu, Y., & Li, X. (2011). Context-Dependent Product Evaluations: An Empirical Analysis of Internet Book Reviews. *Journal of Interactive Marketing*, 25(3), 123–133.
- Huang, A. H., Chen, K., Yen, D. C., & Tran, T. P. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behavior*, 48, 17–27.
- Hutto, C. J., & Gilbert, E. E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14).". *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*. <http://sentic.net/>
- International, E. (2017). *Colour cosmetics in the US* (Issue May).
- Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011a). Language Style Matching Predicts Relationship Initiation and Stability. *Psychological Science*, 22(1), 39–44.
- Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011b). Language Style Matching Predicts Relationship Initiation and

- Stability. *Psychological Science*, 22(1), 39–44.
- Jalilvand, M. R., & Samiei, N. (2012). The effect of electronic word of mouth on brand image and purchase intention: An empirical study in the automobile industry in Iran. *Marketing Intelligence and Planning*, 30(4), 460–476.
- Jiang, S., Cai, S., Olle, G. O., & Qin, Z. (2015). Durable product review mining for customer segmentation. *Kybernetes*, 44(1), 124–138.
- Jiménez, F. R., & Mendoza, N. A. (2013). Too popular to ignore: The influence of online reviews on purchase intentions of search and experience products. *Journal of Interactive Marketing*, 27(3), 226–235.
- John, D. R., Loken, B., Kim, K., & Monga, A. B. (2006). Brand concept maps: A methodology for identifying brand association networks. *Journal of Marketing Research*, 43(4), 549–563.
- Johnson, E. J., Bellman, S., & Lohse, G. L. (2002). Defaults, Framing and Privacy: Why Opting In-Opting Out. *Marketing Letters*, 13(1), 5–15.
- Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, 80(4), 557–571.
- Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*, 33(2), 125–143.
- Kamble, Z., Doan, Q. T., Nguyen, H. N., Zeng, Z., & Zihui, L. L. (2020). The Significance of Online Reviews for Hotel Entrepreneurial Success. In *Entrepreneurial Development and Innovation in Family Businesses and SMEs* (pp. 75–86).
- Keller, K. L. (1993). *Conceptualizing, Measuring, and Managing Customer-Based Brand Equity*. 57(1), 1–22.
- Kelly, G. A. (1991). *The Psychology of Personal Constructs*.
- Khwaja, M. G., & Zaman, U. (2020). Configuring the evolving role of ewom on the consumers information adoption. *Journal of Open Innovation: Technology, Market, and Complexity*, 6(4), 1–13.
- Kim, H. B., Kim, W. G., & An, J. A. (2003). The effect of consumer-based brand equity on firms' financial performance. *Journal of Consumer Marketing*, 20(4–5), 335–351.
- Kim, M., & Lennon, S. (2007). The Effects of Visual and Verbal Information on Attitudes and Purchase Intentions in Internet Shopping. *Psychology & Marketing*, 24(9), 763–785.
- Kim, S. G., & Kang, J. (2018). Analyzing the discriminative attributes of products using text mining focused on cosmetic reviews. *Information Processing and Management*, 54(6), 938–957.
- King, M. F., & Balasubramanian, S. k. (1994). The effects of Expertise, End Goal, and Product Type on Adoption of Preference Formation Strategy. *Journal of the Academy of Marketing Science*, 22.
- Kostyk, A., Niculescu, M., & Leonhardt, J. M. (2017). Less is more: Online consumer ratings' format affects purchase intentions and processing. *Journal of Consumer Behaviour*, 16(5), 434–441.

- Kotler, P., & Armstrong, G. (2020). *Principles of Marketing* (18th Global Edition). Pearson Education Limited.
- Kozinets, R. V., De Valck, K., Wojnicki, A. C., & Wilner, S. J. S. (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of Marketing*, 74(2), 71–89.
- Kübler, R. V., Colicev, A., & Pauwels, K. H. (2019). Social Media's Impact on the Consumer Mindset: When to Use Which Sentiment Extraction Tool? *Journal of Interactive Marketing*, 50, 136–155.
- Kudeshia, C., & Kumar, A. (2017). Social eWOM: does it affect the brand attitude and purchase intention of brands? *Management Research Review*, 40(3), 310–330.
- Lee, P. J., Hu, Y. H., & Lu, K. T. (2018). Assessing the helpfulness of online hotel reviews: A classification-based approach. *Telematics and Informatics*, 35(2), 436–445.
- Lee, S., & Choeh, J. Y. (2018). The interactive impact of online word-of-mouth and review helpfulness on box office revenue. *Management Decision*, 56(4), 849–866.
- Lee, S., & Choeh, J. Y. (2020). Using the social influence of electronic word-of-mouth for predicting product sales: The moderating effect of review or reviewer helpfulness and product type. *Sustainability (Switzerland)*, 12(19).
- Lee, T. Y., & Bradlow, E. T. (2011). Automated Marketing Research Using Online Customer Reviews. *Journal of Marketing Research*, 48(5), 881–894.
- Levinson, P. (2013). *New New Media*. 223. <https://books.google.com/books?id=G009uwAACAAJ&pgis=1>
- Lewsey, J. D., & Thomson, W. M. (2004). The utility of the zero-inflated Poisson and zero-inflated negative binomial models: A case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. *Community Dentistry and Oral Epidemiology*, 32(3), 183–189.
- Li, M., Huang, L., Tan, C.-H., & Wei, K.-K. (2013). Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features. *International Journal of Electronic Commerce*, 17(4), 101–136.
- Li, S. T., Pham, T. T., & Chuang, H. C. (2019). Do reviewers' words affect predicting their helpfulness ratings? Locating helpful reviewers by linguistics styles. *Information and Management*, 56(1), 28–38.
- Li, Xiaolin, Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information and Management*, 56(2), 172–184.
- Li, Xinxin, & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, 19(4), 456–474.
- Liang, T. P., Li, X., Yang, C. T., & Wang, M. (2015). What in Consumer Reviews Affects the Sales of Mobile Apps: A Multifacet Sentiment Analysis Approach. *International Journal of Electronic Commerce*, 20(2), 236–260.
- Liu, J., Cao, Y., Lin, C., Huang, Y., ... M. Z. the 2007 J. C. on, & 2007, U. (2007). Low-quality product review detection in opinion summarization. *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, June, 334–342.
- Liu, Q., & Karahanna, E. (2015). An agent-based modeling analysis of helpful vote on online product reviews. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2015-March, 1585–1595.

- Liu, X., Burns, A. C., & Hou, Y. (2017). An Investigation of Brand-Related User-Generated Content on Twitter. *Journal of Advertising*, 46(2), 236–247.
- Liu, Y. (2006). Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*, 70(3), 74–89.
- Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, 47, 140–151.
- Low, G. S., & Lamb, C. W. (2000). The measurement and dimensionality of brand associations. *Journal of Product & Brand Management*, 9(6), 350–370.
- Lozano, M. B., Martínez, B., & Pindado, J. (2016). Corporate governance, ownership and firm value: Drivers of ownership as a good corporate governance mechanism. *International Business Review*, 25(6), 1333–1343.
- Luca, M. (2011). Revenue : The Case of Reviews , Reputation , and Revenue : The Case of Yelp . com. *Harvard Business Review*, 1–40.
- Ludwig, S., de Ruyter, K., Friedman, M., Brügger, E. C., Wetzels, M., & Pfann, G. (2013a). More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*, 77(1), 87–103.
- Ludwig, S., de Ruyter, K., Friedman, M., Brügger, E. C., Wetzels, M., & Pfann, G. (2013b). More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*, 77(1), 87–103.
- Lynch, Jr., J. G., Marmorstein, H., & Weigold, M. F. (1988). Choices from Sets Including Remembered Brands: Use of Recalled Attributes and Prior Overall Evaluations. *Journal of Consumer Research*, 15(2), 169.
- Magerman, T., Van Looy, B., Baesens, B., & Debackere, K. (2011). Assessment of Latent Semantic Analysis (LSA) text mining algorithms for large scale mapping of patent and scientific publication documents. *FBE Research Report MSI_1114*.
- Magoulas, R., & Swoyer, S. (2020). *AI Adoption in the Enterprise 2020*.
- Mahr, D., Stead, S., & Odekerken-Schröder, G. (2019). Making sense of customer service experiences: a text mining review. *Journal of Services Marketing*, 33(1), 88–103.
- Malhotra, N. K. (1981). A Scale to Measure Self-Concepts, Person Concepts, and Product Concepts. *Journal of Marketing Research*, 18(4), 456.
- Mantonakis, A., Roderio, P., Lesschaeve, I., & Hastie, R. (2009). Order in choice: Effects of serial position on preferences. *Psychological Science*, 20(11), 1309–1312.
- Marchand, A., Hennig-Thurau, T., & Wiertz, C. (2017). Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. *International Journal of Research in Marketing*, 34(2), 336–354.
- Martínez Salinas, E., & Pina Pérez, J. M. (2009). Modeling the brand extensions' influence on brand image. *Journal of Business Research*, 62(1), 50–60.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.
- Moe, W. W., & Trusov, M. (2011). The Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing Research*, 48(3), 444–456.
- Mohammad, S., & Turney, P. (2010). Emotions Evoked by Common Words and Phrases. *Proceedings of the NAACL-HLT, June*, 26–34.

- Moon, S., & Kamakura, W. A. (2017). A picture is worth a thousand words: Translating product reviews into a product positioning map. *International Journal of Research in Marketing*, 34(1), 265–285.
- Mudambi, S., & Schuff, D. (2010). What makes a Helpful online Review? *MIS Quarter*, 28(4), 695–704.
- Munch, J. M., & Swasy, J. L. (1988). Rhetorical question, summarization frequency, and argument strength effects on recall. *Journal of Consumer Research*, 15(1), 69–76.
- Murphy, J., Hofacker, C., & Mizerski, R. (2006). Primacy and recency effects on clicking behavior. *Journal of Computer-Mediated Communication*, 11(2), 173–192.
- Nabi, R. L. (2003). Exploring the framing effects of emotion: Do discrete emotions differentially influence information accessibility, information seeking, and policy preference? *Communication Research*, 30(2), 224–247.
- Nazlan, N. H., Tanford, S., & Montgomery, R. (2018). The effect of availability heuristics in online consumer reviews. *Journal of Consumer Behaviour*, 17(5), 449–460.
- Nelson, P. (1970). *Information and Consumer Behavior*. 78(2), 311–329.
- Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine Your Own Business: Market-Structure Surveillance Through Text Mining. *Marketing Science*, 31(3), 521–543.
- Netzer, Oded, Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521–543.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Personality and Social Psychology Bulletin Lying Words : Predicting Deception From Linguistic Styles. *Personality and Social Psychology Bulletin*, 29(5), 665–675.
- Ngo-Ye, T. L., & Sinha, A. P. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, 61(1), 47–58.
- Nielsen. (2016). *The sweet smell of seasonal success*. Available at: <https://www.nielsen.com/au/en/insights/news/2016/the-sweet-smell-of-seasonal-success.html>
- Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *CEUR Workshop Proceedings*, 718, 93–98.
- Olson, J. C., & Muderrisoglu, A. (1979). The stability of responses obtained by free elicitation: Implications for measuring attribute salience and memory structure. *Advances in Consumer Research*, 6, 269–275.
- Pan, B., Zhang, L., & Law, R. (2013). The Complex Matter of Online Hotel Choice. *Cornell Hospitality Quarterly*, 41(1), 74–83.
- Pan, Y., & Zhang, J. Q. (2011). Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews. *Journal of Retailing*, 87(4), 598–612.
- Pang, J., & Qiu, L. (2016). Effect of online review chunking on product attitude: The moderating role of motivation to think. *International Journal of Electronic Commerce*, 20(3), 355–383.
- Park, D. H., & Lee, J. (2008). eWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications*, 7(4), 386–398.

- Park, E. J., Kim, E. Y., Funches, V. M., & Foxx, W. (2012). Apparel product attributes, web browsing, and e-impulse buying on shopping websites. *Journal of Business Research*, 65(11), 1583–1589.
- Park, H.-J., & Rabolt J., N. (2009). Cultural Value, Consumption Value, and Global Brand Image: A Cross-National Study. *Psychology & Marketing*, 26(8), 714–735.
- Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research*, 50, 67–83.
- Payan, J. M., & McFarland, R. G. (2005). Decomposing Influence Strategies: Argument Structure and Dependence as Determinants of the Effectiveness of Influence Strategies in Gaining Channel Member Compliance. *Journal of Marketing*, 69(3), 66–79.
- Payne, J. W. (1982). Contingent Decision Behavior. *Psychological Bulletin*, 92(2), 382.
- Peladeau, N. (2016). *WordStat: Content Analysis Module for SIM-STAT*.
- Pennebaker, J. W., Boyd, R., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. *Austin, TX: University of Texas at Austin, SEPTEMBER 2015*, 1–22.
- Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS ONE*, 9(12), 1–10.
- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The Development and Psychometric Properties of LIWC2007. *Austin, Tx: LIWC.Net*, 1(2).
- Petty, R. E., Tormala, Z. L., Hawkins, C., & Wegener, D. T. (2001). Motivation to think and order effects in persuasion: The moderating role of chunking. *Personality and Social Psychology Bulletin*, 27(3), 332–344.
- Pindado, J., Requejo, I., & de la Torre, C. (2011). Family control and investment-cash flow sensitivity: Empirical evidence from the Euro zone. *Journal of Corporate Finance*, 17(5), 1389–1409.
- Price, P. C., & Stone, E. R. (2004). Intuitive Evaluation of Likelihood Judgment Producers: Evidence for a Confidence Heuristic. *Journal of Behavioral Decision Making*, 17(1), 39–57.
- Puranam, D., Narayan, V., & Kadiyali, V. (2017). The effect of calorie posting regulation on consumer opinion: A flexible latent dirichlet allocation model with informative priors. *Marketing Science*, 36(5), 726–746.
- Purnawirawan, N., De Pelsmacker, P., & Dens, N. (2012). Balance and Sequence in Online Reviews: How Perceived Usefulness Affects Attitudes and Intentions. *Journal of Interactive Marketing*, 26(4), 244–255.
- Purnawirawan, N., Dens, N., & De Pelsmacker, P. (2012). Balance and Sequence in Online Reviews: The Wrap Effect. *International Journal of Electronic Commerce*, 17(2), 71–98.
- Racherla, P., & Friske, W. (2012). Perceived “usefulness” of online consumer reviews: An exploratory investigation across three services categories. *Electronic Commerce Research and Applications*, 11(6), 548–559.
- Reynolds, T. J., & Gutman, J. (1979). Laddering theory, method, analysis, and interpretation. *Journal of Advertising Research*, 28(1), 11–31.

- Rodríguez-Díaz, M., Rodríguez-Díaz, R., & Espino-Rodríguez, T. (2018). Analysis of the Online Reputation Based on Customer Ratings of Lodgings in Tourism Destinations. *Administrative Sciences*, 8(3), 51.
- Roedder, D. J., Loken, B., Kim, K., & Monga, A. B. (2006). Brand concept maps: A methodology for identifying brand association networks. *Journal of Marketing Research*, 43(4), 549–563.
- Roodman, D. (2009a). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, 9(1), 86–136.
- Roodman, D. (2009b). Practitioners' corner: A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71(1), 135–158.
- Salehan, M., & Kim, D. J. (2015). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, 81, 30–40.
- Saumya, S., Singh, J. P., Baabdullah, A. M., Rana, N. P., & Dwivedi, Y. K. (2018). Ranking online consumer reviews. *Electronic Commerce Research and Applications*, 29(March), 78–89.
- Schnittka, O., Sattler, H., & Zenker, S. (2012). Advanced brand concept maps: A new approach for evaluating the favorability of brand association networks. *International Journal of Research in Marketing*, 29(3), 265–274.
- Selkirk, E. (1996). The prosodic structure of function words. In J. L. (Brown U. Morgan & K. (Brown U. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 187–214). Lawrence Erlbaum Associates, Inc.
- Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80(2), 159–169.
- Sheldon, P., & Bryant, K. (2016). Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in Human Behavior*, 58, 89–97.
- Shugan, S. M. (1980). The Cost of Thinking. *Journal of Consumer Research*, 7(2), 99–111.
- Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Kumar Roy, P. (2017). Predicting the “helpfulness” of online consumer reviews. *Journal of Business Research*, 70, 346–355.
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, 105(1), 131–142.
- Slovic, P. (1972). From Shakespeare to Simon: Speculations and some evidence. *Oregon Research Institute Bulletin*, 1–19.
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality & Social Psychology*, 48(4), 813–838.
- Snizek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, 84(2), 288–307.
- Social Blade. (2017). <https://socialblade.com/>
- Spears, N., & Singh, S. N. (2004). Measuring attitude toward the brand and purchase intentions. *Journal of Current Issues and Research in Advertising*, 26(2), 53–66.
- Spenner, P., & Freeman, K. (2012). To keep your customers, keep it simple. *Harvard*

- Business Review*, 90(5).
- SproutSocial. (2018). *Instagram stats*. <https://sproutsocial.com/insights/instagram-stats/>
- Statista. (2020). <https://www.statista.com/forecasts/997051/sources-of-information-about-products-in-the-us>
- Sun, M. (2012). How Does the Variance of Product Ratings Matter? *Management Science*, 58(4), 696–707.
- Sun, Q., Niu, J., Yao, Z., & Yan, H. (2019). Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level. *Engineering Applications of Artificial Intelligence*, 81(December 2018), 68–78.
- Tang, T. (Ya), Fang, E. (Er), & Wang, F. (2014). Is neutral really neutral? The effects of neutral user-generated content on product sales. *Journal of Marketing*, 78(4), 41–58.
<http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc11&NEWS=N&AN=2014-27450-004>
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
- Tirunillai, S., & Tellis, G. J. (2014). Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation. *Journal of Marketing Research*, 51(4), 463–479.
- Toubia, O., Iyengar, G., Bunnell, R., & Lemaire, A. (2019). Extracting Features of Entertainment Products: A Guided Latent Dirichlet Allocation Approach Informed by the Psychology of Media Consumption. *Journal of Marketing Research*, 56(1), 18–36.
- Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing*, 73(5), 90–102.
- Tybout, A. M., Sternthal, B., Malaviya, P., Bakamitsos, G. A., & Park, S. E. B. (2005). Information accessibility as a moderator of judgments: The role of content versus retrieval ease. *Journal of Consumer Research*, 32(1), 76–85.
- Van Hove, G., & Lievens, F. (2007). Social influences on organizational attractiveness: Investigating if and when word of mouth matters. *Journal of Applied Social Psychology*, 37(9), 2024–2047.
- Veloutsou, C., & Delgado-Ballester, E. (2018). New challenges in brand management. *Spanish Journal of Marketing - ESIC*, 22(3), 255–272.
- Vermeer, S. A. M., Araujo, T., Bernritter, S. F., & van Noort, G. (2019). Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media. *International Journal of Research in Marketing*, 36(3), 492–508.
- Vriens, M., Chen, S., & Schomaker, J. (2019). The evaluation of a brand association density metric. *Journal of Product and Brand Management*, 28(1), 104–116.
- Walker, M. A., Mehl, M. R., Moore, R. K., & Mairesse, F. (2007). Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research*, 30, 457–500.

- Wang, F., Liu, X., & Fang, E. (Er). (2015). User Reviews Variance, Critic Reviews Variance, and Product Sales: An Exploration of Customer Breadth and Depth Effects. *Journal of Retailing*, 91(3), 372–389.
- Wang, W., Feng, Y., & Dai, W. (2018). Topic analysis of online reviews for two competitive products using latent Dirichlet allocation. *Electronic Commerce Research and Applications*, 29(January), 142–156.
- Wang, X., Tang, L. (Rebecca), & Kim, E. (2019). More than words: Do emotional content and linguistic style matching matter on restaurant review helpfulness? *International Journal of Hospitality Management*, 77(August 2018), 438–447.
- Weathers, D., Swain, S. D., & Grover, V. (2015). Can online product reviews be more helpful? Examining characteristics of information content by product type. *Decision Support Systems*.
- Westerman, D., Spence, P. R., & Van Der Heide, B. (2014). Social Media as Information Source: Recency of Updates and Credibility of Information. *Journal of Computer-Mediated Communication*, 19(2), 171–183.
- Wickham, H. (2016). *tidyr: Easily tidy data with spread and gather functions*.
- Wickham, H. (2017). ggplot2 - Elegant Graphics for Data Analysis (2nd Edition). *Journal of Statistical Software*, 77(Book Review 2), 2–5.
- Wickham, H., & Francois, R. (2016). *dplyr: A Grammar of Data Manipulation*.
- Winkelmann, P. D. R. (2008). Econometric Analysis of Count Data. In *Econometric Analysis of Count Data*.
- Wong, C. U. I., & Qi, S. (2017). Tracking the evolution of a destination's image by text-mining online reviews - the case of Macau. *Tourism Management Perspectives*, 23, 19–29.
- Wooldridge, J. M. (2011). *Econometric analysis of cross section and panel data* (Issue November).
- Wu, J. (2017). Review popularity and review helpfulness: A model for user review effectiveness. *Decision Support Systems*, 97, 92–103.
- Xu, P., & Liu, D. (2019). Product engagement and identity signaling: The role of likes in social commerce for fashion products. *Information and Management*, 56(2), 143–154.
- Yang, H., & Cho, S. (2015). *Understanding Brands with Visualization and Keywords from eWOM using Distributed Representation*.
- Yau, K. K. W., Wang, K., & Lee, A. H. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal*, 45(4), 437–452.
- Yazdani, E., Gopinath, S., & Carson, S. (2018). Preaching to the Choir: The Chasm Between Top-Ranked Reviewers, Mainstream Customers, and Product Sales. *Marketing Science*, 37(5), 838–851.
- Yin, G., Wei, L., Xu, W., & Chen, M. (2014). Yin, G., Wei, L., Xu, W., & Chen, M. (2014). Exploring Heuristic Cues for Consumer Perceptions of Online Reviews Helpfulness: The Case of Yelp.com. PACIS 2014 Proceedings, 1–12. Exploring Heuristic Cues for Consumer Perceptions of Online Reviews Helpfulness. *PACIS 2014 Proceedings*, 1–12.
- Zaltman, G. (1997). Rethinking Market Research : Putting People Back In. *Journal of*

Marketing Research, 34(4), 424–437.

Zhang, J. (2019). What's yours is mine: exploring customer voice on Airbnb using text-mining approaches. *Journal of Consumer Marketing*, 36(5), 655–665.

Zhang, Ying, & Lin, Z. (2018). Predicting the helpfulness of online product reviews: A multilingual approach. *Electronic Commerce Research and Applications*, 27, 1–10.

Zhang, Yuchi, Moe, W. W., & Schweidel, D. A. (2017). Modeling the role of message content and influencers in social media rebroadcasting. *International Journal of Research in Marketing*, 34(1), 100–119.

Zhou, S., & Guo, B. (2017). The order effect on online review helpfulness: A social influence perspective. *Decision Support Systems*, 93, 77–87.

Zhu, F., & Zhang, X. (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*, 74(March), 133–148.

4

What are Consumers Saying Online about your Products? Mining the Text of Online Reviews to Uncover Hidden Features

1. Introduction

Online consumer reviews are a type of electronic word-of-mouth (eWOM) communication that can be defined as “**peer-generated product evaluations posted on company or third party websites**”(Mudambi & Schuff, 2010).

A study by the consultancy firm BrightLocal (BrightLocal, 2020) reveals that 87% of consumers read online reviews when evaluating a business and 79% trust online reviews as much as personal recommendations. The same study reveals that including online reviews on the retailer website makes the searchers to see the business as more trustworthy. As far as review ratings is concerned, 94% of consumers say that positive reviews make them more likely to use a business, while 92% claim that negative reviews are less likely to use a business after seeing negative reviews. However, consumers admit that the notion of what makes a review positive is subjective, so they might be influenced by the sentiment of the review to decide if they think it is positive or not. In fact, the survey reveals that 79% of consumers point at review sentiment as the most important factor of online reviews they pay attention to, only and slightly surpassed by factors such **as overall star rating, legitimacy (how “real” the review seems)** and recency. The Statista Global Consumer Survey conducted in the United States in 2020 (*Statista*, 2020) reveal also that 43% of consumers point at online reviews as the most used source of information when they search for product information, only surpassed by search engines (e.g. Google), which was the preferred source of information for 68% of consumers. Online reviews were ahead of sources of information such as online stores, brand websites and social media.

Academic literature has also highlighted the power of online reviews to predict different types of consumer behavior such as information adoption decisions (Cyr et al., 2018; Filieri et al., 2018), purchase intentions (Jiménez & Mendoza, 2013; Kostyk et al., 2017; Park & Lee, 2008) and product sales in product categories such as hardware, books, movies and hotels (Chevalier & Mayzlin, 2006; Chintagunta et al., 2010; Hofmann et al., 2017; Lee & Choeh, 2020; Li et al., 2019; Marchand et al., 2017). In this line, Cao et al. (Cao et al., 2011) claim that online reviews help consumers to better anticipate the performance of products and services and to make better choices. In general, previous literature has focused on studying how non-textual features of online reviews (e.g. review rating and volume of online reviews) influence information adoption decisions (Cyr et al., 2018; Filieri et al., 2018) and purchasing decisions (Chevalier & Mayzlin, 2006; Chintagunta et al., 2010; Hofmann et al., 2017; Li et al., 2019; Marchand et al., 2017). However, less attention has been paid to the study of the textual features of online reviews and how firms can take advantage of them.

Exploring the text of online reviews offers companies the opportunity to expand and deepen their knowledge in aspects such as brand image, brand positioning and consumer preferences (Balducci & Marinova, 2018a; Hartmann et al., 2019; Kübler et al., 2019). However, the content of online reviews is qualitative in nature, which makes it difficult to analyze the data and to extract meaningful insights (Chen et al., 2015). Several text mining techniques are available to explore online reviews, which are usually classified into machine learning approaches and lexicon-based methods (Hartmann et al., 2019; Kübler et al., 2019). The selection of one approach over another relies mainly on two aspects: the specific research objectives of the firm and the skills required to conduct the analysis.

Machine learning algorithms usually perform better than lexicon-based methods in terms of extracting intrinsic content of the text (Hartmann et al., 2019; Kübler et al., 2019), but they require higher expertise and greater levels of computational skills, making it difficult for small and medium firms to implement those text mining techniques based on machine learning algorithms.

Lexicon-based methods, which rely on established dictionaries of words, offer the opportunity to analyze the text of eWOM, in an easier and more intuitive way (Pennebaker et al., 2007). In this line, a report developed by Magoulas and Swoyer (Magoulas & Swoyer, 2020) shows that one of the main reasons why companies do not **adopt further Artificial Intelligence (AI), such as machine learning, is the “lack of skilled**

people/ difficulty **hiring the required roles**". Therefore, companies need more guidance in terms of how to take advantage of this type of methods.

From an academic perspective, the study of textual aspects of eWOM, in particular of online reviews, is starting to gain attention. Most scholars exploring review text have focused on studying how the affective content of online reviews, especially sentiment (positive vs. negative), impacts either on review helpfulness or on product sales. For example, Li et al. (2019) and Guo, Wang and Wu (2020) claim that positive comments on a product leads to higher sales. Ahmad and Laroche (2015) conducted a more exhaustive analysis of review emotions to explore how hope, happiness, anxiety and disgust influence review helpfulness. They revealed that emotions expressed with certainty, regardless of the valence, have a higher impact on review helpfulness. According to the cognitive appraisal theory (Ellsworth & Smith, 1988), happiness and disgust are associated with certainty and hope and anxiety with uncertainty. In a similar line, Wang et al. (2019) found that linguistic aspects have a significant impact on review helpfulness in the hotel industry, especially the use of prepositions and auxiliary verbs. Another small stream of academic research has explored the text of online reviews to uncover hidden topics, with the goal of understanding consumer preferences and perceived product and brand images (Ahani et al., 2019; Al-Obeidat et al., 2018; Moon & Kamakura, 2017; Wang et al., 2018). These papers have been conducted by applying different text mining tools, depending on the research objectives. Some of them have used dictionary-based tools (Ludwig et al., 2013; Wang et al., 2019) while others have relied on machine learning algorithms (Guo et al., 2017; Moon & Kamakura, 2017; Toubia et al., 2019).

In this research, the main objective is to show and to illustrate some methods of text mining with the aim of exploring product image uncovering information contained in online reviews. In this way, we can contribute to literature by suggesting new methodologies to study brand image from the textual content of online reviews. From a practical perspective, this research might serve as a guide for companies that want to analyze the content of online reviews. The proposed text mining methods might have several managerial implications, such as helping companies in the new products development by identifying customer preferred product attributes, increasing customer satisfaction by satisfying customer needs revealed in online reviews and improving product positioning and differentiation. To conduct the empirical application, a set of 62,496 online consumer reviews belonging to a whole category of cosmetics products (131 products) available at a popular US cosmetics online retailer on February the 17th 2017 is used.

2. Theoretical Background

2.1. The Persuasive Power of Language

Narrative and language persuasion have been explored in different research domains, such as psychology, communication, and marketing. In these fields, scholars have agreed that narrative and language have an important persuasive power in consumers (Areni, 2003; Hamby et al., 2015; Holtgraves & Lasky, 1999; Li et al., 2019). As claimed by Tausczik and Pennebaker (2010), language is the way in which people express their internal thoughts and emotions. In the same line, psychologists have revealed that people's personality can be uncovered by analyzing linguistic cues, which include factors such as style, syntax, lexicon and topics discussed (Walker et al., 2007).

Overall, the limited number of studies analyzing the persuasion of online texts have concluded that messages' characteristics have a strong power over different types of consumer behavior, such as purchase intention (M. Kim & Lennon, 2007), conversion rates (Ludwig et al., 2013a), liking and commenting brand posts (De Vries et al., 2012) and social media rebroadcasting (Zhang et al., 2017). In spite of the relevancy of text for consumers, most literature exploring online reviews has focused on the study of non-textual aspects of reviews, such as the review rating or the volume of online reviews (Chevalier & Mayzlin, 2006; Dellarocas et al., 2007; Duan et al., 2008a), and the analysis of textual aspects of online reviews remains still scarce.

2.2. The Use of Online Reviews to Study Brand Image

As revealed in the literature review section of Chapter 3, academics are starting to shift their attention to the study of brand image using the textual content of eWOM in general, and online reviews in particular. Although the study of brand image has traditionally been conducted using **surveys' data**, the large number of online reviews posted by consumers in digital platforms together with the increasing number of methods and tools available for text mining, has made the study of brand image using the textual content of online reviews very important for academics and practitioners.

Even though the study of brand image from online reviews is gaining relevance in academia, it is still a relatively new area of study, meaning that literature is still scarce. Those scholars exploring the textual content of online reviews to analyze brand image usually adopt one text mining technique in their analyses. In this research, we want to put together and to illustrate several techniques that can be used depending on the research question to answer. Every technique provides information about different aspects of brand (or product) image.

2.3. Mining the Text of Online Reviews

The concept of “text mining” refers to the process of extracting useful and meaningful information from unstructured text by means of revealing relationships and patterns present, but not obvious, in a document collection (Magerman et al., 2011; O. Netzer et al., 2012). Compared to the study of structured aspects of online reviews (e.g., rating and volume), literature on the study of online reviews text is still scarce, although researchers are increasingly shifting their attention to explore the textual properties of online reviews. Most scholars studying the text of reviews have focused on the study of how the affective content of online reviews (positive vs. negative) influence purchase behaviors (Eslami & Ghasemaghaei, 2018; Guo et al., 2020; Ludwig et al., 2013). Another stream of research has studied textual aspects of online reviews to uncover hidden topics and to identify consumer preferences and perceived product attributes (Ahani et al., 2019; Al-Obeidat et al., 2018; Moon & Kamakura, 2017; Wang et al., 2018).

The study of online reviews text has been conducted in different product settings, such as hotels, restaurants and pc components and several text mining techniques have been used, such as machine learning techniques (e.g., LDA) and lexicon-based techniques (e.g., Linguistic Inquiry and Word Count). Due to the high relevancy of uncovering the content of eWOM texts, some scholars have conducted extensive research to try to unify

the disjoint literature on text mining in social media (Balducci & Marinova, 2018a; Hartmann et al., 2019; Kübler et al., 2019). For example, Berger et al. (2019) provide a detailed review and discussion on the different methodologies used in text mining analysis. Others, such as Hartmann et al. (2019) and Kübler et al. (2019), compare the performance of different text classification methods, including machine learning algorithms and lexicon-based methods. Overall, scholars claim that there is not a single method that always performs better, and it depends on factors such as the metrics we want to analyze, brand strength and type of good (search vs. experience good). Managers should make a tradeoff between resources available and a deeper understanding of content in social media (Kübler et al., 2019).

From a managerial point of view, several text-mining tools can be found in the market and the selection of one or another might be due to different reasons, such as employees text mining skills and company budget. Following previous literature, text mining tools have been classified into two broad categories: lexicon-based methods and machine learning algorithms for natural language processing (NLP). NLP refers to the study and development of computer systems that can interpret text as humans. Since human communication is vague at times (use of colloquialisms, abbreviations, etc), the analysis for natural language is quite difficult. However, NLP techniques have progressed a lot in the last decade. Table 1 shows some of the most popular text mining tools, which belong either to the category of lexicon-based methods or the one of machine learning methods for NLP. All the tools presented in Table 1, except the LIWC, are open source, so everyone can use them for free in software such as R and Python. In the case of LIWC, either a commercial or an academic paid license is needed. There are other text mining tools available for researchers and marketers, but only some popular examples have been shown in Table 1. For example, marketers could also buy commercial text mining tools, such as Watson Natural Language Classifier by IBM Watson, Google Cloud NLP and Amazon Comprehend. The three tools employ machine learning to unearth insights and extract meaningful information about people, places, or events.

Table 1

Popular text mining methods

Type of method	Tool
Lexicon-based	<ul style="list-style-type: none"> - Valence Aware Dictionary and Sentiment Reasoner (VADER) (Hutto & Gilbert, 2014). It relies on a dictionary that map lexical features to emotion intensities, known as sentiment scores. It is based on word bigrams (pairs of words). The output informs if the text expresses a positive, negative, or neutral opinion. - National Research Council (NRC) lexicon (Salehan & Kim, 2015). Provides the frequency of words in a document belonging to eight basic

	<p>emotions (anger, anticipation, disgust fear, joy, sadness, surprise, and trust). It is based on word unigrams (single words).</p> <ul style="list-style-type: none"> - Finn Arup Nielsen (AFINN) (Nielsen, 2011). It categorizes the sentiment of the text between -5 (negative) and +5 (Positive). It is based in unigrams. - BING (M. Hu & Liu, 2004). The lexicon labels texts as between -1 (negative) and +1 (positive). It is based in unigrams. - Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2015). It categorizes words into around 100 linguistic categories, including style categories (e.g., verbs, adjectives, and pronouns) and psychological processes (e.g., leisure, home and feel).
Machine learning algorithms for Natural Language Processing (NLP)	<ul style="list-style-type: none"> - Supervised machine learning. The machine is trained using data which is already “labelled”. It means that the data is already tagged with the correct answer. Thus, a supervised learning algorithm learns from labelled training data and helps to predict outcomes for unforeseen data. Popular supervised learning algorithms: <ul style="list-style-type: none"> - Support Vector Machines (SVM) - Maximum Entropy - Bayesian Networks - Conditional Random Field - Neural Networks/Deep learning - Unsupervised machine learning. The model does not need to be supervised and it works on its own to discover information. It mainly deals with unlabelled data. Unsupervised learning algorithms allow to perform more complex tasks. <ul style="list-style-type: none"> - Name Entity Recognition (NER) - Latent Semantic Analysis (LSA) - Latent Dirichlet Allocation (LDA) - Dependency Parser - Word2vec word embedding

In this research, the main objective is to show and to illustrate some methods of text mining, based on unsupervised machine learning algorithms, with the aim of exploring product image by uncovering some information contained in online reviews. Texts of online reviews contain much information, so one could explore different types of research questions. In this research, we take some examples of research questions that could be defined in order to study aspects related to product image.

RQ1. Which are the most common words in the category?

RQ2. Which are the words associated to positive and negative emotions?

RQ3. Which are the most common pair of words in the category?

RQ4. How similar are products in the category in terms of textual content?

RQ5. Which are the main topics discussed by consumers in the category?

3. Methodology

3.1. Data

To carry out the research, we used the same database as in Chapter 3, 62,496 online consumer reviews from a US cosmetics retailer website were collected. The database

represents the entire set of online reviews available at the website under the category of **“blush” products on February the 17th 2017**. These reviews belong to a total of 131 products of 44 different cosmetics brands. Cosmetics are classified as a type of experience products according to Girard and Dion (2010). As defined by Nelson (1970), experience products are those whose attribute information cannot be known before use or consumption. In contrast, attribute information of search products (e.g., price, quality, size, and dimension) can be easily evaluated prior to consumption. Since it is very difficult to evaluate experience products, consumers usually rely more on recommendations than in the case of search products (King & Balasubramanian, 1994; Senecal & Nantel, 2004). Most studies focusing on mining the text of online reviews have been conducted in a specific type of experience products, which is the tourism industry (Ahani et al., 2019; Berezina et al., 2016; Guo et al., 2017; Wong & Qi, 2017), and others within search product categories, such as smart phones (Jiang et al., 2015) and wireless mice (Wang et al., 2018). Uncovering hidden aspects of online reviews might be particularly important in the cosmetics category, where prospective consumers might need to know if the product fits his or her physical characteristics before buying it. Therefore, knowing specific product features, such as the skin type or skin colour in which the product fits, or how to apply the product for a better fit, is likely to be quite important for consumers in their decision-making process. Moreover, as revealed by BrightLocal (2020), **the category of “Hair/Beauty” is one of the top-10 industries** in which consumers usually look at reviews. More concretely, it is the 9th industry out of 34 industries studied. In fact, 85% of consumers claim looking at reviews in this industry. **The first two industries are “Restaurants/Cafés” and “Hotels/B&Bs”.** Thus, it is interesting to analyse online reviews in the cosmetics industry to explore aspects related to brand image. Since consumers usually read reviews to evaluate **cosmetics ‘products**, they are likely to be influenced by what other consumers say in those reviews.

Besides collecting the text of individual online reviews, other non-textual information, which allows us to describe the characteristics of the online reviews and products studied in this research, was recorded, such as product price, product average rating, number of reviews of each product and average length (in words) of the reviews.

3.2. Data Analysis

A text mining analysis was conducted in the software R, using packages such as *dplyr* (Wickham & Francois, 2016), *tidyr* (Wickham, 2016), *ggplot2* (Wickham, 2017) and *topicmodels* (Hornik & Grün, 2011). Table 2 shows the data analysis workflow used in

this research and the main issues addressed at each stage. Special attention is paid to explain the topic modelling analysis (LDA), which is the main analysis in this research.

Table 2

Data analysis workflow

Stage	Issues involved
1. Data acquisition	<ul style="list-style-type: none"> - Online reviews were downloaded from one cosmetic online retailer using web scrapping. - Raw data was imported to the statistical software R. - A database of 62,496 online reviews of 131 products belonging to 44 brands was built (all the brands belonging to the “blush” category).
2. Data pre-processing	<ul style="list-style-type: none"> - Tokenization. Break the text into units, in our case words. - Cleaning. Removing non-textual information, such as symbols (e.g., %, & /). - Removing stop words. All common words that do not contribute to the distinctive meaning and context of documents can be removed (e.g., “a”, “the”). Besides, words that are very common in this cosmetics domain to every product have also been removed (e.g., “products”, “blush” and “skin”). Since we are not interested in exploring individual brands or products, we also removed brand names and product names.
3. Text Mining Analysis	<p>Techniques used and research questions to answer:</p> <ol style="list-style-type: none"> 1) Word frequencies. RQ1. Which are the most common words in the category? 2) Sentiment analysis. RQ2. Which are the words associated to positive and negative emotions? 3) Relationship between words. RQ3. Which are the most common pair of words in the category? 4) Product-word correlations. RQ4. How similar are products in the category in terms of textual content? 5) Topic modelling (LDA). RQ6. Which are the main topics discussed by consumers in the category?

Topic Modelling: Latent Dirichlet Allocation (LDA)

A topic model is a type of probability model for discovering the abstract “topics” that occur in a collection of documents (Guo et al., 2017). Latent Dirichlet allocation (LDA) is the most common method to conduct topic modelling and it has been used in several marketing papers to identify, for example, customer preferences and needs (Heng et al., 2018; Tirunillai & Tellis, 2014; Toubia et al., 2019; Zhang et al., 2017). LDA is a Bayesian learning algorithm that extracts “topics” from a large collection of unstructured text documents on the basis of co-occurrence (Toubia et al., 2019). Topics are defined as word distributions that commonly co-occur and thus have a certain probability of appearing in a topic (Berger et al., 2020). Every document (D) is described as a probabilistic mixture of topics (K) and every topic is described as a probabilistic mixture of words (W). Thus, topics may be viewed as groups of words that are semantically related to each

other. Conceptually, LDA assumes that a consumer, when writing an online review, wants to convey one or more topics in a review (e.g., a review about the colour of the product).

4. Results

4.1. Descriptive Statistics

Before conducting the text mining analysis, it is interesting to explore several variables related to products and online reviews in the database, not directly linked to the text of online reviews, to better understand the nature of the data. Table 3 shows the descriptive statistics for some product non-textual variables. The average rating of online reviews in **the “blush” category is 4.5, since most online reviews for each product are 5-star reviews**, while, on average, only 73 reviews of each product are 1-star. In terms of review length, reviews have on average 52.2 words, although there are reviews with just one word and others that have 1,032 words.

Table 3

Descriptive statistics of some non-textual product variables

	N	Mean	St. Dev.	Min	Max
Product price	62,496	29.5	7.7	10	80
Product average rating	62,496	4.5	0.2	3.1	5.0
Product number reviews	62,496	5,403.2	6,636.4	1	16,404
Product 1star Reviews	62,496	73.3	68.5	0	180
Product 2star Reviews	62,496	122.0	121.6	0	318
Product 3star Reviews	62,496	242.4	253.6	0	655
Product 4star Reviews	62,496	858.2	945.7	0	2,411
Product 5star Reviews	62,496	4,107.3	5,259.3	0	12,840
Reviews average length	62,496	52.2	43.3	1	1,032

4.2. Text Mining Analysis

Text mining covers a broad range of tasks (e.g., text clustering, sentiment analysis and entity recognition). Based on the research objectives and the type of data the researcher is dealing with, one task or another should be accomplished. In this research, some basic tasks that can serve as the starting point to obtain relevant insights from the text of online reviews are illustrated. Then, a more complex text mining analysis, LDA, is conducted to extract the main topics discussed in online reviews by consumers. Since the study is **dealing with cosmetics online reviews belonging to the category of “blush”, the findings**

are strongly related to this scenario, but the same text mining process could be followed to explore online reviews in other product settings. Next sub-sections present the analyses and results for each research question.

Word Frequencies: Which are the Most Common Words in the Category?

By exploring word frequencies, the researcher can identify the main words that consumers use in online reviews. Figure 1 shows the top-20 most frequent words in online reviews belonging to the “blush” category. The words “pink”, “day”, “light”, “glow” and “shade” are the 5 most used words in “blush” online reviews.

Figure 1. Top 20 most frequent words in the "blush" category

Sentiment Analysis: Which Words are Associated to a Positive or Negative Sentiment?

A sentiment analysis can be also carried out to identify the words related to positive and negative sentiments. Several dictionaries are available to conduct sentiment analysis (e.g., Bing, NRC and AFINN). Some dictionaries only classify words into two categories: positive or negative. However, we have adopted the National Research Council (NRC) lexicon because it categorizes words into more emotional categories. First, it provides two general categories of sentiment: *positive* and *negative* sentiment. Moreover, this

lexicon adds eight emotions, which adds a level of specificity into the analysis: *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise*, and *trust*. Figure 2 shows the top-10 words for each sentiment and emotion used in online reviews using the NRC lexicon. It classifies some words in several categories at the same time, as it is the case of the word “disappointed” in our online reviews, which belong to the category of *anger*, *disgust*, and *sadness*, all of them expressing *negative* sentiment. In terms of negative emotions, “disappointed” is the word that contributes the most to the *anger* emotion, “dark” the one most contributing to *sadness* and “bad” to *fear*. As far as positive emotions are concerned, “favorite” and “gorgeous” are the words more contributing to *joy*, “wonderful” to *surprise* and “recommend” and “favorite” to *trust*. Finally, “time” is the word more related to *anticipation*.

Figure 2. Top 10 words for each emotion used in online reviews using NRC lexicon

Relationship between Words: Which are the Most Common Pair of Words in the Category?

Another interesting analysis to conduct is to explore which words usually appear together in online reviews to know how consumers associate words one with each other. Figure 3

shows the most common pairs of words that appear together in online reviews of the “blush” cosmetics category. In this case, these words appear together in online reviews more than 200 times. Pairs of words, such as “totally worth”, “absolutely gorgeous”, “4 stars” and “matte finish” are some of the most pairs of words used by consumers. Some words, such as “glow” are related to many other words, such as “nice”, “healthy”, “subtle” and “rosy”.

Figure 3. Most common paired of words in online reviews in the “blush” category (more than 200 times)

Product-Word Correlations: How Similar are Products in the Category in terms of Textual Content?

Text mining also allows the researcher to identify similarities and differences between products in terms of the content of online reviews text. To do that, correlations between products and words can be analyzed to identify which products are more similar one to another in terms of the textual content of their reviews. In the same line, word frequencies between paired of products can be studied to identify frequent words in both set of online reviews and words more frequent in one set than in another. For illustrative purposes, three paired of products with three different levels of correlations have been graphically represented: “P210589” vs. “P384346” (correlation=0.87), “P401101” vs. “P412347” (correlation=0.20) and “P407080” vs. “P416377” (correlation=0.00). Figure

4 shows a comparison of word frequencies between two of the most correlated products in terms of textual content in the category (“P210589”/ “P384346”), meaning that online reviews of the two products are very similar in terms of words. Words close to the red line are those with similar frequency in online reviews of both products. For example, “pink”, “stain” and “day” are words not only with similar frequencies in both set of online reviews but also with high frequency (high frequency end). Words far from the red line are words that are found more in one set of online reviews than in the other. For example, the word “purple” is found in “P384346” but not much in “P210589”. Figure 5 shows a comparison between two medium-correlated products, “P401101” and “P412347”. In contrast to , a smaller number of words are represented in the graph, since both sets of online reviews are less similar and they have fewer common words. The words “shade” and “pigmented” have similar and high frequency in both set of online reviews. On the other hand, the word “vibrant” is found in online reviews of “P416377” but not much in reviews of “P401101”. Figure 6 represents the map for two non-correlated products, so it can be observed that there are not words in common for the two products.

Figure 4. Word frequencies of two high correlated ($\text{cor.} = 0.87$) products in terms of textual content

Figure 5. Word frequencies of two medium correlated products ($\text{cor.} = 0.20$) in terms of textual content

Figure 6. Word frequencies of two non-correlated products ($\text{cor.} = 0.00$) in terms of textual content

Topic Modeling (LDA): Which are the Main Topics Discussed by Consumers in the Category?

One of the main issues when conducting LDA is to determine the optimal number of topics that represent all the text input. To do that, we computed the metrics introduced by Cao et al. (2009), Deveaud et al. (2014), Griffiths and Steyvers (2004) and Arun et al. (2010). Metrics are shown in Figure 7, where the x-axis represents the number of topics and the y-axis represents the log-likelihood of word-topic probability. The optimal

number of topics (K) is that number that minimize the metrics by Cao et al. (2009) and Arun et al. (2010) and maximizes the metrics by Deveaud et al. (2014) and Griffiths and Steyvers (2004). Metrics are estimated using marginal log-likelihood of word-topic probability in the document with fivefold cross-validation. The process to identify the optimal number of topics begins by computing a LDA model with two topics and then gradually increase the number of dimensions until the log-likelihood reaches a maximum, in the case of Griffiths2004 and Devaud2014, or until the log-likelihood reaches a minimum, in the case of CaoJuan2009 and Arun2010. It is observed in that $K=22$ is a number of topics that meets properly the metrics requirements, especially in the case of the Arun2010 and the Deveaud2014 metrics. Adding more topics does not imply a much better log-likelihood of word-probability in the case of Griffiths2004, which log-likelihood value remains quite stable from $K=18$ onwards. In the same line, the CaoJuan2009 metric stabilizes from $K=18$ to $K=54$. When applying LDA, it is important to bear in mind that an excessive number of topics could lead to a model too complex, making interpretations too difficult. Therefore, $K=22$ is selected in this research as the optimal number of topics.

Figure 7. Optimal number of topics (k)

Table 4 shows the 22 topics identified by LDA, the topic label given to each topic id, the top-10 words associated to each topic and the decreasing relevancy of each topic in the “blush” category (topic proportion in %). LDA allows some words to appear in different topics. For example, the word “pink” appears in several topics, such as topic 2 and topic 4. Topic labels are not produced by LDA and it is a task for the researchers to assign labels based on the most frequent words and on the researcher domain knowledge. Thus,

the labelling process is rather subjective and involves some judgment. In this research, the different researchers have participated in the interjudge agreement procedures. The naming of topics was first conducted by one of the researchers and then confirmed by the two other researchers. Thus, to produce the labels, each individual researcher analysed the most frequent words associated to each topic and these words were identified in the original blush online reviews to uncover the context in which they are used by consumers. Since this research deals with online reviews of a very specific product category, which is “blush”, topics might be more related one to another and it is more difficult to see clear differences between them. Nevertheless, topics might be easier to differentiate if the analysis is done comparing different product categories, or with categories in which products are more heterogeneous.

Table 4

Topics, associated words, and topic proportions

Topic id	Name	Top 10 words in decreasing order of probability of belonging to the topic	Topic proportion in “blush” products (%)
1	Dry and matte	dry, matte, apply, bought, pink, day, nice, shade, easy, light	21.46%
2	Glow and soft highlighter	glow, pink, highlighter, light, soft, shimmer, brush, fair, shade, highlight	11.10%
3	Highlighter and bronzer palette	highlighter, bronzer, palette, contour, highlight, shade, light, nice, pigmented, contouring	7.49%
4	Powder cream	cream, powder, day, apply, blend, easy, nice, stain, creamy, fingers	6.78%
5	Pop of product	pop, pigmented, packaging, brush, shade, bought, day, pink, amazing, cute	5.85%
6	Brush	brush, nice, glow, smells, light, pretty, fair, pink, shimmer, coral	4.95%
7	Pink and peach glow	pink, glow, peach, packaging, nice, pigmented, brush, pretty, light, day	4.27%
8	Day usage	day, pink, shade, nice, pigmented, rose, light, bought, price, glow	4.02%
9	Glow and light palette	glow, light, palette, powder, flush, luminous, fair, ambient, shade, pink	3.94%
10	Bronzer glow	bronzer, glow, nice, light, fair, brush, shimmer, time, day, dark	3.81%
11	Stain and tint	stain, tint, day, pink, apply, lasts, blend, time, nice, pretty	3.76%
12	Shimmer bronzer and highlighter	shimmer, bronzer, highlight, glow, highlighter, palette, brush, light, pink, nice	3.74%
13	Soft colour	pink, light, brush, fair, powder, glow, pretty, nice, soft, day	3.64%
14	Duo size	duo, size, pigmented, shade, pink, light, tone, shimmer, amazing, gorgeous	2.78%
15	Easy to apply powder	powder, glow, easy, apply, stick, day, nice, pink, bought, shade	2.51%

16	Day lasting	day, lasts, exposed, pink, pigmented, shade, light, bought, time, hours	2.30%
17	Pigmented shadows	shadows, pigmented, palette, eye, shadow, eyeshadow, blend, eyeshadows, shades, shade	2.09%
18	Pink day shade	pink, shade, day, light, nice, glow, time, fair, easy, bright	1.97%
19	Cocoa and macaroon colours	cocoa, macaroon, blondie, eyes, received, lipstick, multistick, blend, influenster, easy	1.18%
20	Free samples: influenster program	received, influenster, glow, shade, kisses, kink, day, light, soft, packaging	1.03%
21	Shimmer day wear	shimmer, pink, day, bought, time, wear, glow, tone, nice, lasts	0.80%
22	Light shade	shade, light, fair, glow, pink, day, worth, super, tone, nice	0.52%

Looking at results in Table 4, several trends can be identified in the data. First, many topics focus on features related to product colours (e.g., Pink and peach glow, Soft colour, Cocoa and macaroon colours). Second, several topics focus on aspects related to product usage (e.g., day usage, shimmer day wear, day lasting, easy to apply powder, pink day shade). Third, other topics focus on product format (e.g., stain and tint, powder cream, duo size, glow, and light palette). Moreover, different aspects of product main functions or purposes are captured across products (e.g., bronzer glow, shimmer day wear, highlighter and bronzer palette, shimmer bronzer and highlighter). Topic 21 is probably the one that differs more from the rest, since it captures aspects related to a cosmetics program called “Influenster”, which is a collaborative Marketing platform of cosmetics products. Active consumers in the platform receive free samples of different products. Looking at the relevancy of each topic within online reviews in the “blush” category, represented in Figure 8 (fourth column in Table 4), it can be noticed that the topic “dry and matte” is the most discussed within “blush” online reviews, followed by “glow and soft highlighter” and “highlighter and bronzer palette”.

Figure 8. Topic proportion in “blush” online reviews (in %)

5. Conclusions

So far, most research about online reviews has mainly focused on studying aspects of review non-textual features, such as the influence of review rating on review helpfulness (Pan & Zhang, 2011; Wu, 2017) or the impact of review volume on product sales (Chevalier & Mayzlin, 2006; Marchand et al., 2017). However, research on textual features of online reviews is still relatively limited. The present research was therefore conducted to illustrate some of the text mining tools that can be used to uncover hidden aspects of review texts. As stated by literature in language persuasion, messages have a strong power in predicting different types of consumer behavior, such as purchase intention (M. Kim & Lennon, 2007) and conversion rates (Ludwig et al., 2013a). Thus, the study of online consumer texts is quite important to develop companies' marketing strategies.

Although this paper is quite practice-oriented in nature, it contributes to the literature on brand image by illustrating some Natural Language Processing (NLP) methods that can be used to answer some research questions related to the study of brand (or product)

image. By using the unsupervised machine learning algorithms illustrated in this paper, associations with brands and products can be extracted from online reviews.

NLP techniques offer a completely new scenario where any type of online consumer text can be used in order to understand the image of brands, products or services. Online texts have one main advantage over traditional multidimensional scales used to study brand image, eWOM is spontaneous, so consumers usually express their true feelings and perceptions (Marchand et al., 2017; Yang & Cho, 2015). In general, consumers articulate their opinions online because they want to help or warn others or because they want to communicate their status (Kozinets et al., 2010). Therefore, NLP is a very powerful tool to uncover real feelings and perceptions of consumers. In this research, we show some of the research questions that can be answered using NLP techniques.

From a practical point of view, brand managers and retailers could adopt the text mining tools presented in the research for several applications, such as the following ones:

- Get insights about consumer interests and preferences.
- Customer satisfaction. Analyze which positive and negative words are associated to each product and explore which product features are associated to positive and negative sentiments.
- Increase customer loyalty by satisfying consumer interests and preferences.
- Identify customer claims to try to improve the product or service.
- New product development. Get insights about what product characteristics to improve or drop in next product releases.
- Product positioning and differentiation. Companies can analyze how their products are positioned in the market, by identifying which are the main product associations. In this way, companies can either confirm that they are positioned where they want to or, if not, they can develop some strategies to change the product positioning and to try to differentiate from competitors.

In Table 5 we show how the answers to the different research questions defined in this research could be linked to the managerial applications.

Table 5.

Link between research questions and managerial application of findings

Research Question	Application of findings
RQ1. Which are the most common words in the category?	<ul style="list-style-type: none"> - Get insights about consumer interests and preferences. - New product development - Increase customer loyalty by satisfying consumer interests and preferences.
RQ2. Which are the words associated to positive and negative emotions?	<ul style="list-style-type: none"> - Customer satisfaction. Analyze which positive and negative words are associated to each product and explore which product features are associated to positive and negative sentiments. - Identify customer claims to try to improve the product or service. - New product development - Increase customer loyalty by satisfying consumer interests and preferences.
RQ3. Which are the most common pair of words in the category?	Product positioning and differentiation
RQ4. How similar are products in the category in terms of textual content?	Product positioning and differentiation
RQ5. Which are the main topics discussed by consumers in the category?	<ul style="list-style-type: none"> - Get insights about consumer interests and preferences. - New product development - Increase customer loyalty by satisfying consumer interests and preferences.

In this specific example of cosmetics, brand managers could get some valuable insights. **First, “pink”, “light”, “glow” and “day” are the most used words in online reviews in the blush category**, which might reveal the features consumers are looking for when purchasing this type of product. If managers wanted to get more insights on those words, they could explore the words bi-grams or trigrams to know which words they appear together with. By analyzing the most common bigrams in the blush category, it can be

observed that customers are interested in features such as “easily blends”, “healthy glow”, “finish matte” and “super pigmented”. Thus, those product characteristics seem to be quite important for customers. In terms of sentiment, online reviews in the category are quite positive since the product average rating is 4.5. However, several negative words can be identified, such as “disappointed” and “bad”, which are two of the most common. If interested in exploring these words in a greater detail, bi-grams and trigrams could also be explored to try to gain insights of why customers feel disappointed or to which words is linked the negative word “bad”. In terms of positive sentiments, “recommend” is one of the most used words and managers might be interested in knowing the products more associated to this word. From the LDA analysis, it is observed that “Dry and matte” is by far the most discussed topic in the online reviews, which highlights the main customer preferences. Companies could use the previous findings for one or more of the presented managerial applications. For example, since “healthy” products seem to be relevant for consumers, companies could think about launching a new organic product to the market, which satisfy the identified consumer need. Nevertheless, it is important to bear in mind that depending on the research objective of the marketing professionals, one type of text mining analysis or another could be conducted.

5.1. Limitations and Future Research

Although we have highlighted the benefits of using eWOM over traditional survey data to explore brand image and consumer behavior, we could also point out some limitations. For example, surveys can be carefully designed to answer specific research objectives, and it might be the case that those objectives cannot be answered using eWOM. Moreover, companies and scholars have less expertise in analyzing eWOM over questionnaires, which have been used for a long time. Besides, the profile of consumers who write online reviews might not be representative of all the population, so there might be some consumers who do not use the Internet and should be approached by traditional techniques, such as surveys. Ideally, one could combine both techniques to get more insights about consumer behavior and brand image. If applying both techniques, we would be able to compare the findings concerning brand image using either online opinions or traditional methods, such as multidimensional scaling.

In this research, we identify some of the most common techniques for NLP and we focus on answering some of examples of research questions that could be addressed with some of the techniques. However, other research could answer other type of research questions using other type of NLP techniques.

This research analyzes online reviews, but the same techniques could be also applied to other online texts, such as tweets or Instagram posts. In terms of future research, comparisons between textual features of messages in different social media could be addressed. Moreover, instead of analyzing the results of one product category, future research could conduct the analyses in a more disaggregated level, such as brand or product level and compare findings between different brands or products.

References

- Aaker, D. A. (1991). Managing Brand Equity. *Managing Brand Equity*.
- Aaker, D. A. (1996). Building strong brands - aaker,da. *Building Strong Brands*, 13(4), 382–386.
- Ahani, A., Nilashi, M., Yadegaridehkordi, E., Sanzogni, L., Tarik, A. R., Knox, K., Samad, S., & Ibrahim, O. (2019). Revealing customers' satisfaction and preferences through online review analysis: The case of Canary Islands hotels. *Journal of Retailing and Consumer Services*, 51(July), 331–343.
- Ahmad, S. N., & Laroche, M. (2015). How do expressed emotions affect the helpfulness of a product review? Evidence from reviews using latent semantic analysis. *International Journal of Electronic Commerce*, 20(1), 76–111.
- Aiken, L. S., & West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*.
- Al-Obeidat, F., Spencer, B., & Kafeza, E. (2018). The Opinion Management Framework: Identifying and addressing customer concerns extracted from online product reviews. *Electronic Commerce Research and Applications*, 27, 52–64.
- Al., H. et. (2010). *Multivariate Analysis 7 th Edition*.
- Aljukhadar, M., Senecal, S., & Daoust, C.-E. (2012). Using Recommendation Agents to Cope with Information Overload. *International Journal of Electronic Commerce*, 17(2), 41–70.
- Ansari, A., & Mela, C. F. (2003). *E-Customization*. 40(2), 131–145.
- Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*, 57(8), 1485–1509.
- Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2), 277.
- Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1), 29–51.
- Areni, C. S. (2003). The Effects of Structural and Grammatical Variables on Persuasion: An Elaboration Likelihood Model Perspective. *Psychology and Marketing*, 20(4), 349–375.
- Arun, R., Suresh, V., Madhavan, C. E. V., & M.N, N. M. (2010). On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In *Advances in Knowledge Discovery and Data Mining* (pp. 391–403). <http://hdl.handle.net/10722/26088>
- Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. A. (2016). The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors. *Journal of Marketing Research*, 53(3), 297–318.
- Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, 0, 2200–2204.
- Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of Online Consumer Reviews: Readers'

- Objectives and Review Cues. *International Journal of Electronic Commerce*, 17(May 2015), 99–126.
- Baker, W. E. (2001). The Diagnosticity of Advertising Generated Brand Attitudes in Brand Choice Contexts. *Journal of Consumer Psychology*, 11(2), 129–139.
- Baker, W. E., Lutz, R. J., Journal, S., Spring, N., Taylor, P., Baker, W. E., & Lutz, R. J. (2000). Effectiveness An Empirical Test of an Updated Relevance- Accessibility Model of Advertising Effectiveness. *Journal of Advertising*, 29(1), 1–14.
- Bakhshi, S., Kanuparth, P., & Shamma, D. A. (2015). Understanding Online Reviews: Funny, Cool or Useful? *CSCW 2015: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1270–1276.
- Baksi, A. K., & Panda, T. K. (2018). Branding destinations with multisensory brand associations and evaluating its impact on behavioural pattern under the intervention of multiplex phenomenon of relationship-branding. *Management Science Letters*, 8(11), 1169–1182.
- Balducci, B., & Marinova, D. (2018a). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, 46(4), 557–590.
- Balducci, B., & Marinova, D. (2018b). Unstructured data in marketing. In *Journal of the Academy of Marketing Science* (Vol. 46, Issue 4, pp. 557–590). Springer New York LLC.
- Beach, L. R. (1993). Broadening the definition of decision making: The role of prechoice screening option. *Psychological Science*, 4(4), 215–220.
- Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding Satisfied and Dissatisfied Hotel Customers: Text Mining of Online Hotel Reviews. *Journal of Hospitality Marketing and Management*, 25(1), 1–24.
- Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the Tribes: Using Text for Marketing Insight. *Journal of Marketing*, 84(1), 1–25.
- Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive Consumer Choice Processes. *Journal of Consumer Research*, 25(3), 187–217.
- Biehal, G., & Chakravarti, D. (1983). Information Accessibility as a Moderator of Consumer Choice. *Journal of Consumer Research*, 10(1), 1.
- Biswas, D., Grewal, D., & Roggeveen, A. (2010). How the order of sampled experiential products affects choice. *Journal of Marketing Research*, 47(3), 508–519.
- Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), 115–143.
- Brandt, C., De Mortanges, C. P., Bluemelhuber, C., & Van Riel, A. C. R. (2011). Associative Networks: A New Approach to Market Segmentation. *International Journal of Market Research*, 53(2), 187–207.
- Breugelmans, E., Campo, K., & Gijsbrechts, E. (2007). Shelf sequence and proximity effects on online grocery choices. *Marketing Letters*, 18(1–2), 117–133.
- BrightLocal. (2020). *Local Consumer Review Survey*. Available at: <https://www.brightlocal.com/research/local-consumer-review-survey/>
- Brown, C. L., & Krishna, A. (2004). The skeptical shopper: A metacognitive account for the effects of default options on choice. *Journal of Consumer Research*, 31(3), 529–539.

- Bruine de Bruin, W., & Keren, G. (2003). Order effects in sequentially judged options due to the direction of comparison. *Organizational Behavior and Human Decision Processes*, 92(1–2), 91–101.
- Buda, R., & Zhang, Y. (2000). Consumer product evaluation: the interactive effect of message framing, presentation order, and source credibility. *Journal of Product & Brand Management*, 9(4), 229–242.
- Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics Using Stata*. Stata Press Book, StataCorp LP.
- Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data, Second edition. In *Regression Analysis of Count Data, Second Edition*.
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72(7–9), 1775–1781.
- Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, 50(2), 511–521.
- Chaiken, S. (1987). *The heuristic model of persuasion*.
- Chapman, C., & Feit, E. M. (2015). *R for marketing research and analytics*.
- Chen, C. C., & Tseng, Y. De. (2011). Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, 50(4), 755–768.
- Chen, K., Kou, G., Shang, J., & Chen, Y. (2015). Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches. *Electronic Commerce Research and Applications*, 14(1), 58–74.
- Cheng, Y. H., & Ho, H. Y. (2015). Social influence’s impact on reader perceptions of online reviews. *Journal of Business Research*, 68(4), 883–887.
- Cheung, C. M. K., Lee, M. K. O., & Rabjohn, N. (2008). The impact of electronic word-of-mouth. In *Internet Research* (Vol. 18, Issue 3).
- Chevalier, J. A., & Goolsbee, A. (2003). Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics*, 1, 203–222.
- Chevalier, J., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 43(3), 345–354.
- Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science*, 29(5), 944–957.
- Cho, E., Fiore, A. M., & W. Russell, D. (2015). Validation of a Fashion Brand Image Scale Capturing Cognitive, Sensory, and Affective Associations: Testing Its Role in an Extended Brand Equity Model. *Psychology & Marketing*, 32(1), 28–48.
- Chua, A. Y. K., & Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior*, 54, 547–554.
- Chung, C.K., & Pennebaker, J. W. (2007). The psychological function of function words. *Social Communication: Frontiers of Social Psychology*, 343–359.
- Chung, Cindy K., & Pennebaker, J. W. (2007). Revealing dimensions of thinking in open-

- ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality*, 42(1), 96–132.
- Cohn, M. a., Mehl, M. R., & Pennebaker, J. W. (2004). Markers of Linguistic Psychological Change. *Psychological Science*, 15(10), 687–693.
- Cohn, M. a., Mehl, M. R., & Pennebaker, J. W. (2014). Markers of Linguistic Psychological Change Surrounding September 11, 2001. *Psychological Science*, 15(10), 687–693.
- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2014). *Markers of Linguistic Psychological Change Surrounding September*. 15(10), 687–693.
- Cui, G., Lui, H.-K., & Guo, X. (2012). The Effect of Online Consumer Reviews on New Product Sales. *International Journal of Electronic Commerce*, 17(1), 39–58.
- Cyr, D., Head, M., Lim, E., & Stibe, A. (2018). Using the elaboration likelihood model to examine online persuasion through website design. *Information and Management*, April 2017, 1–15.
- Darley, W. K., & Smith, R. E. (1993). Advertising Claim Objectivity: Antecedents and Effects. *Journal of Marketing*, 57(4), 100.
- Davis, D. F., Golcic, S. L., & Marquardt, A. (2009). Measuring brand equity for logistics services. *The International Journal of Logistics Management*, 20(2), 201–212.
- De Vries, L., Gensler, S., & Leeftang, P. S. H. (2012). Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing. *Journal of Interactive Marketing*, 26(2), 83–91.
- Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, 27(4), 293–307.
- Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4), 23–45.
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective Latent Concept Modeling for ad hoc information retrieval. *Document Numerique*, 17(1), 61–84.
- Drèze, X., & Zufryden, F. (2004). Measurement of online visibility and its impact on Internet traffic. *Journal of Interactive Marketing*, 18(1), 20–37.
- Duan, W., Gu, B., & Whinston, A. B. (2008a). Do online reviews matter? - An empirical investigation of panel data. *Decision Support Systems*, 45(4), 1007–1016.
- Duan, W., Gu, B., & Whinston, A. B. (2008b). The dynamics of online word-of-mouth and product sales-An empirical investigation of the movie industry. *Journal of Retailing*, 84(2), 233–242.
- Elberse, A., & Eliashberg, J. (2003). Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures. *Marketing Science*, 22(3), 329–354.
- Ellsworth, P. C., & Smith, C. A. (1988). From appraisal to emotion: Differences among unpleasant feelings. *Motivation and Emotion*, 12(3), 271–302.
- Ert, E., & Fleischer, A. (2016). Mere Position Effect in Booking Hotels Online. *Journal of Travel Research*, 55(3), 311–321.
- Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management*, 55, 62–73.

- Eslami, S. P., & Ghasemaghaei, M. (2018). Effects of online review positiveness and review score inconsistency on sales: A comparison by product involvement. *Journal of Retailing and Consumer Services*, 45(June), 74–80.
- Faircloth, J. B., Capella, L. M., & Alford, B. L. (2001). The Effect of Brand Attitude and Brand Image on Brand Equity. *Journal of Marketing Theory and Practice*, 9(3), 61–75.
- Feldman, J. M., & Lynch, J. G. (1988). Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior. *Journal of Applied Psychology*, 73(3), 421–435.
- Felfernig, A., Friedrich, G., Gula, B., Hitz, M., Kruggel, T., Leitner, G., Melcher, R., Riepan, D., Strauss, S., Teppan, E., & Vitouch, O. (2007). Persuasive recommendation: Serial position effects in knowledge-based recommender systems. *Proceedings of the 2nd International Conference on Persuasive Technology (PERSUASIVE '07)*, 283–295.
- Filieri, R., Hofacker, C. F., & Alguezaui, S. (2018). What makes information in online consumer reviews diagnostic over time? The role of review relevancy, factuality, currency, source credibility and ranking score. *Computers in Human Behavior*, 80, 122–131.
- Filieri, R., & McLeay, F. (2014). E-WOM and Accommodation: An Analysis of the Factors That Influence Travelers' Adoption of Information from Online Reviews. *Journal of Travel Research*, 53(1), 44–57.
- Filieri, R., McLeay, F., Tsui, B., & Lin, Z. (2018). Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services. *Information and Management*, 55(8), 956–970.
- Filieri, R., Raguseo, E., & Vitari, C. (2018). When are extreme ratings more helpful? Empirical evidence on the moderating effects of review characteristics and product type. *Computers in Human Behavior*, 88, 134–142.
- Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., & Freling, T. (2014). How Online Product Reviews Affect Retail Sales: A Meta-analysis. *Journal of Retailing*, 90(2), 217–232.
- Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., & Treinen, M. (2001). What makes web sites credible? A report on a large quantitative study. *Conference on Human Factors in Computing Systems - Proceedings*, 61–68.
- Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 19(3), 291–313.
- Gao, B., Hu, N., & Bose, I. (2017). Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews. *Decision Support Systems*, 95, 1–11.
- Gensler, S., Völckner, F., Egger, M., Fischbach, K., & Schoder, D. (2015). Listen to your customers: Insights into brand image using online consumer-generated product reviews. *International Journal of Electronic Commerce*, 20(1), 112–141.
- Ghose, A., & Ipeirotis, P. G. (2010). Estimating the Helpfulness and Economic Impact of Product Reviews. *Ieee Transactions on Knowledge and Data Engineering*, 1–15.
- Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science*, 31(3), 493–520.

- Girard, T., & Dion, P. (2010). Validating the search, experience, and credence product classification framework. *Journal of Business Research*, 63(9–10), 1079–1087.
- Gobet, F., Lane, P., Croker, S., Cheng, P., Jones, G., Oliver, I., & Pine, J. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5(6), 236–243.
- Godes, D., & Silva, J. C. (2012). Sequential and Temporal Dynamics of Online Opinion. *Marketing Science*, 31(3), 448–473.
- Godes, David, & Mayzlin, D. (2004). Using Online Conversations to Study Word-of-Mouth Communication. *Marketing Science*, 23(4), 545–560.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(SUPPL. 1), 5228–5235.
- Gu, B., Gu, B., & Park, J. (2012). The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products. *Information Systems Research*, 23(1), 182–196.
- Guo, J., Wang, X., & Wu, Y. (2020). Positive emotion bias: Role of emotional content from online customer reviews in purchase decisions. *Journal of Retailing and Consumer Services*, 52(July 2019).
- Guo, Y., Barnes, S. J., & Jia, Q. (2017a). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.
- Guo, Y., Barnes, S. J., & Jia, Q. (2017b). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.
- Hamby, A., Daniloski, K., & Brinberg, D. (2015). How consumer reviews persuade through narratives. *Journal of Business Research*, 68(6), 1242–1250.
- Haridasan, A. C., & Fernando, A. G. (2018). Online or in-store: unravelling consumer's channel choice motives. *Journal of Research in Interactive Marketing*, 12(2), 215–230.
- Harrison-Walker, L. J. (2001). The Measurement of Word-of-Mouth Communication and an Investigation of Service Quality and Customer Commitment As Potential Antecedents. *Journal of Service Research*, 4(1), 60–75.
- Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, 36(1), 20–38.
- Häubl, G., & Trifts, V. (2000). Consumer Decision Making in Online Shopping Environments : The Effects of Interactive Decision Aids Consumer Decision Making in Online Shopping Environments : The Effects of Interactive Decision Aids. *Marketing Science*, 19(1), 4–21.
- Haugtvedt, C. P., & Wegener, D. T. (1994). Message Order Effects in Persuasion: An Attitude Strength Perspective. *Journal of Consumer Research*, 21(1), 205.
- Hauser, J. R., & Koppelman, F. S. (1979). Alternative Perceptual Mapping Techniques : Relative Accuracy and Usefulness. *Journal of Marketing Research*, 16(4), 495–506.
- Henderson, G. R., Iacobucci, D., & Calder, B. J. (1998). Brand diagnostics: Mapping branding effects using consumer associative networks. *European Journal of Operational Research*, 111(2), 306–327.

- Heng, Y., Gao, Z., Jiang, Y., & Chen, X. (2018). Exploring hidden factors behind online food shopping from Amazon reviews: A topic mining approach. *Journal of Retailing and Consumer Services*, 42(January), 161–168.
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, 18(1), 38–52.
- Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective. *Journal of Consumer Research*, 17(4), 454.
- Herrmann, A., Goldstein, D. G., Stadler, R., Landwehr, J. R., Heitmann, M., Hofstetter, R., & Huber, F. (2011). The effect of default options on choice-Evidence from online product configurators. *Journal of Retailing and Consumer Services*, 18(6), 483–491.
- Ho-Dac, N. N., Carson, S. J., & Moore, W. L. (2013). The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter? *Journal of Marketing*, 77(6), 37–53.
- Hofmann, J., Clement, M., Völckner, F., & Hennig-Thurau, T. (2017). Empirical generalizations on the impact of stars on the economic success of movies. *International Journal of Research in Marketing*, 34(2), 442–461.
- Holbrook, M. B. (1978). Beyond Attitude Structure: Toward the Informational Determinants of Attitude. *Journal of Marketing Research*, 15(4), 545.
- Holtgraves, T., & Lasky, B. (1999). Linguistic power and persuasion. *Journal of Language and Social Psychology*, 18(2), 196–205.
- Hornik, K., & Grün, B. (2011). topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software*, 40(13), 1–30.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD'04*, 168.
- Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? the mediating role of online review sentiments on product sales. *Decision Support Systems*, 57(1), 42–53.
- Hu, Y., & Li, X. (2011). Context-Dependent Product Evaluations: An Empirical Analysis of Internet Book Reviews. *Journal of Interactive Marketing*, 25(3), 123–133.
- Huang, A. H., Chen, K., Yen, D. C., & Tran, T. P. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behavior*, 48, 17–27.
- Hutto, C. J., & Gilbert, E. E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14).". *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*.
- International, E. (2017). *Colour cosmetics in the US* (Issue May).
- Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011a). Language Style Matching Predicts Relationship Initiation and Stability. *Psychological Science*, 22(1), 39–44.
- Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011b). Language Style Matching Predicts Relationship Initiation and

- Stability. *Psychological Science*, 22(1), 39–44.
- Jalilvand, M. R., & Samiei, N. (2012). The effect of electronic word of mouth on brand image and purchase intention: An empirical study in the automobile industry in Iran. *Marketing Intelligence and Planning*, 30(4), 460–476.
- Jiang, S., Cai, S., Olle, G. O., & Qin, Z. (2015). Durable product review mining for customer segmentation. *Kybernetes*, 44(1), 124–138.
- Jiménez, F. R., & Mendoza, N. A. (2013). Too popular to ignore: The influence of online reviews on purchase intentions of search and experience products. *Journal of Interactive Marketing*, 27(3), 226–235.
- John, D. R., Loken, B., Kim, K., & Monga, A. B. (2006). Brand concept maps: A methodology for identifying brand association networks. *Journal of Marketing Research*, 43(4), 549–563.
- Johnson, E. J., Bellman, S., & Lohse, G. L. (2002). Defaults , Framing and Privacy :Why Opting In-Opting Out. *Marketing Letters*, 13(1), 5–15.
- Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, 80(4), 557–571.
- Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*, 33(2), 125–143.
- Kamble, Z., Doan, Q. T., Nguyen, H. N., Zeng, Z., & Zihui, L. L. (2020). The Significance of Online Reviews for Hotel Entrepreneurial Success. In *Entrepreneurial Development and Innovation in Family Businesses and SMEs* (pp. 75–86).
- Keller, K. L. (1993). *Conceptualizing, Measuring, and Managing Customer-Based Brand Equity*. 57(1), 1–22.
- Kelly, G. A. (1991). *The Psychology of Personal Constructs*.
- Khwaja, M. G., & Zaman, U. (2020). Configuring the evolving role of ewom on the consumers information adoption. *Journal of Open Innovation: Technology, Market, and Complexity*, 6(4), 1–13.
- Kim, H. B., Kim, W. G., & An, J. A. (2003). The effect of consumer-based brand equity on firms' financial performance. *Journal of Consumer Marketing*, 20(4–5), 335–351.
- Kim, M., & Lennon, S. (2007). The Effects of Visual and Verbal Information on Attitudes and Purchase Intentions in Internet Shopping. *Psychology & Marketing*, 24(9), 763–785.
- Kim, S. G., & Kang, J. (2018). Analyzing the discriminative attributes of products using text mining focused on cosmetic reviews. *Information Processing and Management*, 54(6), 938–957.
- King, M. F., & Balasubramanian, S. k. (1994). The effects of Expertise, End Goal, and Product Type on Adoption of Preference Formation Strategy. *Journal of the Academy of Marketing Science*, 22.
- Kostyk, A., Niculescu, M., & Leonhardt, J. M. (2017). Less is more: Online consumer ratings' format affects purchase intentions and processing. *Journal of Consumer Behaviour*, 16(5), 434–441.

- Kotler, P., & Armstrong, G. (2020). *Principles of Marketing* (18th Global). Pearson Education Limited.
- Kozinets, R. V., De Valck, K., Wojnicki, A. C., & Wilner, S. J. S. (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of Marketing*, 74(2), 71–89.
- Kübler, R. V., Colicev, A., & Pauwels, K. H. (2019). Social Media's Impact on the Consumer Mindset: When to Use Which Sentiment Extraction Tool? *Journal of Interactive Marketing*, 50, 136–155.
- Kudeshia, C., & Kumar, A. (2017). Social eWOM: does it affect the brand attitude and purchase intention of brands? *Management Research Review*, 40(3), 310–330.
- Lee, P. J., Hu, Y. H., & Lu, K. T. (2018). Assessing the helpfulness of online hotel reviews: A classification-based approach. *Telematics and Informatics*, 35(2), 436–445.
- Lee, S., & Choeh, J. Y. (2018). The interactive impact of online word-of-mouth and review helpfulness on box office revenue. *Management Decision*, 56(4), 849–866.
- Lee, S., & Choeh, J. Y. (2020). Using the social influence of electronic word-of-mouth for predicting product sales: The moderating effect of review or reviewer helpfulness and product type. *Sustainability (Switzerland)*, 12(19).
- Lee, T. Y., & Bradlow, E. T. (2011). Automated Marketing Research Using Online Customer Reviews. *Journal of Marketing Research*, 48(5), 881–894.
- Levinson, P. (2013). *New New Media*. Pearson.
- Lewsey, J. D., & Thomson, W. M. (2004). The utility of the zero-inflated Poisson and zero-inflated negative binomial models: A case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. *Community Dentistry and Oral Epidemiology*, 32(3), 183–189.
- Li, M., Huang, L., Tan, C.-H., & Wei, K.-K. (2013). Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features. *International Journal of Electronic Commerce*, 17(4), 101–136.
- Li, S. T., Pham, T. T., & Chuang, H. C. (2019). Do reviewers' words affect predicting their helpfulness ratings? Locating helpful reviewers by linguistics styles. *Information and Management*, 56(1), 28–38.
- Li, Xiaolin, Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information and Management*, 56(2), 172–184.
- Li, Xinxin, & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, 19(4), 456–474.
- Liang, T. P., Li, X., Yang, C. T., & Wang, M. (2015). What in Consumer Reviews Affects the Sales of Mobile Apps: A Multifacet Sentiment Analysis Approach. *International Journal of Electronic Commerce*, 20(2), 236–260.
- Liu, J., Cao, Y., Lin, C., Huang, Y., ... M. Z. the 2007 J. C. on, & 2007, U. (2007). Low-quality product review detection in opinion summarization. *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, June, 334–342.
- Liu, Q., & Karahanna, E. (2015). An agent-based modeling analysis of helpful vote on online product reviews. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2015-March, 1585–1595.
- Liu, X., Burns, A. C., & Hou, Y. (2017). An Investigation of Brand-Related User-

- Generated Content on Twitter. *Journal of Advertising*, 46(2), 236–247.
- Liu, Y. (2006). Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*, 70(3), 74–89.
- Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, 47, 140–151.
- Low, G. S., & Lamb, C. W. (2000). The measurement and dimensionality of brand associations. *Journal of Product & Brand Management*, 9(6), 350–370.
- Lozano, M. B., Martínez, B., & Pindado, J. (2016). Corporate governance, ownership and firm value: Drivers of ownership as a good corporate governance mechanism. *International Business Review*, 25(6), 1333–1343.
- Luca, M. (2011). Revenue : The Case of Reviews , Reputation , and Revenue : The Case of Yelp . com. *Harvard Business Review*, 1–40.
- Ludwig, S., de Ruyter, K., Friedman, M., Brügger, E. C., Wetzels, M., & Pfann, G. (2013a). More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*, 77(1), 87–103.
- Ludwig, S., de Ruyter, K., Friedman, M., Brügger, E. C., Wetzels, M., & Pfann, G. (2013b). More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*, 77(1), 87–103.
- Lynch, Jr., J. G., Marmorstein, H., & Weigold, M. F. (1988). Choices from Sets Including Remembered Brands: Use of Recalled Attributes and Prior Overall Evaluations. *Journal of Consumer Research*, 15(2), 169.
- Magerman, T., Van Looy, B., Baesens, B., & Debackere, K. (2011). Assessment of Latent Semantic Analysis (LSA) text mining algorithms for large scale mapping of patent and scientific publication documents. *FBE Research Report MSI_1114*. https://lirias.kuleuven.be/bitstream/123456789/321960/1/MSI_1114.pdf
- Magoulas, R., & Swoyer, S. (2020). *AI Adoption in the Enterprise 2020*.
- Mahr, D., Stead, S., & Odekerken-Schröder, G. (2019). Making sense of customer service experiences: a text mining review. *Journal of Services Marketing*, 33(1), 88–103.
- Malhotra, N. K. (1981). A Scale to Measure Self-Concepts, Person Concepts, and Product Concepts. *Journal of Marketing Research*, 18(4), 456.
- Mantonakis, A., Roderio, P., Lesschaeve, I., & Hastie, R. (2009). Order in choice: Effects of serial position on preferences. *Psychological Science*, 20(11), 1309–1312.
- Marchand, A., Hennig-Thurau, T., & Wiertz, C. (2017). Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. *International Journal of Research in Marketing*, 34(2), 336–354.
- Martínez Salinas, E., & Pina Pérez, J. M. (2009). Modeling the brand extensions' influence on brand image. *Journal of Business Research*, 62(1), 50–60.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.
- Moe, W. W., & Trusov, M. (2011). The Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing Research*, 48(3), 444–456.
- Mohammad, S., & Turney, P. (2010). Emotions Evoked by Common Words and Phrases. *Proceedings of the NAACL-HLT, June*, 26–34.

- Moon, S., & Kamakura, W. A. (2017). A picture is worth a thousand words: Translating product reviews into a product positioning map. *International Journal of Research in Marketing*, 34(1), 265–285.
- Mudambi, S., & Schuff, D. (2010). What makes a Helpful online Review? *MIS Quarter*, 28(4), 695–704.
- Munch, J. M., & Swasy, J. L. (1988). Rhetorical question, summarization frequency, and argument strength effects on recall. *Journal of Consumer Research*, 15(1), 69–76.
- Murphy, J., Hofacker, C., & Mizerski, R. (2006). Primacy and recency effects on clicking behavior. *Journal of Computer-Mediated Communication*, 11(2), 173–192.
- Nabi, R. L. (2003). Exploring the framing effects of emotion: Do discrete emotions differentially influence information accessibility, information seeking, and policy preference? *Communication Research*, 30(2), 224–247.
- Nazlan, N. H., Tanford, S., & Montgomery, R. (2018). The effect of availability heuristics in online consumer reviews. *Journal of Consumer Behaviour*, 17(5), 449–460.
- Nelson, P. (1970). *Information and Consumer Behavior*. 78(2), 311–329.
- Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine Your Own Business: Market-Structure Surveillance Through Text Mining. *Marketing Science*, 31(3), 521–543.
- Netzer, Oded, Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521–543.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Personality and Social Psychology Bulletin Lying Words : Predicting Deception From Linguistic Styles. *Personality and Social Psychology Bulletin*, 29(5), 665–675.
- Ngo-Ye, T. L., & Sinha, A. P. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, 61(1), 47–58.
- Nielsen. (2016). *The sweet smell of seasonal success*. Available at: <https://www.nielsen.com/au/en/insights/news/2016/the-sweet-smell-of-seasonal-success.html>
- Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *CEUR Workshop Proceedings*, 718, 93–98.
- Olson, J. C., & Muderrisoglu, A. (1979). The stability of responses obtained by free elicitation: Implications for measuring attribute salience and memory structure. *Advances in Consumer Research*, 6, 269–275.
- Pan, B., Zhang, L., & Law, R. (2013). The Complex Matter of Online Hotel Choice. *Cornell Hospitality Quarterly*, 41(1), 74–83.
- Pan, Y., & Zhang, J. Q. (2011). Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews. *Journal of Retailing*, 87(4), 598–612.
- Pang, J., & Qiu, L. (2016). Effect of online review chunking on product attitude: The moderating role of motivation to think. *International Journal of Electronic Commerce*, 20(3), 355–383.
- Park, D. H., & Lee, J. (2008). eWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications*, 7(4), 386–398.

- Park, E. J., Kim, E. Y., Funches, V. M., & Foxx, W. (2012). Apparel product attributes, web browsing, and e-impulse buying on shopping websites. *Journal of Business Research*, 65(11), 1583–1589.
- Park, H.-J., & Rabolt J., N. (2009). Cultural Value, Consumption Value, and Global Brand Image: A Cross-National Study. *Psychology & Marketing*, 26(8), 714–735.
- Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research*, 50, 67–83.
- Payan, J. M., & McFarland, R. G. (2005). Decomposing Influence Strategies: Argument Structure and Dependence as Determinants of the Effectiveness of Influence Strategies in Gaining Channel Member Compliance. *Journal of Marketing*, 69(3), 66–79.
- Payne, J. W. (1982). Contingent Decision Behavior. *Psychological Bulletin*, 92(2), 382.
- Peladeau, N. (2016). *WordStat: Content Analysis Module for SIM-STAT*.
- Pennebaker, J. W., Boyd, R., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. *Austin, TX: University of Texas at Austin, SEPTEMBER 2015*, 1–22.
- Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS ONE*, 9(12), 1–10.
- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The Development and Psychometric Properties of LIWC2007. *Austin, Tx: LIWC.Net*, 1(2).
- Petty, R. E., Tormala, Z. L., Hawkins, C., & Wegener, D. T. (2001). Motivation to think and order effects in persuasion: The moderating role of chunking. *Personality and Social Psychology Bulletin*, 27(3), 332–344.
- Pindado, J., Requejo, I., & de la Torre, C. (2011). Family control and investment-cash flow sensitivity: Empirical evidence from the Euro zone. *Journal of Corporate Finance*, 17(5), 1389–1409.
- Price, P. C., & Stone, E. R. (2004). Intuitive Evaluation of Likelihood Judgment Producers: Evidence for a Confidence Heuristic. *Journal of Behavioral Decision Making*, 17(1), 39–57.
- Puranam, D., Narayan, V., & Kadiyali, V. (2017). The effect of calorie posting regulation on consumer opinion: A flexible latent dirichlet allocation model with informative priors. *Marketing Science*, 36(5), 726–746.
- Purnawirawan, N., De Pelsmacker, P., & Dens, N. (2012). Balance and Sequence in Online Reviews: How Perceived Usefulness Affects Attitudes and Intentions. *Journal of Interactive Marketing*, 26(4), 244–255.
- Purnawirawan, N., Dens, N., & De Pelsmacker, P. (2012). Balance and Sequence in Online Reviews: The Wrap Effect. *International Journal of Electronic Commerce*, 17(2), 71–98.
- Racherla, P., & Friske, W. (2012). Perceived “usefulness” of online consumer reviews: An exploratory investigation across three services categories. *Electronic Commerce Research and Applications*, 11(6), 548–559.
- Reynolds, T. J., & Gutman, J. (1979). Laddering theory, method, analysis, and interpretation. *Journal of Advertising Research*, 28(1), 11–31.

- Rodríguez-Díaz, M., Rodríguez-Díaz, R., & Espino-Rodríguez, T. (2018). Analysis of the Online Reputation Based on Customer Ratings of Lodgings in Tourism Destinations. *Administrative Sciences*, 8(3), 51.
- Roedder, D. J., Loken, B., Kim, K., & Monga, A. B. (2006). Brand concept maps: A methodology for identifying brand association networks. *Journal of Marketing Research*, 43(4), 549–563.
- Roodman, D. (2009a). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, 9(1), 86–136.
- Roodman, D. (2009b). Practitioners' corner: A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71(1), 135–158.
- Salehan, M., & Kim, D. J. (2015). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, 81, 30–40.
- Saumya, S., Singh, J. P., Baabdullah, A. M., Rana, N. P., & Dwivedi, Y. K. (2018). Ranking online consumer reviews. *Electronic Commerce Research and Applications*, 29(March), 78–89.
- Schnittka, O., Sattler, H., & Zenker, S. (2012). Advanced brand concept maps: A new approach for evaluating the favorability of brand association networks. *International Journal of Research in Marketing*, 29(3), 265–274.
- Selkirk, E. (1996). The prosodic structure of function words. In J. L. (Brown U. Morgan & K. (Brown U. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 187–214). Lawrence Erlbaum Associates, Inc.
- Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80(2), 159–169.
- Sheldon, P., & Bryant, K. (2016). Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in Human Behavior*, 58, 89–97.
- Shugan, S. M. (1980). The Cost of Thinking. *Journal of Consumer Research*, 7(2), 99–111.
- Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Kumar Roy, P. (2017). Predicting the “helpfulness” of online consumer reviews. *Journal of Business Research*, 70, 346–355.
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, 105(1), 131–142.
- Slovic, P. (1972). From Shakespeare to Simon: Speculations and some evidence. *Oregon Research Institute Bulletin*, 1–19.
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality & Social Psychology*, 48(4), 813–838.
- Snizek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, 84(2), 288–307.
- Social Blade*. (2017). <https://socialblade.com/>
- Spears, N., & Singh, S. N. (2004). Measuring attitude toward the brand and purchase intentions. *Journal of Current Issues and Research in Advertising*, 26(2), 53–66.
- Spenner, P., & Freeman, K. (2012). To keep your customers, keep it simple. *Harvard*

- Business Review*, 90(5).
- SproutSocial. (2018). *Instagram stats*. Available at: <https://sproutsocial.com/insights/instagram-stats/>
- Statista. (2020). Available at: <https://www.statista.com/forecasts/997051/sources-of-information-about-products-in-the-us>
- Sun, M. (2012). How Does the Variance of Product Ratings Matter? *Management Science*, 58(4), 696–707.
- Sun, Q., Niu, J., Yao, Z., & Yan, H. (2019). Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level. *Engineering Applications of Artificial Intelligence*, 81(December 2018), 68–78.
- Tang, T. (Ya), Fang, E. (Er), & Wang, F. (2014). Is neutral really neutral? The effects of neutral user-generated content on product sales. *Journal of Marketing*, 78(4), 41–58.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
- Tirunillai, S., & Tellis, G. J. (2014). Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation. *Journal of Marketing Research*, 51(4), 463–479.
- Toubia, O., Iyengar, G., Bunnell, R., & Lemaire, A. (2019). Extracting Features of Entertainment Products: A Guided Latent Dirichlet Allocation Approach Informed by the Psychology of Media Consumption. *Journal of Marketing Research*, 56(1), 18–36.
- Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing*, 73(5), 90–102.
- Tybout, A. M., Sternthal, B., Malaviya, P., Bakamitsos, G. A., & Park, S. E. B. (2005). Information accessibility as a moderator of judgments: The role of content versus retrieval ease. *Journal of Consumer Research*, 32(1), 76–85.
- Van Hove, G., & Lievens, F. (2007). Social influences on organizational attractiveness: Investigating if and when word of mouth matters. *Journal of Applied Social Psychology*, 37(9), 2024–2047.
- Veloutsou, C., & Delgado-Ballester, E. (2018). New challenges in brand management. *Spanish Journal of Marketing - ESIC*, 22(3), 255–272.
- Vermeer, S. A. M., Araujo, T., Bernritter, S. F., & van Noort, G. (2019). Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media. *International Journal of Research in Marketing*, 36(3), 492–508.
- Vriens, M., Chen, S., & Schomaker, J. (2019). The evaluation of a brand association density metric. *Journal of Product and Brand Management*, 28(1), 104–116.
- Walker, M. A., Mehl, M. R., Moore, R. K., & Mairesse, F. (2007). Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research*, 30, 457–500.
- Wang, F., Liu, X., & Fang, E. (Er). (2015). User Reviews Variance, Critic Reviews Variance, and Product Sales: An Exploration of Customer Breadth and Depth

- Effects. *Journal of Retailing*, 91(3), 372–389.
- Wang, W., Feng, Y., & Dai, W. (2018). Topic analysis of online reviews for two competitive products using latent Dirichlet allocation. *Electronic Commerce Research and Applications*, 29(January), 142–156.
- Wang, X., Tang, L. (Rebecca), & Kim, E. (2019). More than words: Do emotional content and linguistic style matching matter on restaurant review helpfulness? *International Journal of Hospitality Management*, 77(August 2018), 438–447.
- Weathers, D., Swain, S. D., & Grover, V. (2015). Can online product reviews be more helpful? Examining characteristics of information content by product type. *Decision Support Systems*.
- Westerman, D., Spence, P. R., & Van Der Heide, B. (2014). Social Media as Information Source: Recency of Updates and Credibility of Information. *Journal of Computer-Mediated Communication*, 19(2), 171–183.
- Wickham, H. (2016). *tidyr: Easily tidy data with spread and gather functions*.
- Wickham, H. (2017). ggplot2 - Elegant Graphics for Data Analysis (2nd Edition). *Journal of Statistical Software*, 77(Book Review 2), 2–5.
- Wickham, H., & Francois, R. (2016). *dplyr: A Grammar of Data Manipulation*.
- Winkelmann, P. D. R. (2008). Econometric Analysis of Count Data. In *Econometric Analysis of Count Data*.
- Wong, C. U. I., & Qi, S. (2017). Tracking the evolution of a destination's image by text-mining online reviews - the case of Macau. *Tourism Management Perspectives*, 23, 19–29.
- Wooldridge, J. M. (2011). *Econometric analysis of cross section and panel data* (Issue November).
- Wu, J. (2017). Review popularity and review helpfulness: A model for user review effectiveness. *Decision Support Systems*, 97, 92–103.
- Xu, P., & Liu, D. (2019). Product engagement and identity signaling: The role of likes in social commerce for fashion products. *Information and Management*, 56(2), 143–154.
- Yang, H., & Cho, S. (2015). *Understanding Brands with Visualization and Keywords from eWOM using Distributed Representation*.
- Yau, K. K. W., Wang, K., & Lee, A. H. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal*, 45(4), 437–452.
- Yazdani, E., Gopinath, S., & Carson, S. (2018). Preaching to the Choir: The Chasm Between Top-Ranked Reviewers, Mainstream Customers, and Product Sales. *Marketing Science*, 37(5), 838–851.
- Yin, G., Wei, L., Xu, W., & Chen, M. (2014). Yin, G., Wei, L., Xu, W., & Chen, M. (2014). Exploring Heuristic Cues for Consumer Perceptions of Online Reviews Helpfulness: The Case of Yelp.com. PACIS 2014 Proceedings, 1–12. Exploring Heuristic Cues for Consumer Perceptions of Online Reviews Helpfulness. *PACIS 2014 Proceedings*, 1–12.
- Zaltman, G. (1997). Rethinking Market Research : Putting People Back In. *Journal of Marketing Research*, 34(4), 424–437.

- Zhang, J. (2019). What's yours is mine: exploring customer voice on Airbnb using text-mining approaches. *Journal of Consumer Marketing*, 36(5), 655–665.
- Zhang, Ying, & Lin, Z. (2018). Predicting the helpfulness of online product reviews: A multilingual approach. *Electronic Commerce Research and Applications*, 27, 1–10.
- Zhang, Yuchi, Moe, W. W., & Schweidel, D. A. (2017). Modeling the role of message content and influencers in social media rebroadcasting. *International Journal of Research in Marketing*, 34(1), 100–119.
- Zhou, S., & Guo, B. (2017). The order effect on online review helpfulness: A social influence perspective. *Decision Support Systems*, 93, 77–87.
- Zhu, F., & Zhang, X. (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*, 74(March), 133–148.

General Conclusions

Next sections present the main findings of the thesis and highlight the main theoretical and empirical contributions. Besides, some managerial implications are provided for practitioners. Finally, the limitations of the thesis are discussed and some areas for further research are proposed.

Summary of Findings

Chapter 1 analyses the moderating role of review visibility in explaining the relationship between review characteristics and review helpfulness. This chapter formulates the following two hypotheses:

- H1a. *More visible online reviews when ranked by the “most helpful” mechanism are more likely to be voted as helpful*
- H1a. *More visible online reviews when ranked by the “most recent” mechanism are more likely to be voted as helpful*
- H2a. *Review visibility when online reviews are ranked by the “most helpful” mechanism moderates the impact of review characteristics on review helpfulness*
- H2b. *Review visibility when online reviews are ranked by the “most recent” mechanism moderates the impact of review characteristics on review helpfulness*

This chapter is grounded on the accessibility-diagnosticsity theory (Feldman & Lynch, 1988). The accessibility dimension of the theory is approached using the concept of review visibility, which measures the rank order of online reviews when sorting by the “most helpful” or “most recent” mechanism. The higher the position in the rank, the more visible is the review for consumers. The diagnosticsity dimension of the theory is described as the ability of the information to provide consumers with relevant product information that helps them to understand and evaluate the quality and performance of the product (Filieri, Hofacker, et al., 2018). As claimed by Payne (1982), an input’s diagnosticsity depends on whether it enables a decision maker to discriminate among alternatives and it depends on the characteristics of the input of information. In this research, we incorporate two sets of variables to approach the diagnosticsity dimension of the theory: review non-textual characteristics (e.g., review rating and reviewer physical information disclosure) and review textual characteristics (e.g., subjectivity and confidence). These characteristics are considered as the independent variables in the

empirical models, while review visibility is used as a moderator, which affects the true relationship between review characteristics and review helpfulness and, at the same time, it is independently associated to review helpfulness. Review helpfulness, which is measured by the incremental number of helpful votes within a six-month period is the dependent variable used in the models. A Zero-Inflated Negative Binomial regression is applied since it allows us to understand the possible differences in explaining the excess of online reviews with zero helpful votes and those with a positive count number of helpful votes. Three models are built and compared: the first one does not consider review visibility, so every review is assumed to have the same probability of being viewed; the second one assumes that consumers sort online reviews by the most helpful mechanisms; and the third one assumes that consumers sort online reviews by the most recent mechanisms, which is the default used on the website of analysis to show online reviews. The results show several findings. First, when explaining the helpfulness of online reviews, the process of voting a review as helpful might not be a one-step process but rather different sequential sub-processes. In this line, we observe that the moderating role of review visibility is stronger in explaining the excess of online reviews with zero helpful votes than in explaining the positive number of helpful votes of online reviews. This might mean that, as expected, review visibility is likely to influence the consumer decision of reading a specific online review. However, once a review is read, the role of visibility in explaining the number of helpful votes is weaker. We observe that the relationship between review visibility and the number of helpful votes is additive and not multiplicative. On the one hand, review visibility variables impact on the number of helpful votes that online reviews receive. And, on the other hand, review characteristics by themselves also impact those helpful votes. However, the interaction between visibility and review characteristics is very weak in explaining the number of helpful votes. Moreover, it is observed that results vary depending on what type of information consumers are evaluating: whether they are reading most helpful or most recent online reviews. A possible explanation is that, although the two sorting mechanisms enhance the review visibility likelihood, they capture different aspects of online reviews: the number of helpful votes and the publication recency. When consumers read most helpful online reviews, they already know that other consumers have considered the information provided in those reviews as helpful, which might lead to a social influence effect. This effect might lead consumers to trust more on those reviews. However, when consumers read most recent online reviews, they only know that those reviews are the ones most recently published, but they do not know how helpful those reviews are compared to other reviews of the same product. Thus, the two sets of reviews contain different type of

information and consumers might have different expectations and thus, different behaviours.

Chapter 2 is also grounded in the accessibility-diagnostics theory (Feldman & Lynch, 1988) to study the influence of online reviews on product sales. The accessibility dimension of the theory is approached as in Chapter 1, incorporating the role of review visibility. In this chapter, we analyse how online reviews influence product sales in different cases of review visibility: when we assume that every online review has the same probability of being viewed (the traditional assumption in the literature), when we assume that consumers sort online reviews by the most helpful mechanism, so that most helpful reviews are more likely to be viewed, and when we assume that consumers sort online reviews by the most recent mechanism (default at the website), so that most recent reviews are more likely to be viewed. The diagnostics dimension of the theory is also approached using two sets of review variables: review non-textual variables and review textual variables. Despite chapter 1 where we were doing the study at a review-level to study review helpfulness, every review variable is aggregated to a product-level in Chapter 2 to study the antecedents of product sales. This chapter posits the following hypotheses:

- *H1a. Review non-textual features influence product sales considering different cases of review visibility*
- *H1b. Review textual features influence product sales considering different cases of review visibility*
- *H2a. The influence of review non-textual features on product sales is different depending on the review visibility case considered*
- *H2b. The influence of review textual features on product sales is different depending on the review visibility case considered*

The empirical analysis is conducted using panel data modelling, in particular the system GMM model (Arellano & Bond, 1991). To study the relationship between online reviews and product sales in different cases of review visibility, two sets of independent variables are considered: review non-textual features (e.g., rating and volume) and review textual features (e.g., authentic and analytic). Findings suggest that, both review non-textual and textual features influence product sales not only in the traditional case of review visibility, where every review is assumed to have the same probability of being viewed, but also in the rest of cases, where we assume that consumers sort online reviews either by the most helpful mechanism or by the most recent mechanism. The coefficient of review non-textual features has the same sign in every case of review visibility, but it

changes its magnitude. As far as non-textual features are concerned, findings reveal that their impact on consumer shopping behaviour changes over the different review visibility cases, not only in magnitude but also in sign. Therefore, this suggests that consumers evaluate the textual aspects of online reviews in a different way if they are evaluating most helpful or most recent reviews. Overall, the features of most helpful online reviews have a higher impact on product sales. A possible explanation is that consumers might **experience a “wisdom of the crowd” effect when they evaluate the most helpful online reviews** (Liu and Karahanna, 2015). This effect refers to the fact that consumers might believe that, since many other consumers have voted the information contained in those reviews as helpful, that information about the product might be a better approximation to the truth, so consumers are more likely to rely on it when making a purchase.

Chapter 3 focuses on analysing the textual content of online reviews with the following main research objective:

- *To present a unified and structured procedure to explore brand image and brand positioning using the information contained in the text of online reviews*

To that end, this chapter reviews both the literature on brand image and text mining to find possible gaps and areas for improvement in the study of brand associations from the textual content of online reviews. After that, a unified and structured procedure to explore brand positioning and brand segmentation from the textual content of online reviews is proposed. By **“structured procedure”** we mean a procedure that defines clear stages to conduct the brand positioning analysis using online reviews from the beginning to the end, focusing on the text mining process to extract brand associations. Moreover, **“unified procedure”** means that we have reviewed the literature in text mining, brand positioning and brand segmentation to put ideas together into one unique study, which finally presents an easy-to-follow guide or procedure combining different techniques for text mining, brand positioning and brand segmentation. We **say that it is an “easy-to-follow procedure” because** the text mining analysis is based on a lexicon-based method, the Linguistic Inquiry and Word Count (LIWC) developed by Pennebaker et al. (2007). This method is accessible in terms of price and intuitive to use, since the researcher does not need knowledge on machine learning. Using LIWC, brand associations were extracted from the text of online reviews. Then, these associations were used as input variables in a Principal Component Analysis (PCA), which allowed us to build a brand positioning map. Finally, Hierarchical Clustering was used to conduct a segmentation analysis to group brands with similar associations. Findings regarding brand associations are specific to our product category (blush cosmetics).

Chapter 4 can be seen as an extension of Chapter 3. In this case, rather than using a dictionary-based approach for conducting the text mining analysis, we applied some important unsupervised machine learning algorithms with the goal of exploring product image. Unlike Chapter 3, where we do the analysis at a brand level, we disaggregate it at the product level, to uncover hidden associations regarding each specific product in the blush category. Thus, the main objective of Chapter 4 is the following one:

- To show and to illustrate some methods of text mining with the aim of exploring product image uncovering information contained in online reviews

To that end, this chapter first examines the literature to identify some of the most important machine learning algorithms for text mining. Five tasks were identified as some of the most popular unsupervised machine learning algorithms for text mining: word frequencies, sentiment analysis, relationship between words, product-word correlations and topic modelling. The five identified tasks were then empirically applied to our specific case of online reviews. Amongst those five tasks, all of them except from topic modelling, are relatively straight forward. To perform those tasks, the researcher had to run some specific algorithms for machine learning, but not great effort was needed to interpret the output. However, a greater interpretation effort is needed when dealing with topic modelling. For topic modelling, the Latent Dirichlet Allocation (LDA) model is one of the most popular. By using LDA, we obtained the main topics discussed in our dataset of online reviews. To perform LDA, it is quite important to first establish the main objective of the task. Depending on the goal, we must carefully think about the list of stop words to include in the pre-processing stage of the LDA since this stop words list will influence the final LDA output. If, for example, most common words in the category, **such as “product”, “blush” or “cosmetics”, are not included into the stop words list**, the LDA will give us very general topics, which will not provide us interesting insights. The second important thing when performing LDA is to properly interpret the output. The output of the LDA is quite context dependent. Therefore, to interpret the different topics provided by the model we should get familiar with the specific industry where we are conducting the analysis. Since this task involve some subjectivity in understanding the topics, several researchers should be involved on it to reach a consensus in the interpretation. In our cosmetics setting, 22 topics were identified. Since we are evaluating online reviews for a very specific product category, which is blush, topics are quite similar one each other. Nevertheless, topics might be more different if we were comparing online reviews from different product categories. From the LDA analysis we find the following main results. First, many topics focused on features related to product colours (e.g., Pink and peach glow, Soft colour, Cocoa and macaroon colours). Second,

several topics focus on aspects related to product usage (e.g., day usage, shimmer day wear, day lasting, easy to apply powder, pink day shade). Third, other topics focus on product format (e.g., stain and tint, powder cream, duo size, glow, and light palette).

Contribution to Theory

Overall, this thesis contributes to the eWOM literature by extending the knowledge on how online reviews influence consumer behaviour and how companies can use their textual content to analyse their brand positioning. In this line, we give further insights about the determinants that make consumers adopt the information contained in online reviews, and also the characteristics of online reviews that make consumers buy a product. Moreover, this thesis increases the knowledge on how textual features of online reviews influence first, information adoption and purchase behaviour, and second, how these textual features can be used as inputs to understand brand image and brand positioning.

In Chapter 1, we contribute to the review helpfulness literature by framing our research on the accessibility-diagnostics theory (Feldman & Lynch, 1988). By adopting this theoretical framework, we can incorporate the notion of review visibility, which approaches the accessibility dimension of the theory, to study the relationship between review characteristics, which are the factors that make a review diagnostic, and review helpfulness. When studying the impact of review characteristics on review helpfulness previous literature assumes the characteristic of every review to be equally influential on consumer decision making. However, online reviews are shown in a sequence next to other reviews, and not individually, so the relative visibility of online reviews in the sequence is likely to influence consumer decision-making. In fact, there is evidence to believe that consumers do not evaluate all available information, in this case online reviews, and, instead, they tend to be more influenced by more accessible information (BrightLocal, 2020; Feldman & Lynch, 1988; Q. Liu & Karahanna, 2015). Thus, we contribute to previous literature by incorporating the variable review visibility as a moderator of the relationship between review characteristics and review helpfulness. Moreover, we also contribute to literature by proposing a different perspective to approach the objective. In general, previous literature has overlooked the fact that, in any platform, there are always a lot of online reviews with zero helpful votes. These **“zeros” might be due to a bad review quality**, in the way that consumers do not find the information contained in those reviews as useful but might be also due to the fact that **those online reviews have not been “viewed” by consumers. In this sense, if consumers do not view a specific online review, the only possible outcome is not voting it as helpful.**

By using a Zero Inflated Negative Binomial (ZINB) model, we can analyse not only the determinants that influence the likelihood that online reviews receive helpful votes, but we are also able to analyse the drivers behind the excess of online reviews with zero helpful votes.

In Chapter 2, we contribute to the literature in two main ways. First, in line with Chapter 1, we also frame the research on the accessibility-diagnostics theory (Feldman & Lynch, 1988). By doing so, we can incorporate the notion of review visibility to approach the accessibility dimension of the theory. As previous literature has claimed, the characteristics of online reviews are likely to influence product sales. However, based on the accessibility-diagnostics theory, on the literature about decision making and on some consultancy reports (e.g., BrightLocal, 2020), there is evidence to believe that consumers do not evaluate every available online review for a product. Instead, consumers are likely to evaluate more accessible or visible information, in this case, more visible online reviews. So far, previous literature has assumed every online review to have the same influence on consumer purchasing decisions. In this sense, when studying the effect of online reviews' characteristics (e.g., rating and volume) on product sales, the characteristics of every review are assumed to be equally influential. However, online reviews are shown in a sequence next to other reviews, and not individually, so the relative visibility of online reviews in the sequence is likely to influence which online reviews consumers read. Since consumers are more likely to read most visible online reviews, characteristics of those online reviews are likely to have a greater influence on consumer purchasing behaviour. Thus, in Chapter 2 we contribute to previous literature by incorporating the role of review visibility to study how online reviews influence product sales under different cases of review visibility (when we assume every online review to have the same probability of being viewed, when we assume that most helpful online reviews are more likely to be viewed and when we assume that most recent online reviews are more likely to be viewed). So far, studies have not found consensus when explaining how the characteristics of online reviews, mainly rating and volume, impact product sales. Some scholars, such as **Babić Rosario et al.** (2016), have revealed that the different effects of review variables on product sales depend on factors such as the platform where the online text is published (e.g. social network, blog, reviews' sites, etc.), the type of product (e.g. tangible goods, services, digital products, utilitarian vs. hedonic products, high financial risk vs. low financial risk and mature vs. new products, etc.), the metric used to build the variable (e.g. average, cumulative, incremental, etc.) and the design of the study (e.g. how endogeneity is controlled, the type of control variables, etc.) (**Babić Rosario et al., 2016**). In this research, we posit that review visibility can be

also a factor that might impact the relationship between review characteristics and product sales. In this sense, effects might be different if we assume that consumers read every online review available for a product or if we assume that most visible reviews are more likely to be read.

Moreover, online reviews' literature has until now focused on studying how non-textual features of online reviews influence product sales. In our research, we are interested in not only analysing the impact of non-textual review features but also of some textual review characteristics. Thus, this research extends the knowledge of how different features of online reviews influence consumer shopping decisions.

Chapter 3 and Chapter 4 have a more exploratory focus, because we are interested in exploring brand image and brand positioning, and they are more practice oriented. However, we contribute to previous online reviews' literature by extending the study of textual features of online reviews, which is a quite underexplored area. We show how text mining techniques for Natural Language Processing (NLP) can serve us as a powerful tool to better understand brand image and brand positioning. Brand image has traditionally been explored using multidimensional scaling and qualitative techniques, such as the ZMET developed by Zaltman (1997). However, the use of eWOM, being online reviews a particular case, provide two main advantages over survey data when studying brand image. First of all, eWOM is spontaneous (Marchand et al., 2017; Yang & Cho, 2015). In general, it occurs without direct prompting or influence by marketers and it is usually motivated by a desire to help others, warn others or to communicate status (Kozinets et al., 2010). Because of that, consumers are more likely to express their true brand perceptions and to inform about their actual behaviors. Moreover, there are large amounts of online consumer opinions available online, so brand image can be studied using very big samples of consumers' opinions. This last reason is also an advantage of eWOM over qualitative techniques, which are based on small samples of respondents who participate in in-depth interviews.

In Chapter 3, we show and propose a unified and structured procedure to explore brand image from the text of online reviews to analyse brand positioning and segmentation. Since literature focusing on textual dimensions of online reviews is still scarce, any paper addressing this type of study makes an important contribution to literature. So far, those papers studying textual content of online reviews are quite heterogeneous in terms of research objectives and text mining techniques used. Therefore, we try to bridge this gap by proposing a more structured and easy-to-implement procedure that can be followed by either companies or researchers who want

to analyse the textual content of online reviews to increase their knowledge on brand image and brand positioning. This procedure is based on a lexicon-based text mining technique, the LIWC developed by Pennebaker et al. (2015), but it could be reproduced using any other alternative dictionary-based method.

In Chapter 4, we contribute to the literature on brand image by illustrating some Natural Language Processing (NLP) algorithms, instead of lexicon-based methods as used in Chapter 3, which can be used to answer some research questions related to the study of brand (or product) image. By using the unsupervised machine learning algorithms illustrated in this paper, associations with brands and products can be extracted from online reviews.

Implications for Practice

Although this thesis has several contributions to theory, the study of online reviews has clear managerial implications.

First, companies can extend their knowledge of how consumers process information in an online environment. Our findings suggest that not only the specific characteristics of online reviews influence consumers' information adoption decisions and shopping decisions, but also the way in which online reviews are presented at the reviews' site. Thus, a recommendation derived from our results is that marketers should provide more online tools to allow consumers to organize online reviews according to their preferences. For example, considering that consumers might see the information contained in most helpful reviews as more reliable, firms could make the information contained in those reviews more visible to reduce the cognitive effort of consumers. For example, by incorporating a subsection showing a summary of the characteristics of most helpful reviews. In that subsection, consumers could consult, for example, the profile of the reviewers who write most helpful reviews and also some aspects directly related to the textual content of those online reviews (e.g., subjectivity, level of confidence, etc.). If consumers believe that the firm is helping them in their decision-making process, they will be more likely to have a better online experience, which might be translated into a better attitude towards the firm and a higher purchase intention.

In line with previous implication, since information contained in top-ranked online reviews, in this case most recent and most helpful online reviews, is likely to exert higher influence on consumer shopping behaviour, managers should pay special attention to those online reviews appearing in those top positions. Thus, companies could analyse the

information contained in those reviews to better satisfy consumer needs, to improve current products or to launch new ones. Moreover, considering that consumers use sorting tools to reduce the cognitive effort of managing big amounts of information, managers could incorporate new sorting mechanisms to help consumers in their decision-making. Providing more sorting mechanisms options would make consumers to have a better online consumer experience and would lead to higher customer satisfaction. If more sorting tools available, consumers could select online reviews based on the most preferred criterion. For example, considering the impact that most helpful online reviews have on consumers decision making, online retailers could give the option of **multi-level sorting, so that consumers might be able to sort by “most helpful” and by other option** (e.g., by highest rating, by lowest rating and by most recent). In that way, consumers would be able to better select those reviews of their interest. In line with our findings, the online retailer introduced some changes after we collected the data for the research. For example, they do not longer show online reviews by the default mechanism of most recent order, but by the most helpful mechanism. This corroborates our finding revealing that most helpful online reviews are likely to be more influential on consumer shopping behaviour.

Third, moving on to the study of textual content of online reviews, our research has also clear managerial implications. As revealed by Magoulas and Swoyer (2020), one of the main barriers for Artificial Intelligence (AI) adoption in companies is the lack of skilled people or the difficulty to hire the required roles. One of the methods used for text mining is based on machine learning algorithms, which are a specific application of AI, and therefore, many companies find it difficult to adopt. In this research, we propose to follow a text mining procedure based on a lexicon-based text mining method, the LIWC, which is an easier to implement tool than machine learning algorithms. Therefore, our research might be especially helpful for small and medium companies that do not have the resources to conduct complex machine learning algorithms for text mining and want to uncover hidden features of online reviews' texts (e.g., sentiment expressed, main topics discussed, product/brand attributes elicited, etc.). In our research, we focus on several emotional and psychological variables provided by the LIWC, but other variables given by the tool could be also explored using the same procedure. Besides, other lexicon-based tools could be also used. By using this procedure, companies can better understand how consumers position their brands comparing to **competitors'** brands and could get insights for a better differentiation strategy based on these consumer emotional and psychological associations.

Finally, this research also illustrates some of the most popular machine learning algorithms used for text mining. Companies could adopt those algorithms for several applications, such as getting insights about consumer interests and preferences, analyzing which positive and negative words are associated to each product or which product features are associated to positive or negative sentiments, identifying customer complaints to try to improve the product or service, analyzing brand or product positioning and improving differentiation or getting insights for new product development.

Limitations and Future Research

Since we are focusing on studying a complex phenomenon, which is how online reviews influence different types of consumer behavior and how textual content can be uncovered to study brand image, this thesis only contributes to previous literature in a modest way, and it is necessary to outline the main limitations to be able to raise future lines of research.

The first limitation concerns the database used for this research. This thesis is limited to one category of experience products, which is the category of **“blush” cosmetics**, but future research could be expanded to other categories of cosmetics or even other categories of experience products and search products. The research may be also extended to services, such as hotels or restaurants. In that way, we could know if results in our research could be further generalized to other product or service categories or, instead, they are context dependent. Moreover, this research is carried out using online consumer reviews, but research could be carried out using other types of eWOM, such as social networks and blogs.

Second, review visibility variables used in Chapter 1 and Chapter 2 were proxied by two sorting mechanisms: most recent and most helpful. Future research could use other sorting mechanisms, such as highest rating and lowest rating, to approach review visibility. Results could be compared to the ones in our research. Future research could also use click stream data to know the actual sorting mechanism used by the consumers on the online retailer. Besides, three review textual variables extracted using LIWC were used in Chapter 1 and Chapter 2 to analyse review helpfulness and product sales, but further research might consider the effect of other specific content features, either provided by LIWC or by other lexicon-based or machine learning methods.

In terms of text mining, we use in Chapter 3 the LIWC. Although its use has been widespread in previous text mining literature, future research could use other lexicon-based methods available to study brand positioning. Machine learning methods for text mining could be also used by companies if they need to extract more specific aspects from online reviews texts and if they have the required resources. In this research, we explore some brand associations provided by the LIWC. However, other LIWC variables or aspects hidden in texts could be also analyzed, such as features related to the reviewer writing style (e.g., informal writing, cognitive writing and time focus). Moreover, brand associations are evaluated on a specific online retailer, but it would be interesting to compare associations of brands in different online retailers, to see if brand associations are different depending on the retailer or, on the contrary, they are not platform dependent. Moreover, we could carry out a survey to see if online brand associations are the same as overall brand associations, which could be explored by questionnaires.

Finally, by exploring brand image using the textual content of online reviews, we might be capturing the perceptions of brands of those consumers who usually write online opinions. However, there might be some consumers who do not articulate themselves online. It might be interesting to compare the results with the ones obtained from traditional methods, such as multidimensional scaling or qualitative techniques.

Conclusiones generales

Las siguientes secciones presentan los principales hallazgos de la tesis y destacan las principales contribuciones teóricas y empíricas. Además, se proporcionan algunas implicaciones de gestión para los profesionales. Finalmente, se discuten las limitaciones de la tesis y se proponen algunas áreas para futuras investigaciones.

Resumen de resultados

El capítulo 1 analiza el papel moderador de la visibilidad de las reseñas al explicar la relación entre las características de esas reseñas y su utilidad, que se mide a través del número de votos útiles que las reseñas reciben. Este capítulo formula las siguientes dos hipótesis:

- H1a. *Las reseñas más visibles cuando se ordenan por “más útiles” tienen más probabilidad de ser votadas como útiles*
- H1b. *Las reseñas más visibles cuando se ordenan por “más recientes” tienen más probabilidad de ser votadas como útiles*
- H2a. *La visibilidad de las reseñas cuando se ordenan por “más útiles” modera el impacto de las características de las reseñas en su utilidad*
- H2b. *La visibilidad de las reseñas cuando se ordenan por “más recientes” modera el impacto de las características de las reseñas en su utilidad*

Este capítulo se basa en la teoría de la accesibilidad-diagnostics (Feldman y Lynch, 1988). La dimensión de accesibilidad de la teoría se aborda utilizando el concepto de visibilidad de la reseña, que mide el orden de clasificación de las reseñas cuando se ordenan por el mecanismo de "más útil" o "más reciente". Cuanto más alta sea la posición en el ranking, más visible será la reseña para los consumidores. La dimensión de diagnostics de la teoría se describe como la capacidad de la información para proporcionar a los consumidores información relevante sobre el producto que les ayude a comprender y evaluar la calidad y el rendimiento del producto (Filieri et al., 2018a). Como afirma Payne (1982), la diagnostics de la información depende de si permite al tomador de decisiones discriminar entre alternativas y depende de las características de la información. Para probar las hipótesis, se han considerado dos tipos de variables independientes, las características no textuales de las reseñas (por ejemplo, número de estrellas de la reseña y divulgación de información física del revisor) y las características textuales (por ejemplo, subjetividad y confianza demostrada por el revisor). Por otro

lado, la visibilidad de las reseñas se utiliza como variable moderadora. El número incremental de votos útiles dentro de un período de seis meses es la variable dependiente utilizada en los modelos. Se ha aplicado una regresión Binomial Negativa de Ceros Inflados (Zero Inflated Negative Binomial, ZINB) ya que nos permite entender las posibles diferencias al explicar el exceso de reseñas con cero votos útiles y aquellas con un número positivo de votos útiles. Se han construido y comparado tres modelos: el primero no considera la visibilidad de las reseñas, por lo que se asume que todas las reseñas tienen la misma probabilidad de ser vistas; el segundo asume que los consumidores clasifican las reseñas **según el mecanismo de ordenación de “más útiles”**; y el tercero asume que los consumidores clasifican las reseñas según el mecanismo de **ordenación de “más recientes”, que es el que se utiliza por defecto en el sitio web de análisis**. Los resultados muestran varios hallazgos. Primero, cuando estamos explicando la utilidad de las reseñas, el proceso de votar una reseña como útil puede que no sea un proceso de un solo paso, sino un conjunto de subprocesos secuenciales diferentes. En esta línea, observamos que el papel moderador de la visibilidad de las reseñas es más fuerte cuando explicamos el exceso de reseñas con cero votos útiles que para explicar el número positivo de votos útiles de las reseñas. Esto podría significar que, como se esperaba, es probable que la visibilidad de la reseña influya en la decisión del consumidor de leer una reseña específica. Sin embargo, una vez que se lee la reseña, el papel de la visibilidad a la hora de explicar el número de votos útiles que esa reseña recibe es más débil. Observamos que la relación entre la visibilidad de la reseña y el número de votos útiles es más aditiva y no multiplicativa. Por un lado, la visibilidad de las reseñas impacta en el número de votos útiles que reciben, y por otro, las características de la reseña también impactan en esos votos útiles. Sin embargo, el efecto interactivo entre visibilidad y características es muy débil. Además, se observa que los resultados varían según el tipo de información que los consumidores estén evaluando: si están leyendo las reseñas más útiles o las más recientes. Una posible explicación es que, aunque los dos mecanismos de ordenación influyen en que haya reseñas más visibles que otras, cada uno captura diferentes aspectos de las reseñas: o bien el número de votos útiles o la novedad de su publicación. Cuando los consumidores leen las reseñas más útiles, ya saben que otros consumidores han considerado útil la información proporcionada en esas reseñas, lo que podría generar un efecto de influencia social. Este efecto podría llevar a los consumidores a confiar más en esas reseñas. Sin embargo, cuando los consumidores leen las reseñas más recientes, solo saben que esas reseñas son las más nuevas, pero no saben qué tan útiles son esas reseñas en comparación con otras del mismo producto. Por lo tanto, los dos mecanismos de ordenación influyen en la visibilidad, pero cada uno muestra reseñas con diferentes tipos de información. Esto puede hacer que los consumidores tengan

diferentes expectativas frente a ellas y, por tanto, diferentes comportamiento, como puede ser el comportamiento de voto.

El capítulo 2 también se basa en la teoría de la accesibilidad-diagnóstico (Feldman y Lynch, 1988) para estudiar la influencia de las reseñas en las ventas de productos. La dimensión de accesibilidad de la teoría se aborda como en el capítulo 1, incorporando el papel de la visibilidad de la reseña. En este capítulo, analizamos cómo las reseñas influyen en las ventas de productos en diferentes casos de visibilidad: cuando asumimos que cada reseña tiene la misma probabilidad de ser vista (el supuesto tradicional en la literatura), cuando asumimos que los consumidores ordenan las reseñas mediante el mecanismo de “más útil”, de modo que es más probable que se vean las opiniones más útiles, y cuando suponemos que los consumidores ordenan las reseñas por el mecanismo de “más reciente” (predeterminado en el sitio web), de modo que es más probable que se vean las opiniones más recientes. La dimensión de diagnosticidad de la teoría también se aborda utilizando dos conjuntos de variables de la reseña: las variables no textuales y las variables textuales. Aunque en el capítulo 1 estudiamos la utilidad de la reseña haciendo un análisis a nivel de reseña, en el Capítulo 2 se agregan las variables a nivel de producto para estudiar los antecedentes de las ventas de productos. Este capítulo plantea las siguientes hipótesis:

- H1a. *Las características no textuales de las reseñas influyen en la venta de productos cuando consideramos diferentes casos de visibilidad de las reseñas*
- H1b. *Las características textuales de las reseñas influyen en la ventas de productos cuando consideramos diferentes casos de visibilidad de las reseñas*
- H2a. *La influencia de las características no textuales de las reseñas en la venta de productos es diferente según el caso de visibilidad de reseñas considerado.*
- H2b. *La influencia de las características textuales de las reseñas en la venta de productos es diferente según el caso de visibilidad de reseñas considerado.*

El análisis empírico se ha realizado utilizando un modelo de datos de panel, en concreto el modelo system GMM. (Arellano y Bond, 1991). Para estudiar la relación entre las reseñas y la venta de productos en diferentes casos de visibilidad de las reseñas, se han considerado dos conjuntos de variables independientes: características no textuales de las reseñas (por ejemplo, número medio de estrellas de las reseñas del producto y número de reseñas del producto) y características textuales de las reseñas (por ejemplo, honestidad de las reseñas y si la reseña es más o menos analítica). Los hallazgos sugieren que, tanto las características textuales como las no textuales de la reseña influyen en las ventas de productos no solo en el caso tradicional de visibilidad de la reseña, donde se

supone que cada reseña tiene la misma probabilidad de ser vista, sino también en el resto de los casos, donde asumimos que los consumidores ordenan las reseñas por el **mecanismo de “más útiles” o por el mecanismo de “más recientes”**. El coeficiente de las características no textuales de las reseñas tiene el mismo signo en todos los casos de visibilidad (en cada uno de los modelos), pero varía en magnitud. En lo que respecta a las características no textuales, los resultados revelan que su impacto en el comportamiento de compra de los consumidores cambia en los diferentes casos de visibilidad, no solo en magnitud sino también en signo. Por lo tanto, esto sugiere que los consumidores evalúan los aspectos textuales de las reseñas de una manera diferente si están evaluando las reseñas más útiles o las más recientes. En general, las características de las reseñas más útiles tienen un mayor impacto en las ventas de productos. Una posible explicación es que los consumidores pueden experimentar un efecto de influencia social cuando evalúan las reseñas más útiles (Liu y Karahanna, 2015). Este efecto se refiere al hecho de que los consumidores podrían creer que, dado que muchos otros consumidores han votado la información contenida en esas reseñas como útil, esa información sobre el producto podría ser una mejor aproximación a la verdad. Por lo tanto, la información dada por las reseñas más útiles podría ser más creíble e impactante en el comportamiento de compra del consumidor.

El capítulo 3 se centra en analizar el contenido textual de las reseñas con el siguiente objetivo:

- Presentar un procedimiento unificado y estructurado para explorar la imagen y el posicionamiento de marcas utilizando información textual contenida en las reseñas.

Con ese fin, este capítulo revisa tanto la literatura sobre la imagen de marca como la de minería de textos para encontrar posibles lagunas y áreas de mejora en el estudio de las asociaciones de marca a partir del contenido textual de las reseñas. Posteriormente, se propone un procedimiento estructurado para explorar el posicionamiento y la segmentación de marcas a partir del contenido textual de las reseñas. Se trata de un procedimiento simple y sencillo de implementar que puede ser utilizado por pequeñas y medianas empresas para obtener información sobre el posicionamiento de marcas competidoras en un mercado de referencia. Para analizar el texto de las reseñas, **utilizamos la minería de texto o “text mining” y nos basamos en un enfoque basado en diccionario**, usando el método LIWC (Pennebaker et al., 2015). Utilizando el LIWC, se extraen del texto de las reseñas las asociaciones emocionales y psicológicas que los consumidores tienen con las marcas. Luego, estas asociaciones se utilizan como variables

de entrada en un Análisis de Componentes Principales (Principal Component Analysis, PCA), lo que nos permite construir una mapa de posicionamiento de marcas. Finalmente, se utiliza el método de agrupamiento jerárquico (clustering jerárquico) para realizar un análisis de segmentación y agrupar marcas con asociaciones textuales similares. Aunque el mismo proceso podría extenderse a otras industrias, los resultados de este capítulo son específicos de nuestra categoría de productos (coloretes).

El capítulo 4 puede verse como una extensión del capítulo 3. En este caso, en lugar de utilizar un enfoque basado en un diccionario para realizar el análisis de minería de texto, aplicamos algunos algoritmos importantes de aprendizaje automático sin supervisión. A diferencia del capítulo 3, en lugar de hacer el análisis a nivel de marca, lo desagregamos a nivel de producto, para descubrir aspectos ocultos con respecto a cada producto específico en la categoría de coloretes. Así, el principal objetivo del capítulo 4 es el siguiente:

- Mostrar e ilustrar algunos métodos de minería de texto con el objetivo de explorar la imagen del producto descubriendo información contenida en reseñas

Con ese fin, este capítulo examina primero la literatura en minería de textos para identificar algunos de los algoritmos de aprendizaje automático más importantes que se pueden aplicar. De este modo, se han identificado cinco tareas populares de aprendizaje automático de texto sin supervisión: frecuencias de palabras, análisis de sentimientos, relación entre palabras, correlaciones producto-palabra y “topic modeling”. **Las cinco tareas identificadas se han aplicado empíricamente a nuestro caso específico de reseñas de colorete. Entre esas cinco tareas, todas excepto la de “topic modeling”, son relativamente sencillas de aplicar y no se requiere gran esfuerzo para interpretar el resultado. Sin embargo, para llevar a cabo el “topic modeling” se necesita un mayor esfuerzo de interpretación. Para llevar a cabo el análisis de “topic modeling”, el modelo de Asignación Latente de Dirichlet (en inglés, Latent Dirichlet Allocation o LDA) es uno de los más populares. Al utilizar el modelo LDA, obtuvimos los principales temas discutidos en los textos de las reseñas de nuestra base de datos. Para realizar el análisis LDA, es muy importante establecer primero el objetivo principal de la tarea. Dependiendo del objetivo, se debe pensar cuidadosamente en la lista de “stop words” que se deberán incluir en la etapa de preprocesamiento del LDA, ya que esta lista de “stop words” influirá en el resultado final de LDA. Si, por ejemplo, las palabras más comunes en la categoría, como “producto”, “colorete” o “cosméticos”, no están incluidas en la lista de “stop words”, el análisis LDA nos proporcionará temas muy generales tratados en el texto, y no nos dará insights más relevantes. El segundo aspecto importante al realizar**

LDA es interpretar correctamente los resultados. Los resultados del LDA dependen bastante del contexto, es decir, del tipo de texto que estemos analizando. Por lo tanto, para interpretar los diferentes temas proporcionados por el modelo debemos familiarizarnos con la industria específica donde estamos realizando el análisis. Dado que esta tarea implica cierta subjetividad en la comprensión de los temas, conviene involucrar a varios investigadores para llegar a un consenso en la interpretación. En nuestro entorno de productos de colorete, se identificaron 22 temas. Dado que estamos evaluando reseñas de una categoría de producto muy específica, que es el colorete, los temas son bastante similares entre sí. Sin embargo, los temas podrían ser más diferentes si comparáramos reseñas de diferentes categorías de productos. Del análisis LDA encontramos los siguientes resultados principales. En primer lugar, muchos temas se centran en características del producto relacionadas con los colores. Por otro lado, varios temas se centran en aspectos relacionados con el uso del producto (cómo y cuándo se usa). Por último, otros temas se centran también en discutir el formato del producto (por ejemplo, si el formato es en crema, en polvo, tamaño del producto, etc.).

Contribución teórica

En general, esta tesis contribuye a la literatura de eWOM en dos áreas. Por un lado, ampliando el conocimiento sobre cómo características no textuales y textuales de las reseñas influyen en diferentes comportamientos del consumidor (principalmente en la adopción de información y en el comportamiento de compra). Y, por otro lado, ampliando también el conocimiento de cómo analizar el contenido textual de las reseñas y de cómo las empresas pueden usarlo para analizar el posicionamiento de sus marcas.

En el capítulo 1, contribuimos a la literatura sobre el estudio de la utilidad de las reseñas y podemos basar la investigación en la teoría de la accesibilidad-diagnostics (Feldman y Lynch, 1988). Utilizando este marco teórico, abordamos la dimensión de accesibilidad incorporando la noción de visibilidad de las reseñas. De este modo, estudiamos la relación entre las características de las reseñas, que son los factores que hacen una reseña diagnóstica, y la utilidad de esa reseña, utilizando la visibilidad como variable moderadora. La literatura que estudia el impacto de las reseñas en su utilidad había asumido en su mayoría que las reseñas son independientes entre sí, sin embargo, cada reseña forma parte de una secuencia de reseñas para cada producto. Por lo tanto, en este capítulo contribuimos a la literatura previa incorporando la variable de visibilidad de las reseñas como moderadora de la relación entre las características de las reseñas y su utilidad. Además, proponemos una perspectiva empírica diferente para estudiar esta relación, el uso de una regresión Binomial Negativa de Ceros Inflados (en

inglés Zero Inflated Negative Binomial regression o ZINB). En general, la literatura anterior ha pasado por alto el hecho de que, en cualquier plataforma, encontramos muchas reseñas con cero votos útiles. La razón de que estas reseñas tengan cero votos puede deberse a una mala calidad, de modo que los consumidores no encuentran útil la información contenida en esas reseñas, pero también puede deberse al hecho de que esas reseñas no hayan sido vistas por los consumidores. En este sentido, si los consumidores no ven una reseña, es imposible que la voten como útil. Al utilizar un modelo ZINB, podemos analizar no solo los determinantes que influyen en la probabilidad de que las reseñas reciban un número positivo de votos útiles, sino también los factores que impulsan el exceso de reseñas con cero votos útiles.

En el capítulo 2, contribuimos a la literatura de dos formas principales. Primero, en la misma línea que en el capítulo 1, también enmarcamos la investigación en la teoría de la accesibilidad-diagnosicidad (Feldman y Lynch, 1988). Como ha afirmado la literatura anterior, las características de las reseñas influyen en las ventas de productos. Sin embargo, según las teorías del procesamiento de la información, y en particular la teoría de accesibilidad-diagnosicidad, es muy poco probable que los consumidores evalúen todas las reseñas disponibles para un producto. En cambio, es probable que los consumidores evalúen información más accesible o visible, por lo que las reseñas más visibles podrían tener mayor probabilidad de ser leídas. Dado que es más probable que los consumidores lean las opiniones más visibles, es probable que las características de esas reseñas en línea tengan una mayor influencia en el comportamiento de compra de los consumidores. Por lo tanto, en el capítulo 2 también incorporamos la noción de visibilidad de la reseña para estudiar cómo las reseñas influyen en las ventas de productos en diferentes casos de visibilidad (cuando asumimos que todas las reseñas tienen la misma probabilidad de ser vistas, cuando asumimos que es más probable que se vean las reseñas más útiles y cuando asumimos que es más probable que se vean las reseñas más recientes). Hasta el momento, los estudios no han encontrado consenso al explicar cómo las características de las reseñas, principalmente el rating y el volumen de reseñas, impactan en las ventas de productos. Algunos autores, como **Babić Rosario et al.** (2016), han revelado que los diferentes efectos de las variables de las reseñas en las ventas de productos dependen de factores como la plataforma donde se publica la opinión (por ejemplo, redes sociales, blog, sitios de reseñas, etc.), el tipo de producto (p. por ejemplo, bienes tangibles, servicios, productos digitales, utilitarios frente a productos hedónicos, alto riesgo financiero frente a bajo riesgo financiero y productos maduros frente a nuevos, etc.), la métrica utilizada para construir la variable (por ejemplo, promedio, acumulativo, incremental, etc.) y el diseño del estudio (por ejemplo,

cómo se controla la endogeneidad, el tipo de variables de control, etc.). En esta investigación, sugerimos que la visibilidad de la reseña también puede ser un factor que podría afectar la relación entre las características de la reseña y las ventas de productos. En este sentido, los efectos pueden ser diferentes si asumimos que los consumidores leen todas las reseñas disponibles para un producto o si asumimos que es más probable que se lean las reseñas visibles.

Además, la literatura se ha centrado principalmente hasta ahora en estudiar cómo las características no textuales de las reseñas influyen en las ventas de productos. En nuestra investigación, analizamos no sólo el impacto de las características no textuales de las reseñas en las ventas, sino también el de algunas características textuales. Por lo tanto, este capítulo amplía el conocimiento de cómo las diferentes características de las reseñas influyen en las decisiones de compra de los consumidores.

El capítulo 3 y el capítulo 4 tienen un enfoque más exploratorio, porque estamos interesados en explorar la imagen de marca y el posicionamiento de la marca, y están más orientados a la práctica. Sin embargo, contribuimos a la literatura de imagen de marca y posicionamiento de marca mostrando cómo las técnicas de minería de texto para el procesamiento del lenguaje natural (NLP) pueden servirnos como una herramienta poderosa para comprender mejor la imagen y el posicionamiento de la marca. Tradicionalmente, la imagen de marca se ha explorado tradicionalmente utilizando escalas multidimensional y técnicas cualitativas, como la ZMET desarrollada por Zaltman (1997). Sin embargo, el uso de eWOM, siendo las reseñas online un caso particular de eWOM, aporta varias ventajas sobre los datos de encuestas a la hora de estudiar la imagen de marca. En primer lugar, eWOM es espontáneo (Marchand et al., 2017; Yang y Cho, 2015). En general, ocurre sin que los especialistas en marketing lo impulsen o influyan directamente y suele estar motivado por el deseo de ayudar a los demás, advertir a los demás o comunicar el estado (Kozinets et al., 2010). Debido a eso, es más probable que los consumidores expresen sus percepciones de marca verdadera e informen sobre sus comportamientos reales. Además, hay una gran cantidad de opiniones disponibles online, por lo que la imagen de marca se puede estudiar utilizando muestras muy grandes de opiniones de los consumidores. Esta última es la principal ventaja del eWOM sobre las técnicas cualitativas, que se basan principalmente en pequeñas muestras de personas que participan en encuestas.

En el capítulo 3, mostramos y proponemos un procedimiento unificado para estudiar las asociaciones de marcas a partir del texto de las reseñas con el fin de analizar el posicionamiento y la segmentación de la marcas. Dado que la literatura que estudia las

dimensiones textuales de las reseñas aún es escasa, cualquier artículo que aborde este tipo de estudio constituye una contribución importante a la literatura. Hasta ahora, los artículos que estudian el contenido textual de las reseñas son bastante heterogéneos en términos de objetivos de investigación y técnicas de minería de texto utilizadas. Por lo tanto, en este capítulo proponemos un procedimiento más estructurado y sencillo de implementar que puede ser seguido tanto por empresas como por investigadores que quieran analizar el contenido textual de las reseñas para incrementar sus conocimientos sobre posicionamiento y segmentación de marca. Este procedimiento se basa en una técnica de minería de texto basada en diccionarios, denominada Linguistic Inquiry and Word Count (LIWC), desarrollada por Pennebaker et al. (2015) Sin embargo, el procedimiento podría reproducirse utilizando cualquier otro método de diccionario alternativo de minería de texto.

En el capítulo 4 contribuimos a la literatura sobre la imagen de marca ilustrando algunos algoritmos de procesamiento del lenguaje natural (NLP), en lugar de los métodos basados en el léxico como los utilizados en el capítulo 3, que se pueden usar para responder algunas preguntas de investigación relacionadas con el estudio de la imagen de marca. Al utilizar los algoritmos de aprendizaje automático no supervisados que se ilustran en este documento, las asociaciones con marcas y productos se pueden extraer de los textos de las reviews.

Implicaciones para la práctica

Aunque esta tesis tiene varias contribuciones a la teoría, el estudio de reseñas tiene también claras implicaciones para la gestión.

En primer lugar, las empresas pueden ampliar sus conocimientos sobre cómo los consumidores procesan la información en un entorno online. Los resultados sugieren que no solo ciertas características de las reseñas influyen en las decisiones de adopción de información y en las decisiones de compra de los consumidores, sino también la forma en que estas reseñas se presentan a los consumidores en el sitio web. Por lo tanto, una recomendación derivada de nuestros resultados es que los especialistas en marketing deberían proporcionar más herramientas en sus sitios web para permitir a los consumidores organizar las reseñas de acuerdo con sus preferencias. Por ejemplo, teniendo en cuenta que los consumidores consideran relevantes y de confianza aquellas reseñas votadas como útiles, las empresas podrían hacer que la información contenida en esas reseñas fuera más accesible para los consumidores para reducir el esfuerzo cognitivo derivado de procesar grandes volúmenes de información online. Por ejemplo,

las empresas podrían incorporar un filtro que mostrara un resumen de las características de las reseñas más útiles. Con ese filtro, los consumidores podrían consultar, por ejemplo, el perfil de los revisores que escriben esas reseñas más útiles y también algunos aspectos directamente relacionados con el contenido textual (por ejemplo, el nivel de subjetividad, el nivel de confianza, etc.). Si los consumidores creen que la empresa los está ayudando en su proceso de toma de decisiones, será más probable que tengan una mejor experiencia online, lo que podría traducirse en una mejor actitud hacia la empresa y una mayor intención de compra.

En línea con la implicación anterior, y teniendo en cuenta que las reseñas más visibles es más probable que ejerzan mayor influencia en el comportamiento del consumidor, las empresas deberían de prestar más atención a la información dada en esas reseñas más visibles. Así, las empresas podrían analizar la información contenida en esas reseñas para satisfacer mejor las necesidades de los consumidores, mejorar los productos actuales o lanzar nuevos. Además, considerando que los consumidores usan herramientas de ordenación para reducir el esfuerzo cognitivo de administrar grandes cantidades de información, las empresas podrían incorporar nuevas herramientas de ordenación para ayudar a los consumidores en su toma de decisiones. Proporcionar más opciones de herramientas de ordenación haría que los consumidores tuvieran una mejor experiencia online y conduciría a una mayor satisfacción del cliente. Además, si hubiera más herramientas de ordenación de reseñas disponibles, los consumidores podrían seleccionar qué reseñas leer según su criterio preferido. Por ejemplo, considerando el impacto que tienen aquellas reseñas más útiles en la toma de decisiones de los consumidores, las empresas podrían ofrecer la opción de ordenación en varios niveles, de modo que los consumidores pudieran ordenar al mismo tiempo por reseñas "más útiles" y a la vez, por ejemplo, **por "mayor número de estrellas", "menor número de estrellas" o "más reciente"**. De ese modo, los consumidores podrían seleccionar mejor las reseñas de su interés. En línea con nuestros hallazgos, la empresa de cosméticos de donde hemos obtenido los datos para esta tesis introdujo algunos cambios en su web después de que recopilamos los datos para la investigación. Por ejemplo, ya no muestran las reseñas **por el mecanismo de ordenación predeterminado de " más recientes", sino por el mecanismo de "más útiles"**. Esto corrobora nuestro hallazgo y revela que las reseñas más útiles probablemente influyan más en el comportamiento de compra de los consumidores.

En tercer lugar, en cuanto al estudio del contenido textual de las reseñas, esta tesis también tiene claras implicaciones de gestión. Tal como lo revelan Magoulas y Swoyer (2020), una de las principales barreras para la adopción de la Inteligencia Artificial (IA)

en las empresas es la falta de personas capacitadas o la dificultad para contratar los roles requeridos. Uno de los métodos utilizados para la minería de texto se basa en algoritmos de aprendizaje automático, que son una aplicación específica de la IA y, por lo tanto, muchas empresas tienen dificultades para adoptarlos. En el capítulo 3 de esta tesis, proponemos seguir un procedimiento de minería de texto basado en diccionarios, el método LIWC, que es una herramienta más fácil de implementar que los algoritmos de aprendizaje automático. Por lo tanto, nuestra investigación podría ser especialmente útil para las pequeñas y medianas empresas que no tienen los recursos suficientes para llevar a cabo algoritmos complejos de aprendizaje automático para la minería de texto y desean descubrir características ocultas de los textos de las reseñas u otros tipos de eWOM. En nuestra investigación, nos centramos en varias variables emocionales y psicológicas proporcionadas por el LIWC, pero otras variables dadas por la herramienta también podrían explorarse usando el mismo procedimiento. Además, también se podrían utilizar otras herramientas basadas en diccionarios, diferentes al LIWC. Al utilizar este procedimiento, las empresas pueden comprender mejor cómo los consumidores posicionan sus marcas en comparación con las marcas de la competencia y podrían obtener información para una mejor estrategia de diferenciación basada en estas asociaciones emocionales y psicológicas del consumidor.

Finalmente, esta tesis también ilustra en el capítulo 4 algunos de los algoritmos de aprendizaje automático más populares para la minería de texto. Las empresas podrían adoptar esos algoritmos para varias aplicaciones, como para obtener información sobre los intereses y preferencias de los consumidores, para analizar qué palabras positivas y negativas están asociadas a cada producto o qué características del producto están asociadas a sentimientos positivos o negativos, para identificar quejas de los clientes para tratar de mejorar el producto o servicio, para analizar el posicionamiento de la marca o producto y para obtener insights para el desarrollo de nuevos productos.

Limitaciones e investigación futura

Teniendo en cuenta que esta tesis estudia un fenómeno complejo, que es cómo las reseñas influyen en diferentes tipos de comportamiento del consumidor y cómo se puede analizar e interpretar el contenido textual para estudiar la imagen de marca, esta tesis solo aporta una contribución muy pequeña a la literatura y es necesario esbozar las principales limitaciones para poder plantear futuras líneas de investigación.

La primera limitación hace referencia a la base de datos utilizada para la investigación empírica. Esta tesis se limita a analizar las reseñas de una categoría concreta de

productos cosméticos, que es la categoría de colorete, pero la investigación futura podría ampliarse a otras categorías de cosméticos o incluso a otras categorías de productos de experiencia y productos de búsqueda. La investigación también puede extenderse al estudio de reseñas de servicios, como hoteles o restaurantes. De esa manera, podríamos saber si los resultados de nuestra investigación podrían generalizarse aún más a otras categorías de productos o servicios o, en cambio, si dependen del contexto. Además, esta investigación se lleva a cabo utilizando reseñas de consumidores, pero otros estudios podrían realizarse utilizando otros tipos de eWOM, como comentarios en redes sociales y contenido de blogs.

En segundo lugar, las variables de visibilidad de las reseñas utilizadas en el capítulo 1 y en el capítulo 2 se basa en dos herramientas de ordenación, según las reseñas **“más útiles”** y según las reseñas **“más recientes”**. La investigación futura podría utilizar otras herramientas de ordenación para construir las variables de visibilidad, como la de las reseñas con mayor número de estrellas o con menor número de estrellas. Los resultados podrían compararse con los de esta tesis. Futuros estudios también podrían usar datos de clics para conocer de forma más precisa cómo los consumidores ordenan las reseñas en los sitios web. Además, en el capítulo 1 y capítulo 2 se han utilizado tres variables textuales de las reseñas extraídas mediante el método LIWC para analizar la utilidad de las reseñas y las ventas de productos, pero otras investigaciones podrían considerar el efecto de otras características textuales, ya sea proporcionadas por el LIWC o por otros métodos de minería de texto.

Por último, aunque el uso del LIWC se ha generalizado en la literatura previa sobre minería de textos, futuras investigaciones podrían utilizar otros métodos alternativos basados en diccionarios para estudiar el posicionamiento de marcas. Las empresas también podrían utilizar métodos de aprendizaje automático para la minería de textos si necesitan extraer aspectos más específicos de los textos de las reseñas y si cuentan con los recursos necesarios. En el capítulo 3 de esta tesis, se presta especial interés en explorar las asociaciones de marcas emocionales y psicológicas, que han sido menos exploradas en la literatura anterior. Sin embargo, también se podrían analizar otros aspectos ocultos en los textos, como las características relacionadas con el estilo de escritura del revisor (por ejemplo, escritura informal, escritura cognitiva y enfoque temporal), que también son proporcionadas por el método LIWC. Además, las asociaciones de marcas en esta tesis se evalúan en un minorista online específico, pero sería interesante comparar asociaciones de marcas en diferentes minoristas, para ver si esas asociaciones son diferentes según el minorista o, por el contrario, no dependen de la plataforma. Además, podríamos realizar una encuesta para ver si las asociaciones de

marcas extraídas a través de las reseñas son las mismas que las asociaciones de marcas que los consumidores tienen en general, que podrían explorarse mediante el uso de encuestas.

Finalmente, al explorar la imagen de marca utilizando el contenido textual de las reseñas en línea, podríamos estar capturando las percepciones de las marcas de aquellos consumidores que generalmente escriben opiniones en línea. Sin embargo, puede haber algunos consumidores que no se expresen en línea. Puede ser interesante comparar los resultados con los obtenidos con métodos tradicionales, como el de escalado multidimensional o técnicas cualitativas.

References

- Aaker, D. A. (1991). Managing Brand Equity. *Managing Brand Equity*.
- Aaker, D. A. (1996). Building strong brands - aaker,da. *Building Strong Brands*, 13(4), 382–386.
- Ahani, A., Nilashi, M., Yadegaridehkordi, E., Sanzogni, L., Tarik, A. R., Knox, K., Samad, S., & Ibrahim, O. (2019). Revealing customers' satisfaction and preferences through online review analysis: The case of Canary Islands hotels. *Journal of Retailing and Consumer Services*, 51(July), 331–343.
- Ahmad, S. N., & Laroche, M. (2015). How do expressed emotions affect the helpfulness of a product review? Evidence from reviews using latent semantic analysis. *International Journal of Electronic Commerce*, 20(1), 76–111.
- Aiken, L. S., & West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*.
- Al-Obeidat, F., Spencer, B., & Kafeza, E. (2018). The Opinion Management Framework: Identifying and addressing customer concerns extracted from online product reviews. *Electronic Commerce Research and Applications*, 27, 52–64.
- Al., H. et. (2010). *Multivariate Analysis 7 th Edition*.
- Aljukhadar, M., Senecal, S., & Daoust, C.-E. (2012). Using Recommendation Agents to Cope with Information Overload. *International Journal of Electronic Commerce*, 17(2), 41–70.
- Ansari, A., & Mela, C. F. (2003). *E-Customization*. 40(2), 131–145.
- Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*, 57(8), 1485–1509.
- Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2), 277.
- Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1), 29–51.
- Areni, C. S. (2003). The Effects of Structural and Grammatical Variables on Persuasion: An Elaboration Likelihood Model Perspective. *Psychology and Marketing*, 20(4), 349–375.
- Arun, R., Suresh, V., Madhavan, C. E. V., & M.N, N. M. (2010). On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In *Advances in Knowledge Discovery and Data Mining* (pp. 391–403).
- Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. A. (2016). The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors. *Journal of Marketing Research*, 53(3), 297–318.
- Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, O, 2200–2204.
- Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of Online Consumer Reviews: Readers' Objectives and Review Cues. *International Journal of Electronic Commerce*,

- 17(May 2015), 99–126.
- Baker, W. E. (2001). The Diagnosticity of Advertising Generated Brand Attitudes in Brand Choice Contexts. *Journal of Consumer Psychology*, 11(2), 129–139.
- Baker, W. E., Lutz, R. J., Journal, S., Spring, N., Taylor, P., Baker, W. E., & Lutz, R. J. (2000). Effectiveness An Empirical Test of an Updated Relevance- Accessibility Model of Advertising Effectiveness. *Journal of Advertising*, 29(1), 1–14.
- Bakhshi, S., Kanuparth, P., & Shamma, D. A. (2015). Understanding Online Reviews: Funny, Cool or Useful ? *CSCW 2015: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1270–1276.
- Baksi, A. K., & Panda, T. K. (2018). Branding destinations with multisensory brand associations and evaluating its impact on behavioural pattern under the intervention of multiplex phenomenon of relationship-branding. *Management Science Letters*, 8(11), 1169–1182.
- Balducci, B., & Marinova, D. (2018a). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, 46(4), 557–590.
- Balducci, B., & Marinova, D. (2018b). Unstructured data in marketing. In *Journal of the Academy of Marketing Science* (Vol. 46, Issue 4, pp. 557–590). Springer New York LLC.
- Beach, L. R. (1993). Broadening the definition of decision making: The role of prechoice screening option. *Psychological Science*, 4(4), 215–220.
- Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding Satisfied and Dissatisfied Hotel Customers: Text Mining of Online Hotel Reviews. *Journal of Hospitality Marketing and Management*, 25(1), 1–24.
- Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the Tribes: Using Text for Marketing Insight. *Journal of Marketing*, 84(1), 1–25.
- Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive Consumer Choice Processes. *Journal of Consumer Research*, 25(3), 187–217.
- Biehal, G., & Chakravarti, D. (1983). Information Accessibility as a Moderator of Consumer Choice. *Journal of Consumer Research*, 10(1), 1.
- Biswas, D., Grewal, D., & Roggeveen, A. (2010). How the order of sampled experiential products affects choice. *Journal of Marketing Research*, 47(3), 508–519.
- Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), 115–143.
- Brandt, C., De Mortanges, C. P., Bluemelhuber, C., & Van Riel, A. C. R. (2011). Associative Networks: A New Approach to Market Segmentation. *International Journal of Market Research*, 53(2), 187–207.
- Breugelmans, E., Campo, K., & Gijsbrechts, E. (2007). Shelf sequence and proximity effects on online grocery choices. *Marketing Letters*, 18(1–2), 117–133.
- BrightLocal. (2019). *Local Consumer Review Survey*.
- BrightLocal. (2020). *Local Consumer Review Survey*. <https://www.brightlocal.com/research/local-consumer-review-survey/>
- Brown, C. L., & Krishna, A. (2004). The skeptical shopper: A metacognitive account for the effects of default options on choice. *Journal of Consumer Research*, 31(3), 529–

539.

- Bruine de Bruin, W., & Keren, G. (2003). Order effects in sequentially judged options due to the direction of comparison. *Organizational Behavior and Human Decision Processes*, 92(1–2), 91–101.
- Buda, R., & Zhang, Y. (2000). Consumer product evaluation: the interactive effect of message framing, presentation order, and source credibility. *Journal of Product & Brand Management*, 9(4), 229–242.
- Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics Using Stata*. Stata Press Book, StataCorp LP.
- Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data, Second edition. In *Regression Analysis of Count Data, Second Edition*.
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72(7–9), 1775–1781.
- Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, 50(2), 511–521.
- Chaiken, S. (1987). *The heuristic model of persuasion*.
- Chapman, C., & Feit, E. M. (2015). *R for marketing research and analytics*.
- Chen, C. C., & Tseng, Y. De. (2011). Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, 50(4), 755–768.
- Chen, K., Kou, G., Shang, J., & Chen, Y. (2015). Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches. *Electronic Commerce Research and Applications*, 14(1), 58–74.
- Cheng, Y. H., & Ho, H. Y. (2015). Social influence’s impact on reader perceptions of online reviews. *Journal of Business Research*, 68(4), 883–887.
- Cheung, C. M. K., Lee, M. K. O., & Rabjohn, N. (2008). The impact of electronic word-of-mouth. In *Internet Research* (Vol. 18, Issue 3).
- Chevalier, J. A., & Goolsbee, A. (2003). Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics*, 1, 203–222.
- Chevalier, J., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 43(3), 345–354.
- Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science*, 29(5), 944–957.
- Cho, E., Fiore, A. M., & W. Russell, D. (2015). Validation of a Fashion Brand Image Scale Capturing Cognitive, Sensory, and Affective Associations: Testing Its Role in an Extended Brand Equity Model. *Psychology & Marketing*, 32(1), 28–48.
- Chua, A. Y. K., & Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior*, 54, 547–554.
- Chung, C.K., & Pennebaker, J. W. (2007). The psychological function of function words. *Social Communication: Frontiers of Social Psychology*, 343–359.

- Chung, Cindy K., & Pennebaker, J. W. (2007). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality*, 42(1), 96–132.
- Cohn, M. a., Mehl, M. R., & Pennebaker, J. W. (2004). Markers of Linguistic Psychological Change. *Psychological Science*, 15(10), 687–693.
- Cohn, M. a., Mehl, M. R., & Pennebaker, J. W. (2014). Markers of Linguistic Psychological Change Surrounding September 11, 2001. *Psychological Science*, 15(10), 687–693.
- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2014). *Markers of Linguistic Psychological Change Surrounding September*. 15(10), 687–693.
- Cui, G., Lui, H.-K., & Guo, X. (2012). The Effect of Online Consumer Reviews on New Product Sales. *International Journal of Electronic Commerce*, 17(1), 39–58.
- Cyr, D., Head, M., Lim, E., & Stibe, A. (2018). Using the elaboration likelihood model to examine online persuasion through website design. *Information and Management*, April 2017, 1–15.
- Darley, W. K., & Smith, R. E. (1993). Advertising Claim Objectivity: Antecedents and Effects. *Journal of Marketing*, 57(4), 100.
- Davis, D. F., Golobic, S. L., & Marquardt, A. (2009). Measuring brand equity for logistics services. *The International Journal of Logistics Management*, 20(2), 201–212.
- De Vries, L., Gensler, S., & Leeﬂang, P. S. H. (2012). Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing. *Journal of Interactive Marketing*, 26(2), 83–91.
- Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, 27(4), 293–307.
- Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4), 23–45.
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective Latent Concept Modeling for ad hoc information retrieval. *Document Numerique*, 17(1), 61–84.
- Drèze, X., & Zufryden, F. (2004). Measurement of online visibility and its impact on Internet traffic. *Journal of Interactive Marketing*, 18(1), 20–37.
- Duan, W., Gu, B., & Whinston, A. B. (2008a). Do online reviews matter? - An empirical investigation of panel data. *Decision Support Systems*, 45(4), 1007–1016.
- Duan, W., Gu, B., & Whinston, A. B. (2008b). The dynamics of online word-of-mouth and product sales-An empirical investigation of the movie industry. *Journal of Retailing*, 84(2), 233–242.
- Elberse, A., & Eliashberg, J. (2003). Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures. *Marketing Science*, 22(3), 329–354.
- Ellsworth, P. C., & Smith, C. A. (1988). From appraisal to emotion: Differences among unpleasant feelings. *Motivation and Emotion*, 12(3), 271–302.
- Ert, E., & Fleischer, A. (2016). Mere Position Effect in Booking Hotels Online. *Journal of Travel Research*, 55(3), 311–321.
- Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy:

- The role of personal photos in Airbnb. *Tourism Management*, 55, 62–73.
- Eslami, S. P., & Ghasemaghaei, M. (2018). Effects of online review positiveness and review score inconsistency on sales: A comparison by product involvement. *Journal of Retailing and Consumer Services*, 45(June), 74–80.
- Faircloth, J. B., Capella, L. M., & Alford, B. L. (2001). The Effect of Brand Attitude and Brand Image on Brand Equity. *Journal of Marketing Theory and Practice*, 9(3), 61–75.
- Feldman, J. M., & Lynch, J. G. (1988). Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior. *Journal of Applied Psychology*, 73(3), 421–435.
- Felfernig, A., Friedrich, G., Gula, B., Hitz, M., Kruggel, T., Leitner, G., Melcher, R., Riepan, D., Strauss, S., Teppan, E., & Vitouch, O. (2007). Persuasive recommendation: Serial position effects in knowledge-based recommender systems. *Proceedings of the 2nd International Conference on Persuasive Technology (PERSUASIVE '07)*, 283–295.
- Filieri, R., Hofacker, C. F., & Alguezaui, S. (2018). What makes information in online consumer reviews diagnostic over time? The role of review relevancy, factuality, currency, source credibility and ranking score. *Computers in Human Behavior*, 80, 122–131.
- Filieri, R., & McLeay, F. (2014). E-WOM and Accommodation: An Analysis of the Factors That Influence Travelers' Adoption of Information from Online Reviews. *Journal of Travel Research*, 53(1), 44–57.
- Filieri, R., McLeay, F., Tsui, B., & Lin, Z. (2018). Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services. *Information and Management*, 55(8), 956–970.
- Filieri, R., Raguseo, E., & Vitari, C. (2018). When are extreme ratings more helpful? Empirical evidence on the moderating effects of review characteristics and product type. *Computers in Human Behavior*, 88, 134–142.
- Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., & Freling, T. (2014). How Online Product Reviews Affect Retail Sales: A Meta-analysis. *Journal of Retailing*, 90(2), 217–232.
- Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., & Treinen, M. (2001). What makes web sites credible? A report on a large quantitative study. *Conference on Human Factors in Computing Systems - Proceedings*, 61–68.
- Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 19(3), 291–313.
- Gao, B., Hu, N., & Bose, I. (2017). Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews. *Decision Support Systems*, 95, 1–11.
- Gensler, S., Völckner, F., Egger, M., Fischbach, K., & Schoder, D. (2015). Listen to your customers: Insights into brand image using online consumer-generated product reviews. *International Journal of Electronic Commerce*, 20(1), 112–141.
- Ghose, A., & Ipeirotis, P. G. (2010). Estimating the Helpfulness and Economic Impact of Product Reviews. *Ieee Transactions on Knowledge and Data Engineering*, 1–15.
- Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing Ranking Systems for Hotels on

- Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science*, 31(3), 493–520.
- Girard, T., & Dion, P. (2010). Validating the search, experience, and credence product classification framework. *Journal of Business Research*, 63(9–10), 1079–1087.
- Gobet, F., Lane, P., Croker, S., Cheng, P., Jones, G., Oliver, I., & Pine, J. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5(6), 236–243.
- Godes, D., & Silva, J. C. (2012). Sequential and Temporal Dynamics of Online Opinion. *Marketing Science*, 31(3), 448–473.
- Godes, David, & Mayzlin, D. (2004). Using Online Conversations to Study Word-of-Mouth Communication. *Marketing Science*, 23(4), 545–560.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(SUPPL. 1), 5228–5235.
- Gu, B., Gu, B., & Park, J. (2012). The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products. *Information Systems Research*, 23(1), 182–196.
- Guo, J., Wang, X., & Wu, Y. (2020). Positive emotion bias: Role of emotional content from online customer reviews in purchase decisions. *Journal of Retailing and Consumer Services*, 52(July 2019).
- Guo, Y., Barnes, S. J., & Jia, Q. (2017a). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.
- Guo, Y., Barnes, S. J., & Jia, Q. (2017b). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.
- Hamby, A., Daniloski, K., & Brinberg, D. (2015). How consumer reviews persuade through narratives. *Journal of Business Research*, 68(6), 1242–1250.
- Haridasan, A. C., & Fernando, A. G. (2018). Online or in-store: unravelling consumer's channel choice motives. *Journal of Research in Interactive Marketing*, 12(2), 215–230.
- Harrison-Walker, L. J. (2001). The Measurement of Word-of-Mouth Communication and an Investigation of Service Quality and Customer Commitment As Potential Antecedents. *Journal of Service Research*, 4(1), 60–75.
- Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, 36(1), 20–38.
- Häubl, G., & Trifts, V. (2000). Consumer Decision Making in Online Shopping Environments : The Effects of Interactive Decision Aids Consumer Decision Making in Online Shopping Environments : The Effects of Interactive Decision Aids. *Marketing Science*, 19(1), 4–21.
- Haugtvedt, C. P., & Wegener, D. T. (1994). Message Order Effects in Persuasion: An Attitude Strength Perspective. *Journal of Consumer Research*, 21(1), 205.
- Hauser, J. R., & Koppelman, F. S. (1979). Alternative Perceptual Mapping Techniques : Relative Accuracy and Usefulness. *Journal of Marketing Research*, 16(4), 495–506.
- Henderson, G. R., Iacobucci, D., & Calder, B. J. (1998). Brand diagnostics: Mapping

- branding effects using consumer associative networks. *European Journal of Operational Research*, 111(2), 306–327.
- Heng, Y., Gao, Z., Jiang, Y., & Chen, X. (2018). Exploring hidden factors behind online food shopping from Amazon reviews: A topic mining approach. *Journal of Retailing and Consumer Services*, 42(January), 161–168.
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, 18(1), 38–52.
- Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective. *Journal of Consumer Research*, 17(4), 454.
- Herrmann, A., Goldstein, D. G., Stadler, R., Landwehr, J. R., Heitmann, M., Hofstetter, R., & Huber, F. (2011). The effect of default options on choice-Evidence from online product configurators. *Journal of Retailing and Consumer Services*, 18(6), 483–491.
- Ho-Dac, N. N., Carson, S. J., & Moore, W. L. (2013). The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter? *Journal of Marketing*, 77(6), 37–53.
- Hofmann, J., Clement, M., Völckner, F., & Hennig-Thurau, T. (2017). Empirical generalizations on the impact of stars on the economic success of movies. *International Journal of Research in Marketing*, 34(2), 442–461.
- Holbrook, M. B. (1978). Beyond Attitude Structure: Toward the Informational Determinants of Attitude. *Journal of Marketing Research*, 15(4), 545.
- Holtgraves, T., & Lasky, B. (1999). Linguistic power and persuasion. *Journal of Language and Social Psychology*, 18(2), 196–205.
- Hornik, K., & Grün, B. (2011). topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software*, 40(13), 1–30.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD'04*, 168.
- Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? the mediating role of online review sentiments on product sales. *Decision Support Systems*, 57(1), 42–53.
- Hu, Y., & Li, X. (2011). Context-Dependent Product Evaluations: An Empirical Analysis of Internet Book Reviews. *Journal of Interactive Marketing*, 25(3), 123–133.
- Huang, A. H., Chen, K., Yen, D. C., & Tran, T. P. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behavior*, 48, 17–27.
- Hutto, C. J., & Gilbert, E. E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14).". *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*. <http://sentic.net/>
- International, E. (2017). *Colour cosmetics in the US* (Issue May).
- Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011a). Language Style Matching Predicts Relationship Initiation and Stability. *Psychological Science*, 22(1), 39–44.

- Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011b). Language Style Matching Predicts Relationship Initiation and Stability. *Psychological Science*, 22(1), 39–44.
- Jalilvand, M. R., & Samiei, N. (2012). The effect of electronic word of mouth on brand image and purchase intention: An empirical study in the automobile industry in Iran. *Marketing Intelligence and Planning*, 30(4), 460–476.
- Jiang, S., Cai, S., Olle, G. O., & Qin, Z. (2015). Durable product review mining for customer segmentation. *Kybernetes*, 44(1), 124–138.
- Jiménez, F. R., & Mendoza, N. A. (2013). Too popular to ignore: The influence of online reviews on purchase intentions of search and experience products. *Journal of Interactive Marketing*, 27(3), 226–235.
- John, D. R., Loken, B., Kim, K., & Monga, A. B. (2006). Brand concept maps: A methodology for identifying brand association networks. *Journal of Marketing Research*, 43(4), 549–563.
- Johnson, E. J., Bellman, S., & Lohse, G. L. (2002). Defaults , Framing and Privacy :Why Opting In-Opting Out. *Marketing Letters*, 13(1), 5–15.
- Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, 80(4), 557–571.
- Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*, 33(2), 125–143.
- Kamble, Z., Doan, Q. T., Nguyen, H. N., Zeng, Z., & Zihui, L. L. (2020). The Significance of Online Reviews for Hotel Entrepreneurial Success. In *Entrepreneurial Development and Innovation in Family Businesses and SMEs* (pp. 75–86).
- Keller, K. L. (1993). *Conceptualizing, Measuring, and Managing Customer-Based Brand Equity*. 57(1), 1–22.
- Kelly, G. A. (1991). *The Psychology of Personal Constructs*.
- Khwaja, M. G., & Zaman, U. (2020). Configuring the evolving role of ewom on the consumers information adoption. *Journal of Open Innovation: Technology, Market, and Complexity*, 6(4), 1–13.
- Kim, H. B., Kim, W. G., & An, J. A. (2003). The effect of consumer-based brand equity on firms' financial performance. *Journal of Consumer Marketing*, 20(4–5), 335–351.
- Kim, M., & Lennon, S. (2007). The Effects of Visual and Verbal Information on Attitudes and Purchase Intentions in Internet Shopping. *Psychology & Marketing*, 24(9), 763–785.
- Kim, S. G., & Kang, J. (2018). Analyzing the discriminative attributes of products using text mining focused on cosmetic reviews. *Information Processing and Management*, 54(6), 938–957.
- King, M. F., & Balasubramanian, S. k. (1994). The effects of Expertise, End Goal, and Product Type on Adoption of Preference Formation Strategy. *Journal of the Academy of Marketing Science*, 22.
- Kostyk, A., Niculescu, M., & Leonhardt, J. M. (2017). Less is more: Online consumer

- ratings' format affects purchase intentions and processing. *Journal of Consumer Behaviour*, 16(5), 434–441.
- Kotler, P., & Armstrong, G. (2020). *Principles of Marketing* (18th Global). Pearson Education Limited.
- Kozinets, R. V., De Valck, K., Wojnicki, A. C., & Wilner, S. J. S. (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of Marketing*, 74(2), 71–89.
- Kübler, R. V., Colicev, A., & Pauwels, K. H. (2019). Social Media's Impact on the Consumer Mindset: When to Use Which Sentiment Extraction Tool? *Journal of Interactive Marketing*, 50, 136–155.
- Kudeshia, C., & Kumar, A. (2017). Social eWOM: does it affect the brand attitude and purchase intention of brands? *Management Research Review*, 40(3), 310–330.
- Lee, P. J., Hu, Y. H., & Lu, K. T. (2018). Assessing the helpfulness of online hotel reviews: A classification-based approach. *Telematics and Informatics*, 35(2), 436–445.
- Lee, S., & Choeh, J. Y. (2018). The interactive impact of online word-of-mouth and review helpfulness on box office revenue. *Management Decision*, 56(4), 849–866.
- Lee, S., & Choeh, J. Y. (2020). Using the social influence of electronic word-of-mouth for predicting product sales: The moderating effect of review or reviewer helpfulness and product type. *Sustainability (Switzerland)*, 12(19).
- Lee, T. Y., & Bradlow, E. T. (2011). Automated Marketing Research Using Online Customer Reviews. *Journal of Marketing Research*, 48(5), 881–894.
- Levinson, P. (2013). *New New Media*. 223. <https://books.google.com/books?id=G009uwAACAAJ&pgis=1>
- Lewsey, J. D., & Thomson, W. M. (2004). The utility of the zero-inflated Poisson and zero-inflated negative binomial models: A case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. *Community Dentistry and Oral Epidemiology*, 32(3), 183–189.
- Li, M., Huang, L., Tan, C.-H., & Wei, K.-K. (2013). Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features. *International Journal of Electronic Commerce*, 17(4), 101–136.
- Li, S. T., Pham, T. T., & Chuang, H. C. (2019). Do reviewers' words affect predicting their helpfulness ratings? Locating helpful reviewers by linguistics styles. *Information and Management*, 56(1), 28–38.
- Li, Xiaolin, Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information and Management*, 56(2), 172–184.
- Li, Xinxin, & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, 19(4), 456–474.
- Liang, T. P., Li, X., Yang, C. T., & Wang, M. (2015). What in Consumer Reviews Affects the Sales of Mobile Apps: A Multifacet Sentiment Analysis Approach. *International Journal of Electronic Commerce*, 20(2), 236–260.
- Liu, J., Cao, Y., Lin, C., Huang, Y., ... M. Z. the 2007 J. C. on, & 2007, U. (2007). Low-quality product review detection in opinion summarization. *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, June, 334–342.
- Liu, Q., & Karahanna, E. (2015). An agent-based modeling analysis of helpful vote on

- online product reviews. *Proceedings of the Annual Hawaii International Conference on System Sciences, 2015-March*, 1585–1595.
- Liu, X., Burns, A. C., & Hou, Y. (2017). An Investigation of Brand-Related User-Generated Content on Twitter. *Journal of Advertising*, 46(2), 236–247.
- Liu, Y. (2006). Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*, 70(3), 74–89.
- Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, 47, 140–151.
- Low, G. S., & Lamb, C. W. (2000). The measurement and dimensionality of brand associations. *Journal of Product & Brand Management*, 9(6), 350–370.
- Lozano, M. B., Martínez, B., & Pindado, J. (2016). Corporate governance, ownership and firm value: Drivers of ownership as a good corporate governance mechanism. *International Business Review*, 25(6), 1333–1343.
- Luca, M. (2011). Revenue : The Case of Reviews , Reputation , and Revenue : The Case of Yelp . com. *Harvard Business Review*, 1–40.
- Ludwig, S., de Ruyter, K., Friedman, M., Brügger, E. C., Wetzels, M., & Pfann, G. (2013a). More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*, 77(1), 87–103.
- Ludwig, S., de Ruyter, K., Friedman, M., Brügger, E. C., Wetzels, M., & Pfann, G. (2013b). More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*, 77(1), 87–103.
- Lynch, Jr., J. G., Marmorstein, H., & Weigold, M. F. (1988). Choices from Sets Including Remembered Brands: Use of Recalled Attributes and Prior Overall Evaluations. *Journal of Consumer Research*, 15(2), 169.
- Magerman, T., Van Looy, B., Baesens, B., & Debackere, K. (2011). Assessment of Latent Semantic Analysis (LSA) text mining algorithms for large scale mapping of patent and scientific publication documents. *FBE Research Report MSI_1114*.
- Magoulas, R., & Swoyer, S. (2020). *AI Adoption in the Enterprise 2020*.
- Mahr, D., Stead, S., & Odekerken-Schröder, G. (2019). Making sense of customer service experiences: a text mining review. *Journal of Services Marketing*, 33(1), 88–103.
- Malhotra, N. K. (1981). A Scale to Measure Self-Concepts, Person Concepts, and Product Concepts. *Journal of Marketing Research*, 18(4), 456.
- Mantonakis, A., Roderio, P., Lesschaeve, I., & Hastie, R. (2009). Order in choice: Effects of serial position on preferences. *Psychological Science*, 20(11), 1309–1312.
- Marchand, A., Hennig-Thurau, T., & Wiertz, C. (2017). Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. *International Journal of Research in Marketing*, 34(2), 336–354.
- Martínez Salinas, E., & Pina Pérez, J. M. (2009). Modeling the brand extensions' influence on brand image. *Journal of Business Research*, 62(1), 50–60.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.
- Moe, W. W., & Trusov, M. (2011). The Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing Research*, 48(3), 444–456.

- Mohammad, S., & Turney, P. (2010). Emotions Evoked by Common Words and Phrases. *Proceedings of the NAACL-HLT, June*, 26–34.
- Moon, S., & Kamakura, W. A. (2017). A picture is worth a thousand words: Translating product reviews into a product positioning map. *International Journal of Research in Marketing*, 34(1), 265–285.
- Mudambi, S., & Schuff, D. (2010). What makes a Helpful online Review? *MIS Quarter*, 28(4), 695–704.
- Munch, J. M., & Swasy, J. L. (1988). Rhetorical question, summarization frequency, and argument strength effects on recall. *Journal of Consumer Research*, 15(1), 69–76.
- Murphy, J., Hofacker, C., & Mizerski, R. (2006). Primacy and recency effects on clicking behavior. *Journal of Computer-Mediated Communication*, 11(2), 173–192.
- Nabi, R. L. (2003). Exploring the framing effects of emotion: Do discrete emotions differentially influence information accessibility, information seeking, and policy preference? *Communication Research*, 30(2), 224–247.
- Nazlan, N. H., Tanford, S., & Montgomery, R. (2018). The effect of availability heuristics in online consumer reviews. *Journal of Consumer Behaviour*, 17(5), 449–460.
- Nelson, P. (1970). *Information and Consumer Behavior*. 78(2), 311–329.
- Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine Your Own Business: Market-Structure Surveillance Through Text Mining. *Marketing Science*, 31(3), 521–543.
- Netzer, Oded, Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521–543.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Personality and Social Psychology Bulletin Lying Words : Predicting Deception From Linguistic Styles. *Personality and Social Psychology Bulletin*, 29(5), 665–675.
- Ngo-Ye, T. L., & Sinha, A. P. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, 61(1), 47–58.
- Nielsen. (2016). *The sweet smell of seasonal success*. Available at: <https://www.nielsen.com/au/en/insights/news/2016/the-sweet-smell-of-seasonal-success.html>
- Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *CEUR Workshop Proceedings*, 718, 93–98.
- Olson, J. C., & Muderrisoglu, A. (1979). The stability of responses obtained by free elicitation: Implications for measuring attribute salience and memory structure. *Advances in Consumer Research*, 6, 269–275.
- Pan, B., Zhang, L., & Law, R. (2013). The Complex Matter of Online Hotel Choice. *Cornell Hospitality Quarterly*, 41(1), 74–83.
- Pan, Y., & Zhang, J. Q. (2011). Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews. *Journal of Retailing*, 87(4), 598–612.
- Pang, J., & Qiu, L. (2016). Effect of online review chunking on product attitude: The moderating role of motivation to think. *International Journal of Electronic Commerce*, 20(3), 355–383.

- Park, D. H., & Lee, J. (2008). eWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications*, 7(4), 386–398.
- Park, E. J., Kim, E. Y., Funches, V. M., & Foxx, W. (2012). Apparel product attributes, web browsing, and e-impulse buying on shopping websites. *Journal of Business Research*, 65(11), 1583–1589.
- Park, H.-J., & Rabolt J., N. (2009). Cultural Value, Consumption Value, and Global Brand Image: A Cross-National Study. *Psychology & Marketing*, 26(8), 714–735.
- Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research*, 50, 67–83.
- Payan, J. M., & McFarland, R. G. (2005). Decomposing Influence Strategies: Argument Structure and Dependence as Determinants of the Effectiveness of Influence Strategies in Gaining Channel Member Compliance. *Journal of Marketing*, 69(3), 66–79.
- Payne, J. W. (1982). Contingent Decision Behavior. *Psychological Bulletin*, 92(2), 382.
- Peladeau, N. (2016). *WordStat: Content Analysis Module for SIM- STAT*.
- Pennebaker, J. W., Boyd, R., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. *Austin, TX: University of Texas at Austin, SEPTEMBER 2015*, 1–22.
- Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS ONE*, 9(12), 1–10.
- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The Development and Psychometric Properties of LIWC2007. *Austin, Tx: LIWC.Net*, 1(2).
- Petty, R. E., Tormala, Z. L., Hawkins, C., & Wegener, D. T. (2001). Motivation to think and order effects in persuasion: The moderating role of chunking. *Personality and Social Psychology Bulletin*, 27(3), 332–344.
- Pindado, J., Requejo, I., & de la Torre, C. (2011). Family control and investment-cash flow sensitivity: Empirical evidence from the Euro zone. *Journal of Corporate Finance*, 17(5), 1389–1409.
- Price, P. C., & Stone, E. R. (2004). Intuitive Evaluation of Likelihood Judgment Producers: Evidence for a Confidence Heuristic. *Journal of Behavioral Decision Making*, 17(1), 39–57.
- Puranam, D., Narayan, V., & Kadiyali, V. (2017). The effect of calorie posting regulation on consumer opinion: A flexible latent dirichlet allocation model with informative priors. *Marketing Science*, 36(5), 726–746.
- Purnawirawan, N., De Pelsmacker, P., & Dens, N. (2012). Balance and Sequence in Online Reviews: How Perceived Usefulness Affects Attitudes and Intentions. *Journal of Interactive Marketing*, 26(4), 244–255.
- Purnawirawan, N., Dens, N., & De Pelsmacker, P. (2012). Balance and Sequence in Online Reviews: The Wrap Effect. *International Journal of Electronic Commerce*, 17(2), 71–98.
- Racherla, P., & Friske, W. (2012). Perceived “usefulness” of online consumer reviews: An exploratory investigation across three services categories. *Electronic Commerce*

- Research and Applications*, 11(6), 548–559.
- Reynolds, T. J., & Gutman, J. (1979). Laddering theory, method, analysis, and interpretation. *Journal of Advertising Research*, 28(1), 11–31.
- Rodríguez-Díaz, M., Rodríguez-Díaz, R., & Espino-Rodríguez, T. (2018). Analysis of the Online Reputation Based on Customer Ratings of Lodgings in Tourism Destinations. *Administrative Sciences*, 8(3), 51.
- Roedder, D. J., Loken, B., Kim, K., & Monga, A. B. (2006). Brand concept maps: A methodology for identifying brand association networks. *Journal of Marketing Research*, 43(4), 549–563.
- Roodman, D. (2009a). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, 9(1), 86–136.
- Roodman, D. (2009b). Practitioners' corner: A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71(1), 135–158.
- Salehan, M., & Kim, D. J. (2015). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, 81, 30–40.
- Saumya, S., Singh, J. P., Baabdullah, A. M., Rana, N. P., & Dwivedi, Y. K. (2018). Ranking online consumer reviews. *Electronic Commerce Research and Applications*, 29(March), 78–89.
- Schnittka, O., Sattler, H., & Zenker, S. (2012). Advanced brand concept maps: A new approach for evaluating the favorability of brand association networks. *International Journal of Research in Marketing*, 29(3), 265–274.
- Selkirk, E. (1996). The prosodic structure of function words. In J. L. (Brown U. Morgan & K. (Brown U. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 187–214). Lawrence Erlbaum Associates, Inc.
- Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80(2), 159–169.
- Sheldon, P., & Bryant, K. (2016). Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in Human Behavior*, 58, 89–97.
- Shugan, S. M. (1980). The Cost of Thinking. *Journal of Consumer Research*, 7(2), 99–111.
- Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Kumar Roy, P. (2017). Predicting the “helpfulness” of online consumer reviews. *Journal of Business Research*, 70, 346–355.
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, 105(1), 131–142.
- Slovic, P. (1972). From Shakespeare to Simon: Speculations and some evidence. *Oregon Research Institute Bulletin*, 1–19.
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality & Social Psychology*, 48(4), 813–838.
- Snizek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, 84(2), 288–307.
- Social Blade*. (2017). <https://socialblade.com/>

- Spears, N., & Singh, S. N. (2004). Measuring attitude toward the brand and purchase intentions. *Journal of Current Issues and Research in Advertising*, 26(2), 53–66.
- Spenner, P., & Freeman, K. (2012). To keep your customers, keep it simple. *Harvard Business Review*, 90(5).
- SproutSocial. (2018). *Instagram stats*. Available at: <https://sproutsocial.com/insights/instagram-stats/>
- Statista. (2020). Available at: <https://www.statista.com/forecasts/997051/sources-of-information-about-products-in-the-us>
- Sun, M. (2012). How Does the Variance of Product Ratings Matter? *Management Science*, 58(4), 696–707.
- Sun, Q., Niu, J., Yao, Z., & Yan, H. (2019). Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level. *Engineering Applications of Artificial Intelligence*, 81(December 2018), 68–78.
- Tang, T. (Ya), Fang, E. (Er), & Wang, F. (2014). Is neutral really neutral? The effects of neutral user-generated content on product sales. *Journal of Marketing*, 78(4), 41–58.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
- Tirunillai, S., & Tellis, G. J. (2014). Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation. *Journal of Marketing Research*, 51(4), 463–479.
- Toubia, O., Iyengar, G., Bunnell, R., & Lemaire, A. (2019). Extracting Features of Entertainment Products: A Guided Latent Dirichlet Allocation Approach Informed by the Psychology of Media Consumption. *Journal of Marketing Research*, 56(1), 18–36.
- Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing*, 73(5), 90–102.
- Tybout, A. M., Sternthal, B., Malaviya, P., Bakamitsos, G. A., & Park, S. E. B. (2005). Information accessibility as a moderator of judgments: The role of content versus retrieval ease. *Journal of Consumer Research*, 32(1), 76–85.
- Van Hove, G., & Lievens, F. (2007). Social influences on organizational attractiveness: Investigating if and when word of mouth matters. *Journal of Applied Social Psychology*, 37(9), 2024–2047.
- Veloutsou, C., & Delgado-Ballester, E. (2018). New challenges in brand management. *Spanish Journal of Marketing - ESIC*, 22(3), 255–272.
- Vermeer, S. A. M., Araujo, T., Bernritter, S. F., & van Noort, G. (2019). Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media. *International Journal of Research in Marketing*, 36(3), 492–508.
- Vriens, M., Chen, S., & Schomaker, J. (2019). The evaluation of a brand association density metric. *Journal of Product and Brand Management*, 28(1), 104–116.
- Walker, M. A., Mehl, M. R., Moore, R. K., & Mairesse, F. (2007). Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of*

- Artificial Intelligence Research*, 30, 457–500.
- Wang, F., Liu, X., & Fang, E. (Er). (2015). User Reviews Variance, Critic Reviews Variance, and Product Sales: An Exploration of Customer Breadth and Depth Effects. *Journal of Retailing*, 91(3), 372–389.
- Wang, W., Feng, Y., & Dai, W. (2018). Topic analysis of online reviews for two competitive products using latent Dirichlet allocation. *Electronic Commerce Research and Applications*, 29(January), 142–156.
- Wang, X., Tang, L. (Rebecca), & Kim, E. (2019). More than words: Do emotional content and linguistic style matching matter on restaurant review helpfulness? *International Journal of Hospitality Management*, 77(August 2018), 438–447.
- Weathers, D., Swain, S. D., & Grover, V. (2015). Can online product reviews be more helpful? Examining characteristics of information content by product type. *Decision Support Systems*.
- Westerman, D., Spence, P. R., & Van Der Heide, B. (2014). Social Media as Information Source: Recency of Updates and Credibility of Information. *Journal of Computer-Mediated Communication*, 19(2), 171–183.
- Wickham, H. (2016). *tidyr: Easily tidy data with spread and gather functions*.
- Wickham, H. (2017). ggplot2 - Elegant Graphics for Data Analysis (2nd Edition). *Journal of Statistical Software*, 77(Book Review 2), 2–5.
- Wickham, H., & Francois, R. (2016). *dplyr: A Grammar of Data Manipulation*.
- Winkelmann, P. D. R. (2008). Econometric Analysis of Count Data. In *Econometric Analysis of Count Data*.
- Wong, C. U. I., & Qi, S. (2017). Tracking the evolution of a destination's image by text-mining online reviews - the case of Macau. *Tourism Management Perspectives*, 23, 19–29.
- Wooldridge, J. M. (2011). *Econometric analysis of cross section and panel data* (Issue November).
- Wu, J. (2017). Review popularity and review helpfulness: A model for user review effectiveness. *Decision Support Systems*, 97, 92–103.
- Xu, P., & Liu, D. (2019). Product engagement and identity signaling: The role of likes in social commerce for fashion products. *Information and Management*, 56(2), 143–154.
- Yang, H., & Cho, S. (2015). *Understanding Brands with Visualization and Keywords from eWOM Using Distributed Representation*.
- Yau, K. K. W., Wang, K., & Lee, A. H. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal*, 45(4), 437–452.
- Yazdani, E., Gopinath, S., & Carson, S. (2018). Preaching to the Choir: The Chasm Between Top-Ranked Reviewers, Mainstream Customers, and Product Sales. *Marketing Science*, 37(5), 838–851.
- Yin, G., Wei, L., Xu, W., & Chen, M. (2014). Yin, G., Wei, L., Xu, W., & Chen, M. (2014). Exploring Heuristic Cues for Consumer Perceptions of Online Reviews Helpfulness: The Case of Yelp.com. PACIS 2014 Proceedings, 1–12. Exploring Heuristic Cues for Consumer Perceptions of Online Reviews Helpfulness. *PACIS 2014 Proceedings*, 1–12.

- Zaltman, G. (1997). Rethinking Market Research : Putting People Back In. *Journal of Marketing Research*, 34(4), 424–437.
- Zhang, J. (2019). What's yours is mine: exploring customer voice on Airbnb using text-mining approaches. *Journal of Consumer Marketing*, 36(5), 655–665.
- Zhang, Ying, & Lin, Z. (2018). Predicting the helpfulness of online product reviews: A multilingual approach. *Electronic Commerce Research and Applications*, 27, 1–10.
- Zhang, Yuchi, Moe, W. W., & Schweidel, D. A. (2017). Modeling the role of message content and influencers in social media rebroadcasting. *International Journal of Research in Marketing*, 34(1), 100–119.
- Zhou, S., & Guo, B. (2017). The order effect on online review helpfulness: A social influence perspective. *Decision Support Systems*, 93, 77–87.
- Zhu, F., & Zhang, X. (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*, 74(March), 133–148.