



# Entregable -Proyecto de grupo

29/07/2020 al 21/08/2020

---

## Grupo D

José Luis Manjón, Filipa Teles, y Mónica Villasuso

Data Science - Boot Camp

The Bridge

## Visión General

El objetivo de este proyecto es afianzar y ejercitar los conocimientos adquiridos en la primera mitad del Bootcamp de Data Science, particularmente los relacionados al manejo de APIs y al análisis de datos (EDA) siguiendo todas las etapas del proceso (desde la obtención de los datos hasta las conclusiones derivadas del análisis de los mismos). Como en todo trabajo en equipo, se pretende generar sinergias y hacer más eficiente y motivador el proceso de aprendizaje. La clave está en una adecuada distribución de tareas, roles y responsabilidades según la experiencia de cada miembro, de manera que una vez concluido, permita a cada uno desarrollar mejor y más rápidamente su proyecto individual (a realizarse en el mismo período – Agosto 2020).

Adicionalmente, al plantear este reto sobre un tema tan actual como “**el CoronaVirus (COVID19) en el mundo**”, podemos percibir como el tratamiento, análisis e interpretación de los datos es imprescindible para entender el comportamiento y la evolución del COVID19 y debiera ser una de las herramientas sobre las que apoyarse en la toma de decisiones para el tratamiento freno y erradicación del virus.

Otro asunto a tener en cuenta es la calidad de los datos. Los países han seguido diferentes criterios para determinar cuando un caso o fallecimiento es atribuible al COVID19, complicando las comparaciones entre ellos. En otras ocasiones las modificaciones y/o rectificaciones de los datos dan como resultado extrañas figuras en las graficas de la evolución.

## Objetivos

De acuerdo a los criterios de evaluación establecidos, hemos querido optar por el entregable A+, habiendo seleccionado de este realizar las siguientes opciones:

- 1 Repositorio en GitHub para centralizar el control de versiones del código y recursos.
- 6 (Correlación entre variables representado con un Mapa de calor (*heatmap*)).

NOTA: El detalle con observaciones de cada apartado de los entregables se encuentra en el directorio **/documentation/entregables.xls** También incorporado como anexo al final de este documento para facilitar su revisión)

## Especificaciones

### Software

Python v2.7 o superior

## Hardware

Procesador i5

Memoria RAM 8 GB

Espacio en disco 200 MB (para los gráficos)

## Requerimientos

- Pandas
- Numpy
- Seaborn
- Datetime
- Matplotlib
- Os / Sys
- Json
- Plotly (\*)
- Flask (\*)

(\*) Deben ser instalados previo a la ejecución de import de las librerías

## Steps

### I. Contexto (Investigación)

La investigación (adicional a los datos utilizados de la fuente a través de la url) se realizó sobre los países pertenecientes al grupo D, a saber: España, Portugal, Turquía, Reino Unido y Venezuela. En particular, fue necesario:

- Documentar el período y condiciones en el que cada país del grupo decretó el estado de alarma. Las fuentes consultadas y las fechas resultantes se encuentran en */src/main.ipynb - Apartado 3 (Período de estado de alarma por país)*
- Documentar ajustes a las cifras realizadas por los países para explicar comportamientos "raros" de los datos. (Por ejemplo, picos en las cifras de muertes reportadas por España el 25 de Mayo 2020 y el 19/06/2020).
- Investigación del significado del Stringency Index por país, su interpretación y posible relación con los datos en */src/main.ipynb - Apartado 4.2 (Variación de stringency index por país)*

### II. Obtención de los datos

Los datos fueron obtenidos de la siguiente url <https://ourworldindata.org/coronavirus-source-data>

/src/main.ipynb - Apartado 2

### III. Data Wrangling / Data Mining / Data Cleaning

- Eliminación de datos totales mundiales o sin país o continente (que afecten los análisis globales). Creación de un dataframe **world\_df**
- Eliminación de registros con información previa al inicio de la pandemia que no agreguen valor al análisis.
- Asignación de la posición mundial de cada país (respecto al mundo) para las cifras de casos y muertes por coronavirus.
- Eliminación de columnas que no contengan información.
- Creación de un dataframe específico para el análisis de los países del grupo D (**covid\_grupoD**) y de variables útiles para el análisis (países\_grupoD)

/src/main.ipynb - Apartado 3

### IV. Exploración/ Análisis de datos (EDA)

El análisis de los datos está organizado de la siguiente manera:

- Análisis global Grupo D  
NOTA: gráficos almacenados en **/resources/plots/TOD\_D**
  - Pie Charts  
[..\resources\plots\TOT\\_D\total\\_deaths\\_gD\\_pie.html](..\resources\plots\TOT_D\total_deaths_gD_pie.html)  
[..\resources\plots\TOT\\_D\total\\_cases\\_gDO\\_pie.html](..\resources\plots\TOT_D\total_cases_gDO_pie.html)
  - Ranking  
[..\resources\plots\TOT\\_D\rank\\_gD\\_bar.html](..\resources\plots\TOT_D\rank_gD_bar.html)  
[..\resources\plots\TOT\\_D\rank\\_TC\\_x\\_C\\_D.html](..\resources\plots\TOT_D\rank_TC_x_C_D.html)  
[..\resources\plots\TOT\\_D\rank\\_TCxM\\_x\\_C\\_D.html](..\resources\plots\TOT_D\rank_TCxM_x_C_D.html)  
[..\resources\plots\TOT\\_D\rank\\_TD\\_x\\_C\\_D.html](..\resources\plots\TOT_D\rank_TD_x_C_D.html)  
[..\resources\plots\TOT\\_D\rank\\_TDxM\\_x\\_C\\_D.html](..\resources\plots\TOT_D\rank_TDxM_x_C_D.html)
  - Outliers  
[..\resources\plots\TOT\\_D\new\\_deaths\\_gD\\_box.png](..\resources\plots\TOT_D\new_deaths_gD_box.png)  
[..\resources\plots\TOT\\_D\new\\_deaths\\_gD\\_box.png](..\resources\plots\TOT_D\new_deaths_gD_box.png)
  - Tendencias (gráficos dinámicos de cada variable representando independientemente a cada país del grupo). Se muestran algunos ejemplos  
[..\resources\plots\TOT\\_D\new\\_cases\\_smoothed\\_per\\_million\\_x\\_C\\_D.html](..\resources\plots\TOT_D\new_cases_smoothed_per_million_x_C_D.html)  
[..\resources\plots\TOT\\_D\new\\_deaths\\_per\\_million\\_x\\_C\\_D.html](..\resources\plots\TOT_D\new_deaths_per_million_x_C_D.html)  
[..\resources\plots\TOT\\_D\total\\_tests\\_x\\_C\\_D.html](..\resources\plots\TOT_D\total_tests_x_C_D.html)  
[..\resources\plots\TOT\\_D\new\\_tests\\_x\\_C\\_D.html](..\resources\plots\TOT_D\new_tests_x_C_D.html)  
[..\resources\plots\TOT\\_D\total\\_cases\\_x\\_C\\_D.html](..\resources\plots\TOT_D\total_cases_x_C_D.html)  
[..\resources\plots\TOT\\_D\total\\_cases\\_per\\_million\\_x\\_C\\_D.html](..\resources\plots\TOT_D\total_cases_per_million_x_C_D.html)

- [..\resources\plots\TOT\\_D\total\\_deaths\\_x\\_C\\_D.html](..\resources\plots\TOT_D\total_deaths_x_C_D.html)
    - [..\resources\plots\TOT\\_D\total\\_deaths\\_per\\_million\\_x\\_C\\_D.html](..\resources\plots\TOT_D\total_deaths_per_million_x_C_D.html)
  - Progresión cada 10 días de casos y muertes
    - [..\resources\plots\TOT\\_D\new\\_deaths\\_gD\\_prog.html](..\resources\plots\TOT_D\new_deaths_gD_prog.html)
    - [..\resources\plots\TOT\\_D\new\\_cases\\_gD\\_prog.html](..\resources\plots\TOT_D\new_cases_gD_prog.html)
- Análisis individual por país Grupo D
 

NOTA: Todos incluyen líneas de referencia de estado de alarma por país (color morado). Los gráficos son almacenados en **/resources/plots/ISO**

  - Casos totales y muertes totales
    - [..\resources\plots\ESP\Tot\\_ESP.html](..\resources\plots\ESP\Tot_ESP.html)
    - [..\resources\plots\GBR\Tot\\_GBR.html](..\resources\plots\GBR\Tot_GBR.html)
    - [..\resources\plots\PRT\Tot\\_PRT.html](..\resources\plots\PRT\Tot_PRT.html)
    - [..\resources\plots\VEN\Tot\\_VEN.html](..\resources\plots\VEN\Tot_VEN.html)
    - [..\resources\plots\TUR\Tot\\_TUR.html](..\resources\plots\TUR\Tot_TUR.html)
  - Ranking (evolución diaria de la posición en el ranking )
    - [..\resources\plots\ESP\Rank\\_ESP.html](..\resources\plots\ESP\Rank_ESP.html)
    - [..\resources\plots\GBR\Rank\\_GBR.html](..\resources\plots\GBR\Rank_GBR.html)
    - [..\resources\plots\PRT\Rank\\_PRT.html](..\resources\plots\PRT\Rank_PRT.html)
    - [..\resources\plots\TUR\Rank\\_TUR.html](..\resources\plots\TUR\Rank_TUR.html)
    - [..\resources\plots\VEN\Rank\\_VEN.html](..\resources\plots\VEN\Rank_VEN.html)
  - Casos diarios y muertes diarias
    - [..\resources\plots\ESP\Daily\\_ESP.html](..\resources\plots\ESP\Daily_ESP.html)
    - [..\resources\plots\GBR\Daily\\_GBR.html](..\resources\plots\GBR\Daily_GBR.html)
    - [..\resources\plots\PRT\Daily\\_PRT.html](..\resources\plots\PRT\Daily_PRT.html)
    - [..\resources\plots\TUR\Daily\\_TUR.html](..\resources\plots\TUR\Daily_TUR.html)
    - [..\resources\plots\VEN\Daily\\_VEN.html](..\resources\plots\VEN\Daily_VEN.html)
  - Muertes diarias con líneas verticales que resaltan los momentos de crecimiento (color rojo) o decrecimiento (color verde) de la curva
    - [..\resources\plots\ESP\Deaths\\_ESP.html](..\resources\plots\ESP\Deaths_ESP.html)
    - [..\resources\plots\GBR\Deaths\\_GBR.html](..\resources\plots\GBR\Deaths_GBR.html)
    - [..\resources\plots\PRT\Deaths\\_PRT.html](..\resources\plots\PRT\Deaths_PRT.html)
    - [..\resources\plots\TUR\Deaths\\_TUR.html](..\resources\plots\TUR\Deaths_TUR.html)
    - [..\resources\plots\VEN\Deaths\\_VEN.html](..\resources\plots\VEN\Deaths_VEN.html)
- Variación del Stringency Index por país en **/resources/plots/ISO**
  - [..\resources\plots\TOT\\_D\stringency\\_index\\_gD\\_bar.html](..\resources\plots\TOT_D\stringency_index_gD_bar.html)
- Análisis por continente en **/resources/plots/TOT\_CONT**
  - Muertes y casos (diarios y totales) en gráficos de área por continente
    - [..\resources\plots\TOT\\_CONT\total\\_cases\\_x\\_cont\\_area.html](..\resources\plots\TOT_CONT\total_cases_x_cont_area.html)
    - [..\resources\plots\TOT\\_CONT\total\\_deaths\\_x\\_cont\\_area.html](..\resources\plots\TOT_CONT\total_deaths_x_cont_area.html)
    - [..\resources\plots\TOT\\_CONT\new\\_deaths\\_x\\_cont\\_area.html](..\resources\plots\TOT_CONT\new_deaths_x_cont_area.html)
    - [..\resources\plots\TOT\\_CONT\new\\_deaths\\_x\\_cont\\_area.html](..\resources\plots\TOT_CONT\new_deaths_x_cont_area.html)
- Índice de mortalidad por países del grupo D en **/resources/plots/TOT\_D**

- [..\resources\plots\TOT\\_D\mortality\\_rate\\_gD\\_line.html](..\resources\plots\TOT_D\mortality_rate_gD_line.html)
- Muertes y casos totales en el mundo en [/resources/plots/TOT\\_W](/resources/plots/TOT_W)
  - [..\resources\plots\TOT\\_W\total\\_deaths\\_World\\_pie.html](..\resources\plots\TOT_W\total_deaths_World_pie.html)
  - [..\resources\plots\TOT\\_W\total\\_cases\\_World\\_pie.html](..\resources\plots\TOT_W\total_cases_World_pie.html)
  - [..\resources\plots\TOT\\_W\total\\_deaths\\_World\\_map.html](..\resources\plots\TOT_W\total_deaths_World_map.html)
  - [..\resources\plots\TOT\\_W\total\\_cases\\_World\\_map.html](..\resources\plots\TOT_W\total_cases_World_map.html)
- Progresión cada 10 días de casos y muertes. en [/resources/plots/TOT\\_W](/resources/plots/TOT_W)
  - [..\resources\plots\TOT\\_W\new\\_cases\\_W\\_prog.html](..\resources\plots\TOT_W\new_cases_W_prog.html)
  - [..\resources\plots\TOT\\_W\new\\_deaths\\_W\\_prog.html](..\resources\plots\TOT_W\new_deaths_W_prog.html)
- Correlación
  - Mundial en [/resources/plots/TOT\\_W](/resources/plots/TOT_W)
    - [..\resources\plots\TOT\\_W\W\\_heatmap.png](..\resources\plots\TOT_W\W_heatmap.png)
  - Países del Grupo D en [/resources/plots/TOT\\_D](/resources/plots/TOT_D)
    - [..\resources\plots\TOT\\_D\gD\\_heatmap.png](..\resources\plots\TOT_D\gD_heatmap.png)
  - Por país en </resources/plots/ISO>
    - [..\resources\plots\ESP\ESP\\_heatmap.png](..\resources\plots\ESP\ESP_heatmap.png)
    - [..\resources\plots\GBR\GBR\\_heatmap.png](..\resources\plots\GBR\GBR_heatmap.png)
    - [..\resources\plots\PRT\PRT\\_heatmap.png](..\resources\plots\PRT\PRT_heatmap.png)
    - [..\resources\plots\TUR\TUR\\_heatmap.png](..\resources\plots\TUR\TUR_heatmap.png)
    - [..\resources\plots\VEN\VEN\\_heatmap.png](..\resources\plots\VEN\VEN_heatmap.png)

NOTA: ISO hace referencia a 3 dígitos que identifican a cada país (ej. ESP = España)

## V. Conclusiones

Empezaremos con la más evidente de las conclusiones: El Covid es un fenómeno global, una pandemia que se ha extendido por la práctica totalidad de los países y territorios del Globo. Por ello, en primer lugar les mostraremos dos mapas-mundi donde se puede ver el número de casos y fallecidos en el mundo.

[..\resources\plots\TOT\\_W\total\\_deaths\\_World\\_map.html](..\resources\plots\TOT_W\total_deaths_World_map.html)  
[..\resources\plots\TOT\\_W\total\\_cases\\_World\\_map.html](..\resources\plots\TOT_W\total_cases_World_map.html)

Siguiendo la misma idea, se han realizado varios gráficos de los casos y fallecidos diarios por continente, que se pueden ver en los siguientes enlaces:

[..\resources\plots\TOT\\_CONT\total\\_cases\\_x\\_cont\\_area.html](..\resources\plots\TOT_CONT\total_cases_x_cont_area.html)  
[..\resources\plots\TOT\\_CONT\total\\_deaths\\_x\\_cont\\_area.html](..\resources\plots\TOT_CONT\total_deaths_x_cont_area.html)  
[..\resources\plots\TOT\\_CONT\new\\_deaths\\_x\\_cont\\_area.html](..\resources\plots\TOT_CONT\new_deaths_x_cont_area.html)  
[..\resources\plots\TOT\\_CONT\new\\_deaths\\_x\\_cont\\_area.html](..\resources\plots\TOT_CONT\new_deaths_x_cont_area.html)



## 1) BOXPLOT - OUTLIERS

Tras esta breve introducción, creemos que lo primero que hay que ver es como son los datos que nos proporciona el dataframe. Hemos realizado los diagramas de caja y bigote (boxplot) de los casos y fallecidos diarios, que nos representará gráficamente los datos numéricos a través de sus cuartiles. De esta manera, el diagrama de caja muestra a simple vista la mediana y los cuartiles, pudiendo también representar los valores atípicos o outliers. En nuestro caso, en ambas variables vemos que hay bastantes datos atípicos(outliers) que superan el límite superior (Ls, el bigote, gráficamente hablando) e incluso alguno, como en el caso de España, son extremadamente atípicos.(posteriormente veremos a que son debidos).

Los gráficos de caja o boxplot se pueden ver en los siguientes enlaces:

[..\resources\plots\TOT\\_D\new\\_deaths\\_gD\\_box.png](..\resources\plots\TOT_D\new_deaths_gD_box.png)

[..\resources\plots\TOT\\_D\new\\_deaths\\_gD\\_box.png](..\resources\plots\TOT_D\new_deaths_gD_box.png)

## 2) GRÁFICAS INDIVIDUALES POR PAÍS EN RELACIÓN AL ESTADO DE ALARMA

En las gráficas se ve como durante el estado de alarma se controló el crecimiento de casos (la pendiente de la curva) sin embargo no se observa una disminución radical de los casos luego del mismo (la curva sigue creciendo aunque de forma menos pronunciada).

El estado de alarma sirvió para controlar la letalidad de la epidemia. En las gráficas se aprecia que posteriormente al mismo la curva de muertes totales ha permanecido estable.

Los gráficos en los que se basa este análisis se pueden consultar en los siguientes enlaces:

[..\resources\plots\ESP\Tot\\_ESP.html](..\resources\plots\ESP\Tot_ESP.html)

[..\resources\plots\GBR\Tot\\_GBR.html](..\resources\plots\GBR\Tot_GBR.html)

[..\resources\plots\PRT\Tot\\_PRT.html](..\resources\plots\PRT\Tot_PRT.html)

[..\resources\plots\VEN\Tot\\_VEN.html](..\resources\plots\VEN\Tot_VEN.html)

[..\resources\plots\TUR\Tot\\_TUR.html](..\resources\plots\TUR\Tot_TUR.html)

[..\resources\plots\ESP\Daily\\_ESP.html](..\resources\plots\ESP\Daily_ESP.html)

[..\resources\plots\GBR\Daily\\_GBR.html](..\resources\plots\GBR\Daily_GBR.html)

[..\resources\plots\PRT\Daily\\_PRT.html](..\resources\plots\PRT\Daily_PRT.html)

[..\resources\plots\TUR\Daily\\_TUR.html](..\resources\plots\TUR\Daily_TUR.html)

[..\resources\plots\VEN\Daily\\_VEN.html](..\resources\plots\VEN\Daily_VEN.html)

## 3) “TOTAL CASES” DIARIOS Y “NEW CASES” DIARIOS. GRUPO D

El país que aparentemente ha controlado mejor la cantidad de casos es Portugal, su curva es la que crece más lentamente. España controló muy bien los nuevos casos y logró ralentizar el crecimiento de los casos totales (entre el 15/05 y el 15/07) pero se observa claramente la segunda oleada con un incremento de casos a partir del 15 de julio.

El estado de alarma necesitó un mes para disminuir el crecimiento de la curva de casos totales esto sucedió casi un mes después de su proclamación, A partir de este punto la pendiente de la grafica se reduce.

Al mismo tiempo, se observa que tres semanas después de la supresión del estado de alarma la curva de casos totales vuelve a crecer casi a la misma velocidad que en el inicio del estado de alarma.

Venezuela tiene una curva que hace pensar que está lejos de controlar la epidemia. Los nuevos casos (y por ende el total) están en pleno crecimiento.

Gran Bretaña a pesar de tener en vigor durante más de tres meses algunas recomendaciones para restringir la movilidad de sus ciudadanos, no han reducido sustancialmente el incremento de los casos totales, si bien es cierto que, en estos momentos, no han tenido nuevos repuntes de casos.

Turquía no ha impuesto ningún estado de alarma, únicamente durante un fin de semana prohibió salir de casa. Y aunque al principio el crecimiento de los casos fué muy rápido, ahora se ha suavizado ligeramente.

En cuanto al número de nuevos casos diarios hay que comentar que en el caso de España se puede observar en esa gráfica que en dos días (19 de Abril y 25 de Mayo) estos casos diarios fueron negativos. Esto se explica por las correcciones que efectuó el gobierno en los datos facilitados, eliminando casos que anteriormente habían sido contabilizados.

Estas conclusiones se basan en los gráficos de cada uno de los países del punto anterior y en los siguientes:

[..\resources\plots\TOT\\_D\total\\_cases\\_x\\_C\\_D.html](..\resources\plots\TOT_D\total_cases_x_C_D.html)

[..\resources\plots\TOT\\_D\new\\_cases\\_x\\_C\\_D.html](..\resources\plots\TOT_D\new_cases_x_C_D.html)

#### **4) “TOTAL DEATHS” DIARIOS Y “NEW DEATHS” DIARIOS. GRUPO D.**

Todos los países (especialmente Portugal y España) parecen haber controlado la letalidad de la epidemia pues sus muertes totales están más o menos estabilizadas desde el 15/06:

En Gran Bretaña, aunque el incremento de fallecimientos se ha reducido, todavía no han logrado estabilizarlo y tienen una tasa de fallecimientos diarios muy por encima de los otros cuatro países objeto de nuestro estudio.

NOTA: Se observa claramente el descontrol y el ajuste de cifras realizado por España a mediados de mayo y en Junio. El 25 de Mayo se reduce el número de fallecidos en 1918 personas (curva de la gráfica NEW DEATHS negativa) y a mediados de Junio se vuelven a corregir los datos para añadir cerca de 1500 fallecidos. Esta importante variación se refleja en las distintas gráficas realizadas sobre el número de fallecidos

Los gráficos en los que se basa este análisis son:

[..\resources\plots\TOT\\_D\total\\_deaths\\_x\\_C\\_D.html](..\resources\plots\TOT_D\total_deaths_x_C_D.html)

[..\resources\plots\TOT\\_D\new\\_deaths\\_x\\_C\\_D.html](..\resources\plots\TOT_D\new_deaths_x_C_D.html)

[..\resources\plots\TOT\\_D\new\\_deaths\\_smoothed\\_x\\_C\\_D.html](..\resources\plots\TOT_D\new_deaths_smoothed_x_C_D.html)

#### **5) TOTAL CASES PER MILLION AND TOTAL DEATHS PER MILLION. GRUPO D**



Como los países tienen poblaciones muy dispares, la mejor forma de compararlos es por cada millón de habitantes. A pesar de que Portugal no ha sido uno de los países con más cantidad de casos (ocupa el cuarto lugar de los 5 países), su proporción en relación a su población lo ubica en el segundo lugar (por detrás de España) lo que indicaría que la incidencia también está siendo alta. En cuanto a los fallecidos por millón destacan negativamente Gran Bretaña y España sobre los otros tres países. El que Portugal a pesar de ser el segundo con más casos por millón no esté en los puestos altos de los fallecidos por millón se explica por los distintos índices de mortalidad que hay en cada país.

Los gráficos en los que se basa este análisis son:

[..\resources\plots\TOT\\_D\total\\_cases\\_per\\_million\\_x\\_C\\_D.html](..\resources\plots\TOT_D\total_cases_per_million_x_C_D.html)

[..\resources\plots\TOT\\_D\total\\_deaths\\_per\\_million\\_x\\_C\\_D.html](..\resources\plots\TOT_D\total_deaths_per_million_x_C_D.html)

## 6) TOTAL TEST.

Ni España ni Venezuela están reportando la cantidad de TEST realizados. UK es el país que está haciendo mayor cantidad de Test a su población (en términos absolutos) y su curva de tests crece a un ritmo mucho mayor que en el resto de países del grupo.

Los gráficos en los que se basa este análisis son:

[..\resources\plots\TOT\\_D\total\\_tests\\_x\\_C\\_D.html](..\resources\plots\TOT_D\total_tests_x_C_D.html)

## 7) TOTAL TEST PER THOUSAND

En términos relativos (a la población del país) Portugal supera al resto de países del grupo por ser quien mayor cantidad de test está realizando.

En estos momentos el número de casos se han disparado pero no así el número de fallecidos. En mi opinión esto puede ser debido a varios motivos:

- Han aumentado el número de test que se realizan y ahora se empiezan a detectar más asintomáticos que anteriormente no aparecían en los datos- Se conoce mas sobre la enfermedad y la forma de tratarla.
- El sistema sanitario, en este momento, no está saturado, pudiéndose ofrecer una mejor asistencia a los enfermos que impide que estos fallezcan por falta de medios, evitando el terrible triaje de enfermos.
- Se conoce más sobre el virus y la forma de tratarlo.

Los gráficos en los que se basa este análisis son:

[..\resources\plots\TOT\\_D\total\\_tests\\_per\\_thousand\\_x\\_C\\_D.html](..\resources\plots\TOT_D\total_tests_per_thousand_x_C_D.html)

## 8) RANKING

El análisis del ranking tiene sentido comparando la cantidad de casos o fallecidos por número de habitantes (TOTAL CASES o TOTAL DEATH per million). Siendo así, de los 5 países el único

que va subiendo su posición en el ranking (es decir, que va incrementando la cantidad de casos por estar en pleno crecimiento de los contagios) es Venezuela. Por ejemplo España ocupaba la posición 4 al inicio del estado de alarma, al finalizar este se encontraba en la posición 17 y en la actualizada está por debajo de la posición 25. Para el caso del ranking de muertes totales por millón de habitantes, la situación es análoga a lo comentado anteriormente para los casos totales.

Al ser variables que están relacionadas con los demás países del mundo, las variaciones en el ranking no solo son debidas a la situación que se esté dando en ese país, sino que también dependerá de cómo se comporten los demás.

Los gráficos en los que se basa este análisis y en el que aparecen los cinco países del grupo:

[..\resources\plots\TOT\\_D\rank\\_TCxM\\_x\\_C\\_D.html](..\resources\plots\TOT_D\rank_TCxM_x_C_D.html)

[..\resources\plots\TOT\\_D\rank\\_TD\\_x\\_C\\_D.html](..\resources\plots\TOT_D\rank_TD_x_C_D.html)

[..\resources\plots\TOT\\_D\rank\\_TDxM\\_x\\_C\\_D.html](..\resources\plots\TOT_D\rank_TDxM_x_C_D.html)

También se pueden consultar los gráficos de cada país donde aparecen los rankings que se han calculado para cada uno de ellos y donde aparece el inicio y el final del estado de alarma si en el dicho país fue decretado.

[..\resources\plots\ESP\Rank\\_ESP.html](..\resources\plots\ESP\Rank_ESP.html)

[..\resources\plots\GBR\Rank\\_GBR.html](..\resources\plots\GBR\Rank_GBR.html)

[..\resources\plots\PRT\Rank\\_PRT.html](..\resources\plots\PRT\Rank_PRT.html)

[..\resources\plots\TUR\Rank\\_TUR.html](..\resources\plots\TUR\Rank_TUR.html)

[..\resources\plots\VEN\Rank\\_VEN.html](..\resources\plots\VEN\Rank_VEN.html)

Si lo que se desea es tener la posición actual en el ranking en los países del grupo de estudio se puede obtener en el siguiente gráfico de barras.

[..\resources\plots\TOT\\_D\rank\\_gD\\_bar.html](..\resources\plots\TOT_D\rank_gD_bar.html)

## 9) INDICE DE MORTALIDAD

El índice de mortalidad es la relación existente entre el número de fallecidos por Covid en relación al número de infectados, se expresa en tanto por ciento y nos da una idea de la letalidad del virus, es decir nos indica cuantas personas han fallecido por cada cien enfermos. En los meses de Marzo y hasta mediados de Marzo creció muy rápidamente, sobre todo en Gran Bretaña (en abril superó el 15%) y en España (en Mayo superó el 12%). En Portugal, Turquía y Venezuela, en ese mismo periodo de tiempo también se produce un incremento pero muchísimo menor.

Desde más o menos la primera semana de Junio, en los cinco países se aprecia como este índice va disminuyendo suavemente, coincidiendo con la disminución de nuevos casos y la mejora de las condiciones de ocupación-saturación en los hospitales.

Los gráficos en los que se basa este análisis son:

[..\resources\plots\TOT\\_D\mortality\\_rate\\_gD\\_line.html](..\resources\plots\TOT_D\mortality_rate_gD_line.html)

## 10) COMPARACIÓN DE LOS PAISES GRUPO D ON EL RESTO DEL MUNDO.

Se han realizado una serie de gráficos circulares donde se ha querido establecer una visión general de los casos y de los fallecimientos en el mundo. En los dos primeros se han incluido los países con las cifras más elevadas y al resto se les ha agrupado bajo la denominación de otros.

En el segundo grupo de este tipo de gráficos se ha comparado el grupo de países asignados con el resto del mundo, observando que los cinco países tienen el 5% de los casos totales, pero casi el 10% de fallecimiento.

Por último se ha comparado entre los cinco las cifras de fallecidos y casos totales, observando que en ambos Gran Bretaña y España presentan las cifras más elevadas. Pero aquí se observa una particularidad, mientras que en casos totales España supera a Gran Bretaña (36,5% - 30,5%) en cuanto a los fallecidos se produce un vuelco y Gran Bretaña supera ampliamente a España (53% - 37%). Esto entronca con lo observado anteriormente en el índice de mortalidad (GB 15, ESP 12).

Los gráficos en los que se basa este análisis son:

[..\resources\plots\TOT\\_W\total\\_cases\\_World\\_pie.html](..\resources\plots\TOT_W\total_cases_World_pie.html)  
[..\resources\plots\TOT\\_W\total\\_deaths\\_World\\_pie.html](..\resources\plots\TOT_W\total_deaths_World_pie.html)  
[..\resources\plots\TOT\\_D\total\\_cases\\_gDO\\_pie.html](..\resources\plots\TOT_D\total_cases_gDO_pie.html)  
[..\resources\plots\TOT\\_D\total\\_deaths\\_gDO\\_pie.html](..\resources\plots\TOT_D\total_deaths_gDO_pie.html)  
[..\resources\plots\TOT\\_D\total\\_cases\\_gD\\_pie.html](..\resources\plots\TOT_D\total_cases_gD_pie.html)  
[..\resources\plots\TOT\\_D\total\\_deaths\\_gD\\_pie.html](..\resources\plots\TOT_D\total_deaths_gD_pie.html)

## 11) MATRICES DE CORRELACIÓN.

Muertes Totales es mayor mientras mayor es el GDP per cápita y la densidad de población

Total de muertes por millón aumenta cuando hay mayor expectativa de vida (Esto explica porque países con más viejitos tienen más mortalidad).

En la matriz de correlación de los cinco países de nuestro estudio, no se aprecian correlaciones positivas ni negativas altas entre las variables que directamente miden la incidencia del Covid (casos y fallecidos en sus distintas categorías) salvo en el número de test efectuados y en el índice de mortalidad. En este caso se aprecian correlaciones positivas altas, lo que indicaría que cuando el número de afectados y fallecidos, junto con un alto índice de mortalidad, es alto, se incrementa el número de test realizados.

Sin embargo, otros factores que a priori podrían hacer pensar que tienen influencia sobre el número de casos y el número de fallecidos por Covid, como la diabetes, la densidad de población, el envejecimiento de la población, el índice de pobreza o el stringency index, con los datos disponibles en estos momentos no se ven reflejados en estas matrices. Estos valores presentan índices de correlación muy bajos.

Aquí se puede ver la matriz de correlación del grupo de los cinco países asignados:

[..\resources\plots\TOT\\_D\gD\\_heatmap.png](..\resources\plots\TOT_D\gD_heatmap.png)

Si analizamos las matrices de correlación de cada uno de los cinco países del estudio, y los comparamos entre sí, se observa que en los países donde el estado de alarma no existió o fue muy corto en el tiempo, se da una mayor correlación entre este índice y el número de test realizados. Esto podría hacernos pensar que en esos países optaron por medidas menos restrictivas, compensándolo con un mayor número de test realizados para detectar nuevos casos.

En la matriz de correlación de Gran Bretaña se da un hecho curioso y diametralmente opuesto a lo que sucede en los otros cuatro países. La correlación entre la esperanza de vida y el número de camas de hospital es -1, es decir a menor número de camas en hospitales mayor es la esperanza de vida (esto no habla bien del NHS). En los otros cuatro países la correlación es máxima y positiva.

Los enlaces a esas cinco matrices de correlación son:

[..\resources\plots\ESP\ESP\\_heatmap.png](..\resources\plots\ESP\ESP_heatmap.png)

[..\resources\plots\GBR\GBR\\_heatmap.png](..\resources\plots\GBR\GBR_heatmap.png)

[..\resources\plots\PRT\PRT\\_heatmap.png](..\resources\plots\PRT\PRT_heatmap.png)

[..\resources\plots\TUR\TUR\\_heatmap.png](..\resources\plots\TUR\TUR_heatmap.png)

[..\resources\plots\VEN\VEN\\_heatmap.png](..\resources\plots\VEN\VEN_heatmap.png)

## ANEXOS

En la siguiente tabla se da respuesta, o se indica donde encontrarla, a los distintos puntos de cada una de las opciones propuestas en el documento guía de este proyecto. En nuestro caso hemos optado por realizar las opciones C, B, A y A+.

ENTREGABLES - GRUPO D	
DETALLE	OBSERVACIONES
1. Document all steps. Structure your code to keep it cleaned using good practices.	<b>/documentation</b> <b>/documentation/trello</b> contiene la planificación realizada en TRELLO y el avance medido semanalmente
2. Collect Coronavirus Data . It is mandatory that in each call, it collects the last updated data	<b>/src/main.ipynb</b> Jupyter Notebook . Apartado 2 Los datos se actualizan directamente de la URL
3. Determine and explain if the data is cleaned. If not, then clean it.	<b>/src/main.ipynb</b> Jupyter Notebook . Apartado 3
4. Create an API that returns a Json with the logic explained for your group. The flask server must be executed running the <code>src/api/server.py</code> file.	<b>/src/api</b>
5. Get the jsons generated from your annexed group and plot it. First, try to connect to the private ip of your annexed group. If it is not possible because of physical issues, then simply use what they generate copying it. If your annexed group cannot give you the necessary json, then annotate it, use the json of another group. If there are no jsons from other groups, then use your json from your own API. then annotate it, use the json of another group. If there are	<b>/src/api</b> Tanto el proceso para recibir el archivo (de grupo C) como la entrega del archivo json al grupo A fueron realizados y probados en las instalaciones de THE BRIDGE el día 31 de julio 2020
6. Show different tendencies for each column in your dataset. Show, vertically, the start date and end date of the alarm state in each plot. If there is no alarm state, then show only the start date.	<b>/src/main.ipynb</b> .- Jupyter Notebook. Apartado 4 (EDA) .- Las fechas de estado de alarma se incorporaron en las gráficas por país (apartado 4.2)
7. Answer the questions:	
a. What position do your countries occupy respect to the number of total infected, total deaths and total recoveries?	<b>/src/main.ipynb</b> Jupyter Notebook. Apartado 4 (EDA) .- Apartado 4.1 Rankings
b. What can you conclude about your data study?	.- Apartado 5 Conclusiones generales
c. Are there outliers or some rare data?	.- Apartado 4.1 Outliers
<b>B</b>	<b>/src/main.ipynb</b> .- Jupyter Notebook. Apartado 4 (EDA) .- Las líneas de crecimiento y decrecimiento de las curvas se incorporaron en las gráficas de muertes por país (apartado 4.2)
	2. Create with bars, lines, points and pie charts the daily deaths and infected. .- Apartado 4.1
<b>A</b>	.- Funciones incorporadas en la librería <b>/src/utills/folders_tb.py</b> . .- Los gráficos se graban dentro del directorio <b>resources/plots</b> , a saber: .- Un directorio por cada país del grupo D .- Un directorio (TOT_D) para los gráficos del grupo D .- Un directorio (TOT_C) para los gráficos por continente .- Un directorio (TOT_W) para los gráficos mundiales. Los gráficos interactivos se almacenan en formato .html. Los gráficos estáticos se almacenan en formato .png
	1 Research to save each plot in local files. Save them on different folders for each country
	2 Use distribute modules for each functionality. The jupyter notebooks must not have any loop or function definitions. It only must have the initials imports and the call to necessary functions. Librerías incluidas en el directorio <b>src/utills</b>
	3. Answer the questions: a. Can we conclude that the alarm state has had an effect on the improvement of the daily infected rate? explain why. <b>src/main.ipynb</b> Apartado 5 - Conclusiones generales <b>src/main.ipynb</b> Apartado 4.3 b. How is the progression going each ten days?
<b>A+</b>	1. Use a different github repository adding all group participants with writepermissions. Use that repository to manage the delivery code and resources. <b>github.com/JoseLuisManjon/proy_covid_august</b>
	2. Per country, which are the columns that are more related? find the correlation between columns with a correlation matrix <b>src/main.ipynb</b> Apartado 4.4

