

E.T.S. de Ingeniería Industrial, Informática
y de Telecomunicación

Business Intelligence y el análisis predictivo: COVID 19



Grado en Ingeniería Informática

Trabajo Fin de Grado

Javier López Navas

Jesús Villadangos Alonso y Diego Pérez Rández

Pamplona, 2 de octubre del 2020

upna

Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Índice

Resumen del trabajo	3
Palabras clave.....	3
Business Intelligence	4
¿Qué es el Business Intelligence?	4
¿Para qué sirve el Business Intelligence?	4
¿Por qué interesa el Business Intelligence?	5
Origen.....	5
Herramientas disponibles para el desarrollo del Business Intelligence.....	6
Estructura de un proyecto de Business Intelligence	7
Valortic Servicios Informáticos S.L.	9
Análisis Predictivo	11
¿Qué es el análisis predictivo?	11
¿Cómo se evoluciona al análisis predictivo desde el Business Intelligence?	11
¿Cómo funciona el análisis predictivo?.....	11
Árboles de decisión:	12
Regresión lineal y logística:	13
Motivos por los que usar el análisis predictivo	14
Herramientas para el análisis predictivo en Power BI	15
Forecasting	15
Clustering	15
Árbol de decisión.....	16
Predicción con Redes Neuronales.....	16
Predicción con ARIMA.....	17
Predicción con TBATS.....	17
Desarrollo del proyecto.....	18
Descripción del proyecto y del caso de estudio: COVID-19	18
COVID-19	18
Planificación del proyecto	19
Datos usados para el proyecto.....	19
Microsoft Power BI.....	20
Proyecto de estudio de la pandemia	21
Orígenes de datos	21
Inteligencia de negocio	26

Modelaje y relación de las consultas	38
Graficación y desarrollo del análisis	41
Proyecto de estudio de la economía durante la pandemia del COVID-19.....	56
Orígenes de datos	56
Inteligencia de negocio	60
Modelaje y relación de las consultas	66
Graficación y desarrollo del análisis	67
Proyecto de estudio de la salud durante la pandemia del COVID-19	75
Orígenes de datos	76
Inteligencia de negocio	83
Modelaje y relación de las consultas	91
Graficación y desarrollo del análisis	91
Análisis Predictivo	95
Predicción de Power BI	95
Predicción mediante ARIMA	96
Predicción con redes neuronales	97
Comparación predicciones.....	98
Conclusiones del TFG	99
Bibliografía	100

Resumen del trabajo

En este trabajo se va a realizar un estudio de BI (Business Intelligence) del COVID-19 por medio de la herramienta Power BI de Microsoft.

Gracias al BI podemos transformar el dato en conocimiento por medio de distintas transformaciones que se aplican durante todo el proyecto. Partiendo de la recogida del dato desde distintas fuentes se procede con la normalización del mismo. Esta normalización nos permite tener estructurado y limpios los datos y que no existan erratas que dificulten su visualización. Una vez el dato queda listo, este se carga al programa para poder comenzar con la graficación del mismo.

Es importante preguntarse qué se desea ver para poder así escoger qué herramienta o gráfico se puede ajustar más a la idea que buscamos representar. Por medio de gráficos de líneas, mapas de calor, indicadores KPI y demás herramientas, podemos lograr explotar el dato a un muy alto nivel.

Para complementar este proyecto se ha decidido tratar el análisis predictivo de los datos por medio de distintas técnicas como las redes neuronales o procedimientos estadísticos como las series temporales. El análisis predictivo se encuentra en una fase un tanto madura del business intelligence y se pretende comprobar qué potencial puede tener.

Todo el proyecto se realizará centrándose en el COVID-19 ya que es un tema de actualidad con muchos datos de distintas fuentes.

Palabras clave

- Business Intelligence
- Inteligencia de negocio
- Análisis predictivo
- Power BI
- Microsoft
- COVID-19
- Coronavirus

Business Intelligence

¿Qué es el Business Intelligence?

Se conoce como Business Intelligence o Inteligencia de Negocio al proceso de transformación de datos en información de utilidad para las empresas

¿Para qué sirve el Business Intelligence?

Gracias al Business Intelligence, BI para abreviar, las empresas son más conscientes de sí mismas ya que son capaces de visualizar la información de una manera rápida, ágil y muy visual.

Con distintas herramientas que aplican este proceso se puede lograr a partir de información recogida por la empresa en distintas fuentes (bases de datos, hojas de cálculo, etc.) graficarla para poder realizar un análisis con mayor comodidad y rapidez.

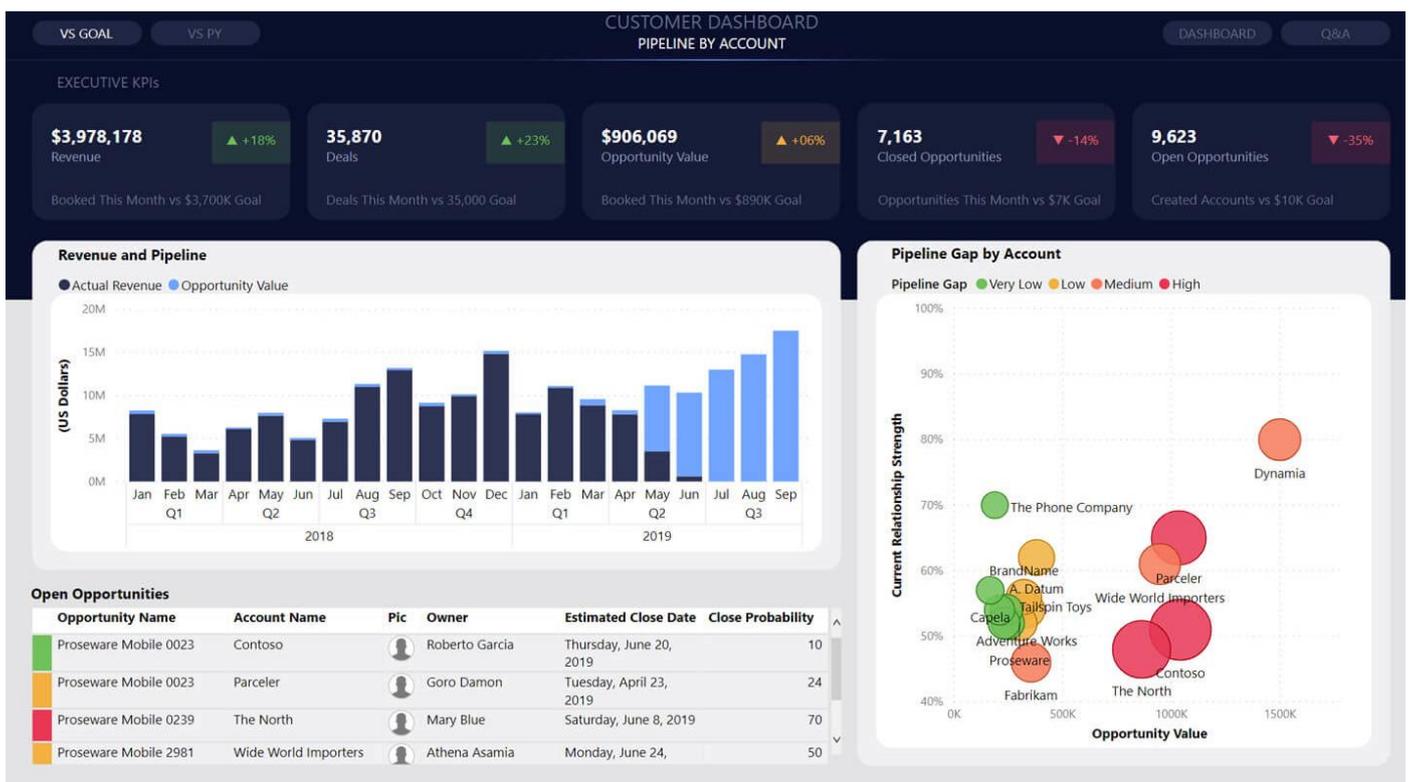


Figura 1. Ejemplo de proyecto de BI

¿Por qué interesa el Business Intelligence?

Ya hemos visto las ventajas que tiene el Business Intelligence, pero ¿en qué se traduce el aplicarlo? ¿mejora nuestro negocio drásticamente? ¿es para todo el mundo? ¿qué podemos esperar si decidimos aplicar Business Intelligence a nuestro negocio?

En primer lugar, cabe decir que el Business Intelligence es interesante para cualquier empresa. Su uso nos aporta grandes ventajas que de otra forma serían más difíciles de lograr.

El control de stock es una de las primeras ventajas que surgen cuando se aplica Business Intelligence. Muchas veces las empresas llevan el control del estocaje mediante programas de ERP si son empresas medianas o grandes, o si son más pequeñas mediante el uso de hojas de cálculo dinámicas.

El problema de estas medidas de control de stock es que no son fáciles la visualización del dato. Es cierto que controlan el stock, pero al analista le interesa más tener una forma rápida de analizar esa información almacenada.

En empresas con una marcada estacionalidad, como por ejemplo una compañía de helados que vende más en verano que en cualquier otra estación, a menudo tienen problemas para lograr optimizar adecuadamente su stock. Empresas con estas características que han aplicado distintas técnicas de Business Intelligence han logrado mejorar su rentabilidad alrededor de un 10%.

Otro gran beneficio del Business Intelligence es la detección de clientes fieles o más rentables de nuestro negocio. Para ello es necesario la captación de datos, los cuales se pueden lograr mediante tarjetas de socio, promociones, etc.

El presupuesto es controlable mediante Business Intelligence ya que podemos controlar las desviaciones y sus causas. También se pueden comparar con históricos de otros años y lograr ver una evolución.

Para las empresas pequeñas puede ser de utilidad ya que se pueden recoger datos de horas de ventas y conocer aquellas en las que más se vende. También para analizar los retrasos de pedidos y sus motivos de demora.

El Business Intelligence tiene un sinnúmero de aplicaciones empresariales que se traducen en un mayor control de nuestro negocio, el cual puede ser transformado en un mayor rendimiento en nuestro negocio si se realizan las medidas y acciones necesarias. El Business Intelligence no viene a revolucionar un negocio que no está dispuesto a cambiar; es una herramienta muy útil para aquel que busca mejorar su negocio.

Origen

El término “Business Intelligence” tiene su origen en un libro de Richard Millard Devens del año 1865, antes del comienzo de lo que actualmente consideramos tecnología. Con este

término se refería a la inteligencia del empresario Sir Henry Furnese, el cual utilizaba la información sobre su entorno para lucrarse y aumentar su negocio.

Con el tiempo, el término fue evolucionando hasta la era tecnológica y en 1958, año en el que Hans Peter Luhn (informático alemán que trabajó para la multinacional IBM) lo definió como “la habilidad para aprehender la interrelación de hechos presentados de tal manera que guían acción hacia un objetivo deseado”. Luhn entendía el Business Intelligence como la habilidad de ser conscientes y con presencia en la mente de todos aquellos datos e información de utilidad para lograr el objetivo propuesto.



Figura 2. Hans Peter Luhn

Hoy en día nos encontramos en la era del Big Data y esto ha supuesto una revolución para los pequeños y medianos empresarios que ahora tienen acceso a la inteligencia de negocio gracias a una más avanzada tecnología. Sir Henry Furnese pudo tejer una red de información gracias a ser una persona de poder e influencia. La inteligencia de negocio era algo reservado a la élite muy lejos del alcance de la persona promedio. Esto ha cambiado drásticamente y ahora cualquier empresario puede hacer suya esta herramienta y beneficiarse de todas sus ventajas.

Herramientas disponibles para el desarrollo del Business Intelligence

Para poder desarrollar y aplicar Business Intelligence podemos usar múltiples herramientas ya existentes muy conocidas y usadas por la industria. Las herramientas de Business Intelligence están pensadas para que la extracción de datos y graficación sean lo más sencillas posibles.

Las distintas herramientas pueden estar pensadas para un público más general o concreto, dependiendo de con qué finalidad se crearon.

Por parte del gigante tecnológico IBM contamos con IBM Cognos, su propia herramienta de Business Intelligence cuya mayor cualidad es que sus análisis están respaldados por inteligencia artificial mediante la detección de patrones en los datos de la empresa en segundo plano y ajeno al usuario. Esta inteligencia artificial puede preguntar al usuario cuestiones acerca de la información

El famoso ERP SAP también tiene sus propias herramientas para desarrollar Business Intelligence. Ellos lo llaman “Suite SAP BusinessObjects Business Intelligence”. Esta herramienta es muy interesante para empresas que ya hagan uso de este ERP por motivos de compatibilidades, rapidez y unidad. La solución de SAP incluye análisis predictivo (tema que trataremos más adelante en este proyecto) y aprendizaje automático mediante machine learning.

Por parte de Oracle, la compañía especializada en el desarrollo de soluciones de nube y locales como Oracle Database o JDeveloper, se ha desarrollado la aplicación Oracle Business Intelligence, una potente aplicación que ofrece múltiples funcionalidades al usuario y que se caracteriza por el descubrimiento de patrones ocultos que logra resultados únicos. Esta herramienta es muy interesante para aquellas empresas que necesiten tratar con una cantidad de volumen muy alta, ya que Oracle Business Intelligence es una herramienta muy robusta.

Para el desarrollo del Business Intelligence en empresas que usan Excel de normal existe una herramienta muy interesante llamada Clear Analytics, la cual no requiere de amplios conocimientos. Clear Analytics es muy utilizada en pequeñas y medianas empresas que buscan soluciones más ágiles que las que pueden ofrecer herramientas como las previamente comentadas.

La última herramienta que vamos a analizar se trata de Microsoft Power BI, la solución de Business Intelligence de Microsoft para el análisis empresarial. Se trata de una herramienta que permite conectarse a datos de diversas fuentes, transformarlos para la visualización del dato y posteriormente su difusión con toda la organización o inserción en sitio web o aplicación. Microsoft Power BI va a ser la herramienta con la que se va a trabajar en este proyecto por esto y otras cualidades que la han hecho idónea para ser la elegida, las cuales se explicarán más adelante.

Estructura de un proyecto de Business Intelligence

Para poder realizar proyecto de Business Intelligence es necesario realizar una serie de pasos previos comunes a todos los proyectos donde se busque realizar este tipo de análisis. Esta estructuración va a seguir un esquema aplicable a cualquier proyecto realizado con Microsoft Power BI.

1. Definir el proyecto

Se trata del primer paso a realizar. Consta de una definición de objetivos claros a lograr con el proyecto, definición de herramientas, medios y demás necesidades que se contemplen para la correcta realización del proyecto. Es de especial importancia la determinación de las fuentes de datos para utilizar en el proyecto, ya que dependiendo de la herramienta con la que se haga el futuro análisis (si es que se usa una herramienta, ya que se puede realizar a mano)

2. Proceso ETL

ETL se denomina al proceso de extracción, transformación y carga (load en inglés, de ahí las siglas) de datos.

Este proceso es sumamente importante ya que es la estructura del proyecto. Un mal proceso de ETL puede dificultar severamente o incluso imposibilitar el desarrollo del proyecto.

- Extracción:

- El proceso de extracción de datos corresponde la extracción de estos desde sus distintas fuentes y una primera comprobación que aseguren que se trate de los datos que se quiere analizar
- Transformación:
 - La transformación es la fase más importante de todo el proceso ETL. Su función es la de transformar y normalizar los datos para que todos cumplan un cierto esquema que permita posteriormente realizar su análisis.
 - Se trata de una parte realmente complicada de hacer en algunos casos, ya sea porque la limpieza de datos que hay que hacer es muy grande, por la preparación de las estructuras o tablas o por cualquier otro motivo.
 - Suele ser habitual encontrar datos comunes en distintas tablas como por ejemplo países. Si en una tabla se encuentra como entrada de país “España” y en otra tabla con la columna de país se encuentra “Spain”, ambos se refieren al mismo país, pero no se pueden cruzar de una manera directa. Como consecuencia, se tienen que preparar una tabla de cruce para poder asociar estos países.
 - La creación de tablas o estructuras también puede conllevar gran dificultad como podría darse en el caso de tener que crear una lista de clientes donde cada provincia tiene sus propios clientes dentro de hojas distintas. Esto obliga a juntar distintas hojas hasta poder crear una tabla común a todas las provincias, con todas las dificultades que esto pueda traer al desarrollo del proyecto.
 - En Microsoft Power BI, como ya veremos más adelante, la transformación se realiza mediante el editor de consultas en el lenguaje Power Query, informalmente conocido como lenguaje M.
- Carga:
 - La carga de datos consiste en, como su nombre indica, la carga de datos en el sistema. Una vez que ya se han construido las tablas en el proceso de transformación, se suben al sistema para posteriormente poder relacionarlas.

3. Modelaje

El modelado de las tablas, como es en el caso de Microsoft Power BI, exige tener que relacionar las tablas entre sí como si de una base de datos se tratara. Para ello debemos elegir qué campos deben relacionarse entre ellos, estableciendo sus multiplicidades (es decir, el número de veces que aparecen en una tabla respecto a la otra) y estableciendo una dirección (la cual nos permite dirigir desde dónde se pueden acceder a esos datos).

4. Formulación

La formulación permite crear medidas que se usarán a la hora de utilizar los gráficos. En Power BI esta formulación se realiza mediante el lenguaje DAX. Este tipo de medidas permite obtener formulas muy precisas que se ajusten a lo que realmente deseamos.

5. Graficación

La graficación consiste en la creación de gráficos y demás elementos visuales que componen un buen análisis de Business Intelligence. Estos gráficos deben ser capaces de mostrar información precisa respecto a los datos obtenidos y, sobre todo, deben de ser de utilidad para el usuario final. Se deben elegir diferentes tipos de elementos dependiendo de lo que se quiera lograr y estructurarlos en distintas pestañas de acuerdo con mantener una mayor organización y estructuración en el proyecto.

6. Publicación y difusión

La publicación y difusión de proyectos realizados en Power BI se podrá realizar de manera online contratando el servicio de Power BI Pro, el cual no solo permite acceder a estos proyectos de manera online si no también embeberlo en otras páginas web

Valortic Servicios Informáticos S.L.

Valortic es una consultora tecnológica especializada en ciencia de datos para el desarrollo del conocimiento de negocio y de capacidades directivas. Se trata de la empresa en la que yo, Javier López Navas, me encuentro realizando prácticas y la cual colabora conmigo en el desarrollo de este proyecto.

Actualmente el Business Intelligence es la rama más importante y en mayor desarrollo de la empresa. La herramienta que se usa en Valortic para el desarrollo del Business Intelligence en la actualidad es Microsoft Power BI.

Esto se debe a que en Valortic se trabaja en el entorno Office 365 y ofrecen soluciones que incorporen las distintas herramientas de Microsoft. Estas otras herramientas son posteriormente enlazables con Microsoft Power BI, lo cual unifica todo un proceso de una manera cómoda y fácil.

Desde Microsoft Power BI se puede hacer una recogida desde varios orígenes como puede ser una base de datos de Access, Excel o soluciones más dinámicas como puede ser una lista de SharePoint que a su vez se alimenta de una Power App (aplicación para móvil de Microsoft diseñada a medida del cliente).

Además de todo esto, Microsoft Power BI es de licencia gratuita, lo que la hace atractiva respecto al resto de herramientas ya que no se encuentra recortada en funcionalidades respecto a sus competidores, los cuales tienen unas tarifas altas.

Microsoft Power BI cuenta con dos planes de pago, los cuales siguen siendo realmente baratos en comparación con la competencia, que en general incorporan la publicación, colaboración y publicación en la nube.

Desde Valortic se busca desarrollar en más profundidad el Business Intelligence y el siguiente paso es el análisis predictivo a partir de los datos. Con este trabajo se pretende estudiar las posibilidades del Business Intelligence y el potencial que posee el análisis predictivo para las empresas.



Figura 3. Logotipo de Valortic

Análisis Predictivo

¿Qué es el análisis predictivo?

El análisis predictivo es un grupo de técnicas estadísticas que forman parte de la analítica avanzada mediante las cuales se busca predecir sucesos futuros a partir de sucesos pasados.

Gracias al análisis predictivo se logra no solo predecir futuros sucesos si no que se abre la posibilidad de mejorar procesos actuales influyentes en dichos sucesos para mejorarlos.

El análisis predictivo no es adivinatorio ya que se encuentra respaldado por múltiples técnicas estadísticas. La modelización, aprendizaje automático y minería de datos son algunas de las técnicas encargadas de otorgar cierto grado de confianza a sus análisis.

¿Cómo se evoluciona al análisis predictivo desde el Business Intelligence?

El análisis predictivo se puede ver como una evolución del Business Intelligence; como el siguiente paso a dar para avanzar en el mundo del análisis empresarial.

En primera instancia surgió el Business Intelligence, que permitió a sus usuarios ser verdaderamente conscientes de su empresa. Les permitió controlar todos los aspectos de su negocio, saber qué iba mal, qué iba bien, qué era rentable y qué no. Con el tiempo, el Business Intelligence se ha hecho hueco en la vida del empresario moderno y se ha asentado para quedarse. Ahora, dicho empresario ya no se hace las preguntas del estilo: ¿Cómo puedo conocer mejor mi negocio? ¿Qué me está yendo bien o mal? Ahora, el empresario ya tiene la respuesta a esas preguntas, por lo que lo lógico es mirar al futuro. Es en el futuro donde entra en juego el análisis predictivo. Se trata de la respuesta a la pregunta ¿Qué va a pasar? Con el análisis predictivo se puede predecir sucesos con una fiabilidad mayor a la que antaño se tenía, todo gracias a sus avanzadas técnicas estadísticas.

¿Cómo funciona el análisis predictivo?

El análisis predictivo hace uso de distintos modelos estadísticos para su funcionamiento. Existen tres grandes modelos con los que el análisis predictivo trabaja: los modelos descriptivos, los modelos prescriptivos y los modelos predictivos.

Los modelos descriptivos son los más usados por las empresas y permiten a partir de datos históricos y actuales que nos permiten identificar tendencias, aventurar qué decisión es la siguiente a tomar.

Los modelos prescriptivos es el conjunto de modelos más complicado de los tres, ya que analiza todos los elementos de una decisión y las variables que implican cada uno de ellos para

establecer la probabilidad de que cada uno se produzca. Son modelos muy usados en compañías en expansión ya que permiten un mayor control de las decisiones a tomar y sus consecuencias.

Los modelos predictivos se subdividen en otros dos modelos, modelos de clasificación o de regresión. Estos modelos analizan datos del pasado para determinar la probabilidad de que ocurra un evento determinado. Se usan en casos en los que se cuenta con objetivo específico y se desea conocer la probabilidad de que tenga lugar en las condiciones previstas.

Los modelos de clasificación permiten predecir cual va a ser la clase de un elemento. Este tipo de modelos suelen ser útiles en análisis predictivo para realizar análisis del estilo de “¿Qué cliente es más probable que vuelva a comprar en mi tienda?”. Estos modelos requieren varias variables de entrada, en nuestro caso podrían ser importe de factura, cantidad de productos comprados, encuestas de satisfacción rellenas, etc. La salida de los modelos de clasificación comprende dos resultados posibles: 0 o 1. Estos valores significan si pertenecen o no a la clase estudiada.

Los modelos de regresión, la otra cara de la moneda del análisis predictivo, nos permiten predecir el valor de un elemento. Su utilidad quizá es más evidente, por ejemplo, podría realizarse una previsión del número de ventas que se va a realizar en una tienda en un determinado mes.

Existen múltiples técnicas de ambos modelos, pero se van a mencionar aquellas técnicas más usadas del modelado de análisis predictivo, el cual engloba los otros dos modelos previos.

Árboles de decisión:

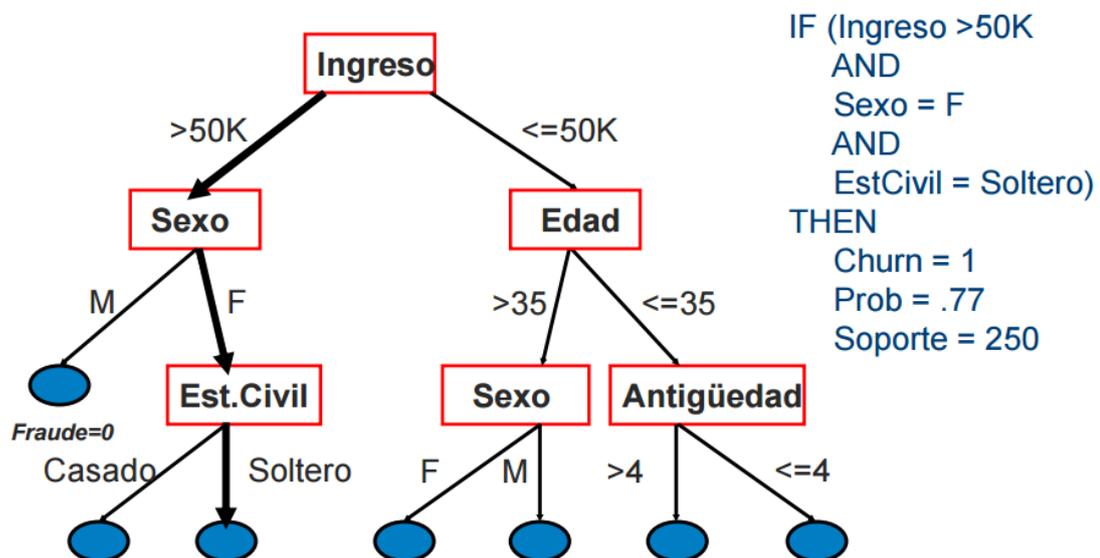


Figura 4. Ejemplo de árbol de decisión

Los árboles de decisión pertenecen a la categoría de modelos de clasificación ya que nos permiten, a partir de los parámetros de entradas, establecer un subconjunto de pertenencia del elemento estudiado.

Esta técnica se llama así por la forma semejante a un árbol, donde cada elección se asemeja a una rama y cada subconjunto de salida a una hoja.

El motivo de su gran uso es su facilidad de comprensión, interpretación, permiten controlar variables que faltan y es de gran utilidad para una selección preliminar de variables.

Regresión lineal y logística:

La regresión lineal es una relación estadística entre dos tipos de variables continuas. La variable de predicción puede ser una sola, en caso de regresión lineal simple, o más de una en caso de regresión lineal múltiple.

La variable de predicción, también conocida como variable independiente, se sitúa normalmente en el eje de abscisas (eje X) y la variable de resultado, o variable dependiente, en el eje de coordenadas (eje Y).

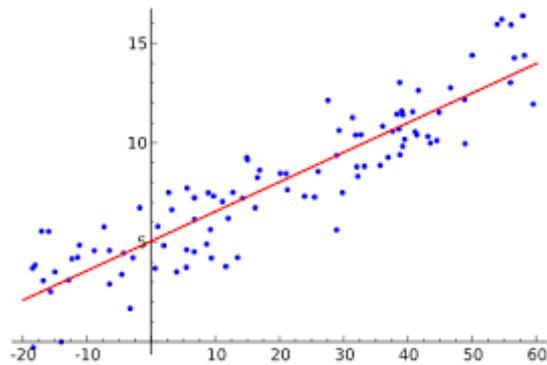


Figura 5. Regresión lineal simple

La regresión lineal pertenece al modelo de regresión, ya que se consigue predecir el valor (resultado representado en la variable del eje Y) respecto a una o unas entradas.

Es importante distinguir una relación estadística como es la regresión lineal de una regresión determinista. En la segunda, la relación entre dos variables está determinada por la otra. Esto se ve claro si tomamos como variables el lado y el volumen de un cubo. Se sabe que el volumen es lado elevado al cubo, por lo que sabiendo el volumen se puede saber el lado y viceversa. Gráficamente esto supondría una línea recta donde todos los resultados caen perfectamente en ella sin salirse. En una relación estadística esto no sucede de tal forma. Un ejemplo de ello podría ser la relación entre la altura y el peso de una persona. Se sabe que cuanto más alto es alguien, más pesado es, sin embargo, no existe una relación directa entre las dos y los resultados pueden variar entre distintos sujetos.

Multiple Regression → $y = m_1x_1 + m_2x_2 + b$

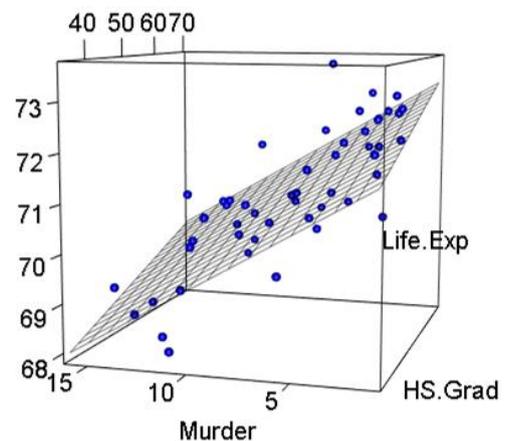


Figura 6. Regresión lineal múltiple

Motivos por los que usar el análisis predictivo

Como dijo Thomas A. Edison: “El valor de una idea radica en el uso de la misma”. Esto se traduce como el valor que posee el análisis predictivo depende de cómo o el para qué lo vayamos a usar.

Las posibilidades con el análisis predictivo son muy variadas, aunque podríamos centrarlas en dos:

1. Predicción de elementos:

La ventaja más evidente del análisis predictivo es la predicción de elementos a partir de datos pasados. Como hemos visto se puede predecir tanto futuros valores o futuras pertenencias a distintos subconjuntos. Se trata de una manera muy eficaz de prever el futuro y de esta manera prepararse para el mismo. El ejemplo claro y recurrente es el del control del estocaje, aunque no es el único: prever un descenso de las ventas, controlar la fidelidad de clientes, etc.

2. Toma de decisiones:

A diferencia de la anterior posibilidad mencionada, donde a partir de la previsión realizábamos modificaciones para prepararnos ante ella, también se puede invertir el proceso y, a partir de una previsión, realizar ciertas modificaciones hasta cambiar dicha previsión. Con esto se pueden evitar crisis o problemas en la empresa. El caso vendría a ser el siguiente: Se prevé cierto suceso que va a ocurrir según el rumbo actual de la empresa gracias al análisis predictivo, se analiza cual es el motivo o causa de esto y se realizan las modificaciones pertinentes para cambiar dicha previsión a una más favorable



Figura 7. Contenidos del análisis predictivo

Herramientas para el análisis predictivo en Power BI

Power BI posee múltiples herramientas para realizar análisis predictivo a partir de los datos. Originalmente Microsoft Power BI era una herramienta muy básica de análisis empresarial, pero con el paso del tiempo Microsoft ha ido añadiendo distintas funcionalidades a la herramienta que le han aportado gran robustez y le han permitido incorporar métodos para el análisis predictivo.

Forecasting

Se trata de la herramienta más conocida para el análisis predictivo de Power BI. Se trata de una función que permite realizar predicciones sobre series temporales. Normalmente es empleada en gráficos de ventas, ya que permite predecir el volumen de ventas venidero en el tiempo. Cuenta con múltiples parámetros de configuración como podrían ser la longitud de previsión (es decir, de cuánto tiempo queremos que realice la previsión), ignorar los últimos datos (en caso de no ser representativos), el intervalo de confianza, estacionalidad, etc.



Clustering

El Clustering es una función propia del aprendizaje no supervisado de Machine Learning que permite a partir de variables continuas realizar un Clustering o análisis de grupos. El Clustering nos permite agrupar elementos parejos en un mismo cluster o grupo, que sea distinto a cualquier otro.

La diferencia entre clasificación y agrupamiento, como es este caso, reside en el hecho de que en el primero conocemos el resultado o clase de antemano y en el agrupamiento no, de ahí el aprendizaje no supervisado propio de la técnica de Clustering.

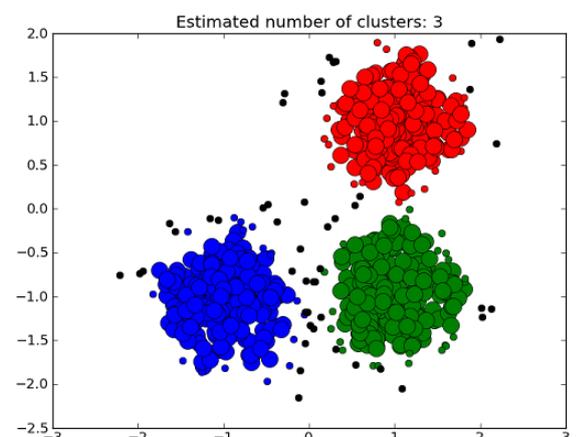


Figura 8. Ejemplo de Clustering

Árbol de decisión

Ya hemos visto los árboles de decisión como técnica dentro del modelo de clasificación del análisis predictivo previamente. En esta implementación se realiza una diferencia entre un árbol de clasificación y uno de regresión, que pertenecería al otro modelo dentro del análisis predictivo. La clave de esto es la variable objetivo: si es categórica o tiene pocos valores posibles entonces se trata de un árbol de clasificación, en caso de ser la variable objetivo de tipo numérico conseguiremos un árbol de regresión.

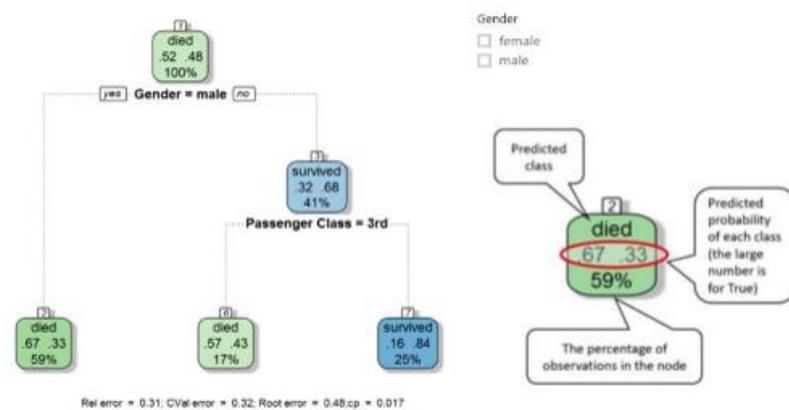


Figura 9. Ejemplo de árbol de decisión

Predicción con Redes Neuronales

Las redes neuronales son software desarrollado capaz de imitar el método de aprendizaje del cerebro humano, el cual lo consigue mediante las neuronas (de ahí su nombre). Estas redes neuronales están formadas de varias capas formadas a su vez por varias neuronas.

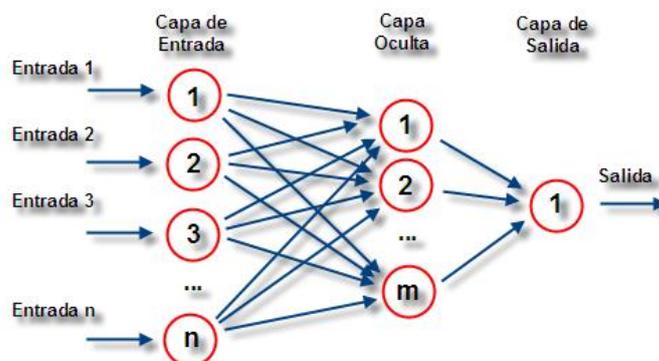


Figura 10. Ejemplo de red neuronal

En análisis predictivo se utilizan estas redes neuronales para aprender sobre datos históricos los cuales ayudan a predecir valores futuros.

Predicción con ARIMA

ARIMA son las siglas de Autoregressive Integrated Moving Average, es decir, modelo autorregresivo integrado de media móvil. Es un modelo estadístico que mediante regresión de los datos es capaz de encontrar patrones que ayuden a realizar predicciones en el futuro. Se trata de un modelo muy usados por los analistas de negocios ya que es capaz de predecir demanda, inventarios, presupuesto, etc.

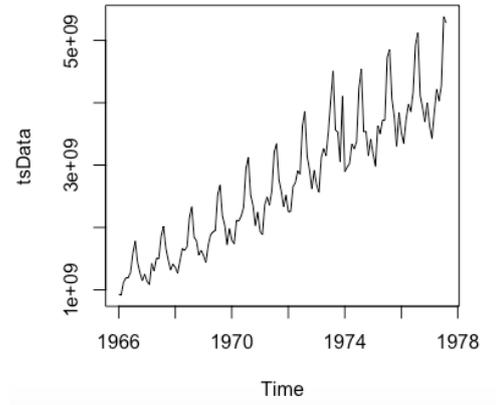


Figura 11. Ejemplo de ARIMA

Predicción con TBATS

El modelo encargado de detectar estacionalidades es sin lugar a duda el modelo TBATS (Trigonometric seasonality, Box-Cox transformation, Arma errors, Trend and Seasonal components). Las estacionalidades pueden ser un factor en algunos negocios y conseguir prever valores a partir de esta estacionalidad es de gran ayuda. TBATS actúa sobre una serie temporal en la cual detecta las estacionalidades existentes si las hubiera.

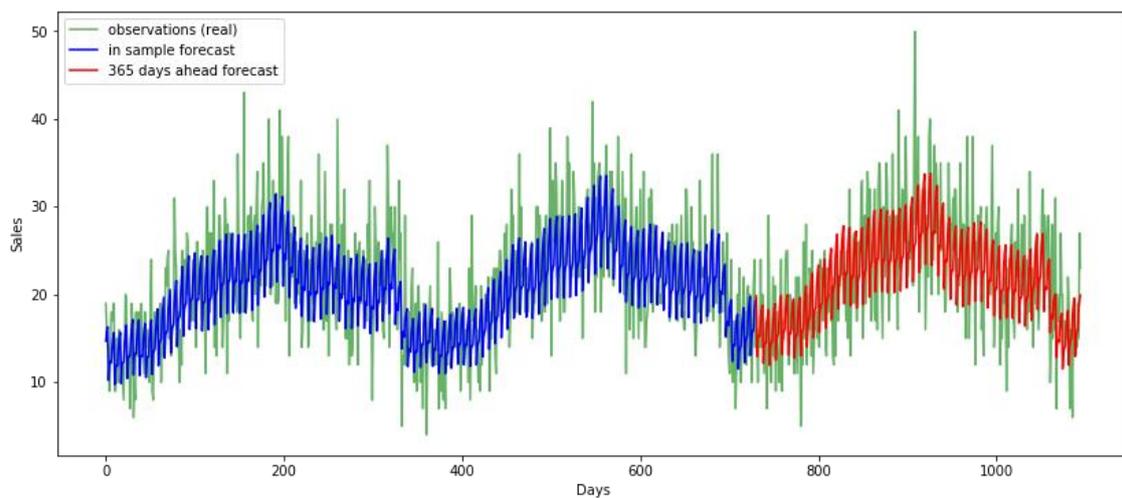


Figura 12. Ejemplo de predicción con TBATS

Desarrollo del proyecto

Descripción del proyecto y del caso de estudio: COVID-19

En este proyecto vamos a realizar un proyecto completo de Business Intelligence al cual vamos a incluirle un desarrollo de análisis predictivo.

El proyecto que se va a realizar tratará sobre la pandemia mundial que está azotando al mundo en este momento: el COVID-19, más informalmente conocido como el coronavirus. El tema se ha elegido por varios motivos: se trata de un tema de actualidad que concierne a todos los ciudadanos, se poseen múltiples datos de fuentes oficiales que actualizan de manera constante y por último y más importante, se puede realizar análisis predictivo de los datos y posteriormente comprobar cómo de acertado o de erróneo han sido dichas predicciones.

La herramienta utilizada para el análisis de datos y predictivo es Microsoft Power BI.

COVID-19

El COVID-19 es una nueva enfermedad provocada por el virus denominado "coronavirus", llamado así por su forma semejante a la de una corona vista desde arriba. Tiene su origen en la ciudad de Wuhan, China, y se transmitió en un primer momento de un animal a un humano. Se sospecha que originalmente la causa podría haber sido por comer murciélago, aunque se sospecha que podría haber sido un pangolín.

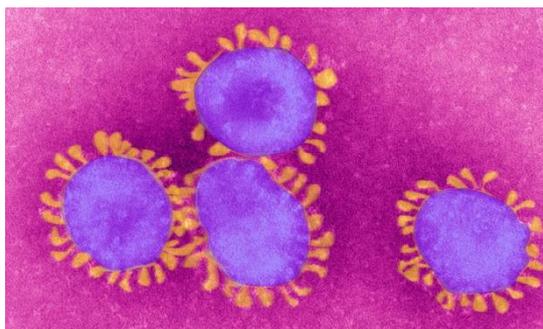


Figura 13. Fotografía del coronavirus

Actualmente casi 70 países se encuentran en cuarentena, lo que ha obligado al encierro de sus ciudadanos en sus hogares y la prohibición de salir de los mismo excepto por causas mayores, compra de alimentos, cuidado de mayores, etc. Las medidas pueden cambiar de un país a otro, así como el alcance de la cuarentena. En la ciudad de Wuhan se levantó la cuarentena después de 76 días el día 8 de abril de 2020 (comenzó el 23 de enero del mismo año).

Se prevé que lleguen distintas olas de infección en las que vuelvan a aumentar el número de casos de infección de manera exponencial. Esta segunda ola podría ser incluso peor que la primera como sucedió con la gripe española, que tuvo tres olas con la segunda como la más letal de las tres. Se estima que esta segunda ola sea en octubre debido a la llegada del otoño, lo que supone temperaturas más bajas (el virus es susceptible a temperaturas altas, aunque no lo elimina).

Se está estudiando la creación de vacunas con dos propósitos: disminuir los efectos dañinos causados en el cuerpo humano por el coronavirus (esta vacuna podría llegar más rápido) y otra que proteja del contagio a los ciudadanos.

Planificación del proyecto

Este proyecto se ha dividido en tres proyectos de Power BI separados. Todos ellos guardan relación con el estudio del COVID-19 pero en distintos ámbitos. Estos proyectos abordarán los siguientes temas:

- Proyecto estudio de la pandemia:
 - En este proyecto se pretende estudiar el efecto de la pandemia en los países a nivel de casos y muertes.
 - Se estudiarán distintos aspectos importantes como pueden ser la realización de tesis por cada país o el número de pacientes ingresados en la UCI.
 - Seguirá una evolución temporal del COVID-19 desde su aparición hasta que se han obtenido los últimos datos
- Proyecto de economía:
 - Se trata de un proyecto más enfocado al aspecto devastador causado por el COVID-19 a nivel económico.
 - Se realizará un estudio de las principales bolsas de valores del mundo y cómo han sido afectadas por la pandemia
 - Existirá un apartado que ayude a comparar las situaciones vividas en países del tercer mundo (o países con índices de pobreza altos) en comparación con países más desarrollados.
 - A nivel nacional se realizará un estudio del paro acompañado de las concesiones de ERES y ERTES.
 - Podremos ver cómo ha afectado esta pandemia a la desaparición de empresas en España, en específico de las PYMES, las cuales predominan en el territorio español.
- Proyecto de salud o de estudio de enfermedades:
 - Un proyecto en el que se tratará el COVID-19 como una enfermedad más, donde podremos compararlo con otras enfermedades recurrentes o importantes de los últimos tiempos.
 - Estudio a nivel nacional del efecto de la gripe española como anterior pandemia al coronavirus y su comparación con el COVID-19

Datos usados para el proyecto

Para el desarrollo de este proyecto se requiere una gran cantidad de datos. Estos datos se van a intentar recoger de fuentes oficiales para intentar asegurar que estos sean lo más reales posibles. En caso de no conseguir fuentes oficiales que proporcionen datos suficientes se recurrirá a otras fuentes no oficiales de gran reputación

Microsoft Power BI

Como se había anticipado, la herramienta elegida para el desarrollo del proyecto es Microsoft Power BI ya que desde Valortic se trabaja con dicha plataforma dado que se encuentra integrada en la suite de Office y es muy potente.

La descarga de la app de escritorio se puede realizar desde la propia store de Microsoft:



Figura 14. Página de la Store de Windows con Power BI

Los requisitos Hardware para poder ejecutar Power BI Desktop son los siguientes:

- Windows 7 y Windows Server 2008 R2 o posterior
- .NET 4.5
- Internet Explorer 9 o posterior
- Memoria (RAM): Al menos 1 GB disponible; se recomienda 1,5 GB o más.
- Pantalla: se recomienda al menos 1440 x 900 o 1600 x 900 (16:9). No se recomiendan las resoluciones inferiores a 1024 x 768 o 1280 x 800, ya que ciertos controles (por ejemplo, para cerrar la pantalla de inicio) solo se muestran en resoluciones superiores a esta.
- Configuración de pantalla de Windows: si la configuración de pantalla está establecida para cambiar el tamaño del texto, las aplicaciones u otros elementos en más de 100 %, quizás no vea ciertos cuadros de diálogo que deben cerrarse o en los que debe actuar para continuar usando Power BI Desktop. Si se produce este problema, compruebe la configuración de su pantalla en Configuración > Sistema > Pantalla en Windows, y use el control deslizante para devolver la configuración de pantalla al 100 %.
- CPU: 1 gigahercio (GHz) o superior; se recomienda un procesador de x86 o x64 bits.

Una vez se ejecute la aplicación nos encontraremos ante la interfaz de Power BI, semejante a la del resto de aplicaciones de la suite de Office:

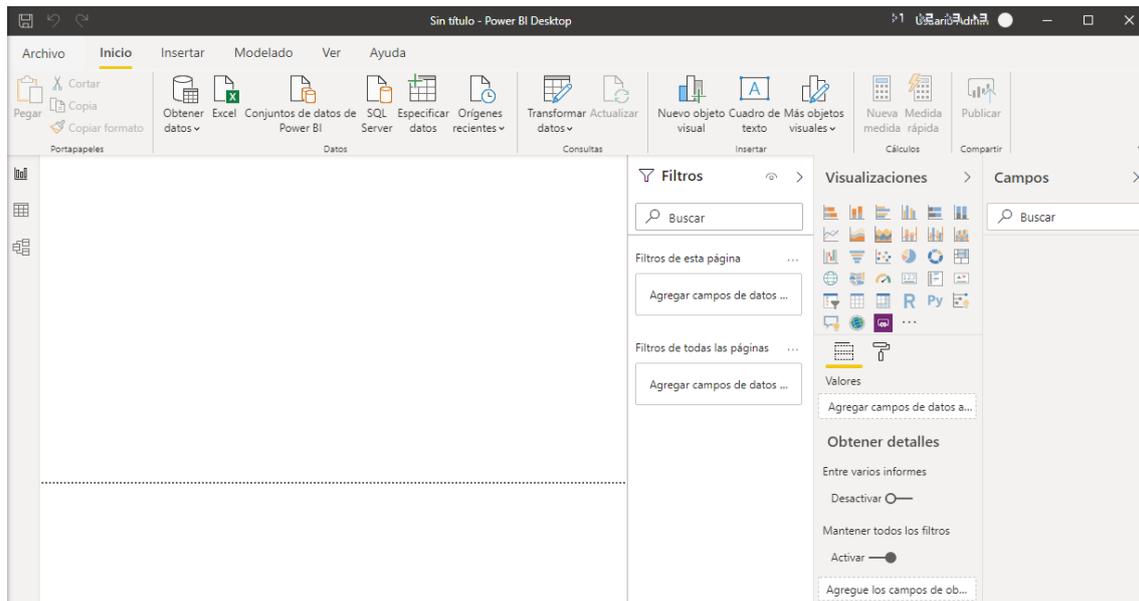


Figura 15. Proyecto de Power BI vacío

Proyecto de estudio de la pandemia

Se va a comenzar explicando el proyecto realizado para el estudio de la pandemia. En este proceso se tratará de explicar detenidamente cómo se realiza la recogida del dato y su transformación hasta llegar a la graficación.

Por ser el primer proyecto se avanzará de manera lenta en las explicaciones para que, en los otros dos proyectos restantes se pueda avanzar más rápido y solo detenerse en aquellos aspectos que quizá no se hayan podido ver en este proyecto o que resulten interesantes o novedosos

Orígenes de datos

Datos mundiales por fecha

El portal de datos abiertos de la Unión Europea da acceso a los datos abiertos publicados por las instituciones y organismos de la Unión Europea. Los usuarios son libres de utilizar y reutilizar todos los datos que encuentren en este catálogo ya sea con fines comerciales o no comerciales.

Dentro de su web poseen una sección dedicada a datos del coronavirus de la enfermedad de COVID-19

<https://data.europa.eu/euodp/es/data/dataset/covid-19-coronavirus-data>

En dicha página encontramos distintos recursos para descargar en varios formatos, como se muestra a continuación:

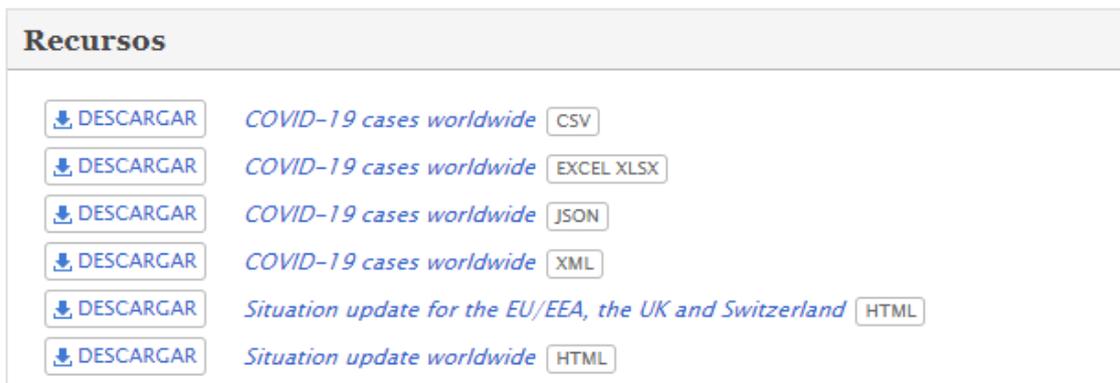


Figura 16. Ejemplos de tipo de descarga

Para este proyecto se utilizará el fichero Excel, aunque perfectamente se podría haber usado ficheros CSV, JSON o XML. El motivo de haber elegido Excel es meramente por fines visuales para poder demostrar más adelante cómo trabaja Power BI.

Otra gran ventaja de Power BI es que permite la obtención de datos desde páginas web, por lo tanto y sabiendo que este fichero Excel va a ser actualizado cada día, se va a optar por obtener su dirección web en vez de descargar el fichero. Esto nos brinda varias ventajas ya que, en caso de descargar el fichero, éste estará actualizado solamente hasta el día en el que dicho fichero se descargó. De esta otra manera, cada vez que se recargue Power BI, éste accederá a la página web, obtendrá el fichero y lo cargará en la aplicación con los datos actualizados.

La dirección web del Excel en la página web de datos abiertos es la siguiente:

<https://www.ecdc.europa.eu/sites/default/files/documents/COVID-19-geographic-disbtribution-worldwide.xlsx>

El contenido del Excel son 10 columnas que contendrán fecha completa (día, mes y año), número de casos detectados, número de fallecidos, país al que corresponden los datos y población total de dicho país de acuerdo con los últimos datos recogidos en 2018. Las fechas para cada país varían ya que cada país afrontó la pandemia en distintas fechas y por lo tanto comenzó a registrar datos en distintas fechas.

	A	B	C	D	E	F	G	H	I	J
1	dateRep	day	month	year	cases	deaths	countries/geold	country	tyte	popData20
2	#####	7	4	2020	38	0	Afghanist,AF	AFG		37172386
3	#####	6	4	2020	29	2	Afghanist,AF	AFG		37172386
4	#####	5	4	2020	35	1	Afghanist,AF	AFG		37172386
5	#####	4	4	2020	0	0	Afghanist,AF	AFG		37172386
6	#####	3	4	2020	43	0	Afghanist,AF	AFG		37172386
7	#####	2	4	2020	26	0	Afghanist,AF	AFG		37172386
8	#####	1	4	2020	25	0	Afghanist,AF	AFG		37172386
9	#####	31	3	2020	27	0	Afghanist,AF	AFG		37172386
10	#####	30	3	2020	8	1	Afghanist,AF	AFG		37172386
11	#####	29	3	2020	15	1	Afghanist,AF	AFG		37172386
12	#####	28	3	2020	16	1	Afghanist,AF	AFG		37172386
13	#####	27	3	2020	0	0	Afghanist,AF	AFG		37172386
14	#####	26	3	2020	33	0	Afghanist,AF	AFG		37172386
15	#####	25	3	2020	2	0	Afghanist,AF	AFG		37172386
16	#####	24	3	2020	6	1	Afghanist,AF	AFG		37172386
17	#####	23	3	2020	10	0	Afghanist,AF	AFG		37172386
18	#####	22	3	2020	0	0	Afghanist,AF	AFG		37172386
19	#####	21	3	2020	2	0	Afghanist,AF	AFG		37172386
20	#####	20	3	2020	0	0	Afghanist,AF	AFG		37172386
21	#####	19	3	2020	0	0	Afghanist,AF	AFG		37172386
22	#####	18	3	2020	1	0	Afghanist,AF	AFG		37172386
23	#####	17	3	2020	5	0	Afghanist,AF	AFG		37172386
24	#####	16	3	2020	6	0	Afghanist,AF	AFG		37172386
25	#####	15	3	2020	3	0	Afghanist,AF	AFG		37172386
26	#####	11	3	2020	3	0	Afghanist,AF	AFG		37172386
27	#####	8	3	2020	3	0	Afghanist,AF	AFG		37172386
28	#####	2	3	2020	0	0	Afghanist,AF	AFG		37172386
29	#####	1	3	2020	0	0	Afghanist,AF	AFG		37172386
30	#####	29	2	2020	0	0	Afghanist,AF	AFG		37172386
31	#####	28	2	2020	0	0	Afghanist,AF	AFG		37172386
32	#####	27	2	2020	0	0	Afghanist,AF	AFG		37172386
33	#####	26	2	2020	0	0	Afghanist,AF	AFG		37172386
34	#####	25	2	2020	1	0	Afghanist,AF	AFG		37172386
35	#####	24	2	2020	0	0	Afghanist,AF	AFG		37172386
36	#####	23	2	2020	0	0	Afghanist,AF	AFG		37172386
37	#####	22	2	2020	0	0	Afghanist,AF	AFG		37172386
38	#####	21	2	2020	0	0	Afghanist,AF	AFG		37172386
39	#####	20	2	2020	0	0	Afghanist,AF	AFG		37172386
40	#####	19	2	2020	0	0	Afghanist,AF	AFG		37172386
41	#####	18	2	2020	0	0	Afghanist,AF	AFG		37172386
42	#####	17	2	2020	0	0	Afghanist,AF	AFG		37172386
43	#####	16	2	2020	0	0	Afghanist,AF	AFG		37172386
44	#####	15	2	2020	0	0	Afghanist,AF	AFG		37172386
45	#####	14	2	2020	0	0	Afghanist,AF	AFG		37172386
46	#####	13	2	2020	0	0	Afghanist,AF	AFG		37172386
47	#####	12	2	2020	0	0	Afghanist,AF	AFG		37172386
48	#####	11	2	2020	0	0	Afghanist,AF	AFG		37172386
49	#####	10	2	2020	0	0	Afghanist,AF	AFG		37172386
50	#####	9	2	2020	0	0	Afghanist,AF	AFG		37172386
51	#####	8	2	2020	0	0	Afghanist,AF	AFG		37172386
52	#####	7	2	2020	0	0	Afghanist,AF	AFG		37172386
53	#####	6	2	2020	0	0	Afghanist,AF	AFG		37172386

Figura 17. Contenido de la tabla Excel

Datos mundiales por fecha específicos

Para poder completar la información que podría faltar en el anterior punto anterior, se encontró un repositorio de GitHub muy completo que añadía datos interesantes como podría ser la población del país, número de fumadores, acceso a facilidades para lavarse las manos o esperanza de vida del país.

Estos datos se encuentran repartidos en 41 columnas con distintas entradas para cada fecha y país. Esto lo hace muy complementario con los datos recibidos del apartado anterior ya que entre los dos se podrá obtener información muy relevante y realizar así un estudio de calidad.

Los datos son accesibles desde la siguiente URL:

<https://github.com/owid/covid-19-data/blob/master/public/data/owid-covid-codebook.csv>

Se informa de las columnas que posee este CSV y de qué información almacenan cada una de ellas. Una pequeña muestra se puede ver en la imagen siguiente, donde se pueden ver las primeras 25 filas con sus descripciones pertinentes

1	column	description
2	iso_code	ISO 3166-1 alpha-3 – three-letter country codes
3	continent	Continent of the geographical location
4	location	Geographical location
5	date	Date of observation
6	total_cases	Total confirmed cases of COVID-19
7	new_cases	New confirmed cases of COVID-19
8	new_cases_smoothed	New confirmed cases of COVID-19 (7-day smoothed)
9	total_deaths	Total deaths attributed to COVID-19
10	new_deaths	New deaths attributed to COVID-19
11	new_deaths_smoothed	New deaths attributed to COVID-19 (7-day smoothed)
12	total_cases_per_million	Total confirmed cases of COVID-19 per 1,000,000 people
13	new_cases_per_million	New confirmed cases of COVID-19 per 1,000,000 people
14	new_cases_smoothed_per_million	New confirmed cases of COVID-19 (7-day smoothed) per 1,000,000 people
15	total_deaths_per_million	Total deaths attributed to COVID-19 per 1,000,000 people
16	new_deaths_per_million	New deaths attributed to COVID-19 per 1,000,000 people
17	new_deaths_smoothed_per_million	New deaths attributed to COVID-19 (7-day smoothed) per 1,000,000 people
18	total_tests	Total tests for COVID-19
19	new_tests	New tests for COVID-19
20	new_tests_smoothed	New tests for COVID-19 (7-day smoothed). For countries that don't report testing data on a daily basis, we assume that tes
21	total_tests_per_thousand	Total tests for COVID-19 per 1,000 people
22	new_tests_per_thousand	New tests for COVID-19 per 1,000 people
23	new_tests_smoothed_per_thousand	New tests for COVID-19 (7-day smoothed) per 1,000 people
24	tests_per_case	Tests conducted per new confirmed case of COVID-19, given as a rolling 7-day average (this is the inverse of positive_rate)
25	positive_rate	The share of COVID-19 tests that are positive, given as a rolling 7-day average (this is the inverse of tests_per_case)

Figura 18. Descripción de las columnas

Datos mundiales totales

También se usarán los datos ofrecidos por la reputada página www.worldometers.info para obtener datos acerca de los totales (nuevos casos, muertes, recuperados, etc.) en todo el mundo.

Esta página es usada por el gobierno del Reino Unido, el gobierno de Tailandia, el New York Times, la BBC y otras tantas entidades y gobiernos de todo el mundo.

La página web específica que nos interesa será la siguiente: <https://www.worldometers.info/coronavirus/#countries>

En esta página solamente contamos con una tabla y es aquella que nos interesa. En este caso no se puede descargar de ninguna forma por lo que más adelante se mostrará cómo obtener dicha información.

All	Europe	North America	Asia	South America	Africa	Oceania						
Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	Active Cases	Serious, Critical	Tot Cases/ 1M pop	Deaths/ 1M pop	Total Tests	Tests/ 1M pop	
World	1,709,593	+10,758	103,506	+822	382,012	1,224,075	49,889	219	13.3			
USA	503,177	+301	18,761	+14	27,314	457,102	10,917	1,520	57	2,538,888	7,670	
Spain	161,852	+3,579	16,353	+272	59,109	86,390	7,371	3,462	350	355,000	7,593	
Italy	147,577		18,849		30,455	98,273	3,497	2,441	312	906,864	14,999	
France	124,869		13,197		24,932	86,740	7,004	1,913	202	333,807	5,114	
Germany	122,171		2,736		53,913	65,522	4,895	1,458	33	1,317,887	15,730	
China	81,953	+46	3,339	+3	77,525	1,089	141	57	2			
UK	73,758		8,958		344	64,456	1,559	1,086	132	316,836	4,667	
Iran	68,192		4,232		35,465	28,495	3,969	812	50	242,568	2,888	
Turkey	47,029		1,006		2,423	43,600	1,667	558	12	307,210	3,643	
Belgium	28,018	+1,351	3,346	+327	5,986	18,686	1,262	2,418	289	102,151	8,814	
Switzerland	24,657	+106	1,003	+1	11,100	12,554	386	2,849	116	178,500	20,625	
Netherlands	23,097		2,511		250	20,336	1,424	1,348	147	101,534	5,926	
Canada	22,148		569		6,013	15,566	557	587	15	383,322	10,156	
Brazil	19,943	+154	1,074	+6	173	18,696	296	94	5	62,985	296	
Portugal	15,472		435		233	14,804	226	1,517	43	140,863	13,815	
Austria	13,713	+153	337	+18	6,604	6,772	246	1,523	37	140,975	15,653	

Figura 19. Contenido tabla HTML

Inteligencia de negocio

ETL Datos mundiales por fecha– Portal de datos Unión Europea

Carga de datos en Power BI

Una vez ya hemos definido qué datos se van a utilizar para el desarrollo del proyecto procedemos a ejecutar Power BI para comenzar la carga de dichos datos.

Para explicar la carga de datos comenzaremos obteniendo los datos del portal de datos abierto de la Unión Europea, que como ya explicamos, se trata de un Excel. Este fichero podríamos descargarlo y traerlo a Power BI como un fichero Excel normal, con la inconveniencia de que dicho Excel estaría actualizado hasta el día que se descargó. Es por ello por lo que copiaremos el link de su ubicación web para que Microsoft Power BI lo recoja cada vez que actualicemos el proyecto.

Comenzaremos pulsando el botón Obtener datos, que desplegará un conjunto de opciones. Dado que nos interesa traer esta información desde una página web, escogeremos la opción titulada como “Web”.

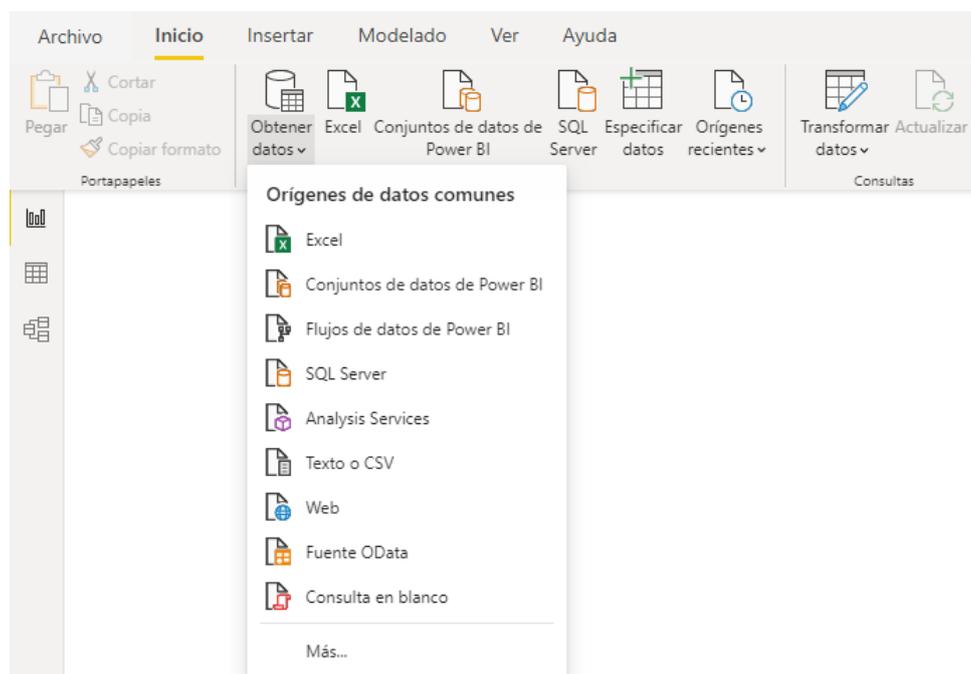


Figura 20. Obtención de datos en Power BI

Una vez seleccionada, se nos mostrará una ventana emergente, en la cual deberemos ingresar la URL donde se encuentra alojado el fichero Excel



Figura 21. Obtención de datos vía web (1)

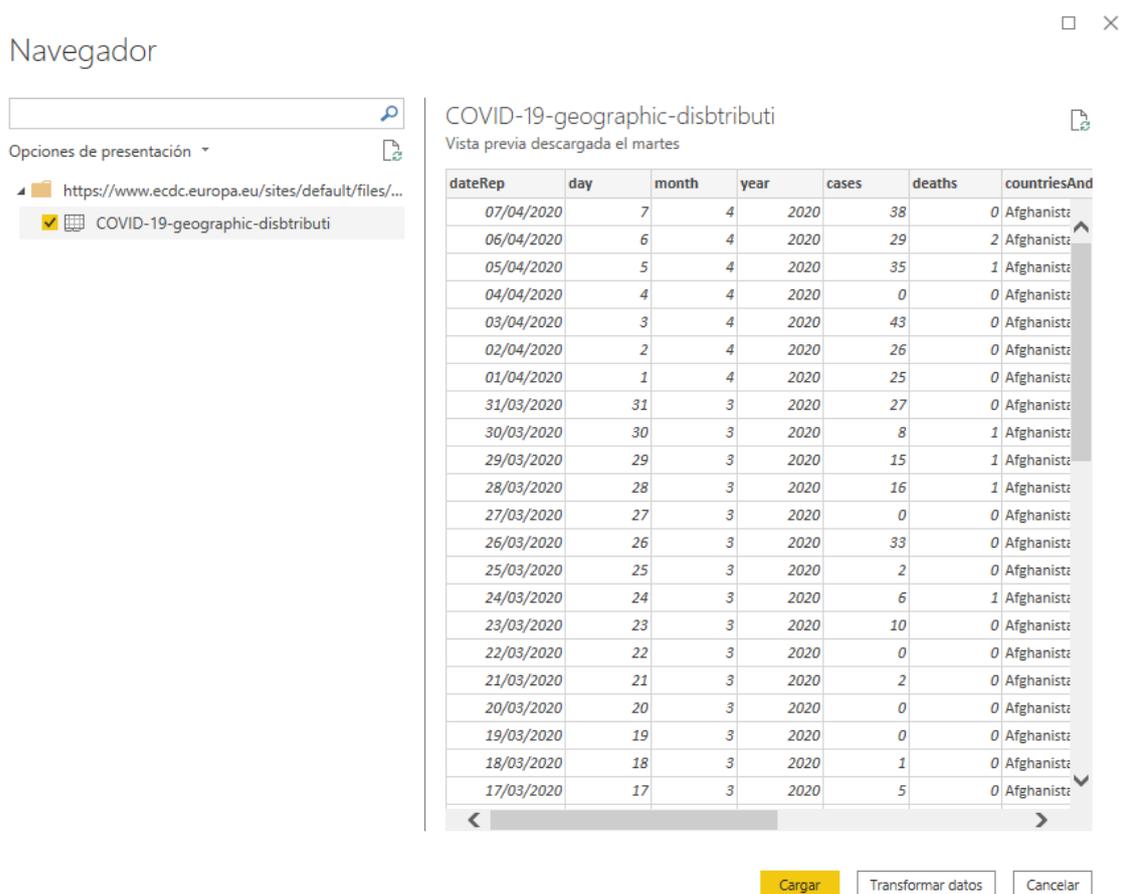


Figura 22. Obtención de datos vía web (2)

Power BI obtiene todo el material de la página web cuyo contenido pueda ser del estilo de una tabla (en este caso como se trata de un fichero Excel encuentra la tabla sin problemas).

Nos muestra una previsualización que permite discernir entre tablas en caso de existir varias y una vez seleccionada la tabla podemos cargarla directamente o transformar los datos de la tabla. Por norma general nunca cargaremos los datos directamente en el sistema. Esta opción se suele utilizar cuando cargamos datos directamente desde otra base de datos, donde los datos ya están preparados, normalizados y modelados. En este caso escogeremos la opción “Transformar

datos”, la cual nos creará otra ventana distinta a la principal de Power BI llamada “Editor de Power Query”.

Transformación de datos

El editor de Power Query es una ventana muy importante dentro de Power BI a la cual se puede acceder de dos formas. La primera, y como ya hemos visto, se realiza en la carga de datos, donde se nos posibilita la opción de “Transformar datos” que nos dirigirá al editor de Power Query. La segunda opción se encuentra desde la ventana principal de Power BI, en el apartado Consultas. El botón “Transformar datos” también se encuentra aquí y cumple exactamente con la misma función: redirigirnos a una nueva ventana en Power Query.



Figura 23. Barra superior de Power BI

Power Query es una tecnología propietaria de Microsoft de conexión de datos, usada en Power BI para transformar los datos para su posterior análisis.

Estas transformaciones se pueden realizar quitando columnas no importantes, cambiando el tipo de los datos, anexando y combinando filas o columnas, limpiando datos incorrectos o incompletos, añadiendo campos calculables, etc.

Microsoft incorpora varias herramientas muy intuitivas para realizar estas transformaciones, que no dejan de ser Macros que ejecutan un script en el lenguaje denominado “M”.

Desde el propio editor de Power Query se pueden editar consultas en dicho lenguaje para poder así crear funciones totalmente personalizadas o editar consultas ya existentes. Esto aporta una gran flexibilidad.

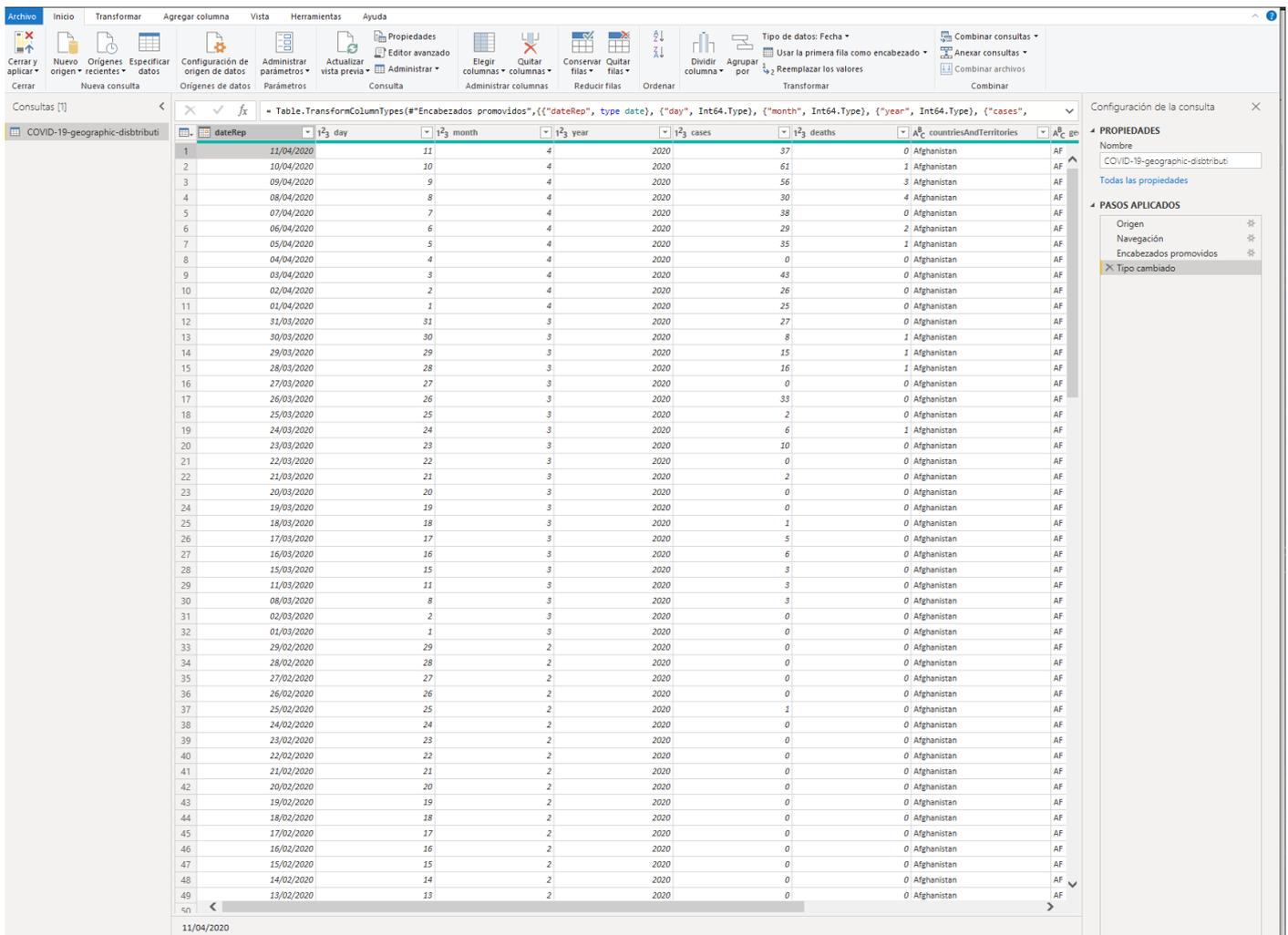


Figura 24. Editor de consultas

Para comprender el funcionamiento del editor de Power Query debemos entender su funcionamiento como una máquina del tiempo: podemos volver al pasado y realizar cambios, pero estos tendrán consecuencias en el futuro. Esto se debe a la existencia de los pasos aplicados, que se pueden encontrar a la derecha de la ventana y que, por lo general, al cargar una nueva tabla, cuenta con 4 pasos ya pre-aplicados: Origen, Navegación, Encabezados promovidos y tipo cambiado.

Estos pasos corresponden con acciones ejecutadas en lenguaje M, las cuales podemos ver pulsando el botón de "Editor avanzado" encontrado en la cabecera del editor de Power Query

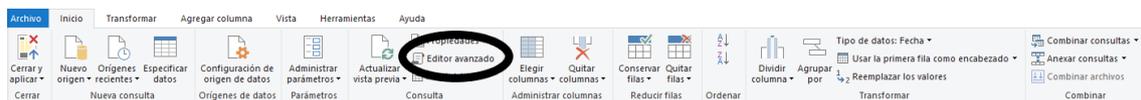


Figura 25. Acceso al editor de consultas

La ventana emergente nos mostrará el código en lenguaje M donde cada acción es nombrada y corresponde con los pasos aplicados mostrados en el editor.



```
let
    Origen = Excel.Workbook(Web.Contents("https://www.ecdc.europa.eu/sites/default/files/documents/COVID-19-geographic-disbtribution-worldwid
    #"COVID-19-geographic-disbtributi_Sheet" = Origen[Item="COVID-19-geographic-disbtributi",Kind="Sheet"]{Data},
    #"Encabezados promovidos" = Table.PromoteHeaders(#"COVID-19-geographic-disbtributi_Sheet", [PromoteAllScalars=true]),
    #"Tipo cambiado" = Table.TransformColumnTypes(#"Encabezados promovidos",{{"dateRep", type date}, {"day", Int64.Type}, {"month", Int64.Type}
in
    #"Tipo cambiado"
```

Figura 26. Editor avanzado

El paso Origen se encarga de recoger la información desde el link proporcionado originalmente y haciendo referencia al mismo, el paso de Navegación trae la tabla incluida en dicha URL.

Estos dos pasos son los más imprescindibles de cualquier consulta y, por norma general, no se deben tocar excepto en casos de creación de funciones donde quizá el origen o tabla a escoger es un parámetro.

El encabezado promovido selecciona la primera fila de la tabla y la promueve a encabezado de la misma. Este paso se obvia en aquellos ficheros de Excel que posean tablas con cabeceras propias. El paso de tipo cambiado reconoce automáticamente el contenido de cada columna de la tabla y adecua el tipo de la misma al detectado. Este paso suele ser revisado en especial ya que nos puede interesar cambiar algunos tipos a otros.

Seleccionando cualquiera de estos pasos la tabla “retrocede en el tiempo” hasta el momento en el que dicho paso tuvo efecto. En este punto se pueden introducir nuevos pasos, aunque habrá que tener precaución ya que pueden alterar y corromper los pasos posteriores. Un ejemplo de ello sería retroceder en una tabla un paso antes de reordenar las columnas y eliminar una de ellas. Al aplicar dicho paso, la tabla eliminaría dicha columna y posteriormente intentará reordenar las columnas. En este paso y dado que la instrucción de reordenar columnas se realizó cuando existía una columna más, al no encontrar dicha columna nos informará de la existencia de un error.

Una vez comprendidos estos conceptos, revisamos las columnas existentes en dicha tabla para establecer si necesitamos todas las columnas o existen algunas que se pudieran eliminar.

	dateRep	day	month	year	cases	deaths	countriesAndTerritories
1	11/04/2020	11	4	2020	37	0	Afghanistan
2	10/04/2020	10	4	2020	62	1	Afghanistan
3	09/04/2020	9	4	2020	56	3	Afghanistan
4	08/04/2020	8	4	2020	30	4	Afghanistan
5	07/04/2020	7	4	2020	38	0	Afghanistan
6	06/04/2020	6	4	2020	29	2	Afghanistan
7	05/04/2020	5	4	2020	35	1	Afghanistan
8	04/04/2020	4	4	2020	0	0	Afghanistan
9	03/04/2020	3	4	2020	43	0	Afghanistan
10	02/04/2020	2	4	2020	26	0	Afghanistan
11	01/04/2020	1	4	2020	25	0	Afghanistan
12	31/03/2020	31	3	2020	27	0	Afghanistan
13	30/03/2020	30	3	2020	8	1	Afghanistan
14	29/03/2020	29	3	2020	15	1	Afghanistan
15	28/03/2020	28	3	2020	16	1	Afghanistan
16	27/03/2020	27	3	2020	0	0	Afghanistan
17	26/03/2020	26	3	2020	33	0	Afghanistan
18	25/03/2020	25	3	2020	2	0	Afghanistan
19	24/03/2020	24	3	2020	6	1	Afghanistan
20	23/03/2020	23	3	2020	10	0	Afghanistan
21	22/03/2020	22	3	2020	0	0	Afghanistan
22	21/03/2020	21	3	2020	0	0	Afghanistan
23	20/03/2020	20	3	2020	0	0	Afghanistan
24	19/03/2020	19	3	2020	0	0	Afghanistan
25	18/03/2020	18	3	2020	1	0	Afghanistan
26	17/03/2020	17	3	2020	5	0	Afghanistan
27	16/03/2020	16	3	2020	6	0	Afghanistan
28	15/03/2020	15	3	2020	3	0	Afghanistan
29	14/03/2020	14	3	2020	0	0	Afghanistan
30	08/03/2020	8	3	2020	3	0	Afghanistan
31	02/03/2020	2	3	2020	0	0	Afghanistan
32	01/03/2020	1	3	2020	0	0	Afghanistan
33	29/02/2020	29	2	2020	0	0	Afghanistan
34	28/02/2020	28	2	2020	0	0	Afghanistan
35	27/02/2020	27	2	2020	0	0	Afghanistan
36	26/02/2020	26	2	2020	0	0	Afghanistan
37	25/02/2020	25	2	2020	1	0	Afghanistan
38	24/02/2020	24	2	2020	0	0	Afghanistan
39	23/02/2020	23	2	2020	0	0	Afghanistan
40	22/02/2020	22	2	2020	0	0	Afghanistan
41	21/02/2020	21	2	2020	0	0	Afghanistan
42	20/02/2020	20	2	2020	0	0	Afghanistan
43	19/02/2020	19	2	2020	0	0	Afghanistan
44	18/02/2020	18	2	2020	0	0	Afghanistan
45	17/02/2020	17	2	2020	0	0	Afghanistan
46	16/02/2020	16	2	2020	0	0	Afghanistan
47	15/02/2020	15	2	2020	0	0	Afghanistan
48	14/02/2020	14	2	2020	0	0	Afghanistan
49	13/02/2020	13	2	2020	0	0	Afghanistan
50							Afghanistan

Figura 27. Transformación de datos

Observamos que la comuna “dateRep”, originalmente oculta en la página web, muestra la fecha en formato día, mes, año. Esta fecha concuerda exactamente con las columnas “day”, “month” y “year” de la tabla, por lo que al ser redundantes procederemos a su eliminación.

Del mismo modo las columnas “countriesAndTerritories”, “geolD” y “countryterritoryCode” hacen referencia al mismo país de formas distintas

A ^B C countriesAndTerritories	A ^B C geolD	A ^B C countryterritoryCode
Afghanistan	AF	AFG

Figura 28. Contenido y tipos de la tabla

Para realizar esta eliminación de columnas, seleccionaremos todas las columnas a eliminar pulsando sobre ellas mientras presionamos la tecla “Control” y pulsaremos en el botón “Quitar columnas” como se muestra a continuación:

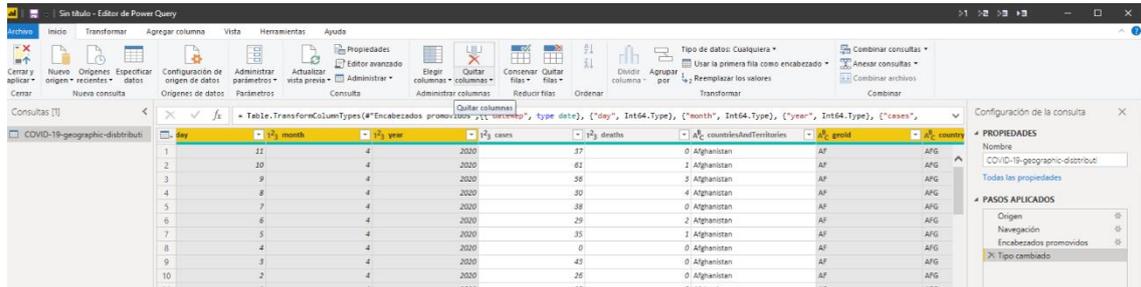
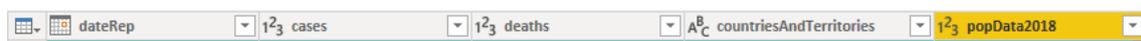


Figura 28. Pasos realizados

El cambio se aplica y nuestra tabla pasa a contener únicamente 5 columnas.

Para concluir con la limpieza de datos de la tabla resultante, debemos comprobar si los tipos aplicados automáticamente por Power BI son correctos o se requiere alguna modificación.



El campo “dateRep” es de tipo fecha, los campos “cases”, “deaths” y “popData2018” son de tipo numérico y por último “countriesAndTerritories” es de tipo texto. Todos los tipos parecen ser los correctos y no se requiere de más operaciones en esta tabla.

Para realizar la carga de la tabla a Power BI pincharemos en el botón “Cerrar y aplicar” del menú de navegación del editor de Power Query.

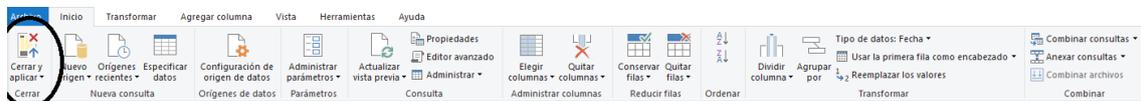


Figura 29. Cerrar y aplicar

Carga de datos en Power BI

Para realizar la carga de datos del Excel alojado en el repositorio de GitHub el procedimiento será igual al anterior, por lo que no deberíamos encontrar ningún problema repitiendo el mismo proceso que en pasos anteriores. No obstante, debemos acordarnos de insertar la URL correcta:

<https://github.com/owid/covid-19-data/raw/master/public/data/owid-covid-data.xlsx>

Desde la previsualización podemos comprobar la veracidad de los datos al tener las columnas con sus correspondientes valores.

The screenshot shows the Power BI Navigator interface. On the left, the 'Navegador' pane displays the data source URL: [https://github.com/owid/covid-19-data/raw/m...](https://github.com/owid/covid-19-data/raw/master/public/data/owid-covid-data.xlsx) and a selected sheet named 'Sheet1'. The main area shows a preview of the data table with the following columns: iso_code, continent, location, date, total_cases, and new_cases. The data rows show records for Afghanistan (AFG) in Asia, with dates from 31/12/2019 to 11/01/2020, and zero cases reported.

iso_code	continent	location	date	total_cases	new_cases
AFG	Asia	Afghanistan	31/12/2019	0	
AFG	Asia	Afghanistan	01/01/2020	0	
AFG	Asia	Afghanistan	02/01/2020	0	
AFG	Asia	Afghanistan	03/01/2020	0	
AFG	Asia	Afghanistan	04/01/2020	0	
AFG	Asia	Afghanistan	05/01/2020	0	
AFG	Asia	Afghanistan	06/01/2020	0	
AFG	Asia	Afghanistan	07/01/2020	0	
AFG	Asia	Afghanistan	08/01/2020	0	
AFG	Asia	Afghanistan	09/01/2020	0	
AFG	Asia	Afghanistan	10/01/2020	0	
AFG	Asia	Afghanistan	11/01/2020	0	

Below the table, a message states: "Los datos de la vista previa se han truncado debido a límites de tamaño." At the bottom right, there are three buttons: 'Cargar' (highlighted in yellow), 'Transformar datos', and 'Cancelar'.

Figura 30. Carga de tabla

Antes de cargar los datos, por buena práctica es recomendable siempre darle a Transformar datos para asegurarnos los tipos y obtener solo las columnas necesarias.

Transformación de datos

Parece ser que en este caso, al tratarse de un Excel muy trabajado, no había que realizar ninguna clase de limpieza de datos. Lo único que ha habido que realizar es el promover encabezados y el cambio de tipos (si es necesario):

iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths
AFG	Asia	Afghanistan	31/12/2019	0	0	0	null
AFG	Asia	Afghanistan	01/01/2020	0	0	0	null
AFG	Asia	Afghanistan	02/01/2020	0	0	0	null
AFG	Asia	Afghanistan	03/01/2020	0	0	0	null
AFG	Asia	Afghanistan	04/01/2020	0	0	0	null
AFG	Asia	Afghanistan	05/01/2020	0	0	0	null
AFG	Asia	Afghanistan	06/01/2020	0	0	0	0
AFG	Asia	Afghanistan	07/01/2020	0	0	0	0
AFG	Asia	Afghanistan	08/01/2020	0	0	0	0
AFG	Asia	Afghanistan	09/01/2020	0	0	0	0
AFG	Asia	Afghanistan	10/01/2020	0	0	0	0
AFG	Asia	Afghanistan	11/01/2020	0	0	0	0
AFG	Asia	Afghanistan	12/01/2020	0	0	0	0
AFG	Asia	Afghanistan	13/01/2020	0	0	0	0

Figura 31. Transformando datos

ETL Datos Mundiales – Worldometers

Carga de datos en Power BI

La carga de datos para la información recogida en la página de Worldometers se realizará de una manera algo distinta ya que no se trata de ningún Excel o archivo descargable. Se trata de una tabla escrita en HTML perteneciente a la propia página web.

Comenzamos del mismo modo, en el escritorio de Power BI, accediendo desde el botón “Obtener datos”: “Web”

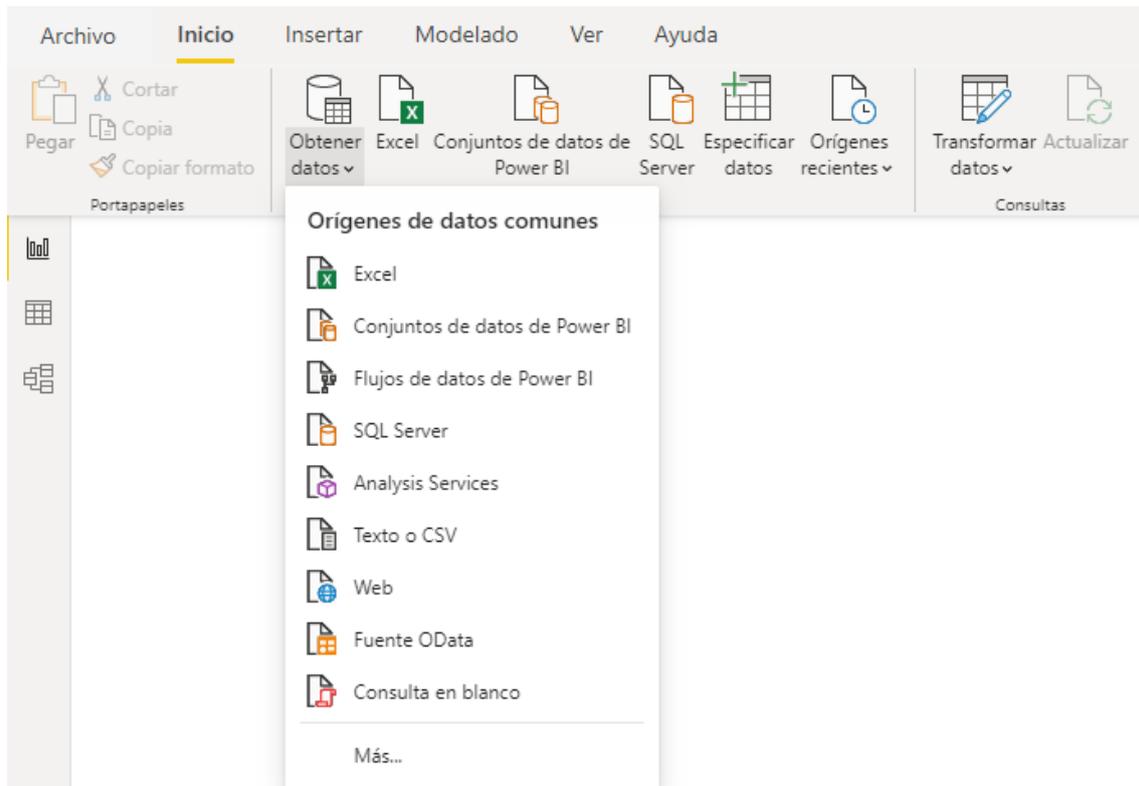


Figura 32. Obtención de datos

Insertamos la URL de la página (<https://www.worldometers.info/coronavirus/#countries>) para proceder a visualizar qué contenido ha sido capaz de recoger Power BI desde la web. Se trata de tres tablas distintas por lo que debemos escoger cuál es la que nos interesa. Según el contenido de cada una de ellas, la tabla denominada “Table 0” es la indicada, por lo que procedemos a realizar la transformación de los datos de dicha tabla en el editor de Power Query.

Navegador

https://www.worldometers.info/coronavirus/#c...

- Document
- Table 0
- Table 1

Vista de tabla Vista web

Table 0

Country, Other	Total Cases	New Cases	Total Deaths	New Deaths
World	2,013,935	16,075	127,587	
Europe	948,291	11,003	84,429	
North America	656,863	781	27,786	
USA	614,246	360	26,064	
Asia	324,801	3,45	11,874	
Spain	177,633	3,573	18,579	
Italy	162,488	null	21,067	
France	143,303	null	15,729	
Germany	132,210	null	3,495	
UK	93,873	null	12,107	
China	82,295	46	3,342	
Iran	76,389	1,512	4,777	
Turkey	65,111	null	1,403	
South America	58,330	641	2,535	
Belgium	33,573	2,454	4,44	
Netherlands	27,419	null	2,945	
Canada	27,063	null	903	
Switzerland	26,023	87	1,19	
Brazil	25,684	422	1,552	
Russia	24,490	3,388	198	
Portugal	17,448	null	567	

Agregar tabla mediante ejemplos
Cargar
Transformar datos
Cancelar

Figura 33. Navegador de tablas

Transformación de datos

Una vez los datos se hayan cargado en el editor de Power Query podemos observar que los datos traídos son los correctos y que, como es costumbre, ya se han realizado los pasos de “Origen”, “Navegación” y “Tipo cambiado”.

El contenido de la tabla son 12 columnas en las que se puede ver información del país, número total de casos hasta la fecha, número de casos nuevos diarios, total de muertes, muertes en el último día, total de recuperados, total de casos todavía activos, total de pacientes en estado crítico (UCI), total de casos por cada millón de habitantes, total de muertes por cada millón de habitantes, número total de pruebas PCR realizadas y número total de pruebas PCR realizados por cada millón de habitantes.

	Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	Active Cases	Serious
1	World	2,013,935	16,075	127,587	227,587	987,491,765	1,394,583	
2	Europe	948,291	11,003	84,429	711,255,377	610,485		
3	North America	656,863	781	27,786	69,50,009	579,068		
4	USA	614,246	360	26,064	17,38,820	549,362		
5	Asia	324,801	3,45	11,874	150,158,359	154,568		
6	Spain	177,633	3,573	18,579	824,70,853	88,201		
7	Italy	162,488	null	21,067	null,37,130	104,291		
8	France	143,303	null	15,729	null,28,805	98,769		
9	Germany	132,210	null	3,495	null,72,800	56,115		
10	UK	93,873	null	12,107	null,16	81,422		
11	China	82,295	46	3,342	1,77,816	1,137		
12	Iran	76,389	1,512	4,777	94,48,933	21,679		
13	Turkey	65,111	null	1,403	null,4,799	58,909		
14	South America	58,330	641	2,535	24,21,617	34,178		
15	Belgium	33,573	2,454	4,44	283,7,107	22,026		
16	Netherlands	27,419	null	2,945	null,250	24,224		
17	Canada	27,063	null	903	null,8,235	17,925		

Figura 34. Pasos aplicados

La tabla, al proceder directamente de un HTML publicado en web, no posee columnas redundantes. Esto no la exime de transformaciones ya que el editor de PowerQuery ha realizado unas transformaciones de tipo no adecuadas. “Total Cases”, “Total Recovered”, “Active Cases” y “Total Test” están detectados como tipo texto cuando deberían ser numérico. El resto de las columnas, obviando la primera que contiene el nombre del país, son de tipo decimal cuando deberían ser de tipo numérico. Esto se debe a la connotación americana de separar los miles con comas.

Estas transformaciones las realizaremos en dos pasos. El primer paso será sustituir las “,” de las columnas numéricas por “.”. De forma que, al aplicar un cambio a tipo entero, el editor de Power Query reconozca los miles. El segundo paso será cambiar el tipo a todas las columnas menos la primera a tipo numérico.

Para realizar el primer paso marcamos todas aquellas columnas detectadas como decimales o aquellas que tengan “,” en sus números. El siguiente paso será hacer click derecho: reemplazar valores. Se nos mostrará una pantalla como la siguiente en la que indicamos que queremos reemplazar las “,” por “.”.

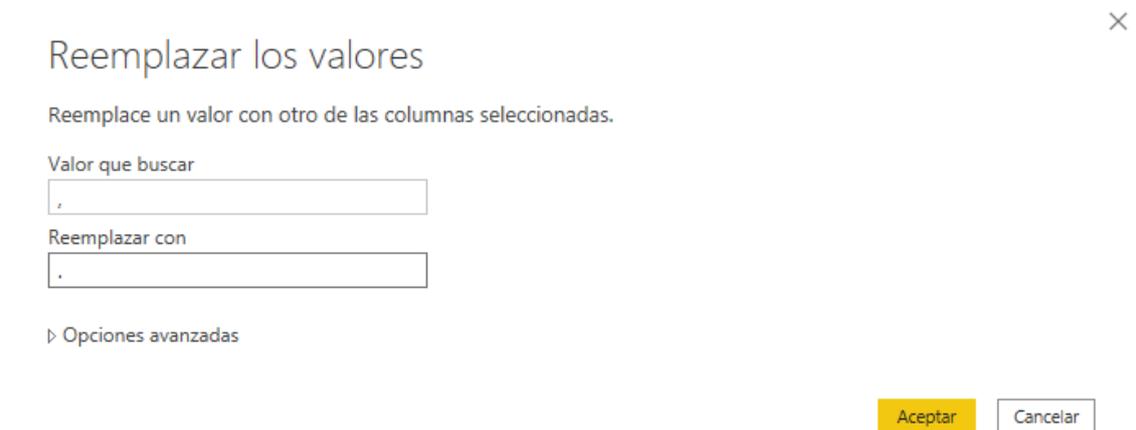


Figura 35. Reemplazar valores

Ahora todos los números son separados por puntos en los miles, de forma que el editor de Power Query es capaz de detectarlos. Seleccionamos todas las columnas menos la primera y aplicamos un cambio de tipo al entero.

	Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	Active Cases
1	World	2013935	16075	127587	987	491765	1394583
2	Europe	948291	11003	84429	711	253377	610485
3	North America	656863	781	27786	99	50009	579068
4	USA	614246	360	26064	17	38820	549362
5	Asia	324801	3450	11874	150	158359	154568
6	Spain	177633	3573	18579	324	70853	88201
7	Italy	162488	null	21067	null	37130	104291
8	France	143303	null	15729	null	28805	98769
9	Germany	132210	null	3495	null	72600	56115
10	UK	93873	null	12107	null	Error	81422
11	China	82295	46	3342	1	77816	1137
12	Iran	76389	1512	4777	94	49933	21679
13	Turkey	65111	null	1403	null	4799	58909
14	South America	58330	641	2535	24	21617	34178

Figura 36. Selección de columnas

Una vez realizado este paso, aplicamos los cambios correspondientes en nuestro proyecto.

Modelaje y relación de las consultas

El proceso ETL ha sido completado por lo que el siguiente paso a dar es el modelaje de las consultas creadas por el proceso anterior. El modelaje es la relación que las consultas tienen entre ellas como si de una base de datos se tratara. En una base de datos se debe crear claves primarias, claves secundarias, asignar si la relación es 1 a N, etc.

En Power BI esto se realiza de una manera muy similar. Para comenzar, debemos dirigirnos al panel de “Modelo” que se encuentra en el lateral izquierdo y cuyo símbolo son tres consultas conectadas entre sí.

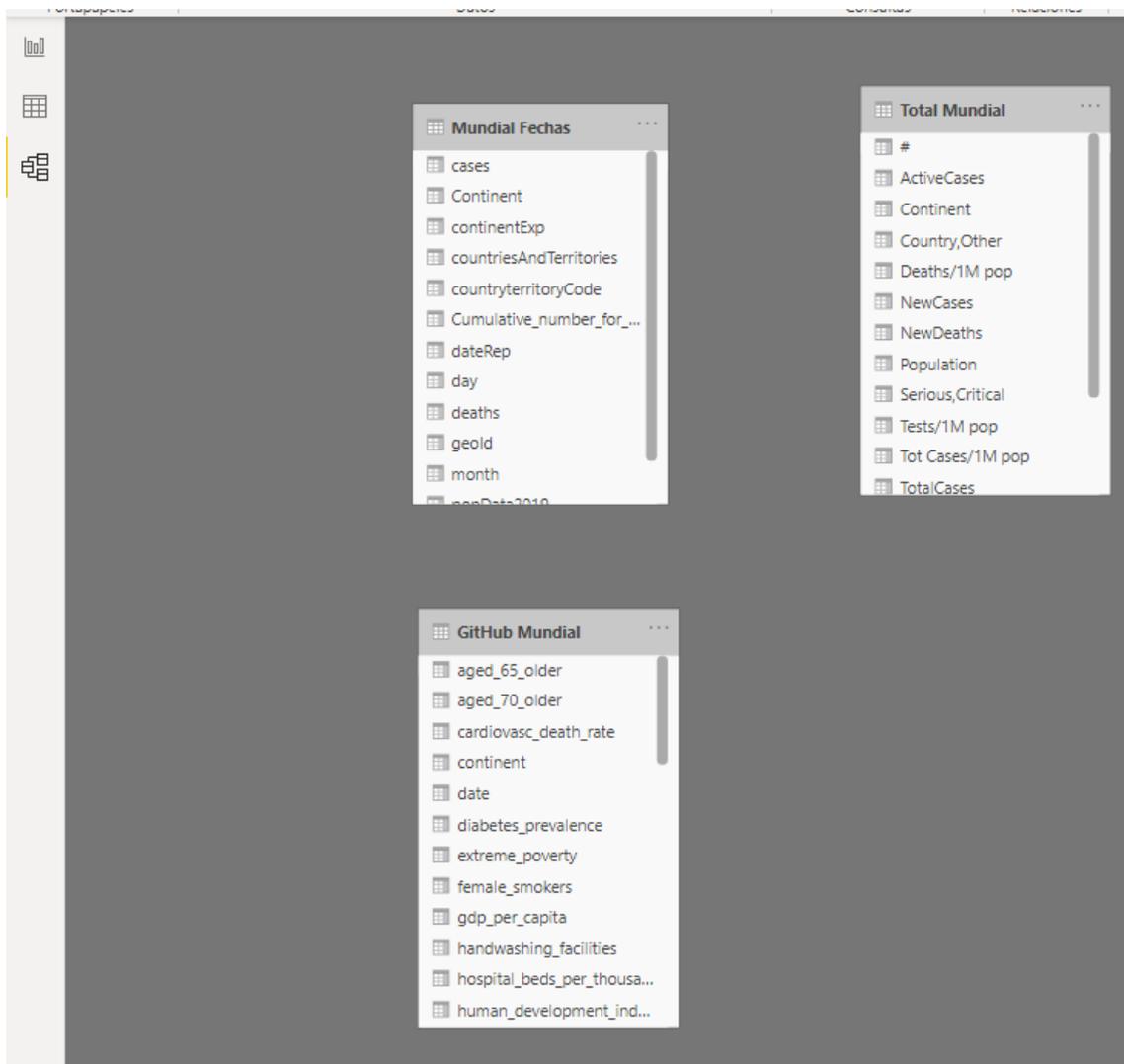


Figura 37. Modelaje de consultas

Una vez hayamos accedido a esta sección de Power BI, podemos ver que existen las tres consultas que habíamos creado en el editor de Power Query

La idea que debemos tener en mente es la de conectar estas tablas para que así, mediante los gráficos, podamos interactuar con distintos datos alojados en tablas distintas.

Debemos comenzar creando una tabla auxiliar para las fechas que nos haga de tabla maestra. De este modo podremos utilizar dicha tabla siempre que queramos referirnos a una fecha y todas las tablas estarán relacionadas a ella. Para ello, nos dirigimos a la pestaña Modelado → Nueva tabla e ingresamos la siguiente fórmula:

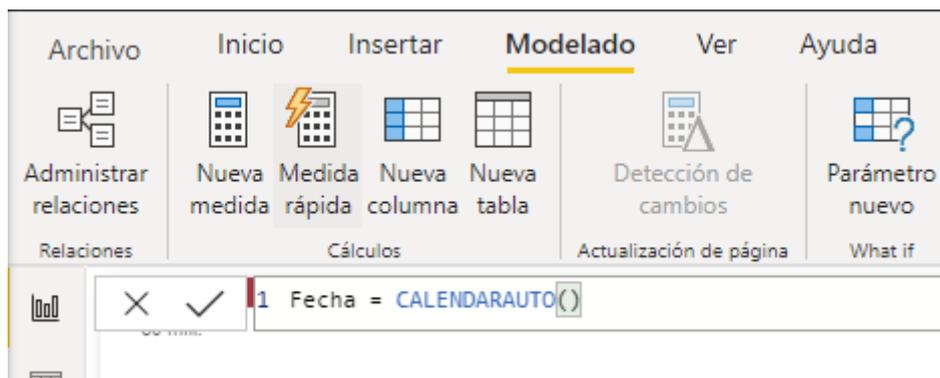


Figura 38. Creación de tabla

Automáticamente aparece en el modelo una nueva tabla llamada Fecha con el campo “Date” que contiene una entrada diaria de este año.

Una vez realizado esto debemos coger este campo “Date” y arrastrarlo a su correspondiente campo en el resto de las tablas, que creará una relación de tipo 1..N (ya que hay varias entradas para un mismo día)

El siguiente paso es relacionar la tabla “Total Mundial” con la tabla “Mundial Fechas” por medio de los países. Realizando el mismo proceso con los campos “countriesAndTerritories” y “Country, Other”.

El resultado final sería el siguiente:

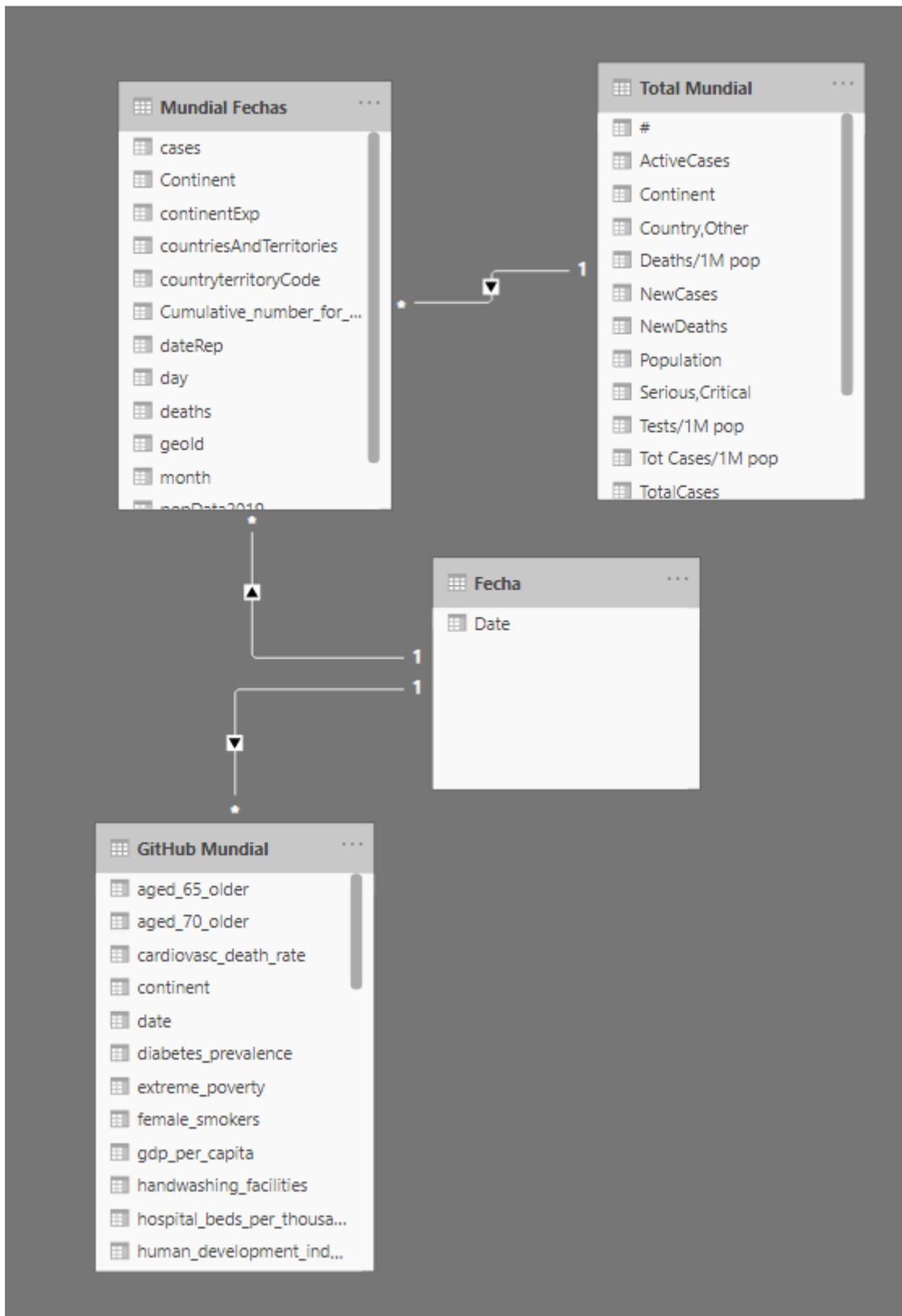


Figura 39. Relación en el modelo

Graficación y desarrollo del análisis

A esta altura del proyecto se puede comenzar a realizar la visualización de los datos recogidos y realizar así el análisis de Business Intelligence y su posterior análisis predictivo.

En Power BI existen las páginas al igual que en Excel, esto nos va a permitir separar el análisis según lo que se quiera analizar en cada hoja. En este caso, aprovecharemos cada hoja para tratar distintos temas relacionados con el COVID-19.

Todo el desarrollo se realizará en la parte de informe de Power BI, la cual posee las herramientas para realizar el análisis de los datos. Aquí encontraremos las visualizaciones, estas están formadas por gráficos de líneas, de barras, listas, matrices, etc.

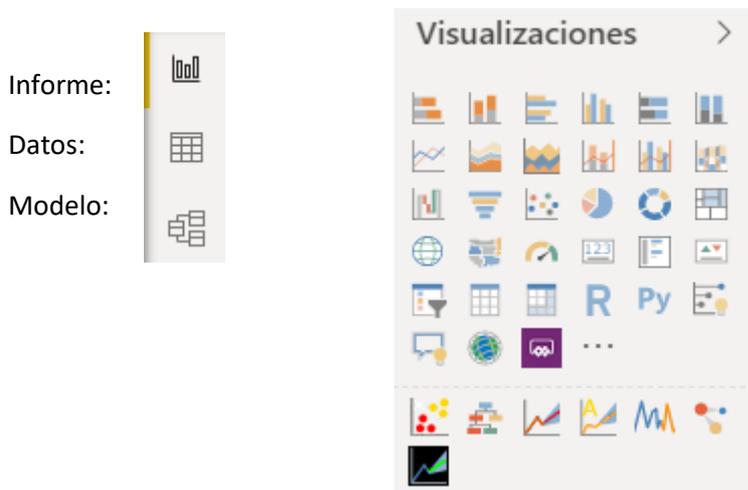


Figura 40. Visualizaciones existentes

Página 1 – Efectos mundiales

En esta página se pretende reflejar en un mapa mundial el número de casos y muertes y que estos sean filtrables por fechas, continentes o países.

Comenzaremos creando filtros temporales que nos permitan seleccionar distintos intervalos de tiempo para la observación de los datos. Para ello se van a crear un filtro de rango de fechas y

una lista de meses. Para crear el filtro con el rango de fechas utilizaremos la visualización “Segmentación de datos”

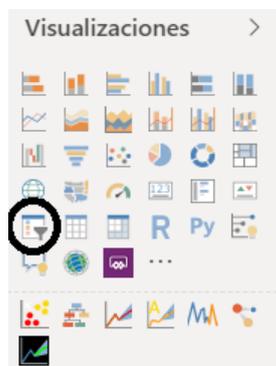


Figura 41. Selección de segmentación de datos

Se mostrará una imagen vacía en la pantalla (la visualización) que necesita que se le pase un campo de tipo fecha para funcionar. Para ello, debajo de las visualizaciones se muestra la ventana “Campos”. Esta ventana hace referencia a los campos de la visualización que en ese momento se esté usando. Desde el panel “Campos” (distinto a la ventana de campos, este contiene los datos que hemos recogido con Power Query) arrastraremos el campo “dateRep” de la consulta “Mundial fechas” hasta la ventana de campos.

Si utilizamos este campo “dateRep” en lugar del campo “Date” de la tabla “Fecha” es debido a que este último contiene muchas más fechas que entradas en la primera tabla, lo que haría que nuestro rango de fechas pueda filtrar en fechas sin datos.

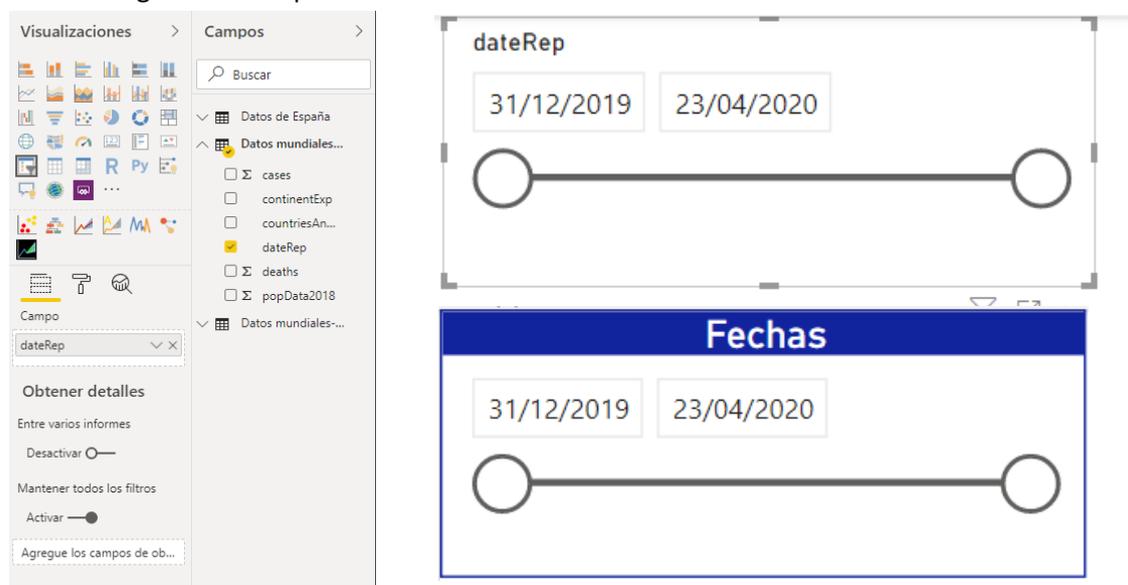


Figura 42. Proceso de creación filtro de fecha

El resultado es la imagen superior con fechas de título “dateRep”. Para darle un formato más bonito como el de la segunda imagen, cambiaremos desde la ventana “Formato”, que se

encuentra al lado de la ventana “Campos” que contiene el campo “dateRep”, el título y diseño del gráfico.

El siguiente paso es crear un filtro por meses. Para ello sí que utilizaremos la tabla “Fecha” y aprovecharemos la jerarquía existente dentro del campo “Date”, la cual nos permite escoger entre toda la fecha, el año, el trimestre, el mes o el día (a nivel de filtro). Crearemos una nueva segmentación únicamente con la jerarquía de mes añadida. Para añadir una segmentación anual, realizamos los mismos pasos añadiendo la jerarquía de años en lugar de meses en una nueva segmentación.

Para terminar con los filtros de selección, añadiremos una segmentación con el campo “continentExp” de la tabla “Mundial Fechas”, la cual nos brinda los continentes en los que se encuentra el país de cada entrada de la tabla.

Una vez terminados los filtros y ajustados los colores a gusto personal, estos quedan tal que así:

The image shows a dashboard with three filter panels. The top panel, titled "Fechas", has an orange header and contains two date input fields: "31/12/2019" and "29/08/2020", with a range slider below them. The middle panel is split into two columns: "Años" with checkboxes for "2019" and "2020", and "Meses" with checkboxes for "enero", "febrero", "marzo", "abril", "mayo", "junio", "julio", "agosto", "septiembre", "octubre", "noviembre", and "diciembre". The bottom panel, titled "Continentes", has an orange header and checkboxes for "Africa", "America", "Asia", "Europe", and "Oceania".

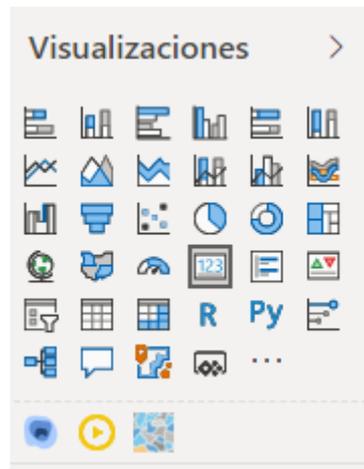
Para conseguir un aspecto visual más bonito, se intenta siempre hacer encajar las “cajas” que contienen los filtros con el tamaño de aquellas adyacentes a ella. Es por ello por lo que, aunque el filtro de años solamente tenga dos entradas (2019 y 2020), esta se alargue hacia abajo considerablemente.

Además de ello, el orden en el que aparecen tiene una razón de ser. El filtro más usado en proyectos así suele ser el intervalo de tiempo, por lo que ha sido colocado el primero. Seguidamente se ha colocado los años y meses para respetar la jerarquía existente entre ellos y por último los continentes en la parte baja de la página ya que no pertenecen al aspecto temporal de los filtros.

Figura 43. Filtros de la página

En este momento queda realizar las visualizaciones sobre las que afectan los filtros anteriores. La idea es mostrar información suficiente para que el usuario sienta que no le falta nada pero sin llegar a confundirlo con demasiadas visualizaciones.

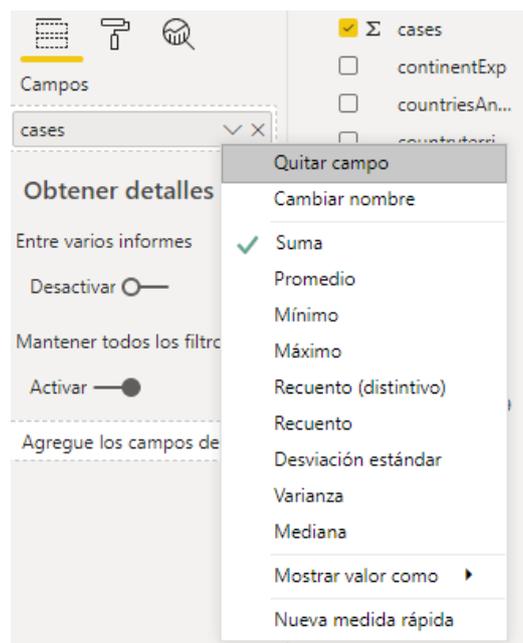
Comenzaremos con tres tarjetas que nos muestren el número de infectados, el número de muertos y qué porcentaje de fallecidos hay. Para esto último haremos uso por primera vez de las medidas de DAX, el lenguaje en el que se programa Power BI.



Filtro 44. Visualización de tarjeta

Seleccionamos de las visualizaciones la tarjeta y le añadimos el campo “cases” de la tabla “Mundial Fechas”. Como podemos comprobar, nos devuelve la suma de todas las entradas de la tabla anterior del campo “cases”. Por defecto Power BI siempre devuelve la suma en campos numéricos y el recuento en campos de texto. Esto se puede confirmar y cambiar pulsando en el desplegable que aparece en el campo de la tarjeta como se muestra a continuación:

Figura 45. Detalles del campo



Para añadir una tarjeta con el número de muertos repetimos el proceso cambiando el campo a "deaths". Como ya se había avisado, entra por defecto la suma del campo a la hora de representarlo en la tarjeta. Esto es lo que se llama una medida implícita dentro de Power BI. Para nuestra última tarjeta, el porcentaje de fallecidos, requeriremos realizar una medida explícita con DAX.

Para la realización de medidas puede ser aconsejable crearse una nueva tabla a la que se le crearán las medidas, con el fin de no "manchar" la tabla original con múltiples medidas que al fin y al cabo, trabajan sobre esos campos. Para crear una nueva tabla nos dirigiremos a la ventana "Modelado" y a continuación sobre "Nueva tabla"

Figura 46. Crear tablas



Al hacerlo se nos mostrará una línea de texto sobre la que escribir. Desde aquí se pueden crear tablas completas con datos, pero en nuestro caso solamente queremos una tabla completamente vacía sobre la que podremos trabajar.

Escribiremos pues lo siguiente: "Medidas = ". De esta forma le decimos mediante DAX a Power BI que queremos una tabla llamada Medidas que se encuentre vacía. Esta aparecerá junto a las otras tablas y no hay que relacionarla en el modelo de datos.

Para crear una medida pulsaremos click derecho del ratón sobre la tabla de medidas y en el desplegable siguiente sobre "Nueva medida". En la entrada de texto siguiente podremos escribir la medida que queramos, que en nuestro caso solo queremos una proporción entre el número de muertes y el de casos.

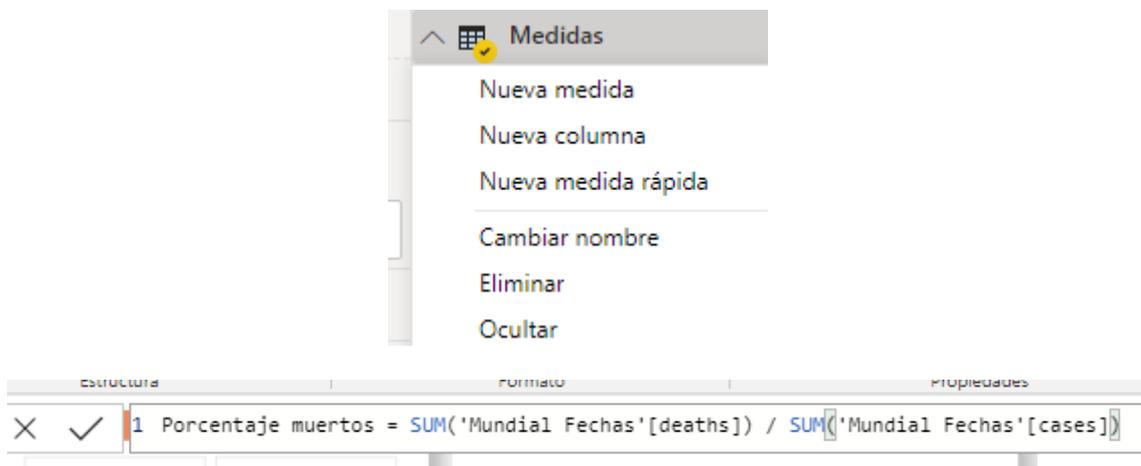
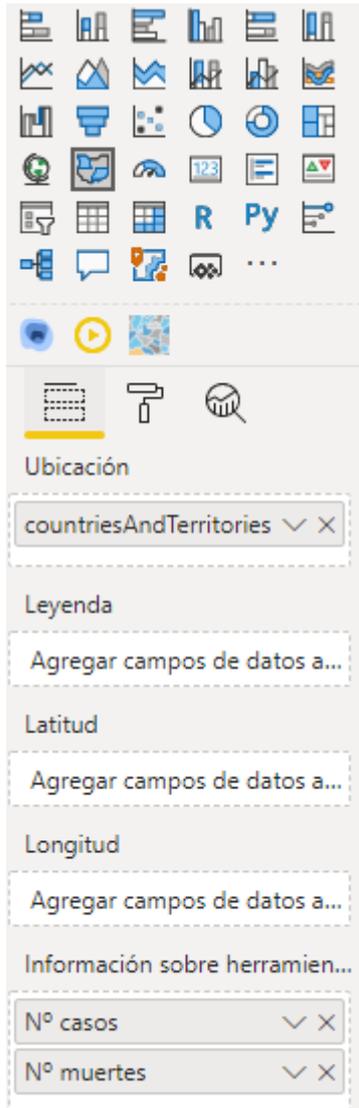


Figura 47. Creación de medida

La medida es muy sencilla de entender, pero es obvio que nos devuelve un valor que va a estar comprendido entre el 0 y el 1, ya que nunca va a haber más muertes que casos y lo que nosotros queremos es verlo como porcentaje. Para ello, haciendo click sobre la medida nos dirigimos a “Herramientas de medición” y podemos pulsar sobre el símbolo del porcentaje “%”



Figura 48. Modelar como porcentaje



Una vez hecho esto, ya tenemos las tres tarjetas listas y funcionales. Solamente faltaría el mapa de calor del mundo.

Dentro de las visualizaciones contamos con una para ello que se llama Mapa coroplético, que es capaz de mostrar gracias a los mapas de Bing, un mapa en Power BI. Necesariamente hay que añadir la localización que queremos ver, por lo que en el campo “Ubicación” arrastraremos la columna “countriesAndTerritories” de la tabla “Mundial Fechas”. Como podremos ver el mapa del mundo reconoce estos países y los marca de otro color en el mapa.

En el campo “Información sobre herramientas” añadiremos las columnas “cases” y “deaths” ya que queremos que cuando pinchemos sobre un mapa nos salga información sobre estos datos.

Por último, para crear el mapa de calor nos dirigimos al rodillo de formato y en “colores de datos” seleccionamos que el color predeterminado sea una escala de colores (con los colores que nosotros elijamos).

Figura 49. Contenidos de la visualización

Una vez montada la página por completo se puede visualizar el resultado final que quedará semejante al siguiente:

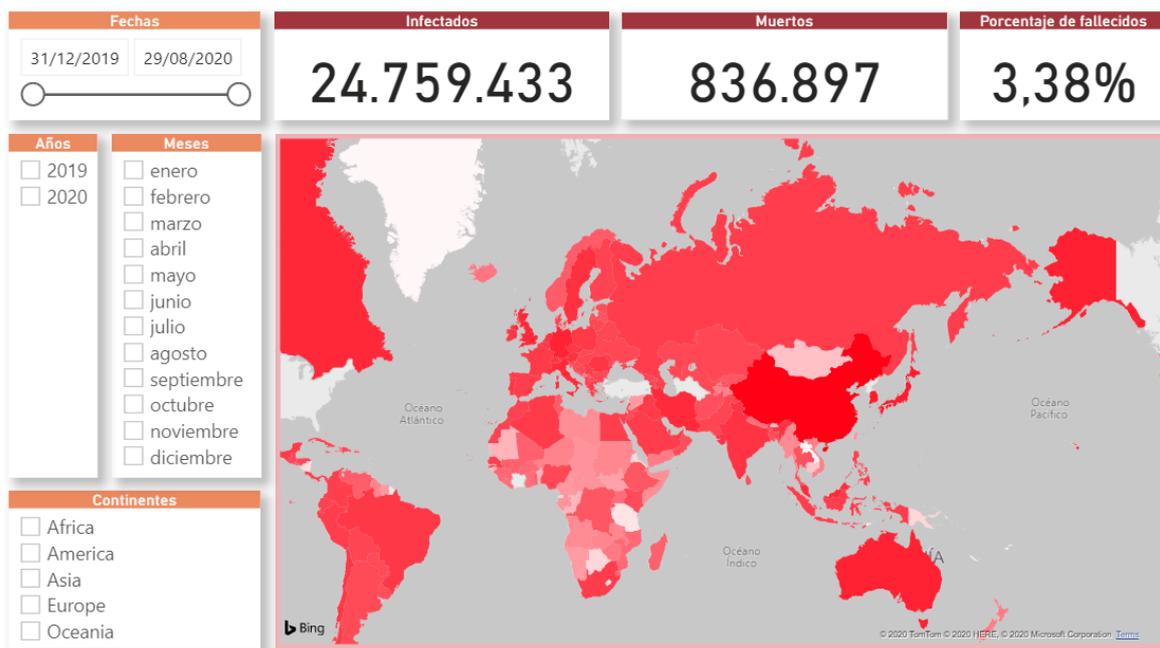


Figura 50. Resultado página 1

En esta página podemos comprobar como las visualizaciones (de color rojo) son afectadas por los filtros (color naranja). De este modo se puede visualizar de una manera muy interactiva, interesante y fácil el paso del COVID-19 por los distintos países.

Página 2 – Evolución temporal del COVID-19

La siguiente página del primer proyecto se trata de una animación en la que se pretende mostrar la evolución de distintos países en el tiempo con respecto al número de casos de COVID-19 que han padecido.

Para ello necesitaremos únicamente dos visualizaciones: un gráfico de líneas y una visualización específica para que itere sobre las fechas.

Comenzando por el gráfico de líneas, se trae a esta nueva página desde la ventana de visualizaciones. Para su funcionamiento requiere de valores en el eje, en los valores y depende de cómo se desee el gráfico, en la leyenda.

Para el eje se utilizará la tabla Fecha creada con el campo "Date" (sin jerarquía), ya que queremos que se vea de manera lineal en el tiempo esta evolución. En los valores volcaremos el campo "cases" de la tabla "Mundial Fechas" y automáticamente veremos una evolución en el gráfico del número de casos a nivel mundial. Como lo que se desea es comparar varios países y no ver la evolución mundial, añadiremos a la leyenda el campo "countriesAndTerritories".

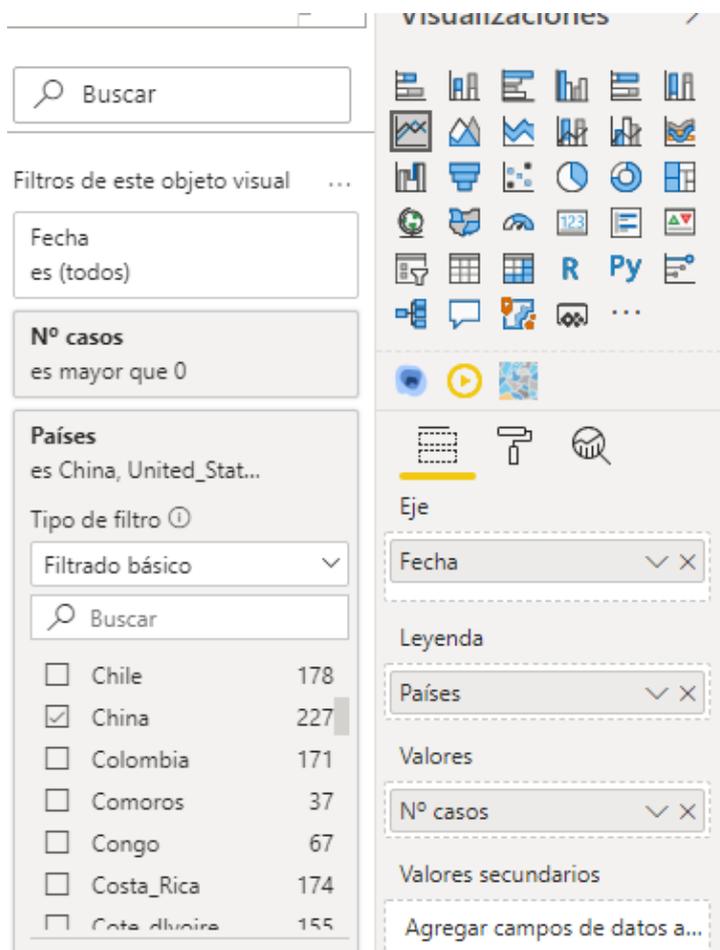
En este momento se visualizará una línea por cada país existente en la tabla "Mundial Fechas", lo cual emborrona bastante visión y no deja realizar un análisis claro de los datos.

Se deciden qué países analizar, los cuales en este caso serán los siguientes:

- China: Por ser el país originario de la pandemia
- Italia: Uno de los países europeos más azotados por la pandemia
- España: Similar a Italia y además ser el país del que se está realizando este análisis
- India: Como el país con más crecimiento de casos en este momento
- Estados Unidos: El país con más casos del mundo de COVID-19

Una vez decididos los países pasaremos a filtrarlos desde la ventana de “Filtros de este objeto visual” que se encuentra a la izquierda de las visualizaciones. Esta ventana permite realizar todo tipo de filtrados: filtrados básicos (como el que vamos a hacer) o filtrados más avanzados (top N, mayor que, está en blanco, etc)

Figura 51. Filtros de la visualización



En el filtro básico seleccionaremos únicamente los países que se han mencionado anteriormente para que así sean los únicos que aparezcan en el gráfico de líneas.

Se pueden comprobar ya que únicamente aparecen 5 países en la leyenda del gráfico y 5 líneas.

Por último solamente nos faltaría conseguir animar el gráfico para que se pueda apreciar la evolución simultánea del número de casos en cada país.

Para ello necesitaremos de una visualización que no se encuentra de manera predeterminada en Power BI, por lo que habrá que traerla desde la AppSource.

Para traer nuevos objetos visuales pulsaremos en los tres puntos que se encuentran en la

ventana “Visualizaciones” → “Obtener más objetos visuales”

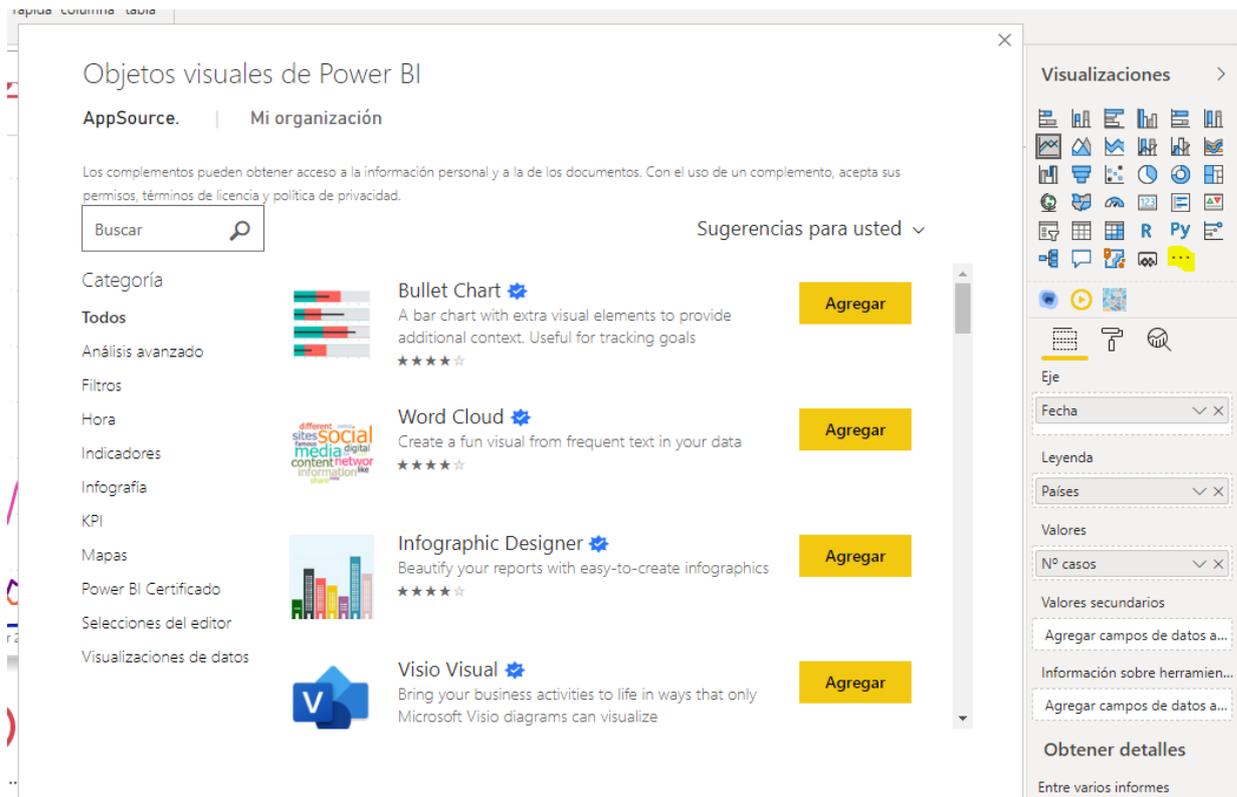


Figura 52. Objetos visuales de Power BI

Esta visualización lo que realiza es una iteración sobre el campo fecha que le pasemos de forma automática. Si probáramos con el campo “date” de la tabla “Fecha” y le diésemos al “Play” no veríamos el gráfico de líneas si no solamente un punto, el punto que represente el número de casos en ese día concreto. Teniendo esto en cuenta y que lo que queremos mostrar son el conjunto de días anteriores a ese día se van a realizar ciertos ajustes.

El primero es crear una tabla auxiliar de fecha de la misma manera que se creó la primera (con la fórmula de CALENDARAUTO()) y no se ligará a nada. Con esto se pretende tener una tabla que no es influenciada por ninguna otra, y su campo “date” se lo pasaremos a la nueva visualización.

A continuación creamos una medida como la siguiente:

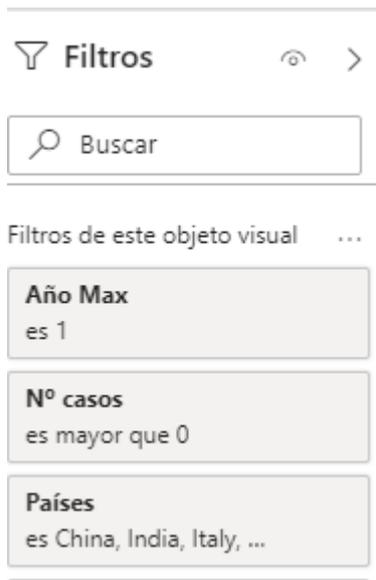
```

1 Año Max =
2 IF(
3     MAX(Fecha[Date]) <= MAX('Fecha LIBRE'[Date]),
4     1,
5     0
6 )

```

Figura 53. Medida para la visualización

Con esta medida lo que queremos es que se nos devuelva un 1 si la fecha de la tabla “Fecha” es menor que la de la tabla “Fecha LIBRE” en ese momento. Como vamos a iterar sobre la tabla “Fecha LIBRE” con la nueva visualización, este valor cambiará de manera dinámica para el resto de las fechas.



Por último, al gráfico de líneas se le debe añadir en el campo “Información sobre herramientas” esta medida. Una vez realizado esto en los filtros del gráfico se le dirá que muestre aquellos “Año Max” cuyo valor sea 1.

Ahora cuando presionemos “Play”, el gráfico responderá de la manera que deseamos ya que avanzará sobre la tabla “Fecha LIBRE” y todas aquellas fechas de la tabla “Fecha” que sean menores, tendrán valor 1.

Figura 54. Filtros de la visualización

Página 3 – Tabla COVID-19

Siguiendo con la idea de la tabla vista en la página de “Worldometers” de la que se han extraído los datos de la tabla “Total Mundial”, se va a intentar replicar la idea haciéndola algo más visible e intuitiva para el usuario medio.

Una tabla de datos no es una visualización muy recomendable para la representación de datos numéricos ya que aunque sí que muestran los distintos valores, no hay ninguna relación visual entre ellos.

En esta página se va a mostrar cómo intentar reducir este efecto para convertir la tabla en algo más llamativo para la vista.

Comenzamos arrastrando la visualización de la tabla a la nueva página y añadiéndoles los valores que deseamos ver. Todos ellos van a ser de la tabla “Total Mundial” y serán estas columnas:

- “Country, Others”: para traer los países a la tabla
- “TotalCases”: nos mostrará el acumulado de casos del país
- “TotalDeaths”: similar al anterior, pero con las muertes
- “NewCases”: campo que indica el número de casos nuevos respecto al día anterior
- “NewDeaths”: al igual que anteriormente, número de muertes nuevas respecto al día anterior
- “TotalTests”: el número de tests realizados por cada país
- “Serious, Critical”: corresponde al número de ingresos en la unidad de cuidados intensivos (UCI) o la correspondiente unidad

Actualmente la tabla debería encontrarse con datos pero, como he indicado anteriormente, poco llamativos a la vista.

Lo primero que se podría hacer es intentar jugar con los colores de los datos para que, por ejemplo, los países tengan un fondo de cierto color que el resto de los valores no. De esta forma se le consigue dar un aspecto de matriz a la tabla que indique que la primera columna, la de países es la que manda sobre el resto.

Para realizar esto nos dirigiremos al valor del “País” y pulsaremos sobre la flecha descendente y en el formato condicional, sobre color de fondo. De ahí se elige el color que se desee para esta columna.

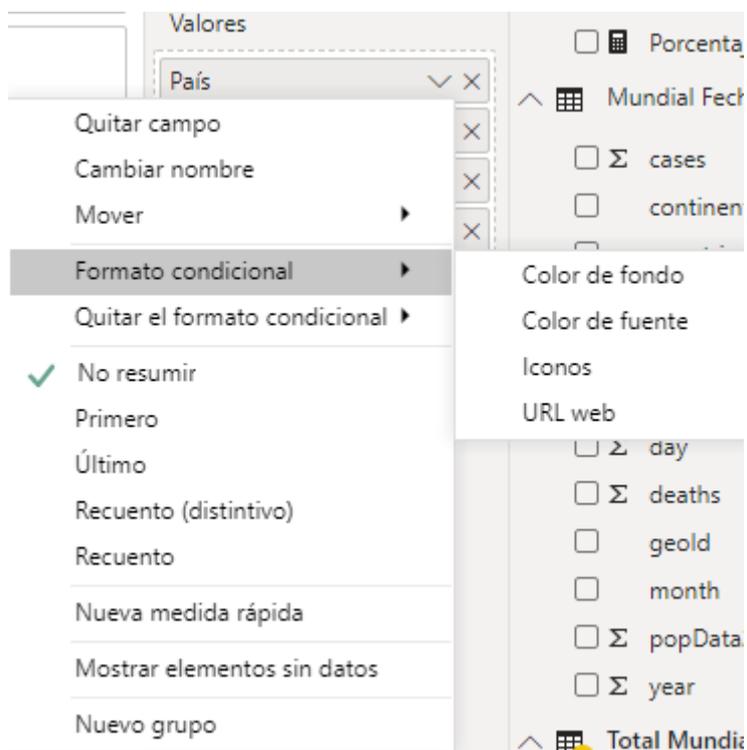


Figura 55. Formato condicional

Con los campos de “Casos nuevos” y “Muertes nuevas” se ha decidido realizar algo muy semejante cambiando en lugar del color de fondo, el color de la fuente. Si se tratara de un campo que pudiera tener valores tanto negativos como positivos, se podría configurar los colores para que, por ejemplo, los valores positivos fueran de color verde y los negativos de color rojo.

Por último, respecto a los campos “Casos totales” y “Muertes totales” se va a añadir a la tabla un gráfico de barras horizontal que ayude a comparar entre los datos de sus propias columnas los valores que manejan. Además, como la tabla puede ordenarse de mayor a menor, se logrará un efecto descendente en este gráfico de barras que lo hará muy llamativo a la vista.

Para lograr esto, el proceso es el mismo tanto para los casos como para las muertes. Nos dirigiremos a “Formato condicional” (al igual que anteriormente) y ahora sobre “barra de datos”. La ventana emergente será muy parecida a la que salía cuando se realizó el mapa de calor para la primera página de este proyecto. En esta ventana podremos elegir el tipo de gráfico de barras que nos salga, el color de la misma y desde qué valores queremos que empiece y acabe.

Una vez terminado todo esto, el resultado final de la página sería semejante al siguiente:

Tabla comparativa entre distintos países

País	Casos totales	Casos nuevos	Muertes totales	Muertes nuevas	Total teses	Ingresos en UCI
USA	6.257.938	+367	188.902	+2	83.353.338	15.079
Brazil	3.952.790		122.681		14.352.484	8.318
India	3.769.523	+3.415	66.460		44.337.201	8.944
Russia	1.000.048		17.299		36.900.000	2.300
Peru	657.129		29.068		3.233.034	1.493
South Africa	628.259		14.263		3.705.408	539
Colombia	624.069		20.052		2.777.107	1.493
Mexico	606.036	+6.476	65.241	+827	1.360.123	3.275
Spain	470.973		29.152		9.210.337	823
Argentina	428.239		8.913		1.277.751	2.314
Chile	413.145		11.321		2.458.762	962
Iran	376.894		21.672		3.256.122	3.709

Figura 56. Resultado de la tabla

Como se puede ver, se trata de una tabla mucho más visual que la original en la que, a pesar de analizarse los mismos datos, estos se asimilan mucho más rápido.

Página 4 – Estudios de casos

En esta cuarta página, la de estudio de casos, se pretende realizar un estudio un poco más específico que abarque también otros aspectos de la pandemia, como podría ser la realización de teses que comprueben si existe infección con COVID-19.

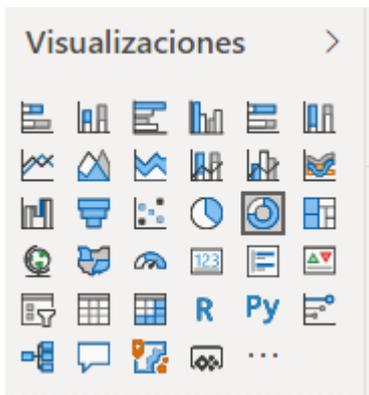
Por la parte de filtros esta página va a resultar muy similar a la primera, la de efectos mundiales de la pandemia. Con esto se quiere decir que se van a utilizar los mismos filtros de fechas y continentes, los cuales son:

- Filtro de selección entre fechas
- Filtro de meses
- Filtro de años
- Filtro de continentes

Como provienen de las mismas tablas, se pueden volver a copiar en esta misma página o realizar un duplicado de la página anterior y eliminar las graficaciones que no se vayan a utilizar.

Comenzaremos con la realización del gráfico de anillos. El gráfico de anillos, muy similar al gráfico de tarta, permite visualizar de manera sencilla porcentajes sobre un total. En este caso se va a realizar una representación del número de casos por mes.

Figura 57. Visualización anillo



Para empezar, se trae el gráfico de anillos desde la pestaña de visualizaciones, el cual permite agregarle tres parámetros: leyenda, detalles y valores. En este caso solamente vamos a utilizar dos: leyenda y valores.

Ambos campos van a ser traídos de la tabla “Mundial Fechas” y serán el campo “month” para la leyenda y el campo “cases” para el parámetros valores.

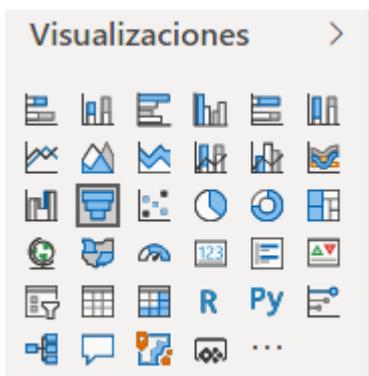
Inmediatamente se creará el gráfico de anillos que mostrará cómo se han repartido el número de casos en los distintos meses de una manera muy representativa.

Para continuar analizando los casos de COVID-19, se ha decidido que es interesante comparar el número de casos y muertes por continente. Su realización se hará mediante el uso de una matriz, visualización que ya se ha usado previamente y se encuentra a la derecha de la visualización de la tabla.

En esta matriz añadiremos al campo filas los continentes que se encuentran en “continentExp” y en el campo valores los campos “cases” y “deaths” respectivamente. Como se realizó en la página 3, podemos añadir un formato condicional al número de casos para incluir una gráfica de barras lateral que nos compare el número de casos entre los distintos continentes.

La siguiente visualización por realizar va a ser un embudo que refleje el top 10 países con teses más realizadas por millón de habitantes. De esta manera conseguiremos información de qué países se han involucrado más en la realización de teses independientemente del número de personas que habiten en dicho país. Es una manera de normalizar los datos y poder comparar los distintos países entre sí.

Figura 58. Filtro embudo



Una vez elegido el embudo, añadimos al grupo el campo “Country, Other” de la tabla “Total Mundial” y al campo valores “Test/1M pop” de la misma tabla.

Como podemos comprobar, se muestran una cantidad muy superior a 10 países en este filtro, lo que dificulta su interpretación y hace del gráfico algo muy poco llamativo. Para subsanar este problema se hará uso de los filtros del propio gráfico, que se encuentran a la izquierda de este mismo.

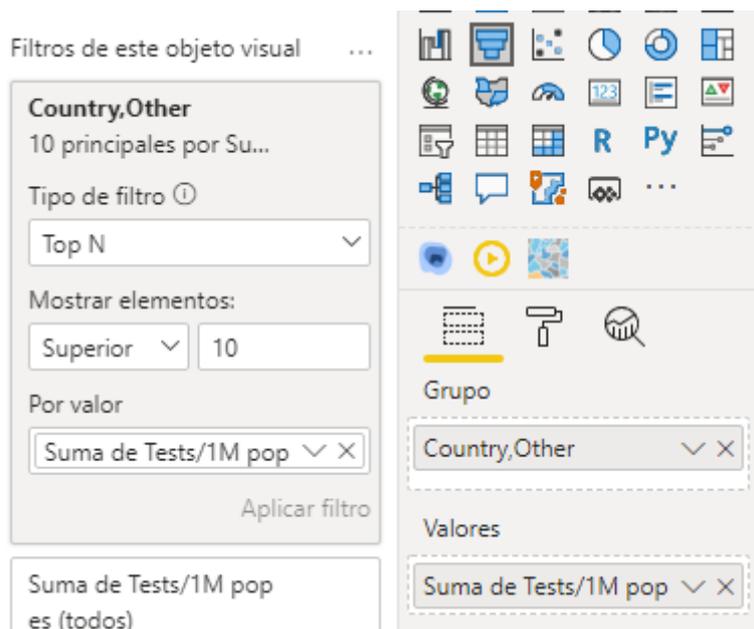


Figura 59. Filtro visualización

Como se muestra en la imagen superior, se realizará un filtro del tipo “Top N”, el cual nos mostrará los valores inferiores o superiores que se le indiquen con respecto a un valor en específico. En este caso se quiere un top 10 (superior) de la suma de valores de “Test/1 M pop”.

Con este pequeño ajuste el filtro vuelve a un estado más sencillo y por lo tanto más visual en el que se puede comparar fácilmente los distintos países con más teses realizados.

El resultado final de la página es el siguiente:

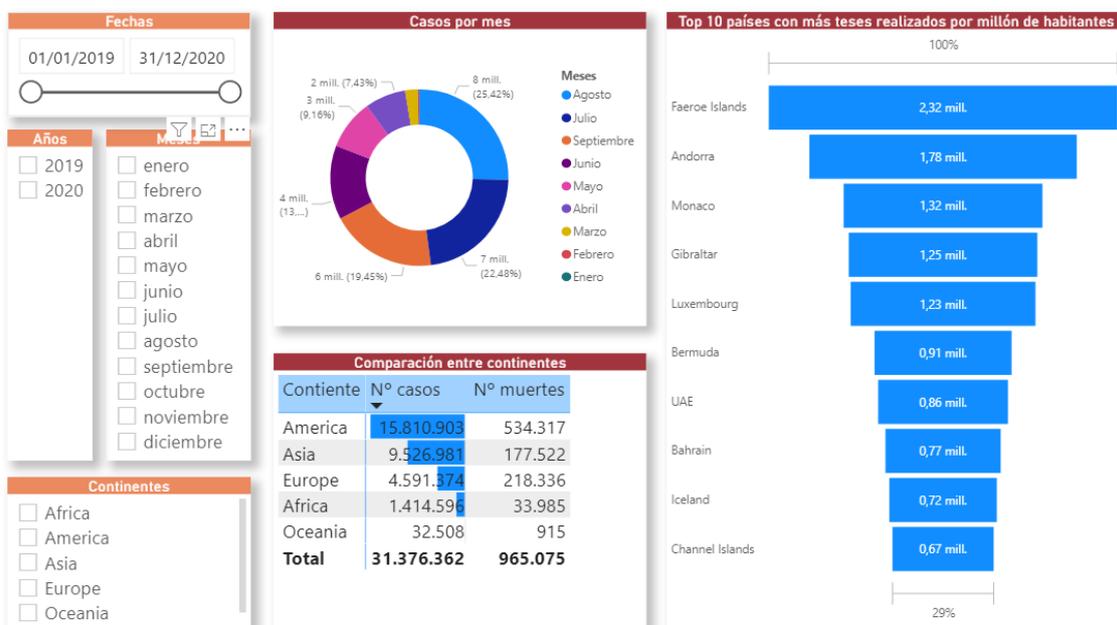


Figura 60. Resultado de la página

Para concluir con este primer informe de Power BI se ha querido realizar una evolución temporal de los niveles de restricción que han ido imponiendo los países conforme avanzaba la pandemia.

Gracias al campo “stringency_index” de la tabla “GitHub Mundial”, podemos saber en las distintas fechas qué índice de restricción existía en el país (de 0 a 100).

Se crea un mapa coroplético al que se añade como ubicación el campo “location” de la tabla “GitHub Mundial” y como información sobre herramientas el campo “stringency”. Es importante comprobar que no realice la suma de este campo si no que realice la media. Esto se debe a que hay múltiples entradas (una por día) y realizar la suma nos daría valores muy superiores a 100. Es por ello por lo que realizando el promedio volvemos a normalizar los valores entre el 0 y el 100.

Dentro del mapa realizaremos un esquema de color como en la primera página, que quede similar al de la siguiente imagen:

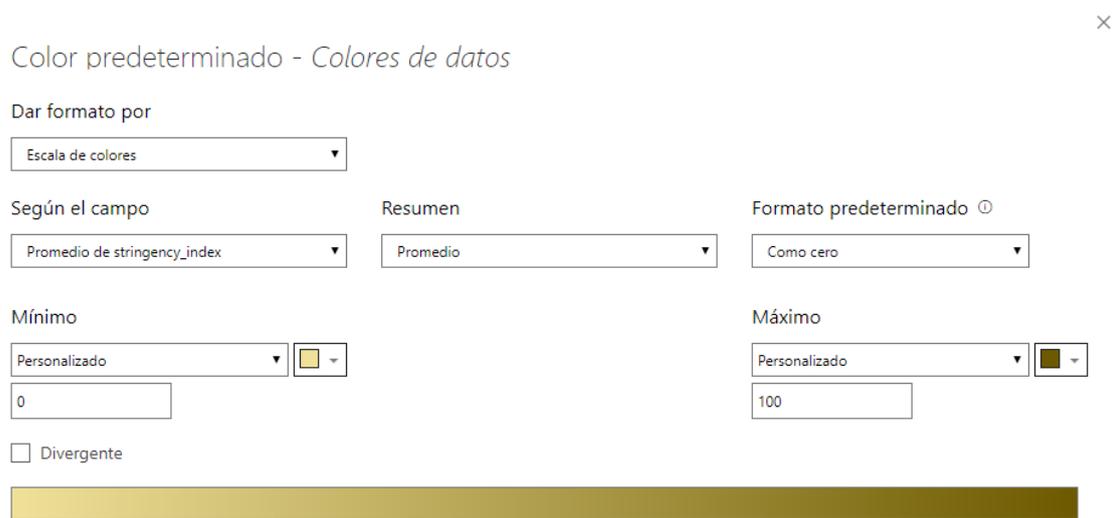


Figura 61. Mapa de colores

Como se puede ver, se realiza este esquema de colores de acuerdo al promedio de “stringency_index” con un esquema de colores que va desde el valor 0 hasta el valor 100.

Una vez conseguido todo esto, falta añadir la animación que permita ver la evolución de los países en el tiempo. Siguiendo la visualización traída de la “store” la añadimos a la página junto al campo “date” de la misma tabla “GitHub Mundial”.

Por último, para poder guiarnos en qué fecha estamos en ese momento, añadiremos una tarjeta con el mismo campo. Esta tarjeta cambiará de fecha en cuanto la visualización anterior avance a la siguiente fecha registrada.

Con todo lo anterior realizado, la página queda tal que así:

Evolución de medidas restrictivas de manera global

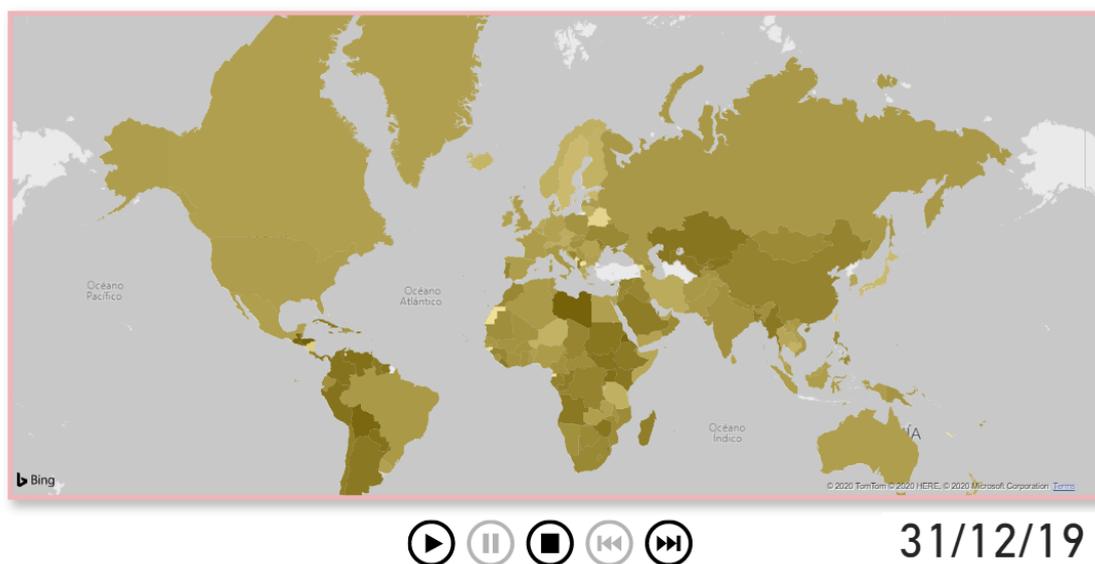


Figura 62. Resultado mapa de color

Proyecto de estudio de la economía durante la pandemia del COVID-19

Una vez concluido el anterior proyecto en el cual se ha estudiado la pandemia mundial del COVID-19, se va a realizar un segundo proyecto para estudiar los aspectos económicos mundiales y nacionales que hayan podido verse afectados por la pandemia.

Al tratarse del segundo proyecto de este trabajo y de que en el primero se ha estado explicando durante una gran parte del tiempo el funcionamiento de la herramienta de Power BI, en este segundo se avanzará de una forma más rápida y ligera que en el primero.

Además de todo ello, hay orígenes de datos reutilizados para la realización de este proyecto que no requerirán de explicación puesto que son los mismos.

Estos son los siguientes:

- Tabla “PANDEMIA – Mundial Fechas”: En el proyecto anterior correspondía a la tabla “Mundial Fechas”, con información por país y fecha de los avances del COVID-19
- Tabla “GitHub – Mundial”: El origen de datos sacados del repositorio de Git-Hub que contiene información por fechas y países con varios aspectos importantes complementarios a los de la tabla anterior

Orígenes de datos

BOLSA – Evolución de la bolsa

Para poder realizar un estudio del aspecto económico mundial, es imprescindible analizar cómo ha evolucionado la bolsa durante el tiempo que ha afectado la pandemia. Como en el mundo existen decenas de bolsas, se ha realizado una selección de aquellas bolsas principales a nivel mundial que podrían reflejar mejor cómo ha afectado a la economía esta pandemia.

Estas bolsas son las siguientes:

- Dow Jones – EE.UU.
- NASDAQ – EE.UU.
- S&P 500 – EE.UU.
- NIKKEI 255 – Japón
- IBEX35 – España

Para conseguir información de las mismas nos valdremos de la página web de Yahoo habilitada para el estudio de finanzas <https://es.finance.yahoo.com>

Dentro de esta página conseguiremos información de las distintas bolsas con los siguientes links:

- <https://es.finance.yahoo.com/quote/%5EIBEX/history/>
- <https://es.finance.yahoo.com/quote/%5EIXIC?p=%5EIXIC>
- <https://es.finance.yahoo.com/quote/%5EN225/history?p=%5EN225>
- <https://es.finance.yahoo.com/quote/%5EGSPC/history?p=%5EGSPC>
- <https://es.finance.yahoo.com/quote/%5EDJI/history?p=%5EDJI>

Los datos que encontraremos están estructurados de la siguiente manera: para cada día existen registros de cómo abrió la bolsa, qué máximo obtuvo así como el mínimo y cómo cerro. También el volumen que se manejó en ese día.

Figura 63. Datos bolsa

Divisa en EUR [↓ Descargar](#)

Fecha	Abrir	Máx.	Mín.	Cierre*	Cierre ajust.**	Volumen
04 sept. 2020	7015,70	7133,40	6954,70	6989,70	6989,70	462.539.120
03 sept. 2020	7065,20	7159,60	6971,20	7006,00	7006,00	246.671.100
02 sept. 2020	6973,10	7046,10	6963,90	6996,90	6996,90	208.505.300
01 sept. 2020	7038,50	7042,00	6903,00	6956,90	6956,90	226.079.800
31 ago. 2020	7174,30	7198,30	6960,00	6969,50	6969,50	205.260.600
28 ago. 2020	7114,70	7162,90	7055,10	7133,00	7133,00	193.354.900
27 ago. 2020	7114,10	7130,20	7029,20	7090,70	7090,70	162.183.000
26 ago. 2020	7088,40	7136,50	7070,40	7123,00	7123,00	122.064.500
25 ago. 2020	7147,90	7215,10	7103,20	7108,40	7108,40	189.748.500
24 ago. 2020	7044,20	7120,80	7037,30	7109,10	7109,10	147.789.200
21 ago. 2020	7016,70	7025,90	6900,50	6982,10	6982,10	171.116.500
20 ago. 2020	7011,40	7034,50	6974,90	6993,30	6993,30	155.668.600
19 ago. 2020	7039,10	7094,30	7002,80	7094,30	7094,30	132.230.000
18 ago. 2020	7065,70	7165,60	7024,40	7043,50	7043,50	142.775.200

Con respecto al paro se ha encontrado un fichero Excel del servicio público de empleo estatal que nos permite comparar mes a mes el número de parados con respecto al año anterior. Esto nos servirá para poder comparar el año anterior a la pandemia con el de la misma y ver así el impacto de la misma en la contratación de trabajadores.

Este Excel se puede encontrar en la siguiente dirección:

https://www.sepe.es/SiteSepe/contenidos/que_es_el_sepe/estadisticas/datos_avance/xls/empleo/evolparo.xls

Dentro nos encontramos una tabla como la siguiente:

Figura 64. Datos paro

Evolución del paro registrado

	AÑO 2019			AÑO 2020		
	PARO REGISTRADO	% DE VARIACIÓN		PARO REGISTRADO	% DE VARIACIÓN	
		MENSUAL	ANUAL		MENSUAL	ANUAL
ENERO	3.285.761	2,61	-5,49	3.253.853	2,85	-0,97
FEBRERO	3.289.040	0,10	-5,22	3.246.047	-0,24	-1,31
MARZO	3.255.084	-1,03	-4,89	3.548.312	9,31	9,01
ABRIL	3.163.566	-2,81	-5,17	3.831.203	7,97	21,1
MAYO	3.079.491	-2,66	-5,31	3.857.776	0,69	25,27
JUNIO	3.015.686	-2,07	-4,63	3.862.883	0,13	28,09
JULIO	3.011.433	-0,14	-3,94	3.773.034	-2,33	25,29
AGOSTO	3.065.804	1,81	-3,65	3.802.814	0,79	24,04
SEPTIEMBRE	3.079.711	0,45	-3,83			
OCTUBRE	3.177.659	3,18	-2,37			
NOVIEMBRE	3.198.184	0,65	-1,68			
DICIEMBRE	3.163.605	-1,08	-1,21			

ERES y ERTES

De mano del estudio del paro, este año ha sido gran noticia la concesión de EREs (expedientes de regulación de empleo) y ERTES (expedientes de regulación de empleo temporal) por parte de la Dirección General de Trabajo.

Estas medidas no son nuevas y llevan existiendo desde hace varias décadas, no obstante sí que es novedad que se hayan concedido este gran número de EREs y ERTES.

En la siguiente página podremos encontrar información que aprovecharemos para realizar nuestros propios gráficos:

<https://www.epdata.es/datos/trabajadores-afectados-ere-graficos/450>

Empresas inscritas en la seguridad social

El COVID-19 no sólo ha afectado a los trabajadores, a las empresas también les ha afectado muy duramente. Teniendo en cuenta que en España la mayoría de las empresas son de menos de 5 trabajadores, una pandemia ha logrado que miles de estas empresas desaparezcan.

Para lograr un estudio de las empresas inscritas en la seguridad social recurriremos a la secretaría de estado de empleo y economía social en la siguiente dirección, que nos permite obtener los datos que necesitamos:

<http://www.mites.gob.es/estadisticas/emp/welcome.htm>

Dentro del Excel que nos devuelve la página, nos interesa aquella pestaña denominada “EMP-E1”. De aquí podremos comparar año tras año el número de empresas en la seguridad social de acuerdo al tamaño de las mismas.

EMPRESAS INSCRITAS EN LA SEGURIDAD SOCIAL								EMP
EMP- E1. EMPRESAS, SEGÚN TAMAÑO DE LA EMPRESA								
AÑOS (1)	TOTAL	De 1 a 2 trabajadores	De 3 a 5 trabajadores	De 6 a 9 trabajadores	De 10 a 49 trabajadores	De 50 a 249 trabajadores	De 250 a 499 trabajadores	Más de 499 trabajadores
2015	1.286.565	713.587	286.164	124.547	137.178	21.010	2.222	1.857
2016	1.312.345	718.275	294.121	129.519	143.941	22.238	2.306	1.945
2017	1.326.161	717.403	297.673	133.380	149.850	23.357	2.435	2.063
2018	1.346.629	720.549	302.539	138.043	156.242	24.559	2.535	2.162
2019	1.340.415	714.284	301.576	137.417	157.113	25.154	2.633	2.238
2018:								
Jul.	1.325.845	714.189	293.985	133.286	155.120	24.729	2.464	2.072
Ago.	1.311.299	709.934	289.870	130.650	152.202	24.170	2.425	2.048
Sep.	1.335.474	720.275	297.162	133.741	154.618	25.006	2.545	2.127
Oct.	1.320.683	716.072	292.526	130.091	152.529	24.772	2.549	2.144
Nov.	1.324.883	716.237	294.412	131.463	153.315	24.730	2.578	2.148
Dic.	1.346.629	720.549	302.539	138.043	156.242	24.559	2.535	2.162
2019:								
Ene.	1.326.961	713.369	295.843	133.803	154.738	24.508	2.575	2.125
Feb.	1.324.997	715.655	294.639	131.141	153.962	24.865	2.583	2.152
Mar.	1.337.695	720.840	297.713	132.950	155.981	25.387	2.619	2.205
Abr.	1.338.348	720.455	297.438	133.067	156.844	25.680	2.645	2.219
May.	1.344.083	720.223	298.987	134.905	159.015	26.058	2.680	2.215
Jun.	1.349.325	719.788	300.902	136.527	161.057	26.245	2.613	2.193
Jul.	1.328.629	709.368	295.551	134.469	158.711	25.818	2.561	2.151
Ago.	1.323.342	707.837	293.909	133.655	157.562	25.653	2.570	2.156
Sep.	1.327.869	712.069	294.617	133.811	156.799	25.745	2.650	2.178
Oct.	1.318.915	710.459	292.025	131.272	154.816	25.488	2.650	2.205
Nov.	1.332.983	713.037	297.350	134.249	157.765	25.623	2.715	2.244
Dic.	1.340.415	714.284	301.576	137.417	157.113	25.154	2.633	2.238
2020:								
Ene.	1.318.325	707.742	293.648	132.138	154.861	25.050	2.674	2.212
Feb.	1.324.427	710.752	294.792	132.284	156.063	25.576	2.696	2.264
Mar. (2)	1.238.554	675.961	271.711	120.103	142.028	23.974	2.580	2.197
Abr.	1.233.187	670.679	271.210	120.074	142.387	24.066	2.559	2.212
May.	1.259.417	682.944	278.501	123.213	145.581	24.388	2.577	2.213
Jun.	1.273.339	688.018	283.134	126.208	147.459	23.930	2.495	2.095
Jul.	1.282.346	689.104	286.488	128.434	149.628	24.123	2.468	2.101

(1) Los datos anuales están referidos a último día hábil del año y los mensuales a último día hábil del mes.

Figura 65. Datos empresas inscritas

Inteligencia de negocio

ETL BOLSA – Evolución de la bolsa

Carga de datos en Power BI

La idea que seguir para la utilización de estos datos es la siguiente:

1. Traer los datos de una de las bolsas a Power BI
2. Transformar los tipos al tipo que corresponde
3. Añadir una columna que identifique el país al que corresponde esa bolsa
4. Realizar los mismos pasos con el resto de bolsa
5. Anexar todas las consultas en una única tabla
6. Deshabilitar el resto de las tablas con información de las bolsas exceptuando la del paso número 5

Como se ha realizado en ocasiones anteriores, partiremos de la obtención de datos a través de la web. La URL que se incluirá dependerá de la bolsa que estemos trayendo a Power BI en este momento (los links ya han sido incluidos en este informe).

Transformación de datos

Las columnas que nos interesan para la realización del estudio son la de “Date” que será de tipo fecha, “Open” y “Volume” de tipo entero y una última columna que no se encuentra originalmente entre los datos pero que añadiremos nosotros por medio de la opción “Columna personalizada”.

Esta columna personalizada se va a utilizar para que, una vez anexemos todas las consultas, poder diferenciar qué valores corresponden a qué bolsa. Las columnas personalizadas se escriben en lenguaje M, aunque en este caso únicamente se quiere añadir "IBEX 35" a todas las columnas.

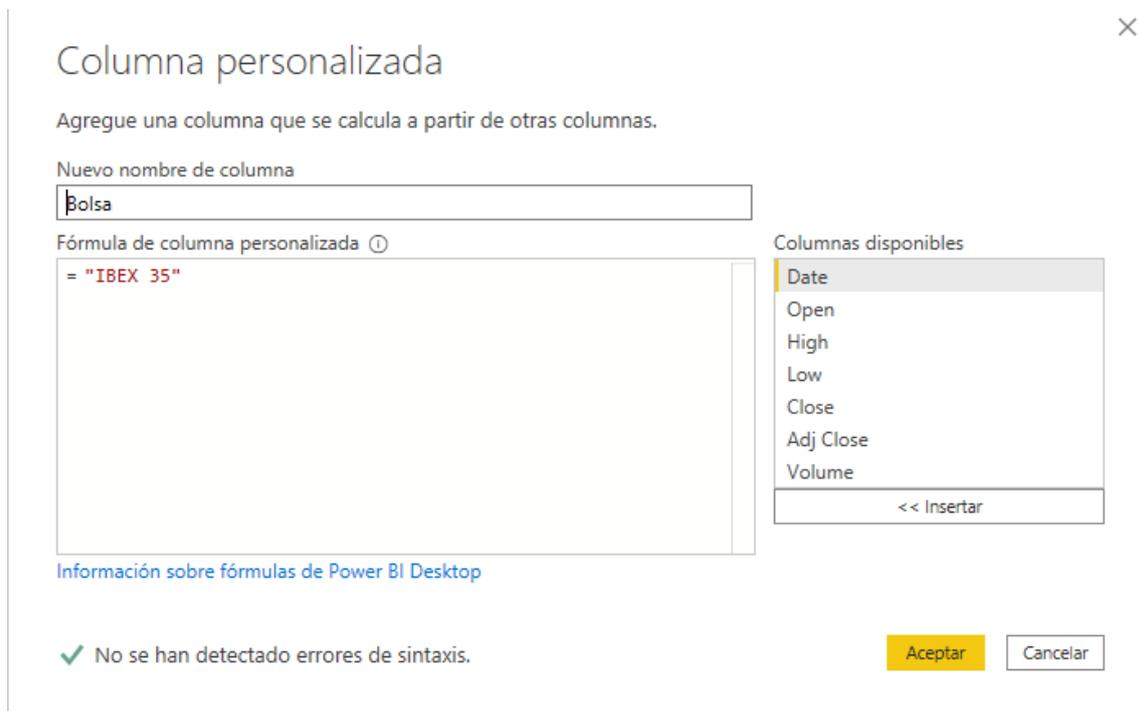


Figura 66. Creación columna personalizada

De esta manera sencilla, todas las entradas de la tabla poseen una nueva columna llamada "Bolsa" que identifican la bolsa a la que pertenecen.

Este proceso lo repetiremos con todas las bolsas cambiando en la columna bolsa el texto de la columna por el de la bolsa que estemos implementando en ese momento.

Una vez tengamos todas las tablas traídas en Power BI desactivaremos su carga para evitar que nos aparezcan en el proceso de graficación y modelaje.

El último paso que falta es la anexión de todas las tablas en una única. Comenzaremos por duplicar una de las tablas que contengan información de la bolsa, la que queramos. Se renombra la tabla a "BOLSA – Evolución de la bolsa" y procederemos a anexar las tablas.

En este proceso únicamente tenemos que indicar qué tablas son aquellas que deseamos anexar y estas se traerán automáticamente detectando las columnas que coinciden (todas).

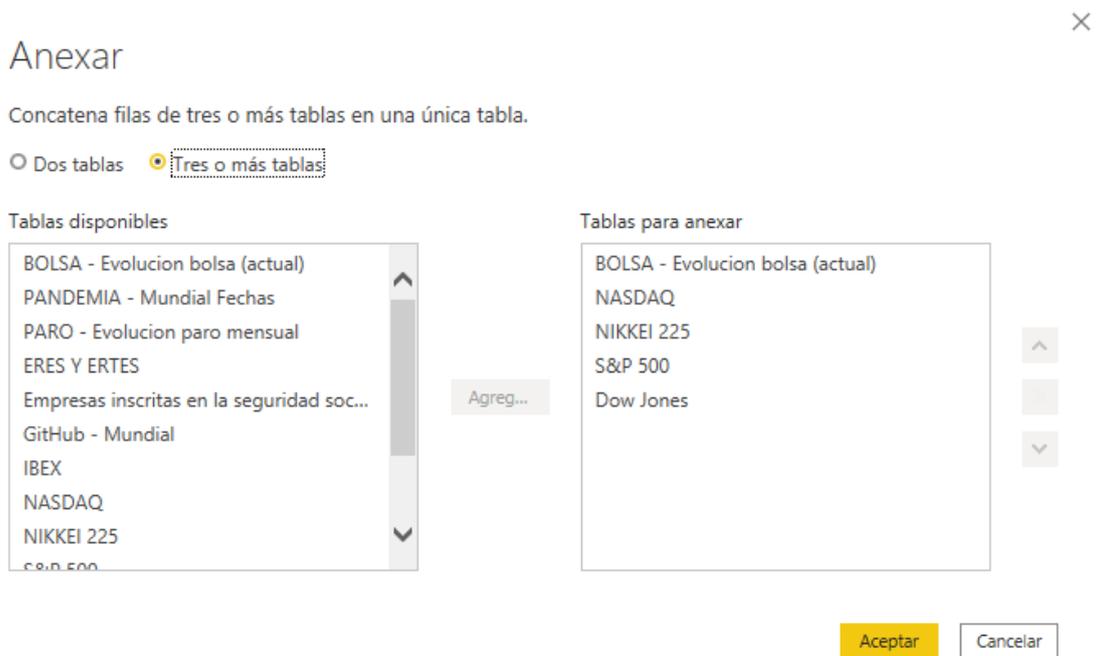


Figura 67. Anexar consultar

Una vez realizado este proceso podemos comprobar cómo tenemos en una única tabla toda la información de todas las bolsas, que se puede identificar por la columna “Bolsa”.

ETL PARO – Evolución paro mensual

Carga de datos en Power BI

Siguiendo la idea de tener todos los datos de forma online, obtendremos los datos (en este caso un fichero Excel) de la siguiente página:

https://www.sepe.es/SiteSepe/contenidos/que_es_el_sepe/estadisticas/datos_avance/xls/empleo/evolparo.xls

De este modo cualquier actualización que realicen sobre el fichero la reflejaremos en el proyecto automáticamente al refrescar los datos.

Transformación de datos

Los datos que trae este Excel están muy sucios, por lo que se van a limpiar hasta el punto de poder trabajar con ellos:

	A ^B _C Column1	A ^B _C Column2	A ^B _C Column3	A ^B _C Column4	A ^B _C Column5	A ^B _C Column6	A ^B _C Column7	A ^B _C Column8
1		null	null	null	null	null	Servicio Público de Empleo Estatal	null
2		null	null	null	null	null		null
3		null	Evolución del paro registrado	null	null	null	null	null
4		null						
5		null	null	AÑO 2019	null	null	AÑO 2020	null
6		null	null	PARO REGISTRADO	% DE VARIACIÓN	null	PARO REGISTRADO	% DE VARIACIÓN
7	Cod Nomb Mes		null	MENSUAL	ANUAL	null	MENSUAL	ANUAL
8	1	ENERO	3.285.761	2,61	-5,49	3.253.853	2,85	-0,9
9	2	FEBRERO	3.289.040	0,10	-5,22	3.246.047	-0,24	-1,3
10	3	MARZO	3.255.084	-1,03	-4,89	3.548.312	9,31	9,01
11	4	ABRIL	3.163.566	-2,81	-5,17	3.831.203	7,97	21,1
12	5	MAYO	3.079.491	-2,66	-5,31	3.857.776	0,69	25,2
13	6	JUNIO	3.015.686	-2,07	-4,63	3.862.883	0,13	28,0
14	7	JULIO	3.011.433	-0,14	-3,94	3.773.034	-2,33	25,2
15	8	AGOSTO	3.065.804	1,81	-3,65		null	null
16	9	SEPTIEMBRE	3.079.711	0,45	-3,83		null	null
17	10	OCTUBRE	3.177.659	3,18	-2,37		null	null
18	11	NOVIEMBRE	3.198.184	0,65	-1,68		null	null
19	12	DICIEMBRE	3.163.605	-1,08	-1,21		null	null

Figura 68. Limpieza de datos (1)

Como se puede ver, nos interesan a partir de la línea 8 los datos y en específico las columnas que indican el paro registrado, no aquellas que marquen el % de variación. No obstante hay que tener cuidado para no borrar la fila que contiene el año al que corresponde la columna.

Realizando la eliminación de las columnas superiores llegamos a un punto tal que así:

	A ^B _C Column2	A ^B _C Column3	A ^B _C Column4	A ^B _C Column5	A ^B _C Column6	A ^B _C Column7	A ^B _C Column8
	null	AÑO 2019	null	null	AÑO 2020	null	null
	null	PARO REGISTRADO	% DE VARIACIÓN	null	PARO REGISTRADO	% DE VARIACIÓN	null
	null	null	MENSUAL	ANUAL	null	MENSUAL	ANUAL
	ENERO	3.285.761	2,61	-5,49	3.253.853	2,85	-0,97
	FEBRERO	3.289.040	0,10	-5,22	3.246.047	-0,24	-1,31
	MARZO	3.255.084	-1,03	-4,89	3.548.312	9,31	9,01
	ABRIL	3.163.566	-2,81	-5,17	3.831.203	7,97	21,1
	MAYO	3.079.491	-2,66	-5,31	3.857.776	0,69	25,27
	JUNIO	3.015.686	-2,07	-4,63	3.862.883	0,13	28,09
	JULIO	3.011.433	-0,14	-3,94	3.773.034	-2,33	25,29
	AGOSTO	3.065.804	1,81	-3,65		null	null
	SEPTIEMBRE	3.079.711	0,45	-3,83		null	null
	OCTUBRE	3.177.659	3,18	-2,37		null	null
	NOVIEMBRE	3.198.184	0,65	-1,68		null	null
	DICIEMBRE	3.163.605	-1,08	-1,21		null	null

Figura 69. Limpieza de datos (2)

El siguiente paso para aplicar va a ser la transposición de la tabla para poder limpiar los datos de una manera mejor. Una vez se ha alcanzado un nivel alto de limpieza, la tabla queda algo parecida a esta:

ABC 123 Personalizado	ABC 123 Column4	ABC 123 Column5	ABC 123 Column6	ABC 123 Column7	ABC 123 Column8	ABC 123 Column9	ABC 123 Column10
	null	ENERO		ABRIL	MAYO	JUNIO	JULIO
AÑO 2019-PARO REGISTRADO...	3.285.761	3.289.040	3.255.084	3.163.566	3.079.491	3.015.686	3.011.433
AÑO 2019-% DE VARIACIÓN...	2,61	0,10	-1,03	-2,81	-2,66	-2,07	-0,14
AÑO 2019-% DE VARIACIÓN...	-5,49	-5,22	-4,89	-5,17	-5,31	-4,63	-3,94
AÑO 2020-PARO REGISTRADO...	3.253.853	3.246.047	3.548.312	3.831.203	3.857.776	3.862.883	3.773.034
AÑO 2020-% DE VARIACIÓN...	2,85	-0,24	9,31	7,97	0,69	0,13	-2,33
AÑO 2020-% DE VARIACIÓN...	-0,97	-1,31	9,01	21,1	25,27	28,09	25,29

Figura 70. Limpieza de datos (3)

En este punto volveremos a transponer la tabla, eliminaremos las columnas que no deseemos y limpiaremos los encabezados para poder tener finalmente la tabla sobre la que trabajaremos:

A ^B _C Mes	A ^B _C Año	1.2 Valor	Fecha
ENERO	2019	3285761	01/01/2019
ENERO	2020	3253853	01/01/2020
FEBRERO	2019	3289040	01/02/2019
FEBRERO	2020	3246047	01/02/2020
MARZO	2019	3255084	01/03/2019
MARZO	2020	3548312	01/03/2020
ABRIL	2019	3163566	01/04/2019
ABRIL	2020	3831203	01/04/2020
MAYO	2019	3079491	01/05/2019
MAYO	2020	3857776	01/05/2020
JUNIO	2019	3015686	01/06/2019
JUNIO	2020	3862883	01/06/2020
JULIO	2019	3011433	01/07/2019
JULIO	2020	3773034	01/07/2020
AGOSTO	2019	3065804	01/08/2019
SEPTIEMBRE	2019	3079711	01/09/2019
OCTUBRE	2019	3177659	01/10/2019
NOVIEMBRE	2019	3198184	01/11/2019
DICIEMBRE	2019	3163605	01/12/2019

Figura 70. Limpieza de datos (4)

ETL ERES y ERTES

Carga de datos en Power BI

En este caso ha resultado imposible obtener los datos directamente desde la página web, por lo que han tenido que ser descargados (en formato CSV) y traídos a Power BI como tal.

Para ello, en obtener datos seleccionaremos desde “fichero de texto/CSV” y añadiremos la ruta en la que se ha descargado los datos de la página de la que se ha hablado al hablar de EREs y ERTES en el punto anterior.

Transformación de datos

La transformación de datos de normal en un archivo CSV no requiere de muchas acciones ya que suelen estar separados por comas y son datos bastante limpios. No obstante, en este caso el fichero ha sido separado por punto coma (“;”) y ha recibido anotaciones de manera manual al final del fichero que dificultan su transformación y que en definitiva, hacen que la limpieza de datos sea algo más costosa.

La transformación que se va a realizar será primordialmente la de dividir columnas mediante un limitador, que será el punto coma. De este modo lograremos separar en tres columnas los datos. Una vez divididas estas columnas, basta con eliminar las filas con datos “basura” y promover encabezados.

Como en el caso de la bolsa, añadiremos una columna que nos determine si el dato corresponde a un ERTE o ERE y combinaremos todos ellos en una única tabla como ya se ha explicado anteriormente.

ETL Empresas inscritas en la seguridad social

Carga de datos en Power BI

Al igual que ha sucedido en el caso anterior, ha sido imposible traer estos datos directamente desde la web, por lo que han tenido que ser descargados para ser introducidos en Power BI.

La carga de datos desde un Excel se puede realizar mediante “Obtener datos” o directamente con el botón de su derecha “Excel”, que permite elegir ficheros de Excel para traer a Power BI. Una vez se ha traído, basta con seleccionar la pestaña que deseamos traer de todo el fichero Excel, ya que contiene decenas de tablas.

Transformación de datos

Los datos de esta pestaña de Excel están muy sucios y no se pueden utilizar directamente en Power BI. Requieren de una limpieza ya que contienen además varios datos que no deseamos, por lo que podremos eliminarlos.

AÑOS (1)	TOTAL	De 1 a 2 trabajadores	De 3 a 5 trabajadores	De 6 a 9 trabajadores	De 10 a 49 trabajadores
Nov.	2320803	733037	297250	124249	1249007
Dic.	1240413	724284	102176	102176	127417
2020:					
Ene.	2328125	707742	293648	123238	1224917
Feb.	1242427	710752	294792	123294	1229499
Mar. (2)	1228541	675561	272111	120203	1200266
Abr.	1233187	670678	272120	120074	1200215
May.	1225417	682944	278501	123213	1200179
Jun.	1273339	688018	283134	126208	1257019
(1) Los datos anuales están referidos a último día hábil del año y los m.					
EMP. E1. EMPRESAS, SEGÚN TAMAÑO DE LA EMPRESA					
Variaciones sobre igual periodo año anterior					
AÑOS (1)	TOTAL	De 1 a 2 trabajadores	De 3 a 5 trabajadores	De 6 a 9 trabajadores	De 10 a 49 trabajadores
2015	2.465060756	0.518077759	5.116588290	7.259918875	7.0
2016	2.003785273	0.658862641	2.780573178	1.892067212	4.9
2017	1.952772946	-0.121401597	2.307066233	2.96102981	4.1
2018	1.544602347	0.438328184	1.634879803	2.496020391	4.2
2019	-0.46144855	-0.869475914	-0.12830807	-0.453481886	0.5
2018:					
Jun.	1.862772175	0.779300138	2.128729839	3.784476239	5.0
Jul.	1.17700514	0.361259051	1.183633573	2.656440054	3.6
Ago.	0.980513812	0.124011144	0.801330621	2.369412118	3.5
Sep.	1.469303245	0.271464866	1.906500759	3.906355524	3.5
Oct.	1.085494768	0.110350281	1.301159849	2.413418827	3.4
Nov.	1.001335622	0.197530864	0.983728014	2.034274538	3.4
Dic.	1.544602347	0.438328184	1.634879803	2.496020391	4.2
2019:					
Ene.	1.120049868	0.189200259	1.185018844	2.407811192	3.6
Feb.	1.3528513	0.226887051	2.66842481	2.844335517	4.2
Mar.	1.217888835	0.284064009	2.307721201	2.521210504	3.6
Abr.	0.789912037	0.034407406	0.805077311	1.788880820	2.7
May.	0.657270551	-0.200781736	0.802748427	2.050773651	2.5
Jun.	0.493708181	-0.42539181	0.805160871	1.517630088	2.6
Jul.	0.209979296	-0.675015199	0.532680039	0.887565086	2.3
Ago.	0.918402387	-0.29537857	1.193188241	2.30009827	3.5
Sep.	-0.565040073	-1.19267079	-0.8546521	0.052193971	1.4
Oct.	-0.121870136	-0.78680974	-0.171266638	0.90762694	1.4
Nov.	0.1117474	-0.446759488	0.99792128	2.115212765	2.5
Dic.	-0.46144855	-0.869475914	-0.12830807	-0.453481886	0.5
2020:					
Ene.	-0.650201965	-0.789321516	-0.743475817	-1.244656718	0.0
Feb.	-0.043018965	-0.08510651	0.05167853	0.87138097	1.3
Mar. (2)	-7.421330684	-6.225930858	-8.73914878	-9.66301221	-8.9
Abr.	-7.857522857	-6.90886725	-8.817972149	-9.7642541	-9.2
May.	-6.295164561	-5.176035756	-6.851802921	-6.66819628	-8.4
Jun.	-5.631408397	-4.413799619	-5.904921983	-7.558212939	-8.4

Figura 71. Datos sin transformar

Para ser concretos, solamente nos interesan los datos de los años 2018, 2019 y 2020, por lo que el resto de las filas que no contengan información relativa a estos años sobra.

Las transformaciones que hay que realizar son bastante sencillas y se basan prácticamente en eliminar filas y quitar columnas. La transformación más especial que se ha realizado ha sido la creación de una columna a partir del mes y el año para más adelante poder transformarlo a tipo fecha y así lograr una marca temporal para cada entrada.

El resultado final quedaría tal que así:

Fecha	Año	Mes	TOTAL	De 1 a 2 trabajadores	De 3 a 5 trabajadores	De 6 a 9 trabajadores
01/06/2018	2018	Jun.	1342696	722863	298509	134486
01/07/2018	2018	Jul.	1325845	714189	293985	133286
01/08/2018	2018	Ago.	1311299	709934	289870	130650
01/09/2018	2018	Sep.	1335474	720275	297162	133741
01/10/2018	2018	Oct.	1320683	716072	292526	130091
01/11/2018	2018	Nov.	1324883	716237	294412	131463
01/12/2018	2018	Dic.	1346629	720549	302539	138043
01/01/2019	2019	Ene.	1326961	713369	295843	133803
01/02/2019	2019	Feb.	1324997	715655	294639	131141
01/03/2019	2019	Mar.	1337695	720840	297713	132950
01/04/2019	2019	Abr.	1338348	720455	297438	133067
01/05/2019	2019	May.	1344083	720223	298987	134905
01/06/2019	2019	Jun.	1349325	719788	300902	136527
01/07/2019	2019	Jul.	1328629	709368	295551	134469
01/08/2019	2019	Ago.	1323342	707837	293909	133655
01/09/2019	2019	Sep.	1327869	712069	294617	133811
01/10/2019	2019	Oct.	1318915	710459	292025	131272
01/11/2019	2019	Nov.	1332983	713037	297350	134249
01/12/2019	2019	Dic.	1340415	714284	301576	137417
01/01/2020	2020	Ene.	1318325	707742	293648	132138
01/02/2020	2020	Feb.	1324427	710752	294792	132284
01/03/2020	2020	Mar.	1238554	675961	271711	120103
01/04/2020	2020	Abr.	1233187	670679	271210	120074
01/05/2020	2020	May.	1259417	682944	278501	123213
01/06/2020	2020	Jun.	1273339	688018	283134	126208

Figura 72. Datos limpios

Modelaje y relación de las consultas

Para el modelaje de este proyecto se va a incluir una tabla mediante lenguaje DAX para unificar todas las fechas. Para ello, en la pestaña “Modelado” podremos acceder a “Nueva tabla” y añadiremos la siguiente línea de código DAX:

```
Fechas = CALENDAR(DATE(2018, 1, 1),DATE(2020,12,12))
```

Automáticamente se crea una entrada para cada día desde el año 2018 hasta el final del 2020. Esta tabla es muy importante para unificar todos los datos y va a ser la tabla de unión entre las otras.

La única tabla que no se va a unir a esta tabla de Fechas va a ser la de GitHub – Mundial, que se unirá a la tabla “PANDEMIA – Mundial Fechas” por medio de la columna de países.

El resultado final del modelaje sigue una estructura de estrella en la que el eje central es la tabla de Fechas.

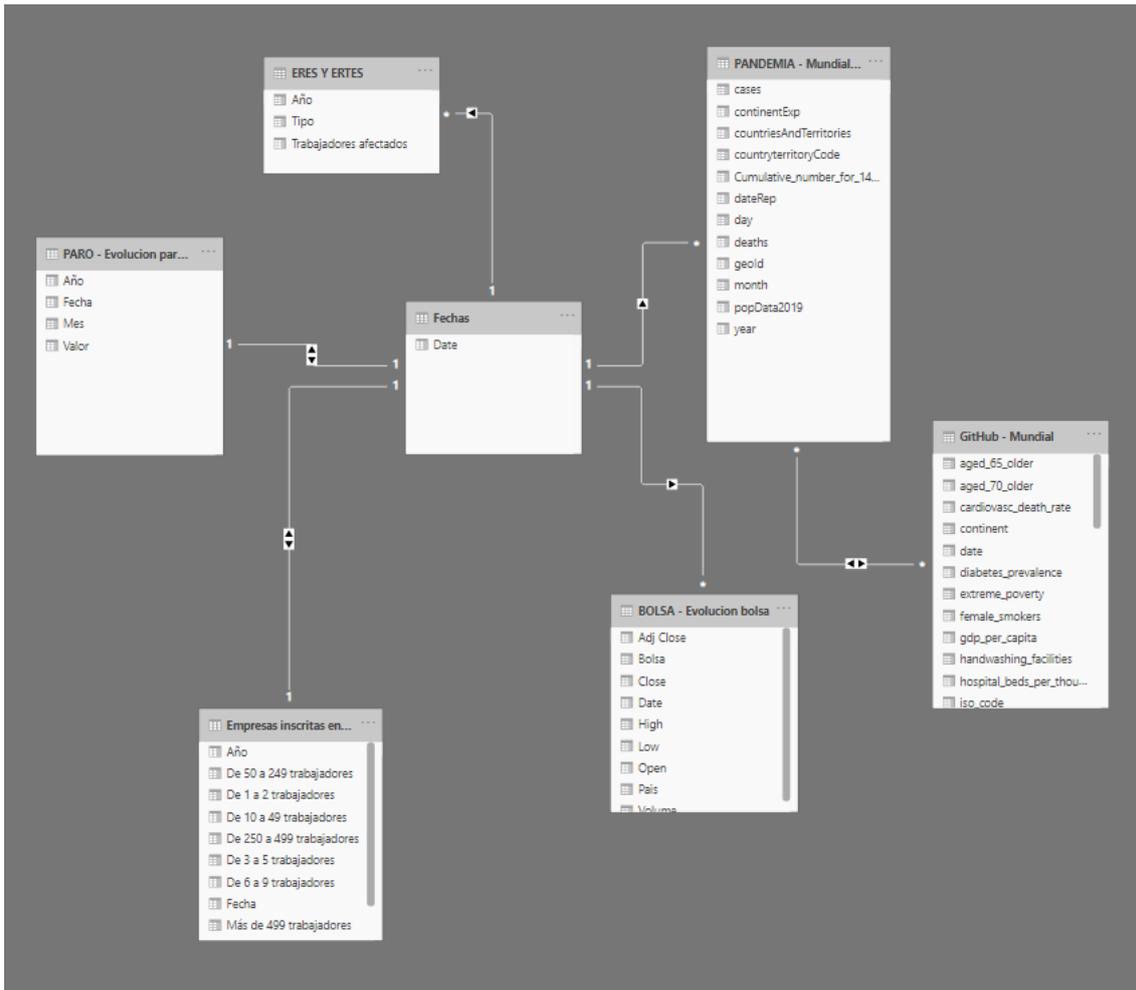


Figura 73. Modelado del proyecto

Graficación y desarrollo del análisis

Ya se han realizado todas las cargas de datos, se han limpiado los datos y estos pueden ser utilizados correctamente en la graficación gracias al modelaje realizado en el punto anterior. En este momento y con las ideas claras sobre qué se desea ver se puede proceder con la primera página.

Se vuelve a recalcar que no se va a parar a explicar filtros ya realizados en páginas anteriores para evitar repetir contenido que ya ha sido explicado.

Página 1 – Bolsa

En la página de la bolsa se va a comenzar incluyendo filtros para las fechas. En estos filtros nos interesan un filtro de intervalos de fechas, de años, uno de trimestres y de meses. Estos filtros

de fechas son ideales para realizar estudios en el tiempo como va a suceder con la evolución de la bolsa, ya que nos permite controlar de una manera fácil y eficaz el intervalo de tiempo que se desee estudiar.

Aprovecharemos la tabla central “Fechas” con su único campo “Date” para crear todos estos filtros. En el caso del filtro de fechas por intervalo se usará aquella entrada “Date” de la tabla “BOLSA – Evolución de la bolsa” para contar únicamente con aquellos valores de fecha de los que se tiene registro.

Los siguientes filtros para realizar van a ser los países y bolsas. Estos dos filtros ya vienen preparados de una manera sencilla gracias a cómo se han creado los datos. Contamos con una entrada “País” y otra “Bolsa” que nos relaciona qué bolsas corresponden a qué país.

Con respecto al filtro de países, el único cambio especial a realizar será desde el menú “Formato”, que se activará la selección única. Este cambio se va a realizar con la finalidad de no afectar al gráfico final, ya que realizar un estudio de todas las bolsas a la vez en un único gráfico de líneas no es útil ni claro para la vista.

Una vez terminados los filtros de la página, comenzamos con aquellos gráficos o indicadores que van a reaccionar ante dichos filtros. Los dos primeros serán unos indicadores del número de casos de COVID-19 en el país (que como requiere selección única, siempre habrá uno) y otro con el volumen medio de la bolsa, que nos ayude a comprender el tamaño de la bolsa entre los distintos países.

Los siguientes indicadores van a corresponder al estudio de la bolsa en estos intervalos de tiempo y a analizar las caídas que han podido tener en esas franjas de tiempo.

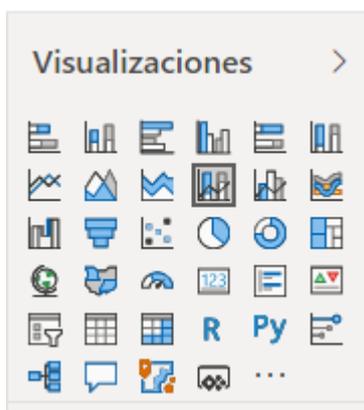
Se procede a explicar por parejas la creación de estos indicadores ya que ha sido necesaria la utilización de medidas en algunas de ellas:

- Fecha máximo de apertura y mínimo de apertura:
 - Se trata de analizar en qué día hubo un máximo o mínimo dentro del intervalo total de fechas por analizar
 - Fecha máxima apertura =

```
CALCULATE(
MAX('BOLSA - 68volución bolsa'[Date]),
FILTER(
'BOLSA - 68volución bolsa',
'BOLSA - 68volución bolsa'[Open]= MAX('BOLSA -
68volución bolsa'[Open])
)
)
```
 - Se diferencian en que la fecha mínima de apertura lleva un MIN en la última línea de código en lugar de un MAX
- Diferencia y Porcentaje de diferencia:
 - Para calcular la diferencia entre el máximo alcanzado en bolsa y el mínimo, la medida es la siguiente:
 - $Diferencia\ bolsa = MAX('BOLSA - Evolucion\ bolsa'[Open]) - MIN('BOLSA - Evolucion\ bolsa'[Open])$

- La realización del porcentaje se vale de esta nueva medida para su creación y la aprovecha, siendo su fórmula:
 - $\text{Porcentaje diferencia} = [\text{Diferencia bolsa}] / \text{MAX}(\text{'BOLSA - Evolucion bolsa' [Open]})$
- Mínimo de apertura y Máximo de apertura:
 - Estas dos no requieren de ninguna medida en DAX, únicamente añadir a los indicadores el campo “Open” y filtrarlo por el mínimo o máximo, dependiendo qué se desee estudiar.

La última visualización por estudiar quizá es la más importante de la página y es el gráfico de líneas y barras que nos va a comparar cómo azotó la pandemia al país que se estudia y como la bolsa ha sido afectada por la misma.



Esta visualización se llama “Gráfico de columnas apiladas y de líneas” y nos permite tener en un mismo gráfico ambos estudios.

El eje que compartirán será el campo “Date” de la columna “Fecha” y los valores de las columnas y las líneas vendrán definidos por dos medidas distintas:

Figura 74. Visualización gráfico

```
Total casos por pais =
CALCULATE(
    SUM('PANDEMIA - Mundial Fechas'[cases]),
    FILTER(
        'PANDEMIA - Mundial Fechas',
        'PANDEMIA - Mundial Fechas'[countriesAndTerritories] = SELECTEDVALUE('BOLSA - Evolucion bolsa'[Pais])
    )
)
```

Esta medida reflejará los casos existentes en el país que se estudia e irá en el valor de las columnas de nuestro gráfico.

```
Valor de apertura en bolsa = SUM('BOLSA - Evolucion bolsa'[Open])
```

Esta medida simplemente permitirá ver la evolución de la bolsa. Se añadirá a los valores de las líneas para terminar así con el gráfico y que este ya sea totalmente funcional y permita analizar los distintos países y bolsas, conjunto al efecto del COVID-19 en ese país.

La página final nos muestra un resultado tal que el siguiente:



Figura 75. Resultado página

Página 2 – ERTES y Paro

En esta segunda página del proyecto se va a compartir el espacio para analizar por un lado los ERES y ERTES concedidos por el gobierno, así como un estudio del paro a nivel nacional. Son dos temas que están estrechamente relacionados por lo que parece lógico analizarlos de manera conjunta.

Comenzando por el primero, realizaremos una matriz que nos servirá no sólo para visualizar valores, si no también para poder utilizarlo de filtro más adelante. Las tablas y matrices son a veces utilizadas como filtro ya que con pulsar en una columna (o fila), logramos filtrar el resto de los gráficos.

En este caso añadiremos todos los valores de la tabla “ERES Y ERTES” a la matriz. Estos son los campos “Año” a las filas, “Tipo” a las columnas y por último “Trabajadores afectados” a los valores. Podemos crear un formato condicional en los ERES y ERTES para añadir un gráfico de barras a la matriz y poder comparar de cierta manera el número de ERES y ERTES concedidos en el tiempo

Continuamos con un gráfico de líneas en el que se espera poder representar de una manera más visual esta descompensación con respecto a otros años en el número de ERES y ERTES.

En este caso no hace falta añadir el tipo al gráfico ya que por defecto realiza la suma de ambas, para estudiar el total de concesiones de este tipo, y si se desea filtrar por uno de los dos podemos recurrir a la matriz anteriormente creada y filtrar por ERES o ERTES según deseemos.

La última visualización de ERES y ERTES va a ser Treemap. Un treemap tiene la misma función que un gráfico de tarta o de anillos con la salvedad de que este refleja mejor las proporciones cuando existen varios grupos para comparar. La idea es comparar para cada año el número de ERES y ERTES concedidos. Añadimos al campo grupo el valor “Año” y a los valores “Trabajadores afectados” para realizar el gráfico.

Pasando a realizar el estudio del paro, comenzaremos por un gráfico de columnas apiladas y de líneas como el que se ha realizado en la primera página para el estudio de la bolsa. En este gráfico se pretende estudiar el número de casos de COVID-19 (bajo la columna “cases” de la tabla “PANDEMIA – Mundial fechas”), en comparación con el número de parados (la suma del campo “Valor” de la tabla “PARO – Evolución paro mensual”) por fecha (campo “Date” de la tabla “Fechas”).

Con todo esto realizado logramos visualizar ya en el gráfico como el número de parados ha ascendido conforme lo ha hecho el número de casos de COVID-19.

Por último realizaremos una matriz que comparará el número de parados del 2019 y 2020 tanto en número como en porcentaje de parados con respecto a la población en España. Para ello añadiremos a las filas los meses de la tabla “Fechas” y a las columnas el de año de la misma. En los valores volcaremos el campo “Valor” de la tabla “PARO – Evolución paro mensual” para saber el número de parados y para el porcentaje realizaremos dos medidas como las siguientes:

```

1 Poblacion españa =
2 CALCULATE(
3     AVERAGE('GitHub - Mundial'[population]),
4     FILTER(ALL('GitHub - Mundial'), 'GitHub - Mundial'[location] = "Spain" ),
5     |CROSSFILTER('GitHub - Mundial'[location], 'PANDEMIA - Mundial Fechas'[countriesAndTerritories], None))

```

Figura 76. Creación medidas

```

1 Porcentaje de parados = SUM('PARO - Evolucion paro mensual'[Valor]) / [Poblacion españa]

```

En la primera medida calculamos la población de España que se encuentra en la tabla “GitHub – Mundial”, pero como se encuentra en cada registro de la fecha necesitamos que nos coja la medida de dichas entradas sin tener en cuenta ningún filtro que pueda haber afectando a la tabla, ya que si no en el año 2019 no tendríamos valor de población porque no hay entradas con esas fechas.

La segunda medida simplemente calcula el porcentaje con respecto al número de parados.

El resultado final de la página es el siguiente

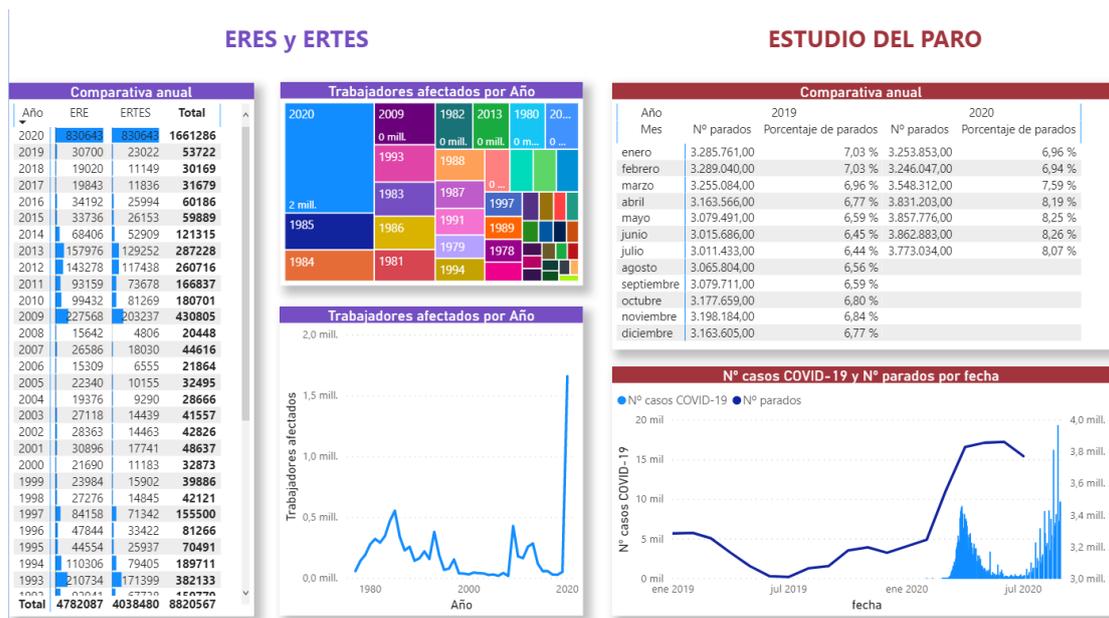


Figura 77. Resultado de la página

Página 3 – Efectos en empresas

En esta nueva página se va a realizar un estudio para averiguar qué efectos ha tenido la pandemia en las empresas, ya que como se ha podido comprobar en las noticias, cientos de empresas han tenido que echar el cierre a su negocio por culpa de la pandemia.

Comenzaremos comparando las empresas inscritas en la seguridad social de acuerdo al tamaño de la empresa. Esta comparación se realizará por medio de una matriz en la que las filas van a ser la jerarquía de Fechas del campo “Date” de la tabla “Fechas”, pero únicamente los años y meses. Para ello, en lugar de traer todo el campo, desplegamos la jerarquía creada y seleccionamos estos dos valores que deseamos traer a las filas de la matriz. En los valores por otro lado, traeremos de forma ordenada, según la cantidad que representan, los distintos campos de recuento (por ejemplo “De 1 a 2 trabajadores” o “De 3 a 5 trabajadores”).

Como se ha explicado anteriormente, con esto no solo logramos comparar ambos años por meses si no que logramos convertir esta matriz en un filtro para el resto de las visualizaciones que vayan a existir dentro de la página.

Dado que la matriz anterior no es muy entendible a nivel visual, vamos a acompañarla de un gráfico de anillos que nos permita averiguar qué cantidad de empresas existen inscritas según el tamaño de empresas. Traeremos la visualización y añadiremos a los valores los campos de recuento de la tabla “Empresas inscritas en la seguridad social” como hemos realizado para la matriz. En este momento se ha formado la visualización y podemos comprobar cómo la gran cantidad de empresas (más del 75 %) son empresas con menos de 6 empleados).

Acompañando a las dos visualizaciones anteriores, el gráfico de anillos y la matriz, vamos a realizar un gráfico de tarta que nos sirva de comparativa del número de empresas desglosable en el tiempo. Los campos para utilizar en este gráfico son el de la tabla “Fecha” → “Date” y de la tabla “Empresas inscritas en la seguridad social” → “TOTAL”. Este gráfico se puede expandir para analizar en profundidad una fecha específica.

Continuaremos con un indicador del número de empresas inscritas, para poder tener de manera más visual el dato, ya que es importante. El campo que utilizaremos es “Total” de la misma tabla de empresas.

Para finalizar con la página, se va a realizar un gráfico de columnas apiladas y de línea que, siguiendo la idea de las páginas anteriores, nos compare el número de casos con el número de empresas inscritas.

El resultado final de la página será el siguiente:

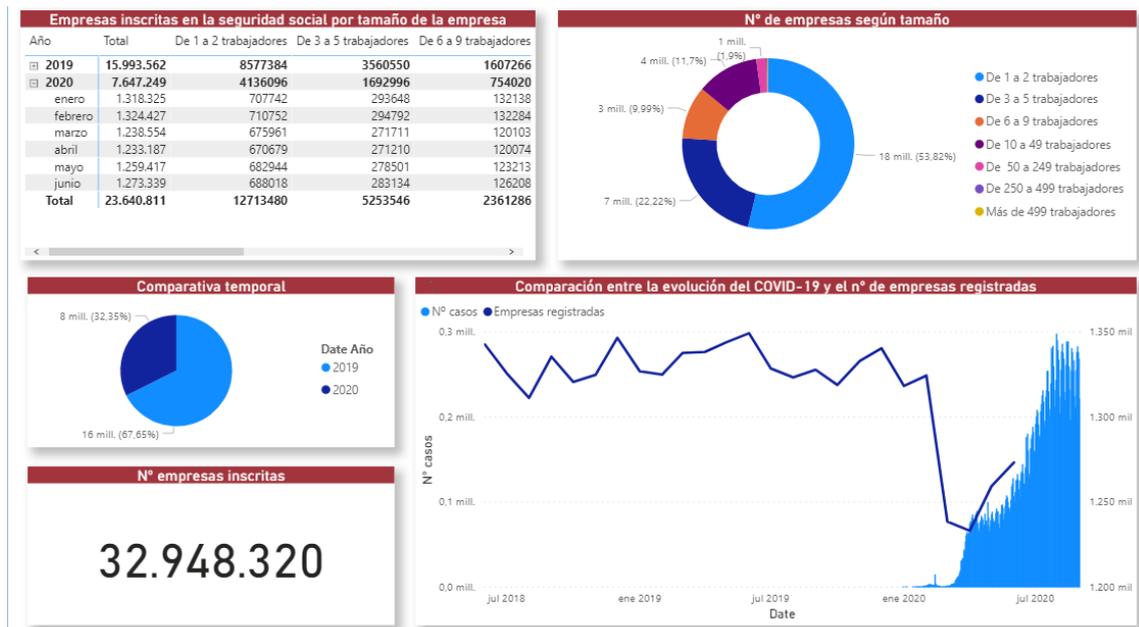


Figura 78. Resultado de la página

Página 4 – Países pobres

Para finalizar con este proyecto, se va a realizar un estudio de los países pobres (aquellos con un índice de pobreza superior al 5%). Este estudio pretende analizar si estos países han sido afectados de una manera más severa debido a la falta de acceso a recursos o si por el contrario han tenido una evolución acorde a la del resto de países.

Comenzaremos creando una tabla en la que mostraremos el país, su número de casos y el índice de pobreza del país.

Utilizaremos los campos “countriesAndTerritories” para los países y “cases” de la tabla “PANDEMIA – Mundial Fechas”, el campo “extreme_poverty” de la tabla “GitHub – Mundial”.

En este caso no quedaría estético añadir un formato condicional ya que el número de casos de Bangladesh es muy superior al del resto de países, por lo que no quedaría estético.

Debemos aplicar como filtro del objeto visual aquellos países cuya extrema pobreza supera al 5%. Este tipo de filtro lo repetiremos continuamente en distintas visualizaciones de la página. De ser en todas, podríamos aplicar el filtro a toda la página, pero debido a que existen algunos casos en los que no deseamos que esto suceda, no lo haremos así.

Para ayudar a localizar los países pobres, vamos a añadir un mapa coroplético que nos los represente, utilizando los países como ubicación, añadiendo en la leyenda el indicador de extrema pobreza y aplicando como filtro de la página aquellos que superen el 5%.

A continuación se van a añadir tres indicadores distintos, que también llevarán el filtro de países con pobreza superior al 5%. Estos indicadores son (todos sobre la tabla “GitHub – Mundial”):

1. Teses realizados:
 - “new_tests”: campo que nos indica el número de teses realizado para cada país. Aquí actúa como suma de todos ellos.
2. Edad media:
 - “median_age”: edad media de la población, que por lo general será joven debido a una esperanza de vida también baja
3. Esperanza de vida:
 - “life_expectancy”: indicador de la esperanza de vida media de aquellos países pobres.

Por último se desea comparar los países pobres contra los ricos, o aquellos que tengan el índice de pobreza inferior al 5%.

Se requiere crear dos medidas nuevas, las cuales son:

Total covid pobres =

```
CALCULATE(
  SUM('PANDEMIA - Mundial Fechas'[cases]),
  FILTER(
    'GitHub - Mundial',
    'GitHub - Mundial'[extreme_poverty] >= 5
  )
)
```

Total covid ricos =

```
CALCULATE(
  SUM('PANDEMIA - Mundial Fechas'[cases]),
  FILTER(
    'GitHub - Mundial',
    'GitHub - Mundial'[extreme_poverty] < 5
  )
)
```

Con estas medidas creadas, podemos volcarlas en el gráfico de líneas para poder comparar la evolución temporal de los países ricos contra los países ricos por número de casos.

El resultado final de la página es el siguiente:

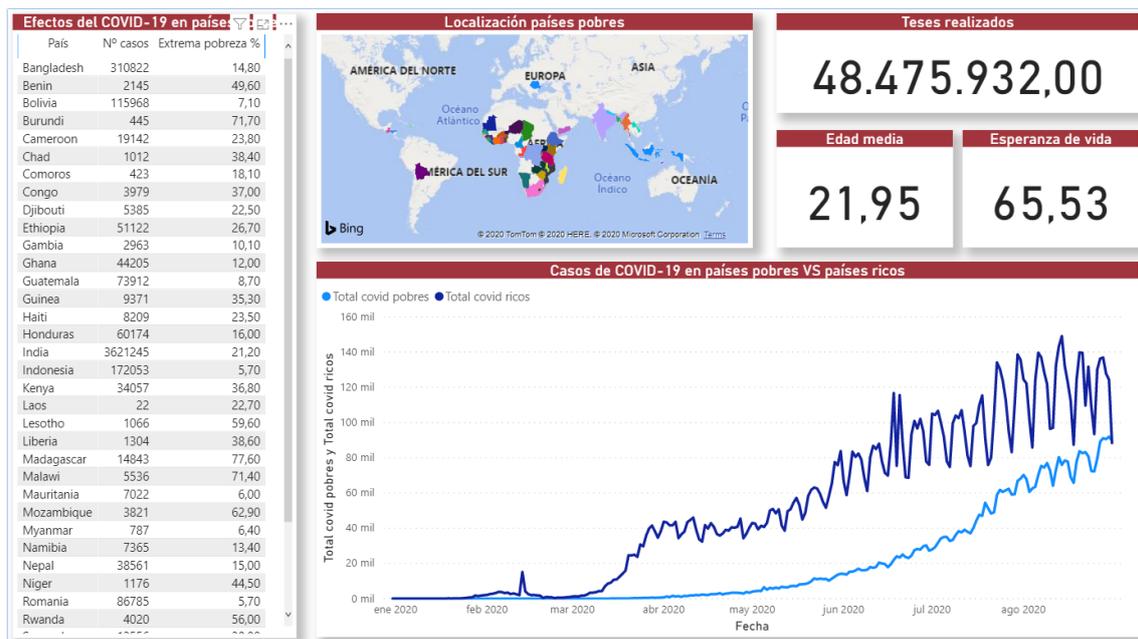


Figura 79. Resultado de la página

Proyecto de estudio de la salud durante la pandemia del COVID-19

El último proyecto que se va a realizar en este trabajo se encuentra relacionado con el tema de la salud, estudio de la enfermedad y comparación con otras enfermedades. La idea que sigue el proyecto es la de estudiar la gripe española, como la anterior pandemia que azotó a varios países, posteriormente compararla con el COVID-19 y por último comparar el COVID-19 con otras enfermedades que han sido famosas en los últimos años.

Este proyecto es algo distinto a los otros dos previamente realizados ya que los datos provienen de dos páginas distintas en la que una de ellas nos brinda los datos en distintos CSV que requieren pocas modificaciones y en la otra página se trata de tablas HTML ya preparadas.

Esto afecta al modelaje del proyecto, que es inexistente debido a la falta de necesidad de relacionar tablas debido al origen de los datos.

Las dos páginas principales son las siguientes:

https://www.elconfidencial.com/espana/2020-06-08/exceso-muertes-coronavirus-espana-gripe-1918_2626504/

[https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(20\)30484-9/fulltext](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30484-9/fulltext)

La primera nos servirá para tratar el tema de la gripe española y la segunda para comparar los datos del COVID-19 con otras enfermedades.

En los orígenes de datos, aquellas tablas que comiencen por “GRIPE - ...” corresponderán a la de datos sacados de la primera página y aquellas que empiecen por “ENFERMEDADES - ...” a la segunda.

Orígenes de datos

GRIPE – número total de defunciones

Comenzaremos estudiando el número total de defunciones que hubo en el año de la gripe española (1918) con respecto al año anterior, 1917, todo esto estudiado por grupos de edad repartidos en distintas franjas.

Los datos originales vienen en un CSV como el siguiente (la imagen ha sido preparada para mejor visualización)

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Column13	Column14	Column15
Grupo	Menos de 1 año	1-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	Más de 60 años
Año 1917	93446	72350	15760	8125	10881	12684	11758	11063	10918	12685	12742	16165	18860	156891
Año 1918	112097	108646	28839	17614	27574	32518	36393	34130	26865	24465	20825	23258	23721	176944

Figura 80. Datos del CSV

GRIPE – número de muertes según la causa

Seguidamente obtendremos otro CSV con el número de muertes según la causa entre los años 1917 y 1918, que nos permitirá observar el crecimiento en el número de muertes con enfermedades relacionadas con la gripe o aspectos respiratorios.

Como se puede comprobar, los datos vienen bastante preparados con la salvedad de problemas con los acentos, que se arreglarán más adelante.

X.1	Año 1917	Año 1918	Variación
Gripe	7479	147114	1867
Otras enfermedades del aparato respiratorio	30467	51330	685
Neumonía	14206	19714	388
Bronquitis aguda	23495	32416	38
Fiebre tifoidea	5499	7400	346
Septicemia puerperal	1886	2534	344
Bronquitis crónica	11610	14318	233
Diarrea y enteritis	45165	55591	231
Otros accidentes puerperales	1145	1383	208
Enfermedades desconocidas o mal definidas	11637	14049	207
Tuberculosis de los pulmones	28514	34046	194
Meningitis simple	18812	22414	191
Debilidad congénita y vicios de conformación	16504	19172	162
Otras tuberculosis	5136	5955	159
Viruela	2565	2969	158
Otras enfermedades	87585	100016	142
Senilidad	22501	25467	132
Suicidios	1242	1397	125
Enfermedades orgánicas del corazón	38277	42799	118
Sarampión	5082	5597	101

Figura 81. Datos del CSV

GRYPE – muertes mensuales en cada año

La primera comparación entre el coronavirus y la gripe española de 1918 viene de la mano de este CSV en el que se compara para ambas pandemias un número aproximado de muertes mensuales en el año anterior de la pandemia y el mismo.

Mes	Coronavirus 2020	Gripe 1918
Enero (-1)	44596	41876
Febrero (-1)	37711	39177
Marzo (-1)	37041	42803
Abril (-1)	34181	39798
Mayo (-1)	33853	33049
Junio (-1)	31858	33275
Julio (-1)	33528	41333
Agosto (-1)	31633	38782
Septiembre (-1)	29875	34776
Octubre (-1)	32672	38761
Noviembre (-1)	34623	37559
Diciembre (-1)	36054	44533
Enero	42567	44593
Febrero	36479	37647
Marzo	56755	43783
Abril	60305	37474
Mayo	34832	35435
Junio		44668
Julio		45068
Agosto		44165
Septiembre		53076
Octubre		163422
Noviembre		96205
Diciembre		50222

Figura 82. Datos del CSV

GRYPE – muertes en los tres meses de mayor impacto de la pandemia

Una pequeña tabla de 3x3 nos permite obtener información sobre el número de muertes en los tres meses de mayor impacto de ambas pandemias así como del promedio de los cuatro años anteriores a las mismas.

Indicador	Gripe 1918	Coronavirus 2020
Promedio de los cuatro años anteriores	109456	105198
Muertes en el año de la epidemia	312703	151892

Figura 83. Datos del CSV

GRYPE – fallecimientos por 1000 habitantes en España

Uno de los CSV más importantes e interesantes de este proyecto se trata de este mismo. Nos muestra información desde el año 1882 hasta el 2020 acerca de la tasa de mortalidad en los distintos años.

Esto nos permite comparar eventos históricos importantes y su impacto en la mortalidad de los habitantes de España, ya que se reflejará en los datos.

Año	Tasa mortalidad
1882	3138
1883	3266
1884	3056
1885	3797
1886	2926
1887	3273
1888	3009
1889	3083
1890	3252
1891	3173
1892	3094
1893	3007
1894	3077
1895	2904
1896	2956
1897	2839
1898	2819
1899	2884
1900	2891

Figura 84. Datos del CSV

GRIPE – evolución mensual de las defunciones

Siguiendo el esquema ofrecido por el CSV anterior, en este nos encontraremos el número de muertes en lugar de por año por mes, de los años 1918 hasta 1920

Mes	Muertes
01/01/1918	44593
02/01/1918	37647
03/01/1918	43783
04/01/1918	37474
05/01/1918	35435
06/01/1918	44668
07/01/1918	45068
08/01/1918	44165
09/01/1918	53076
10/01/1918	163422
11/01/1918	96205
12/01/1918	50222
01/01/1919	45826
02/01/1919	39277
03/01/1919	44287
04/01/1919	38960
05/01/1919	37423

Figura 85. Datos CSV

GRUPE – diferencia en puntos porcentuales de la tasa de mortalidad

Se trata del último archivo que trataremos con respecto a la gripe española. En este se obtiene para cada provincia de España un valor correspondiente a la tasa de mortalidad de 1917, de 1918 y la diferencia que existe entre estos dos valores.

provincia_dw	diferencia	provincia_1918	muerres_1917	muerres_1918
Araba	1452	Álava	1957	3409
Albacete	1314	Albacete	2246	356
Alicante	1265	Alicante	1836	3101
Almería	2091	Almería	2276	4367
Ávila	1685	Ávila	2498	4183
Badajoz	857	Badajoz	2546	3403
Baleares	865	Baleares	1686	2551
Barcelona	982	Barcelona	2364	3346
Burgos	2177	Burgos	2347	4524
Cáceres	61	Cáceres	3106	3716
Cádiz	586	Cádiz	2795	3381
Las Palmas	114	Canarias	134	1454
Santa Cruz de Tenerife	114	Canarias	134	1454
Castellón	974	Castellón	1857	2831
Ciudad Real	1144	Ciudad Real	2266	341
Córdoba	794	Córdoba	2729	3523
A Coruña	1454	La Coruña	1952	3406
Cuenca	913	Cuenca	2271	3184
Girona	692	Gerona	1856	2548
Granada	1068	Granada	2526	3594
Guadalajara	1199	Guadalajara	2158	3357
Gipuzkoa	868	Guipúzcoa	1764	2632
Huelva	1004	Huelva	2276	328
Huesca	1343	Huesca	2061	3404
Jarón	804	Jarón	2861	3665
León	1993	León	2244	4237
Lleida	863	Lérida	1953	2816
La Rioja	1814	Logroño	2117	3931
Lugo	1026	Lugo	1974	30

Figura 86. Datos CSV

Utilizando una página web que no ha sido mencionada previamente (<https://elglobal.es/politica/el-coronavirus-en-contexto-con-otras-enfermedades/>) podemos obtener una pequeña comparativa entre epidemias globales.

Si bien es cierto que los datos del COVID-19 no son muy acertados, esto se puede arreglar más adelante por medio de consultas DAX o en lenguaje M. Es interesante obtener los datos de las otras enfermedades para poder realizar así una comparación.

Comparativa de epidemias globales

Enfermedad	Casos	Muertes	Tasa letalidad
2019-nCov	2.4506	492	2%
SARS	8.437	813	10%
MERS	2.494	858	35%
Ébola	34.453	15.158	44%

Figura 87. Tabla HTML

Comenzando con la segunda página web de la que habíamos hablado previamente, tenemos una tabla que compara datos muy interesantes acerca de algunas enfermedades, como puede ser el periodo de incubación de la enfermedad o el intervalo entre el comienzo de los síntomas y máxima infección.

Table 1 Characteristics of SARS-CoV-2, SARS-CoV, and pandemic influenza

	SARS-CoV-2	SARS-CoV	Pandemic influenza 1918	Pandemic influenza 2009	Interpretation
Transmissibility, R_0	2-5	2-4	2-0	1-7	SARS-CoV-2 has the highest average R_0
Incubation period, days	4-12	2-7	Unknown	2	Longer incubation period; SARS-CoV epidemics form slower
Interval between symptom onset and maximum infectivity, days	0	5-7	2	2	SARS-CoV-2 is harder to contain than SARS-CoV
Proportion with mild illness	High	Low	High	High	Facilitates undetected transmission
Proportion of patients requiring hospitalisation	Few (20%)	Most (>70%)	Few	Few	Concern about capacity in the health sector

Figura 88. Tabla HTML

La última tabla que se traerá al informe de esta página nos permitirá, de forma parecida a la tabla de “Comparación 1”, comparar entre varias enfermedades la cantidad de fallecimientos que han existido, así como la media de edad para cada enfermedad.

Table 3 Mortality from influenza and coronaviruses^{30, 31}

	Number of deaths (adjusted to year 2000 population)	Mean age at death (years)	Years of life lost (adjusted to year 2000 population)
2009 influenza pandemic	7500–44 100 [*] ; 8500–17 600 [†]	37·4	334 000–1 973 000; 328 900–680 300
1968 influenza pandemic	86 000 [‡]	62·2	1 693 000
1957 influenza pandemic	150 600 [‡]	64·6	2 698 000
1918 influenza pandemic	1 272 300 [‡]	27·2	63 718 000
1979–2001 average influenza A H3N2 season	47 800	75·7	594 000
2003 SARS-CoV	774	Unknown	Unknown
2012 MERS-CoV	858	>65·0	Unknown
2019 SARS-CoV-2	302 059 [§]	Unknown	Unknown

Figura 89. Tabla HTML

Inteligencia de negocio

ETL GRIPE – número total de defunciones

Carga de datos en Power BI

Para la realización de la carga de datos en Power BI utilizaremos la importación desde ficheros de texto/CSV. Todas las tablas que comienzan por “GRIPE - ...” son ficheros CSV que se cargarán de la misma manera, por lo que este paso se omitirá en esas tablas.

Transformación de datos

La tabla recién traída a Power Query únicamente requiere de cambiar los valores que deberían representar a la “ñ” por la propia “ñ”. Una vez cambiadas, traspondremos la tabla para dejar en las columnas el grupo de edad, los valores del año 1917 y en otra columna los de 1918.

La tabla queda así preparada para promover los encabezados de la tabla y arreglar los tipos. Para terminar y para más tarde poder ordenar los grupos de edad añadiremos una columna índice para terminar con la transformación de los datos.

La tabla final queda tal que así:

A ^B C Grupo	1 ² ₃ Año 1917	1 ² ₃ Año 1918	1 ² ₃ Índice
Menos de 1 año	93446	112097	1
1-4	72350	108646	2
5-9	15760	28839	3
10-14	8125	17614	4
15-19	10881	27574	5
20-24	12684	32518	6
25-29	11758	36393	7
30-34	11063	34130	8
35-39	10918	26865	9
40-44	12685	24465	10
45-49	12742	20825	11
50-54	16165	23258	12
55-59	18860	23721	13
Más de 60 años	156891	176944	14

Figura 89. Tabla limpiada

ETL GRIPE – número de muertes según la causa

Transformación de datos

La tabla original con la que se parte sufre de problemas similares a la anterior con caracteres especiales como la “ñ” o los acentos, que podemos comenzar arreglándolos. En este caso no hace falta trasponer la tabla, ya que viene correctamente formada, por lo que podemos promover los encabezados. Una vez los tipos son arreglados (entre ellos arreglar la columna “Variación”, que había que cambiarla de tipo a porcentual).

La tabla final sería la siguiente:

	A ^B _C Causa	1 ² ₃ Año 1917	1 ² ₃ Año 1918	% Variación
1	Gripe	7479	147114	1867,00 %
2	Otras enfermedades del aparato respiratorio	30467	51330	68,50 %
3	Neumonía	14206	19714	38,80 %
4	Bronquitis aguda	23495	32416	38,00 %
5	Fiebre tifoidea	5499	7400	34,60 %
6	Septicemia puerperal	1886	2534	34,40 %
7	Bronquitis crónica	11610	14318	23,30 %
8	Diarrea y enteritis	45165	55591	23,10 %
9	Otros accidentes puerperales	1145	1383	20,80 %
10	Enfermedades desconocidas o mal definidas	11637	14049	20,70 %
11	Tuberculosis de los pulmones	28514	34046	19,40 %
12	Meningitis simple	18812	22414	19,10 %
13	Debilidad congénita y vicios de conformación	16504	19172	16,20 %
14	Otras tuberculosis	5136	5955	15,90 %
15	Viruela	2565	2969	15,80 %
16	Otras enfermedades	87585	100016	14,20 %
17	Senilidad	22501	25467	13,20 %
18	Suicidios	1242	1397	12,50 %
19	Enfermedades orgánicas del corazón	38277	42799	11,80 %
20	Sarampión	5082	5597	10,10 %

Figura 90. Tabla limpiada

ETL GRIPE – muertes mensuales en cada año

Transformación de datos

Posiblemente una de las tablas que menos transformaciones requieren, ya que basta con promover los encabezados y cambiar los tipos. Para futuro orden se añadirá una columna índice, la cual se usará más adelante a la hora de graficar.

A ^B _C Mes	1 ² ₃ Coronavirus 2020	1 ² ₃ Gripe 1918	1 ² ₃ Índice
Enero (-1)	44596	41876	1
Febrero (-1)	37711	39177	2
Marzo (-1)	37041	42803	3
Abril (-1)	34181	39798	4
Mayo (-1)	33853	33049	5
Junio (-1)	31858	33275	6
Julio (-1)	33528	41333	7
Agosto (-1)	31633	38782	8
Septiembre (-1)	29875	34776	9
Octubre (-1)	32672	38761	10
Noviembre (-1)	34623	37559	11
Diciembre (-1)	36054	44533	12
Enero	42567	44593	13
Febrero	36479	37647	14
Marzo	56755	43783	15
Abril	60305	37474	16
Mayo	34832	35435	17
Junio	null	44668	18
Julio	null	45068	19
Agosto	null	44165	20
Septiembre	null	53076	21
Octubre	null	163422	22
Noviembre	null	96205	23
Diciembre	null	50222	24

Figura 91. Tabla limpiada

ETL GRIPE – muertes en los tres meses de mayor impacto de la pandemia

Transformación de datos

Debido a la naturaleza de ser una tabla tan pequeña, esta tampoco requiere de ningún paso especial que no sean los habituales: promover encabezados y cambiar tipos

A ^B _C Indicador	1 ² ₃ Gripe 1918	1 ² ₃ Coronavirus 2020
Promedio de los cuatro años anteriores	109456	105198
Muertes en el año de la epidemia	312703	151892

Figura 92. Tabla limpiada

ETL GRIPE – fallecimientos por 1000 habitantes en España

Transformación de datos

La tabla de fallecimientos por 1000 habitantes, que va desde 1882 hasta el 2020, tampoco ha requerido de grandes transformaciones.

Esto principalmente se debe a que los CSV que utilizamos para este proyecto ya están bastante tratados previamente.

Los cambios que se realizan serán los propios en estos CSV: promover encabezados y cambiar tipos:

1.3 Año	1.2 Tasa mortalidad
1882	31,38
1883	32,66
1884	30,56
1885	37,97
1886	29,26
1887	32,73
1888	30,09
1889	30,83
1890	32,52
1891	31,73
1892	30,94
1893	30,07
1894	30,77
1895	29,04
1896	29,56
1897	28,39
1898	28,19

Figura 93. Tabla limpiada

ETL GRIPE – evolución mensual de las defunciones

Transformación de datos

La tabla originalmente parece bastante limpia excepto por el fallo aparente de que la columna “Mes” contiene la fecha en formato americano. Esto importa ya que marca primero el mes, posteriormente el día y por último el año. Esto se puede arreglar de distintas maneras, pero la elegida en este caso ha sido la de dividir la columna bajo el delimitador “/” en tres columnas: “Día”, “Mes” y “Año”. Una vez tengamos todas las columnas separadas las fusionaremos de

nuevo en una columna con el formato de fecha europeo y de este modo, al cambiar el tipo obtendremos la fecha real que se pretendía obtener.

Fecha	Muertes
01/01/1918	44593
01/02/1918	37647
01/03/1918	43783
01/04/1918	37474
01/05/1918	35435
01/06/1918	44668
01/07/1918	45068
01/08/1918	44165
01/09/1918	53076
01/10/1918	163422
01/11/1918	96205
01/12/1918	50222

Figura 94. Tabla limpiada

ETL GRIPE – evolución mensual de las defunciones

Transformación de datos

Esta tabla sufre de problemas con caracteres especiales que deberemos arreglar mediante el remplazamiento de valores. Una vez estos caracteres han sido arreglados, se promueven encabezados y arreglan tipos para, aparentemente, terminar con la tabla.

Si se realiza una prueba en un mapa coroplético podremos observar como existen ciertas provincias que son localizadas en América Latina en lugar de España, por lo que debemos continuar arreglando la tabla.

Crearemos una columna duplicada de las provincias en la que intentaremos dar más información para que el mapa localice correctamente estas provincias conflictivas, las cuales son (con su nueva información):

- Córdoba → Córdoba, Andalucía
- Granada → Granada, Spain
- Guadalajara → Guadalajara, Castilla la Mancha
- Toledo → Toledo, Spain
- Zaragoza → Zaragoza, Spain

Con todo esto y realizando de nuevo la prueba en el mapa, esta vez sí nos reconoce todas las provincias de manera correcta:

A ^B _C provincia_dw	1.2 diferencia	A ^B _C provincia_1918	1.2 muertes_1917	1.2 muertes_1918	A ^B _C Provincias
Araba	14,52	Álava	19,57	34,09	Araba
Albacete	13,14	Albacete	22,46	35,6	Albacete
Alicante	12,65	Alicante	18,36	31,01	Alicante
Almería	20,91	Almería	22,76	43,67	Almería
Ávila	16,85	Ávila	24,98	41,83	Ávila
Badajoz	8,57	Badajoz	25,46	34,03	Badajoz
Baleares	8,65	Baleares	16,86	25,51	Baleares
Barcelona	9,82	Barcelona	23,64	33,46	Barcelona
Burgos	21,77	Burgos	23,47	45,24	Burgos
Cáceres	6,1	Cáceres	31,06	37,16	Cáceres
Cádiz	5,86	Cádiz	27,95	33,81	Cádiz
Las Palmas	1,14	Canarias	13,4	14,54	Las Palmas
Santa Cruz de Tenerife	1,14	Canarias	13,4	14,54	Santa Cruz de Tenerife
Castellón	9,74	Castellón	18,57	28,31	Castellón
Ciudad Real	11,44	Ciudad Real	22,66	34,1	Ciudad Real
Córdoba, Andalucía	7,94	Córdoba	27,29	35,23	Córdoba

Figura 95. Tabla con correcciones ya hechas

ETL ENFERMEDADES – Comparación 1

Carga de datos en Power BI

La carga de datos en Power BI se realizará mediante el método de Obtener datos desde web, lo que nos permitirá ingresar la URL desde la cual podemos acceder a la página. Una vez ingresada elegiremos la tabla de HTML que recoja la información que nosotros deseamos.

Transformación de datos

Nada más traer la tabla a Power Query se nos añade un paso que no solemos ver y es “Tabla extraída a partir de HTML”, la cual nos transforma el código HTML en la tabla que deseamos ver.

Una vez realizado este paso podremos promover los encabezados y arreglar los tipos para que correspondan. Por último normalizaremos los nombres de las enfermedades para que en el resto de tabla se llamen del mismo modo y no andar viendo distintas tablas que hablan de las mismas enfermedades llamándolas ligeramente distinto (ej: 2019-nCov transformarlo a COVID-19)

A ^B _C Enfermedad	1.2 Casos	1.2 Muertes	% Tasa letalidad
COVID-19	22150585	780654	3,52 %
SARS	8437	813	10,00 %
MERS	2494	858	35,00 %
Ébola	34453	15158	44,00 %

Figura 96. Tabla limpiada

ETL ENFERMEDADES – Comparación 2

Carga de datos en Power BI

A partir de aquí, el resto de las tablas que se traerán siguen el mismo método de carga a Power BI, el cual es muy similar al de la anterior tabla estudiada. A través de la obtención de datos vía web, ingresamos la URL de la página web que deseamos. En este momento, al tener la página varias tablas distintas, debemos elegir cuales deseamos traer. Para ahorrar pasos en el futuro escogeremos aquellas correspondientes a “ETL ENFERMEDADES – Comparación 2”, “ETL ENFERMEDADES – Comparación 3” y “ETL ENFERMEDADES – mortalidad asociada a la edad”

Transformación de datos

Una vez traída la tabla al editor de Power Query, podemos observar rápidamente como la tabla viene bastante preparada salvo por el detalle de que viene sin traducir. Por medio de varios pasos de remplazamiento de valores, conseguiremos traducir la tabla así como unificar los nombres de las enfermedades tal cual se ha realizado en la tabla anterior.

Por último, traspondremos la tabla con la finalidad de dejar las enfermedades en una única columna y respetar la estructura de la tabla anterior.

A _C ^B Enfermedad	A _C ^B Transmisibilidad - ratio de reproductividad	A _C ^B Periodo de incubación (en días)	A _C ^B Intervalo entre detección de síntomas y máxima i
COVID-19	2.5	4-12	0
SARS	2.4	2-7	5-7
1918 Gripe pandémica	2.0	Desconocido	2
Gripe pandémica de 2009	1.7	2	2
Interpretacion	SARS-Cov-2 tiene el mayor ratio de reproductividad	Incubación más larga; epidemias de SARS se forman más tarde	COVID-19 es más difícil de mantener que SARS

Figura 97. Tabla limpiada

ETL ENFERMEDADES – Comparación 3

Transformación de datos

En este caso nos encontramos en la misma tesitura que anteriormente: los datos se encuentran sin traducir. Comenzaremos traduciendo estos datos así como normalizando algunos de ellos. También debemos normalizar el nombre de las enfermedades, las cuales en este caso ya se encontraban en su propia columna.

El resultado final sería el siguiente:

A ^B _C Enfermedad	1.2 N° muertes	1 ² ₃ Media de edad de muerte	1 ² ₃ N° vidas perdidas
2009 Gripe pandémica	44100	37	680300
1968 Gripe pandémica	86000	62	1693000
1957 Gripe pandémica	150600	64	2698000
1918 Gripe pandémica	1272300	27	63718000
Gripe promedio de 1979-2001 estacionaria H3N2	47800	75	594000
SARS	813	null	null
MERS	858	65	null
COVID-19	836897	null	null

Figura 98. Figura limpiada

Modelaje y relación de las consultas

Como ya se ha avisado anteriormente, en este proyecto no existe ninguna relación necesaria para el correcto funcionamiento ya que todas son tablas independientes. No obstante se podría relacionar las tablas que comienzan por “ENFERMEDADES - ...” mediante la columna “Enfermedad” que poseen varias de ellas, pero no se ve como algo necesario ya que no afecta al funcionamiento del proyecto.

Graficación y desarrollo del análisis

Comenzando con la última graficación del trabajo correspondiente a este proyecto se prevé la creación de 3 páginas distintas:

- La gripe española
- Gripe española VS COVID-19
- COVID-19 VS Otras enfermedades

El estudio de la gripe española se debe a que ha sido la última gran pandemia global, por lo que es interesante realizar el estudio sobre la misma que nos permita comparar qué estragos causó y cómo se comparan con los del COVID-19.

Página 1 – La gripe española

En esta primera página vamos a realizar 4 gráficos que se van a corresponder a 4 tablas distintas.

El primer gráfico que se va a realizar es una tabla con el número de muertes según la causa entre 1917 y 1918. Para ello utilizaremos la tabla “GRIPE – número de muertes según la causa”. Utilizaremos todos los campos disponibles:

- Causa
- Año 1917 (nº de muertes)
- Año 1918 (nº de muertes)
- Variación (entre cada año, en %)

Para realizar una variación visual con respecto a las barras horizontales que hemos hecho en proyectos anteriores, en este se va a realizar una degradación de colores. Para esto realizaremos un formato condicional respecto al color de fondo y lo ajustaremos según la variación con los colores que deseemos. Debemos tener presente qué colores elegimos para que no hagan invisible o dificultoso la visualización del texto que encierra y en caso de ser así, cambiarlo del mismo modo.

El segundo gráfico que se añadirá será un estudio del número total de defunciones, independientemente de la causa, según el grupo de edad. En este caso se utilizará la tabla “GRIPE – número total de defunciones”, pero antes de ello, debemos realizar un pequeño ajuste.

Este pequeño ajuste se trata de ordenar la columna “Grupo” de acuerdo a la columna “Índice” para que a la hora de graficar el grupo aparezca ordenado según se esperaba. Empezaremos yendo a la ventana de Datos, la que se encuentra encima de la de Modelo y seleccionaremos la tabla que hemos mencionado. Una vez ahí, elegiremos la columna grupo y en la parte superior escogeremos “Ordenar por columna” y seleccionaremos la columna “Índice”. De este modo cuando grafiquemos se encontrará ordenada según el valor del índice pero con los valores del grupo.

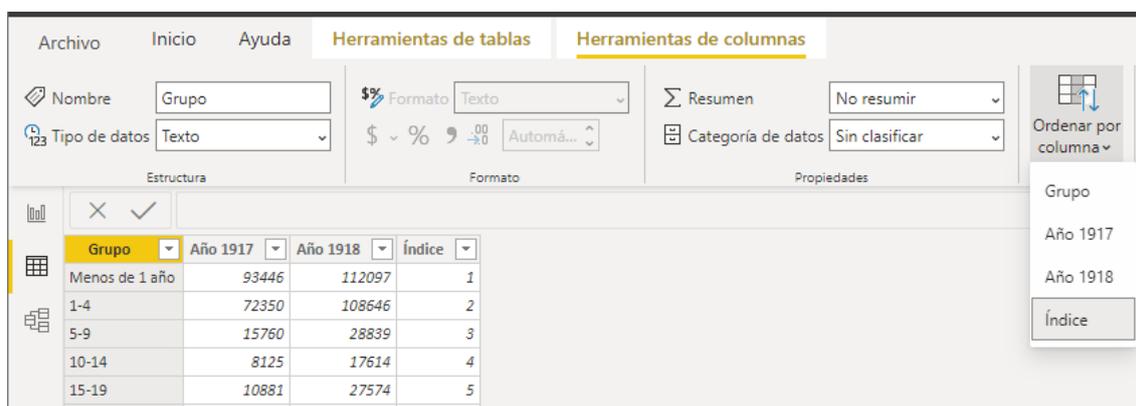


Figura 99. Ordenar por columna

Podemos continuar con la graficación creando un gráfico de columnas apiladas en la que añadiremos al eje la columna “Grupo” y en los valores las otras dos columnas “Año 1917” y “Año 1918”. Con esto ya habremos terminado con el segundo gráfico.

El tercer gráfico de esta página será el mapa coroplético con los datos de la tabla “GRIPE – diferencia en puntos porcentuales de la tasa de mortalidad”. Se recuerda que en la preparación del dato habíamos encontrado problemas con las provincias ya que algunas no habían sido situadas en su lugar correcto. No obstante, se subsanó este error y se dejó lista la tabla para la graficación. Añadiremos a la ubicación la columna “Provincia” que hemos creado y a la información sobre herramientas la columna “diferencia”. Por último arreglaremos los colores del mapa para crear un mapa de calor según la diferencia en puntos porcentuales de las muertes de cada año.

El último gráfico que se realizará en esta página es un gráfico de líneas con la evolución mensual de las defunciones entre 1918 y 1920. En este caso utilizaremos la tabla “GRIPE – evolución mensual de las defunciones” de la siguiente manera:

- Eje: escogeremos la columna “Fecha”

- Valores: la columna “Muertes”

Con todo esto el gráfico y como consecuencia la página, queda terminada y luce así:



Figura 100. Resultado página

Página 2 – Gripe española VS COVID-19

Una vez estudiada y analizada la gripe española, pasaremos a realizar un estudio en paralelo con el COVID-19, empezando por la realización de dos gráficos iguales en los que se estudiará en uno la gripe y en el otro el COVID-19.

Estos gráficos son gráficos de barras apiladas, ambos usando la tabla “GRIPE – muertes en los tres meses de mayor impacto”, usando en el eje la columna “Indicador” y en los valores variará según el grafico de “Coronavirus 2020” a “Gripe de 1918”.

El siguiente gráfico que realizar se trata de un gráfico de líneas en el que se estudiará el número de fallecimientos por 1000 habitantes en España desde 1882, utilizando para ello la tabla “GRIPE – fallecimientos por 1000 habitantes en España”. En el eje se usará la columna “Año” y en los valores “Tasa mortalidad”.

Por último se van a comparar las muertes mensuales en el año anterior y el de la pandemia de ambos casos, la gripe española y el COVID-19. Se usará la tabla “GRIPE – muertes mensuales en el año anterior y el de cada pandemia”. En el eje se insertarán los meses de la columna “Mes” y los valores de “Gripe de 1918” y “Coronavirus 2020”.

El resultado final es el siguiente:

Comparativa: Gripe española VS COVID-19

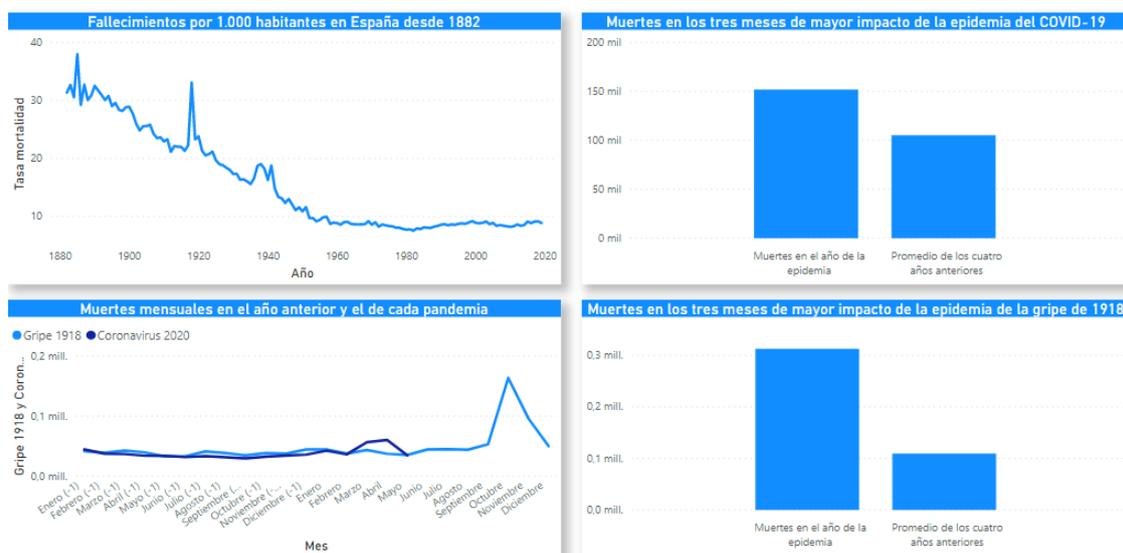


Figura 101. Resultado página

Página 3 – COVID-19 VS Otras enfermedades

Para finalizar con el proyecto se realizará una comparativa del COVID-19 con otras enfermedades que se han dado a conocer en los últimos años.

Se van a utilizar 3 tablas o matrices distintas correspondientes a otras 3 tablas traídas anteriormente. Esto se debe a que los datos se han traído para ser mostrados como tablas o matrices y realizar otro tipo de gráfico no tendría sentido.

Comenzando con la tabla "ENFERMEDADES – Comparación 1" traeremos a las filas de la matriz la columna "Enfermedad" y en los valores las columnas "Casos", "Muertes" y "Tasa letalidad".

La siguiente tabla "ENFERMEDADES – Comparación 2" se plasmará en una tabla en la que se traerán todos los datos de las columnas ordenadas primeramente por la columna "Enfermedad".

La última tabla, correspondiente a "ENFERMEDADES – Comparación 3" sigue la misma estructura que la anterior, trayendo todos los datos ordenándose primero bajo la columna "Enfermedad". En este caso a la columna "Nº de muertes" se le puede traer como formato condicional un gráfico de barras para comparar los valores.

Por último y para dar un aspecto uniforme a los datos se asignarán colores iguales a cada tabla

El resultado final de la página quedaría tal que así:

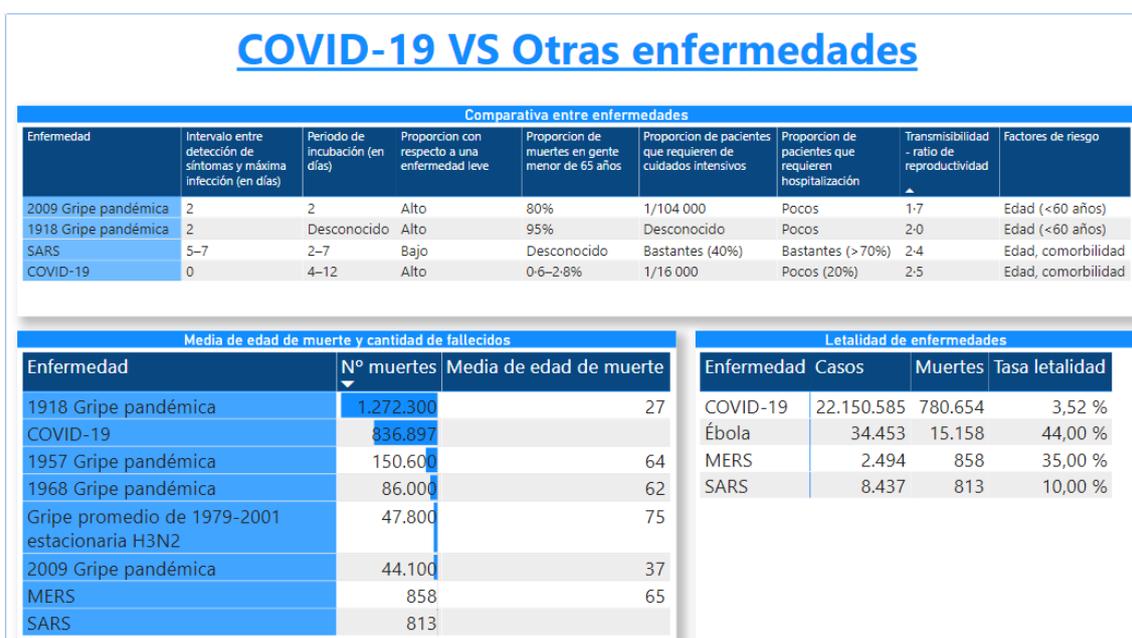


Figura 102. Resultado página

Análisis Predictivo

Se ha dejado para el final de este trabajo el análisis predictivo por varios motivos. Principalmente porque se ha decidido no incluirlo en ningún proyecto debido a la naturaleza de cada uno de ellos. No existe ninguna necesidad de que haya una página dedicada al estudio de los distintos métodos de análisis predictivo en unos proyectos destinados para un público general.

El análisis predictivo sí puede ser usado en proyectos actualmente como una herramienta más de apoyo. Puede que por el momento aún deba madurar estos desarrollos debido a las limitadas opciones con las que contamos ahora mismo en la mayoría de ellas, que dejan poco espacio a la personalización y ajuste.

Se ha realizado un análisis predictivo para el número de casos a nivel mundial reflejados en un gráfico de líneas. Este gráfico es de los únicos, junto con el de barras, que pueden llegar a admitir análisis predictivo, no obstante dependerá de la herramienta que usemos para la misma. Usaremos la tabla "Mundial Fechas" y los campos "dateRep" (fecha) y "cases" (casos) para todos los análisis.

Predicción de Power BI

La primera herramienta que usaremos será la incluida con Power BI en el gráfico de líneas. Para activarla deberemos situarnos en la pestaña de "Analytics" que se encuentra a la derecha de la de "Formato". Esta funcionalidad posee varias características de personalización que muchas otras herramientas no poseen y con las que podremos trastear para ajustar la previsión.

Comenzaremos decidiendo cuantos puntos sobre el eje, es decir, días, queremos prever. Para evitar realizar previsiones erróneas debido a causas puntuales se pueden evitar ignorar los últimos puntos. Por último deberemos ajustar el intervalo de confianza de la previsión, el cual nos indicará cómo de fiable es la previsión que se muestra. Por último deberemos indicar qué estacionalidad posee este gráfico. Por defecto se muestra una estacionalidad con detección automática, la cual es de gran utilidad para usuarios no expertos. La estacionalidad nos permite indicar cada cuantos puntos (días) se repiten ciertos hechos, como podrían ser en este caso picos de infectados.



Figura 103. Predicción con Power BI

Como podemos observar por la línea negra, parece ser que es una predicción que sigue con bastante lógica la naturaleza del gráfico. Las zonas sombreadas mediante gris corresponden al intervalo de confianza designado e indican sobre qué valores pueden andar los máximos y mínimos calculados por la previsión.

Predicción mediante ARIMA

Como se ha hablado en la introducción de este trabajo, las siglas ARIMA corresponden al modelo autorregresivo integrado de promedio móvil. Es un tipo de modelo estadístico ideal para series temporales, es decir para estimaciones futuras que vienen determinadas por datos del pasado y no por variables independientes.

Es debatible la utilidad de una predicción mediante ARIMA en la previsión del número de casos ya que estos se determinan por varios factores que poco tienen que ver con el número de casos pasados (si no, no existirían rebotes una vez los casos hayan casi llegado a cero). No obstante y con el fin de comparar los resultados obtenidos con las otras herramientas se va a realizar un

gráfico del mismo. Para ello utilizaremos de la tienda la herramienta “Forecasting with ARIMA”. Esta herramienta permite cierta personalización con respecto a la estacionalidad y personalización de las variables utilizadas en el modelo.

Con el fin de ser lo más justos posibles se van a dejar todos los valores en automático para así comparar como de precisos son estos modelos estadísticos.



Figura 104. Predicción con ARIMA

Predicción con redes neuronales

Las redes neuronales son de gran utilidad a la hora de realizar todo tipo de cálculo predictivo gracias a la inteligencia neuronal y a los modelos entrenados por expertos. Con ellas no se pueden modificar prácticamente los parámetros debido a que todos ellos han tenido que ser ajustados previamente por los desarrolladores y cualquier cambio haría que la red tuviera que volver a ser preparada para su uso.

Los resultados esperados son positivos y para comprobarlo se usará la herramienta “Forecast Using Neural Network by MAQ Software”

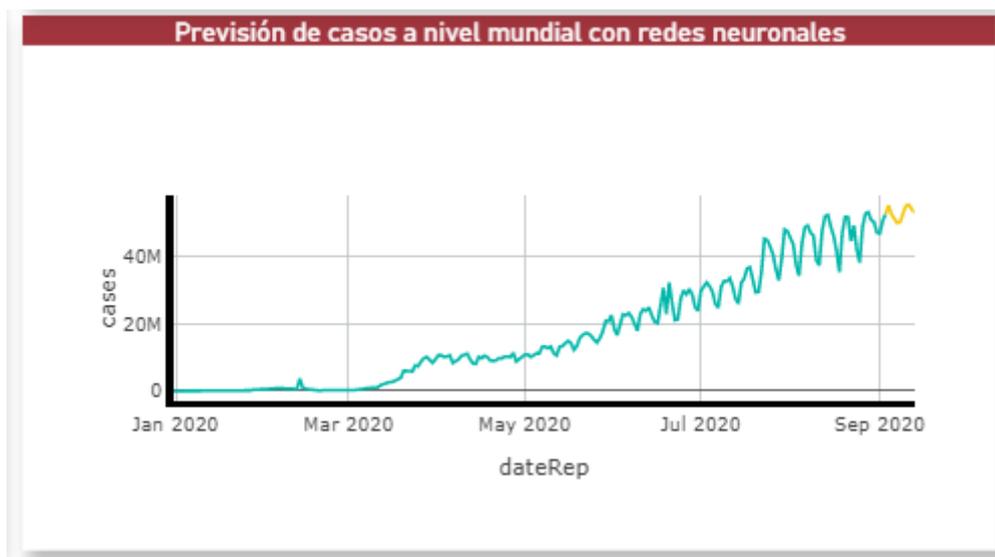


Figura 105. Predicción con redes neuronales

Comparación predicciones

Estas han sido las herramientas utilizadas para el análisis predictivo. A pesar de existir otras más, como la regresión lineal o el “exponential smoothing”, estas no han dado resultados lo suficientemente buenos como para ser mostrados y comparados.

En la siguiente tabla se va a intentar comparar los datos que se han predicho en los distintos gráficos para posteriormente actualizar el proyecto y compararlo con los datos reales que han sucedido en dichos días previstos.

Herramienta	Pico de subida 1	Pico de bajada	Pico de subida 2
Previsión Power BI	4 de septiembre, 55.682.822 casos	8 de septiembre, 44.216.975 casos	11 de septiembre, 57.210.816 casos
Previsión con ARIMA	4 de septiembre, 55.623.323 casos	8 de septiembre, 49.550.541 casos	11 de septiembre, 56.395.799 casos
Previsión con redes neuronales	4 de septiembre, 55.542.350 casos	8 de septiembre, 50.467.040 casos	11 de septiembre, 55.706.910 casos

Resultado final	Pico de subida 1	Pico de bajada	Pico de subida 2
	4 de septiembre, 58.677.558 casos	8 de septiembre, 43.253.333 casos	11 de septiembre, 65.882.093 casos

Tabla 1. Resultados de la predicción

Como podemos comprobar, todas han acertado en los días que se producirían estos picos, no así con el número de casos que habría. En todos los casos se han quedado cortos a la hora de estimar, tanto por arriba como por abajo. Esto se podría ajustar determinando un intervalo de confianza algo menor, lo cual permitiría a las herramientas arriesgarse más a la hora de realizar las previsiones.

Conclusiones del TFG

En este trabajo hemos podido comprobar la utilidad de una herramienta como Power BI para el Business Intelligence. Este tipo de herramienta es la ideal a la hora de interpretar cualquier clase de dato, no únicamente datos empresariales, como hemos podido comprobar.

Además de ello, gracias a las múltiples herramientas que posee y que permite crear a los desarrolladores independientes, Power BI es un programa en constante crecimiento y evolución. Esto permitirá mantenerlo vivo durante bastante tiempo y permitirle acomodarse a los tiempos modernos.

El Business Intelligence, ya de manera más general, es un tipo de estrategia empresarial que se le está empezando a dar importancia en estos últimos años y no es para menos. Permite reconocer el valor de los datos que tenemos de una manera muy rápida y sencilla a cualquier persona. Personalmente opino que cualquier empresa independientemente de su tamaño debería tener un proyecto de Business Intelligence que le ayude a reconocer su estado y evitar así caer en errores que serían solucionables gracias a estos proyectos.

Por último, el análisis predictivo puede ser la herramienta perfecta para el ámbito empresarial y aunque esta deba ser perfeccionada aún, su potencial es enorme. Actualmente puede ser implantado en proyectos, sabiendo qué herramienta de análisis usar.

Espero que este trabajo os haya resultado interesante y hayáis podido comprobar la utilidad de las herramientas anteriormente mencionadas.

Bibliografía

<https://www.iep.edu.es/que-es-el-business-intelligence/>

<https://www.universidadviu.es/business-intelligence-cuales-aspectos-clave/>

<http://info.simetrical.com/blog/el-origen-de-business-intelligence>

<https://www.captio.net/blog/algunos-ejemplos-practicos-de-uso-de-business-intelligence>

<https://softwarepara.net/herramientas-business-intelligence/>

<https://powerbi.microsoft.com/es-es/what-is-power-bi/>

<https://www.valortic.es>

<http://www.bigdata-social.com/que-es-el-analisis-predictivo/>

<https://artyco.com/que-analisis-predictivo-y-como-utilizarlo-marketing/>

<https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/qu-son-los-procesos-etl>

<https://www.baoss.es/analisis-predictivo-que-es/>

<https://www.pbusergroup.com/blogs/pablo-moreno/2018/06/11/analisis-predictivo-con-power-bi>

<https://www.analiticaweb.es/la-prediccion-del-dato-redes-neuronales-artificiales/>

<http://sitiobigdata.com/2019/10/25/que-es-la-regresion-lineal/>

https://www.cienciadedatos.net/documentos/25_regresion_lineal_multiple

[Predictive Analytics – El impacto de la predicción en una organización](#)

<https://jarroba.com/que-es-el-clustering/>

<https://medium.com/intive-developers/forecasting-time-series-with-multiple-seasonalities-using-tbats-in-python-398a00ac0e8a>

<https://www.redaccionmedica.com/secciones/sanidad-hoy/coronavirus-origen-evolucion-porque-no-es-igual-sars-mers-1429>

<https://www.lavanguardia.com/vida/junior-report/20200326/48103758469/coronavirus-covid-19-escola-pejoan-josep-ferre-que-es-origen-sintomas-transmision-proteccion.html>

https://es.wikipedia.org/wiki/Cuarentena_por_la_pandemia_de_enfermedad_por_coronavirus_de_2019-2020

https://as.com/diarioas/2020/04/03/actualidad/1585905299_777674.html

[Manual de Power BI – Diego Pérez Rández](#)

<https://www.isciii.es/Paginas/Inicio.aspx>

<https://www.worldometers.info/coronavirus/#countries>

<https://support.office.com/es-es/article/power-query-descripción-general-y-aprendizaje-ed614c81-4b00-4291-bd3a-55d80767f81d>

<https://www.ecdc.europa.eu/sites/default/files/documents/COVID-19-geographic-disbtribution-worldwide.xlsx>

<https://www.worldometers.info/coronavirus/#countries>

https://covid19.isciii.es/resources/serie_historica_acumulados.csv

<https://github.com/owid/covid-19-data/raw/master/public/data/owid-covid-data.xlsx>

[https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(20\)30484-9/fulltext](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30484-9/fulltext)

<https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/GPFFAQ/WHWPJI&version=81.0#>

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GPFFAQ>

<https://elglobal.es/politica/el-coronavirus-en-contexto-con-otras-enfermedades/>

<https://github.com/owid/covid-19-data/raw/master/public/data/owid-covid-data.xlsx>

<https://es.finance.yahoo.com/quote/%5EIBEX/history/>

<https://es.finance.yahoo.com/quote/%5EIXIC?p=^IXIC>

<https://es.finance.yahoo.com/quote/%5EN225/history?p=%5EN225>

<https://es.finance.yahoo.com/quote/%5EGSPC/history?p=%5EGSPC>

<https://es.finance.yahoo.com/quote/%5EDJI/history?p=%5EDJI>

<https://www.sepe.es/HomeSepe/que-es-el-sepe/estadisticas/datos-avance/datos.html>

https://www.sepe.es/SiteSepe/contenidos/que_es_el_sepe/estadisticas/datos_avance/xls/empleo/evolparo.xls

<https://www.mincotur.gob.es/es-ES/IndicadoresyEstadisticas/Paginas/Estadisticas.aspx>

<https://www.epdata.es/datos/trabajadores-afectados-ere-graficos/450>

<http://www.mitramiss.gob.es/estadisticas/emp/welcome.htm>

https://www.elconfidencial.com/espana/2020-06-08/exceso-muertes-coronavirus-espana-gripe-1918_2626504/

https://www.elconfidencial.com/espana/2020-06-08/exceso-muertes-coronavirus-espana-gripe-1918_2626504/

<https://elglobal.es/politica/el-coronavirus-en-contexto-con-otras-enfermedades/>