1. Presentación del taller:

Conocer las bases de datos que van a analizar.

<https://www.datos.gov.co/Minas-y-Energ-a/Accidentes-el-ctricos-a-os-2010-2020/4c8y-i7iw>

descargar y guardar como csv delimitado por comas

Explicar cómo usar Git y cómo crear un repositorio para almacenar su trabajo.

<https://github.com/dsuarezm0425/Taller-SQL-Python-Numpy-Matplotlib-y-Seaborn>

2. Carga y limpieza de los datos:

Crear la base de datos en MySql

CREATE DATABASE accidentes\_electricos

USE accidentes\_electricos

Cargar la tabla

Click derecho, Tabla data import wizard – create table - next

SELECT \* FROM accidentes\_electricos.accidentes\_de\_origen\_electrico;

Usar pandas para inspeccionar las primeras filas con df.head() y entender la estructura de los datos.

Limpiar los datos eliminando duplicados (df.drop\_duplicates()) y manejando los valores nulos (df.dropna()).

import mysql.connector

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

# Conectar a la base de datos MySQL

conn = mysql.connector.connect(

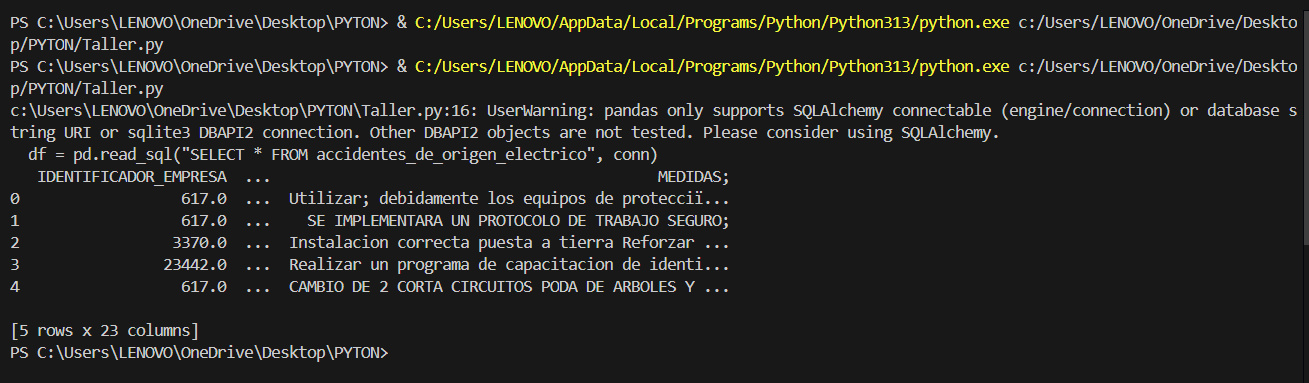
    host="localhost",

    user="root",

    password="0919",

    database="accidentes\_electricos",

)



3. Exploración de los datos:a. Visualizar distribuciones básicas con matplotlib y seaborn (ejemplo: histogramas, gráficos de cajas).

b. Crear una lista de preguntas basadas en los datos

4. Formulación de hipótesis:

a. Las integrantes del equipo deben formular una o dos hipótesis basadas en sus preguntas

5. Entregables del día:

a. Subir los datos limpiados a Git.

b. Incluir el código para cargar, limpiar y explorar los datos en el repositorio.

c. Escribir el primer borrador del pdf con título, objetivos, preguntas y las primeras observaciones.

Realizar un análisis más profundo usando pandas, numpy, y crear visualizaciones avanzadas.1. Análisis de los datos:

a. Usar pandas para realizar cálculos específicos como la suma y el producto de ciertos campos de

interés.

b. Crear filtros para responder las preguntas planteadas. Por ejemplo, filtrar por diagnóstico o

por unidad funcional.

2. Visualización de los datos:

a. Crear gráficos de barras, gráficos de líneas, gráficos de dispersión y gráficos de cajas usando

matplotlib y seaborn

3. Hypothesis testing:

a. Realizar un análisis que permita confirmar o refutar las hipótesis planteadas el primer día.

4. SQL y NoSQL:

a. Si están trabajando con bases de datos SQL o NoSQL, deben escribir los scripts o dumps para

consultas importantes que hayan utilizado en su análisis

5. Documentación y Reporte:

a. Escribir las primeras conclusiones del análisis en el informe en PDF.

b. Agregar capturas de pantalla del código y gráficos generados.

c. Subir el archivo .pbix de power BI

6. Entregables del día:

a. Subir el análisis y las visualizaciones al repositorio de Git.

b. Incluir el código de análisis y visualización con comentarios explicativos.

c. Escribir las primeras conclusiones en el informe

Día 3: Refinamiento, Conclusiones y Entrega

1. Refinamiento del análisis:

a. Revisar el análisis realizado y hacer ajustes si es necesario.

b. Realizar análisis más detallados o pruebas adicionales si las preguntas o hipótesis no se

resolvieron por completo.

2. Finalización del reporte:

a. Redactar la conclusión final del análisis, resumiendo los hallazgos más importantes.

b. Asegurarse de que todo esté claramente documentado en el informe en PDF, incluyendo:

i. Título y objetivo del análisis.

ii. Preguntas y hipótesis.

iii. Pasos realizados.

iv. Resultados y hallazgos.

v. Conclusiones finales.

3. Revisión de los entregables:

a. Comprobar que el repositorio de Git contiene el código bien organizado.

b. Verificar que todos los pasos estén documentados y que se haya usado numpy, pandas, matplotlib y seaborn en el análisis.

c. Revisar que los scripts SQL o dumps NoSQL estén correctos y bien documentados.

4. Entrega final:

a. Subir todos los archivos finales a GitHub, asegurándose de que todo esté organizado

(código, análisis, gráficos, informes).

b. Comprimir y entregar el repositorio junto con el informe en PDF

5. Entregables del día:

a. Subir el repositorio final a Git.

b. Entregar el archivo .pbix (si se utiliza Power BI).

c. Entregar el informe en PDF con todos los detalles