

## Práctica 1: IBM Cognos Analytics

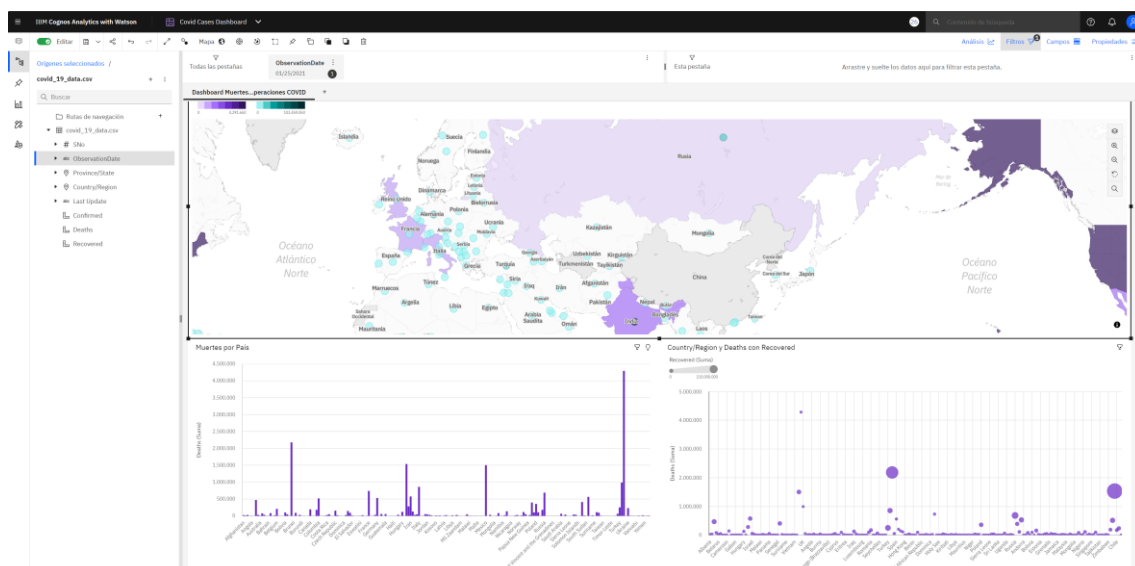
Para esta práctica se ha optado por realizar un dashboard que indique por cada día de la pandemia, los valores mas representativos del estado de la misma y tratar de extraer conclusiones una ve los datos estén representados.

Explorando por [kaggle.com](https://www.kaggle.com) nos encontramos con el dataset [Novel Corona Virus 2019](https://www.kaggle.com/covid19). Este dataset contiene información diaria del avance del virus, con detalle diario de los casos nuevos, muertes y recuperados.

Se debe de tener en cuenta que el dataset tiene varios ficheros. Para realizar el estudio que se va a presentar, se utilizó el fichero covid\_19\_data.csv el cual se compone de los siguientes campos:

- **Sno** – Número de serie asociado al reporte
- **ObservationDate** – Fecha de observación asociada al reporte, en formato MM/DD/YYYY
- **Province/State** – Provincia o estado de la observación (puede esta vacío)
- **Country/Region** – País donde se ha realizado la observación.
- **Last Update** – Marca de tiempo con respecto al UTC en el que el registro actualizó su valor.
- **Confirmed** – Número acumulado de casos dados hasta la fecha
- **Deaths** - Número acumulado de muertes hasta la fecha
- **Recovered** - Número acumulado de casos recuperados hasta la fecha

Con este fichero descargado, lo cargamos dentro de la plataforma y creamos el siguiente Dashboard:



El Dashboard se resume en 3 partes:

1. Un mapa que representa en color morado las muertes que se han tenido por país y con puntos azules variables en tamaño, los recuperados (a mayor tamaño del punto, mayor número de recuperados).
2. Una gráfica que nos dice los muertos por país/estado
3. Un diagrama de burbujas que nos dice, basados en para país y sus muertes, los recuperados del virus. De nuevo, a mayor tamaño, mayor número de recuperados.

Los datos vienen dados a modo de serie temporal, lo cual hace que toda operación de agregación sea descartada, ya que no se estaría dando una visión real del número de infectados en las últimas 4 semanas si así se desease.

Este dashboard por tanto permitirá poder comparar muertes/recuperados, y controlar de una manera intuitiva usando el mapa los números por país.

Una de las primeras cosas que llaman la atención es que cuando buscamos las primeras fechas que se registran en el dataset, casi todos los reportes son dados desde China, por Enero del 2020. Esto es lógico, dado que el virus tiene como origen Wuhan, ciudad china. A medida que avanzamos en el tiempo, se observa como los distintos países a los que viajaron las personas que entraron en contacto con el virus, aumentan los casos, pero no los muertos, hasta unas semanas mas tarde.

También podemos observar cómo existen períodos (ciclos) en los datos, ya que, dependiendo de la fecha, los casos en los países suben y bajan con intervalos de tiempo entre 1 y 2 meses. Esto se asume que es al endurecimiento de las restricciones en el caso de períodos alto de contagio, lo que hacía que bajasen los contactos entre personas y por tanto los contagios.

Por último, hemos de destacar que es posible que algunos datos del dataset no sean del todo correctos, ya que, para algunas fechas, los datos de contagios acumulados por provincia superan el total de los contagios de la pandemia, lo cual creemos que no es correcto.