

Ejercicio 1

****Objetivo**:** Utilizar PySpark para analizar los datos globales de casos confirmados de COVID-19.

****Dataset**:** <https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv>

****Instrucciones**:**

1. Carga el dataset en un DataFrame de PySpark.
2. Realiza las siguientes operaciones:
 - a) Muestra los 10 países con más casos confirmados en la última fecha disponible.
 - b) Calcula el total de casos globales para la última fecha disponible.
3. Muestra los resultados de cada operación.

Ejercicio 2

****Objetivo**:** Utilizar PySpark para analizar los datos de viajes en taxi amarillo de Nueva York.

****Dataset**:** <https://d37ci6vzurychx.cloudfront.net/trip-data/yellow_tripdata_2023-01.parquet>

****Instrucciones**:**

1. Carga el dataset en un DataFrame de PySpark.
2. Realiza las siguientes operaciones:
 - a) Calcula el número total de viajes y el ingreso total para todo el mes.
 - b) Encuentra la hora del día con más viajes.
 - c) Calcula la tarifa promedio por viaje.
3. Muestra los resultados de cada operación.

Ejercicio final

Objetivo

Realizar un análisis completo de datos utilizando PySpark y MySQL.

Instrucciones

1. ****Datos****
 - Busca y descarga un conjunto de datos de interés de una fuente en línea.

- Documenta la fuente y describe brevemente el contenido.

2. **MySQL**

- Carga los datos en MySQL.

3. **PySpark**

- Procesa los datos con PySpark:
 - Limpieza y transformación.
 - Análisis exploratorio con DataFrames.
 - Aplica al menos dos técnicas avanzadas (ej. Window Functions, UDFs).

4. **Análisis**

- Extrae al menos tres informaciones significativos.

5. **Visualización**

- Crea al menos una visualización significativa de tus resultados.
- Exporta los resultados de tu análisis a un archivo CSV y luego impórtalos en una herramienta externa como Excel, Google Sheets o Google Looker Studio para crear la visualización.
- Incluye una captura de pantalla o imagen de tu visualización en el informe.