**Memoria**

EDA: Terremotos 1900-2022

Las herramientas principales que se ocuparon en el EDA

* Python 3.11.5
* Kaggle: Datasets
* Librerías Panda, Matplotlib, Zipfile, Folium, Plotly, Seaborn, geopy, numpy
* ChatGPT 4.0

Busque el tema el Kaggle en el siguiente enlace:

<https://www.kaggle.com/datasets/jahaidulislam/significant-earthquake-dataset-1900-2023>.

1. Aunque el conjunto de datos abarca información desde 2023, no se disponía del año completo para ese periodo específico. Por ende, dicho año no fue incluido en el análisis. Después de importar el archivo CSV a Python (solicité asistencia a ChatGPT para recordar cómo importar un archivo Zip), procedí a realizar las consultas iniciales sobre la cantidad de columnas y filas presentes en el conjunto de datos.
2. Tras obtener la información resultante, llevé a cabo la limpieza de los datos al notar que algunas columnas no contenían información completa en las 37,331 filas. Estas columnas incompletas fueron excluidas para iniciar el análisis exploratorio.
3. Se utilizaron gráficas con la librería Plotly para visualizar los datos totales de terremotos, incluyendo:

a. El total de terremotos desde 1900 hasta 2022.

b. La magnitud máxima por década.

c. El promedio de magnitudes por década.

1. Después de trabajar con el conjunto de datos completo, se llevó a cabo una limpieza de los países, considerando únicamente los países y excluyendo islas y archipiélagos.
2. Durante el proceso de limpieza del archivo, se unificaron nombres, agrupando, por ejemplo, estados dentro de países para una mayor coherencia.
3. Posterior a la limpieza, se extrajo del archivo únicamente los 30 principales países que registraron terremotos desde 1900 hasta 2022.
4. Utilizando la librería Nominatim, se implementó un bucle para obtener las coordenadas geográficas de los 30 países principales.
5. Con la asistencia de ChatGPT, se desarrolló un bucle utilizando la librería Folium para ubicar las coordenadas de los 30 principales países. Se crearon mapas de calor y mapas de ubicación. Tuve un problema con las columnas de latitud y longitud debido a la especificidad de los puntos en el Dataset. Por esta razón, se aplicó el bucle utilizando el nombre del país en lugar de las coordenadas.
6. Se elaboró una representación gráfica de los 30 países con el mayor número de terremotos.
7. Se generó un gráfico que muestra el promedio de magnitudes basado en los 30 países principales.
8. Se creó un gráfico que representa la magnitud más alta registrada mundialmente en los terremotos ocurridos desde 1900 hasta 2022.
9. En el código implementado, se utilizaron diversos métodos y técnicas para llevar a cabo el análisis y la manipulación de datos. Algunos de los métodos clave incluyen:
10. Bucles: Se emplearon bucles para iterar a través de datos.
11. Funciones: Se definieron funciones para modularizar y reutilizar código.
12. Máscaras: Se emplearon máscaras para filtrar y seleccionar datos de interés basándose en condiciones específicas.
13. Variables y Dataframe: Se crearon varias variables para almacenar y manipular datos temporales, y se utilizaron Dataframe para organizar y estructurar conjuntos de datos.
14. Listas: Se utilizaron listas para almacenar y manipular conjuntos ordenados de datos.
15. Diccionarios: Se emplearon diccionarios para asociar claves y valores, facilitando la organización y acceso eficiente a la información.
16. Tuplas: Se usaron tuplas para estructurar datos inmutables en casos donde la modificación de valores no era necesaria.
17. Programación funcional: Se aplicaron conceptos de programación funcional, como el uso de funciones de orden superior y expresiones lambda, para realizar operaciones de manera más concisa y legible.

En conjunto, la combinación de estos métodos y técnicas contribuyó a la implementación eficiente y estructurada del código, permitiendo un análisis detallado de los datos sobre terremotos.