

Universidad de Costa Rica
Escuela de Ciencias de la Computación e Informática
CI-0163 - Análisis de Grandes Volúmenes de Datos
Tarea 01 | Ciclo I 2022
Estudiante: Carlos Espinoza Peraza B92786

Primera Parte

Selección del Dataset

Para este trabajo se seleccionó un dataset relacionado al conteo de casos y muertes de COVID-19 en los Estados Unidos durante la primera parte de la pandemia **(21/01/20 - 26/07/20)**. Este conjunto de datos muestra específicamente cual es el conteo por Estados.

Selección de preguntas relacionadas al dataset

1. ¿Cuál fue el Estado que más se vio afectado en términos de cantidad de muertes?
2. ¿Durante cuáles fechas se dieron picos en la cantidad de casos confirmados?
3. ¿Cuál fue el pico de la pandemia y dónde se registraron estos casos?

Problemas con el dataset

1. Se encontraron valores vacíos en la columna **Admin2**. Para resolver este problema, dado que aquí se encuentran los valores de los estados, se procedió a poner la misma información que en la columna de **Province_State** mediante el uso de la condición **if**.

Atributos importantes para resolver las preguntas propuestas

- **Confirmados**: Este atributo permite llevar un acumulado de los casos que han sido confirmados a lo largo del tiempo, con este se puede saber cuál ha sido el grado de contagio de contagio cuando es contrastado con diferentes fechas.
- **Muertes**: Este atributo permite llevar el acumulado de bajas que podría mostrar que tan letal fue el virus dependiendo del lugar y momento.
- **Tiempo**: Permite contrastar los atributos anteriores para poder tener un mayor entendimiento de cuando fue que realmente se vio la peor cara de la pandemia dentro de esta muestra.

Transformaciones al Dataset

1. Eliminación de la columna **Country_Region**: claramente al ser un dataset que toma en cuenta los datos de Estados Unidos es innecesario tener un atributo que presenta un único valor **US**.
2. Eliminación de la columna **code3**: Al realizar una comparación se pudo encontrar que este atributo era igual a **UID** por lo que se procedió a su eliminación.