

Proyecto: Análisis y Visualización de Datos

Objetivo:

El objetivo del proyecto es realizar un análisis de un conjunto de datos de elección libre, **preferible con algún enfoque en geociencias**, utilizando las librerías de Python *numpy*, *pandas*, *matplotlib*, *seaborn*, entre otras, para identificar tendencias y observaciones importantes de los datos.

Partes del Proyecto:

1. Elección del conjunto de datos a utilizar:

- Puede ser de formato csv, Excel, o de alguno de los vistos en clase.

2. Preparación del Entorno de Trabajo:

- Crear un repositorio en **GitHub** que almacene los datos, notebook del proyecto, y demás información relevante.
- Generar un nuevo ambiente virtual para el desarrollo del proyecto, instalando las librerías necesarias.

3. Carga y Preprocesamiento de los Datos:

- Cargar el conjunto de datos utilizando *pandas* o la librería más adecuada para el tipo de archivo que será importado.
- Realizar limpieza inicial de datos: Corregir tipos de datos, modificar nombres de columnas, etc.

4. Análisis Exploratorio de Datos (EDA):

- Realizar un análisis exploratorio de los datos, como describir estadísticas básicas, visualizar la distribución de las variables y detectar posibles outliers.
- Crear funciones para automatizar algunas de las tareas repetitivas del EDA.

5. Visualización de Datos:

- Utilizar *matplotlib* y *seaborn* para crear visualizaciones que muestren:
 - La distribución de variables numéricas.
 - Comportamiento de variables categóricas.

- Tendencias de alguna variable a lo largo del tiempo (gráficos de líneas).
- Mapas de calor para visualizar variaciones de variables de interés.
- Otros gráficos que considere pertinente añadir.

Entregables:

❖ Informe final en un cuaderno de Jupyter Lab que incluya:

- Introducción y descripción del proyecto.
- Metodología utilizada para el análisis de datos.
- Resultados principales y visualizaciones.
- Conclusiones y recomendaciones basadas en los hallazgos del análisis.
- Referencias bibliográficas

❖ Presentación del Proyecto:

- Preparar una presentación de 10 minutos, la cual será llevada a cabo el **jueves 27 de junio**.

❖ Conjunto de datos utilizado

❖ Enlace del repositorio en GitHub.

Recursos necesarios:

🚀 **Lenguaje de programación:** Python

🚀 **Librerías:** *numpy, pandas, matplotlib, seaborn, etc.*

🚀 **IDE:** Jupyter Lab

Evaluación:

La evaluación del proyecto se basará en los siguientes criterios:

- 🚀 Claridad y orden del código.
- 🚀 Uso de celdas de Markdowns para todas las secciones del Notebook.
- 🚀 Complejidad y profundidad del análisis.
- 🚀 Calidad de las visualizaciones.
- 🚀 Claridad y calidad del informe final y la presentación.