

Dataset: Avance del CORONA VIRUS a nivel mundial 2019 - 2021

Luis Alberto Caro Saldivia

01 Abril de 2021

Descripción

En esta actividad se trabajó con algunos datasets sobre la contaminación del **CORONA VIRUS** en *Europa, Africa, Reino Unido, Norteamérica, Asia y Latinoamérica* desde los años 2019 y 2021. Los datos se obtuvieron mediante algoritmos de *webscraping* utilizando la librería **BeautifulSoup** recopilando las estadísticas de cantidad de personas contagiadas, confirmadas, muertas, su procedencia y totales.

Imagen Contexto

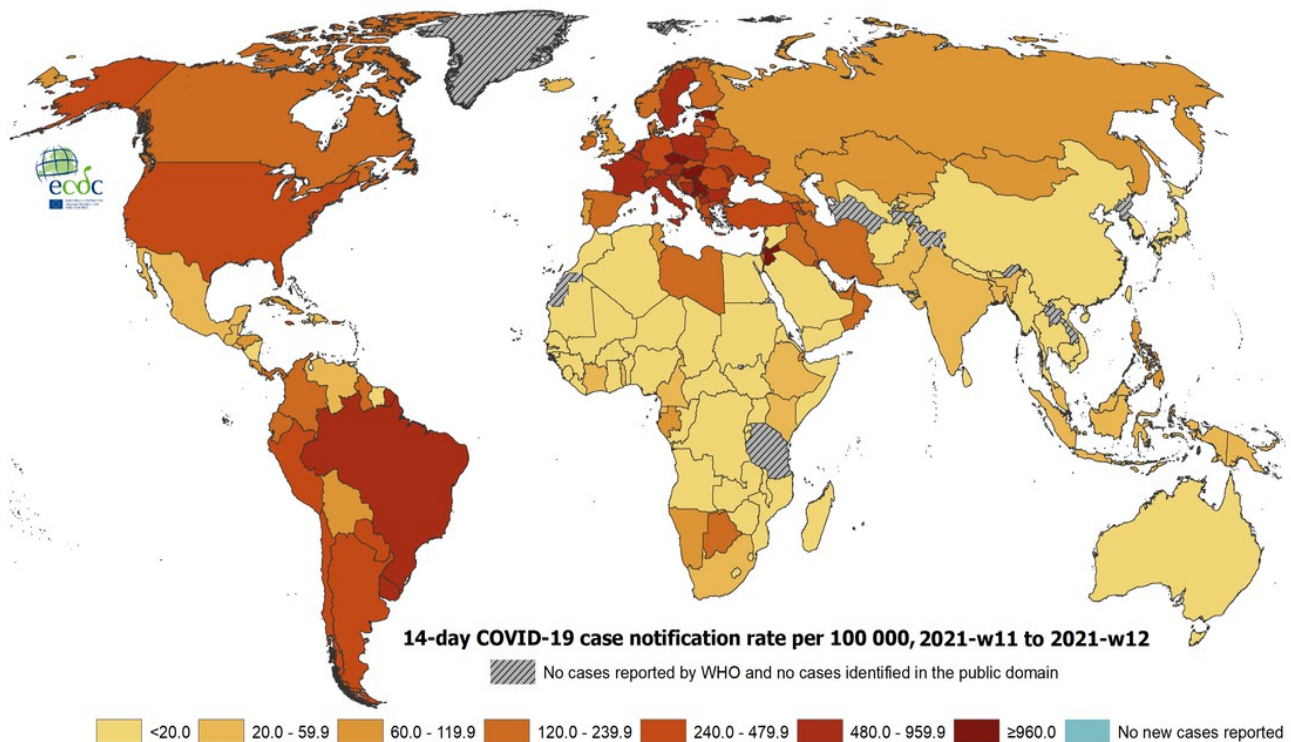


Figura 1: Mapa contaminación mundial de COVID 19

Fuente: <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>

Contexto

La pandemia mundial que estamos viviendo desde finales del 2019 ha transformado al mundo. No solo ha tenido severas repercusiones sanitarias si no que ha transformado el curso de la historia de todos los países. La economía mundial se ha visto tremendamente afectada generando una gran recesión mundial y afectado a todas las sociedades de diferentes continentes. Grandes pérdidas de vidas humanas con mucho sufrimiento y un intento por ganar el juego a la naturaleza para controlar este virus es fuente de mucha investigación y trabajo colaborativo a nivel mundial. Dado que esta pandemia ha puesto al mundo y sus instituciones gubernamentales en *Jaque Mate*, me interesa conocer su avance y su control sanitario a nivel mundial. Esas son las razones principales de la elección de este dataset.

Contenido

El dataset que se ha obtenido de <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases> posee las siguientes características:

- **Índice:** Se utiliza para identificar el país o región de origen
- **País:** Nombre del país estudiado
- **Contagiados:** Número de personas contagiadas por *Corona Virus*
- **Confirmados:** Número de personas confirmadas (POSITIVOS) con *Corona Virus*
- **Muertos:** Número de personas muertas por *Corona Virus*
- **Total Contagiados:** Suma total de personas contagiadas
- **Total Confirmados:** Suma total de personas confirmadas
- **Total Muertos:** Suma total de personas fallecidas

Esta información se puede consultar en el archivo **data_export.csv**. El *Script* que realiza esta operación es **ws_01.py**.

Además, se han obtenido varios archivos de estadísticas relacionadas del impacto del *Corona Virus* a nivel mundial, los cuales se pueden consultar en: <https://www.ecdc.europa.eu/en/publications-data/data-covid-19-vaccination-eu-eea>.

Los archivos son los siguientes:

- data.csv
- data.json
- data.xlsx
- data.xml

El *Script* que realiza esta operación es **ws_02.py**.

Agradecimiento

Toda esta información es registrada periódicamente y está disponible desde el portal del Centro Europeo para la Prevención y Control de Enfermedades (<https://www.ecdc.europa.eu/europa>). Todos los datasets son datos abiertos y públicos. Se agradece este repositorio de abundantes datos sanitarios actualizados sobre la pandemia del Corona Virus.

Para la extracción de los datasets señalados se utilizó el lenguaje de programación **Python** (<https://python.org>) y la librería de *webscraping* **BeautifulSoup** (<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>)

Inspiración

La razón fundamental en la utilización de estos datasets está relacionada por el gran impacto a nivel mundial que ha tenido este virus y que ha sido fuente de tantas desgracias humanas, tanto de muertes como de recesión económica. Creo que el mundo no estaba preparado para este escenario y la naturaleza nos ha demostrado de manera cruda pero real, sin distinción de ninguna clases, lo débiles que somos como especie. De hecho, los animales son claramente más resistentes que todos nosotros con estos virus.

Creo que siempre se puede sacar algo positivo de algo tan malo. Debemos aprender a cuidar los ecosistemas que tenemos y del cual vivimos. Esta situación claramente nos hará más responsables y más humildes. En términos prácticos, los datasets obtenidos se pueden utilizar para entender estadísticamente la evolución de esta pandemia y como un caso real de estudio para poner en práctica lo aprendido de técnicas de *webscraping*.

Licencia

He seleccionado para estos datasets la licencia **CC0: Public Domain Licence**, pues en el portal <https://creativecommons.org/share-your-work/public-domain/cc0/> se declara lo siguiente:

CC0 : Sin derechos reservados

CC0 permite a los científicos, educadores, artistas y otros creadores y propietarios de contenido protegido por derechos de autor o base de datos renunciar a esos intereses en sus obras y, por lo tanto, colocarlas de la manera más completa posible en el dominio público, de modo que otros puedan construir libremente, mejorar y reutilizar las obras para cualquier propósito sin restricciones bajo derechos de autor o leyes de bases de datos.

Creo que estos datasets cumplen con estas características pues su utilización busca un bien mayor el cual permite a todos los países y centros de salud la utilización de estos datos para sus investigaciones que permitan erradicar este mortal virus.

Código fuente e Informe final

Todo el material referenciado en este documento y los *Scripts* en **Python** se pueden acceder en:

<https://github.com/skalberto/mitaller>

Recursos

1. Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data
2. Mitchel, R. (2015). Web Scraping with Python: Collecting Data from the Modern Web. O'Reilly Media, Inc. Chapter 1. Your First Web Scrap