

Laboratorio de Datos - TP 1

Integrantes:

- Dorogov Cristina
- Pucciarelli Francisco Lautaro
- Salto Julian

Para poder correr los scripts de Python, es necesario:

- Python 3.6 o superior.

Que en el environment (donde ejecutan los scripts) tenga los módulos:

- Pandas
- Numpy
- Duckdb
- Seaborn
- Matplotlib
- Pathlib (viene incluido en los paquetes básicos de Python)
- Sklearn

Descripción General

El archivo llamado Script_Main.py contiene el código correspondiente a

- Limpieza y Procesamiento de Datos
- Consultas SQL
- Gráficos
- Toma de decisiones
- Selección de pixeles
- Modelos KNN y Árboles de Decisión

El código está separado por secciones. Puede ejecutarse todo junto de corrido (esto puede tomar unos segundos), o ir ejecutándose por secciones. En caso de ir por la segunda opción, por favor ir en el orden en el que aparecen; ya que algunas secciones generan archivos que son utilizados en secciones posteriores.

El código está super comentado. No solo se encuentran los comentarios correspondiente a qué hace cada parte del código, sino también las distintas ideas y decisiones que se fueron tomando en el camino (los comentarios funcionaron como

cuaderno de laboratorio, a lo largo del desarrollo de cada etapa).

Aclaraciones y Comentarios Adicionales

Antes de ejecutar el código se deben cambiar los path para poder acceder a los archivos utilizados. Específicamente en las líneas:

- 40
- 444
- 649