

MÁSTER U. DE CIENCIA DE DATOS

Tipología y Ciclo de Vida de los Datos. Aula 1

(M2.851)

PRÁCTICA 1

Alumnos:

Olast Arrizibita Iriarte - oarrizibita@uoc.edu

Enrique Pérez Balbuena - eperezbal@uoc.edu

1.- Contexto

Debido al estado de alarma que sufre el planeta como consecuencia de la pandemia provocada por el Covid-19 y viendo cómo los números por contagio y fallecidos se van multiplicando sin cesar a nivel mundial, hemos pensado que podríamos recolectar información sobre las cifras que está provocando la crisis del coronavirus en cada país.

2.- Título de dataset

Las cifras del Covid-19 a nivel mundial

3.- Describir el dataset

El dataset contiene básicamente los datos por países de ocho variables en una determinada fecha. En nuestro caso esa fecha es el 1 de abril, que es el día que hemos tomado la muestra. Los datos se actualizan de un día para otro, por eso hay algunas cifras como la de incrementos, que se hace respecto al día anterior.

4.- Imagen que representa el dataset



5.- Contenido

Las nueve variables que tenemos son las siguientes:

- **País:** Nombre del país al que hace referencia los datos.
- **Activos:** es el número de infectados menos (los fallecidos más los recuperados).
- **Inc. Recuperados:** Incremento del número de recuperados respecto al día anterior.
- **Recuperados:** Número de pacientes recuperados hasta el día de hoy.
- **Inc. Fallecidos:** Incremento del número de fallecidos respecto al día anterior.
- **Fallecidos:** Número de personas fallecidas hasta el día de hoy.
- **Inc. Confirmados:** Incremento del número de pacientes confirmados por contagio respecto al día anterior.
- **Confirmados:** Número de personas confirmadas por contagio hasta el día de hoy.

- **Confirmados / 100 mil hab.:** Número de casos confirmados de personas contagiadas por cada 100.000 habitantes.

6.- Agradecimientos

Agradecer a 'DatosMacro' (www.datosmacro.com) que pertenece al grupo del periódico digital 'Expansión' por permitirnos el acceso a sus datos. También agradecer su trabajo de recogida de dichas cifras tanto a través de sus fuentes propias como las utilizadas del Ministerio de Sanidad y del CSSE (JHU).

7.- Inspiración

Nos parece que este conjunto de datos pueden ayudar a otras personas a entender mejor el alcance de esta pandemia puesto que los datos son a nivel mundial, son cifras de todos los países que tienen algún caso oficial.

8.- Licencia

La licencia por la que hemos optado ha sido la de: 'Released Under CC BY-NC-SA 4.0 License'

Nos ha parecido la más adecuada ya que de esta forma reconocemos el trabajo ajeno, otorgando el crédito apropiado, además no permitimos su uso comercial y si sufre algún tipo transformación el trabajo, deberá de ser distribuido bajo la misma licencia y en los mismos términos que el original.

9.- Código

Adjuntamos en este archivo, el fichero de Notebook 'cifras_covid_19_mundial.ipynb' en código python

10.- Dataset

Adjuntamos en este archivo, el fichero 'cifras_covid_19_mundial.csv' con los datos

Contribuciones	Firma
Investigación previa	OA, EP
Redacción de las respuestas	OA, EP
Desarrollo código	OA, EP

Bibliografía

[1] Datos Macro. *COVID-19 – Crisis del coronavirus*. [En línea]. Actualización 1 de abril de 2020. Disponible en: <https://datosmacro.expansion.com/otros/coronavirus> [Fecha acceso: 1 de abril de 2020]

[2] Lawson, R. (2015). *Web Scraping with Python*. Packt Publishing Ltd. Chapter 2. Scraping the Data.

[3] Mitchel, R. (2015). *Web Scraping with Python: Collecting Data from the Modern Web*. O'Reilly Media, Inc. Chapter 1. Your First Web Scraper.