

Caso COVID-19 Fase 1

Autor

Martín Vladimir Alonso Sierra Galvis

Gestión de Datos, Maestría en Ciencia de Datos
Pontificia Universidad Javeriana Cali

Versión 1.0
Santiago de Cali, marzo de 2023

Tabla de contenido

Contexto	3
Caso	3
Alcance	3
Recursos	4
Consideraciones	4

Contexto

El SARS-COV-2, responsable de la enfermedad COVID-19, es un virus de la familia de los coronavirus que apareció en China a finales del año 2019 y se expandió a gran velocidad a lo largo y ancho del globo, sorprendiendo a la mayoría de países del planeta y causando una pandemia a inicios de 2020. La gran preocupación con respecto a este virus es su alto índice de contagio, lo que le permitió tal expansión en un período muy corto de tiempo; y su alto grado de afectación en pacientes con comorbilidades previas, al punto de ocasionar la muerte rápidamente.

Caso

Consciente del peligro latente que representa el nuevo coronavirus, las principales instituciones se han tomado seriamente la tarea de recopilar la mayor cantidad de datos posibles sobre casos confirmados, decesos y pacientes recuperados y almacenarlos en bases de datos. Lo anterior con el fin de ver la evolución de la enfermedad, entender su comportamiento y, así, tratar de crear mecanismos de contención. Ahora bien, las instituciones desean proveer estos datos a todo el público, por lo que piensan que la mejor manera de mostrarlos es por medio de gráficas, de forma que sea fácil de entender tanto por expertos como por personas del común. Para ello contratan a un equipo de ingenieros y científicos de datos cuya misión es realizar todo el proceso de limpieza, preparación y visualización de datos, específicamente de aquellos relacionados a casos confirmados de contagio.

Alcance

Como ya se explicó, este es un caso de tipo programado, por lo que su resolución se hará siguiendo una serie de pasos o fases. Para esta unidad, deben desarrollar la **Fase #1** que consiste en utilizar la base de datos de la Universidad Johns Hopkins y el módulo de preparación y limpieza de datos pandas de forma que puedan responder a las siguientes **preguntas de interés**.

1. ¿En cuál mes se presentó el mayor número de contagios?
2. ¿En ese mismo mes, cuál fue el país que reportó más contagios?
3. ¿Cuál es el país con el menor número de casos reportados hasta la fecha?

Recursos

El dataset que deben utilizar en el caso es el de la Universidad Johns Hopkins, mismo que pueden encontrar en el siguiente enlace.

https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv

Recuerden que el dataset lo pueden importar en el código Python, usando para ello el enlace y el módulo **pandas**.

Consideraciones

Es importante leer las condiciones de uso para trabajar con el dataset de la Universidad Johns Hopkins. Dichas condiciones las pueden encontrar en el siguiente enlace.

<https://github.com/CSSEGISandData/COVID-19>