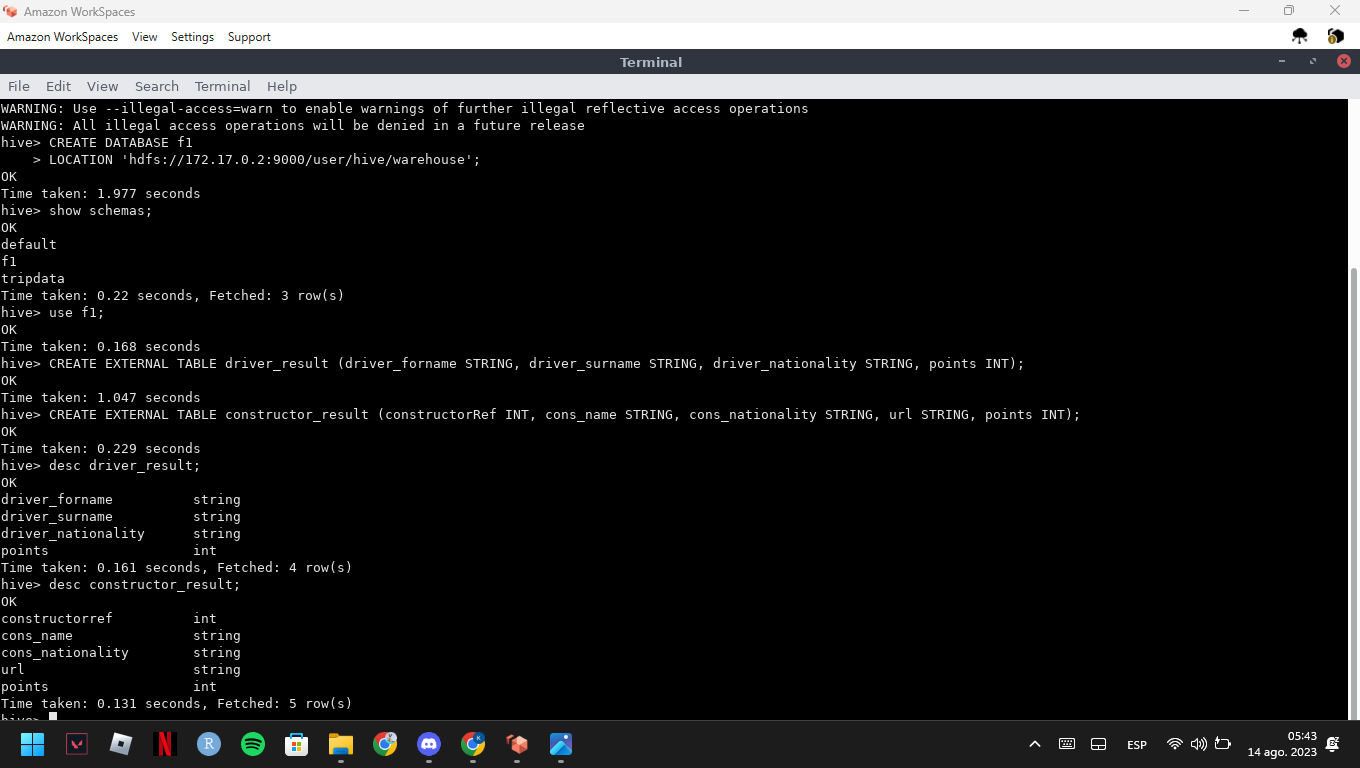
**Clase 8**

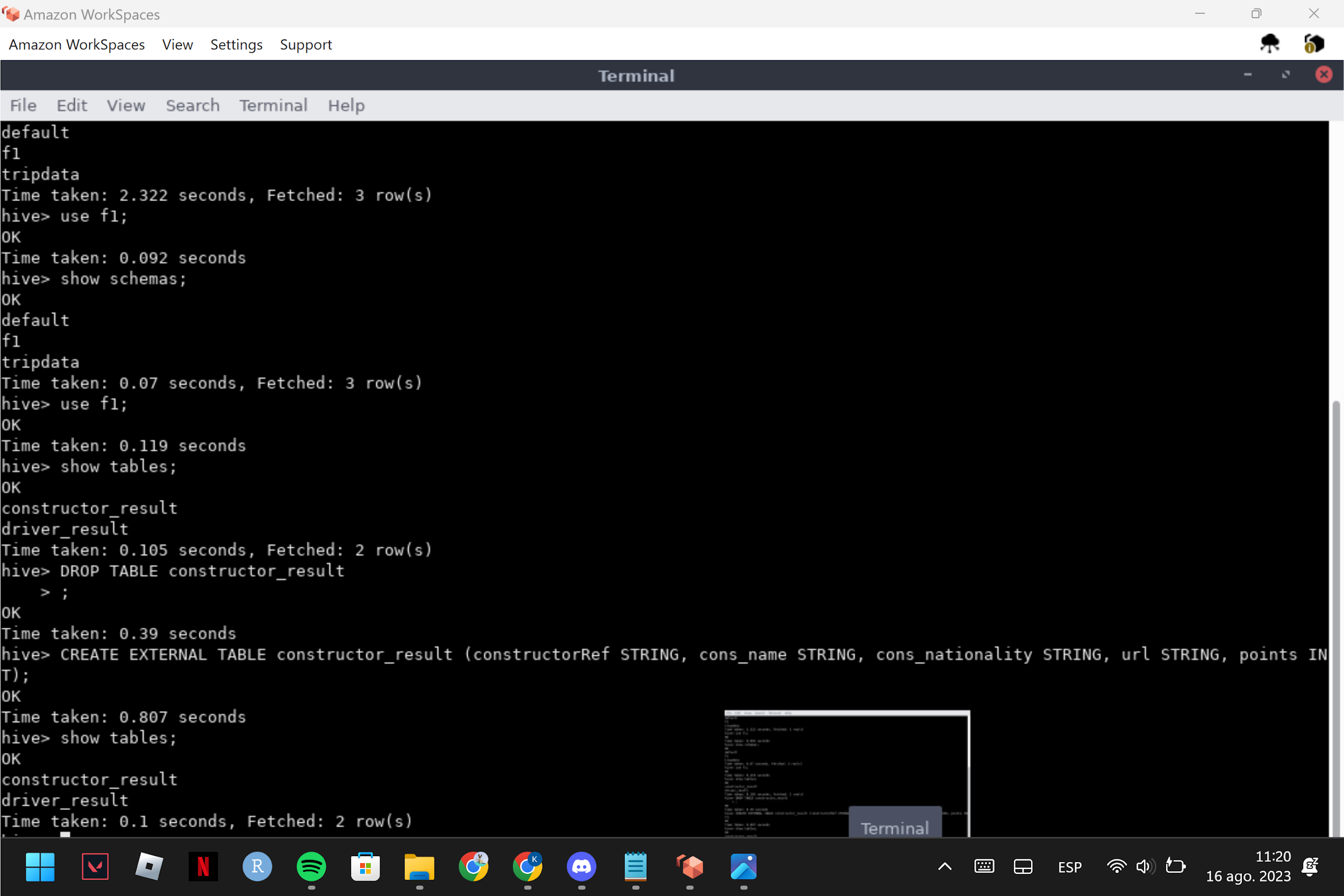
Consigna: Por cada ejercicio, escribir el código y agregar una captura de pantalla del resultado obtenido.

Diccionario de datos: https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020?se lect=results.csv

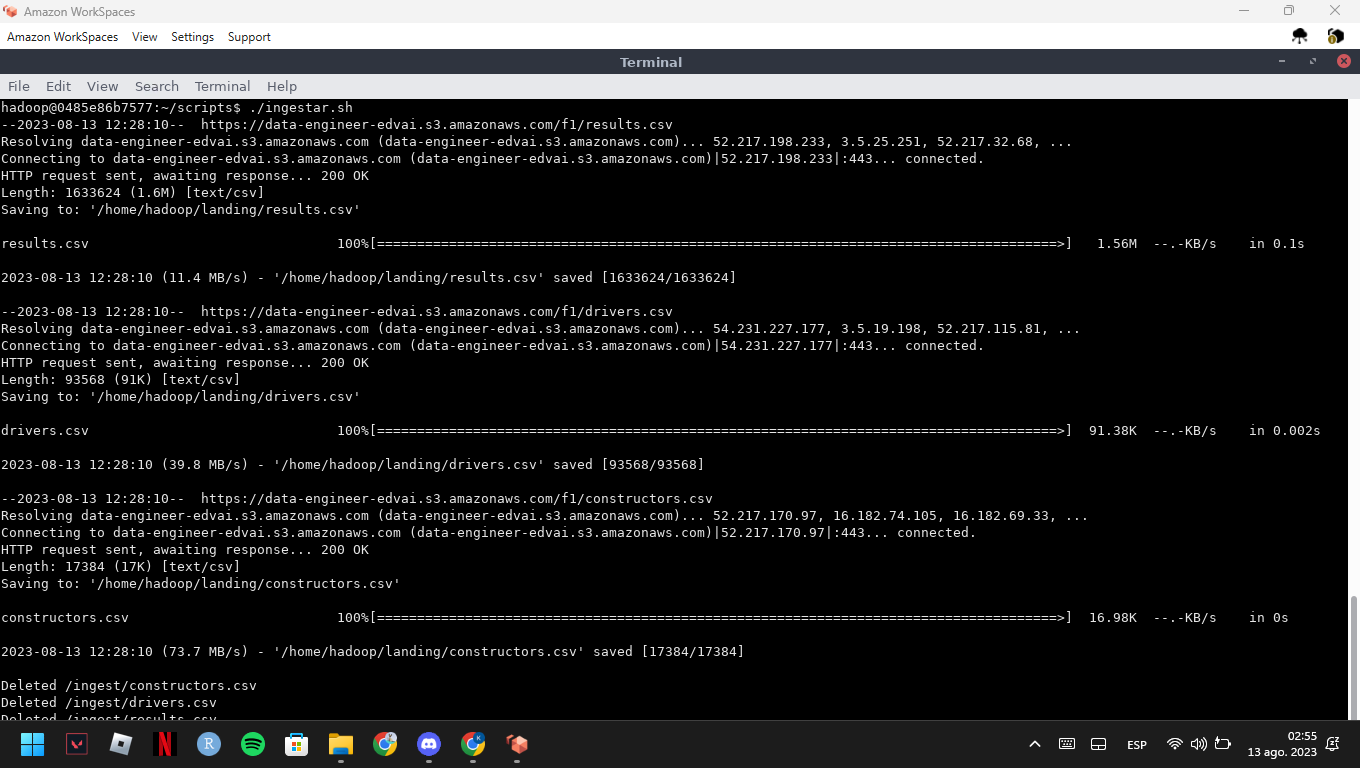
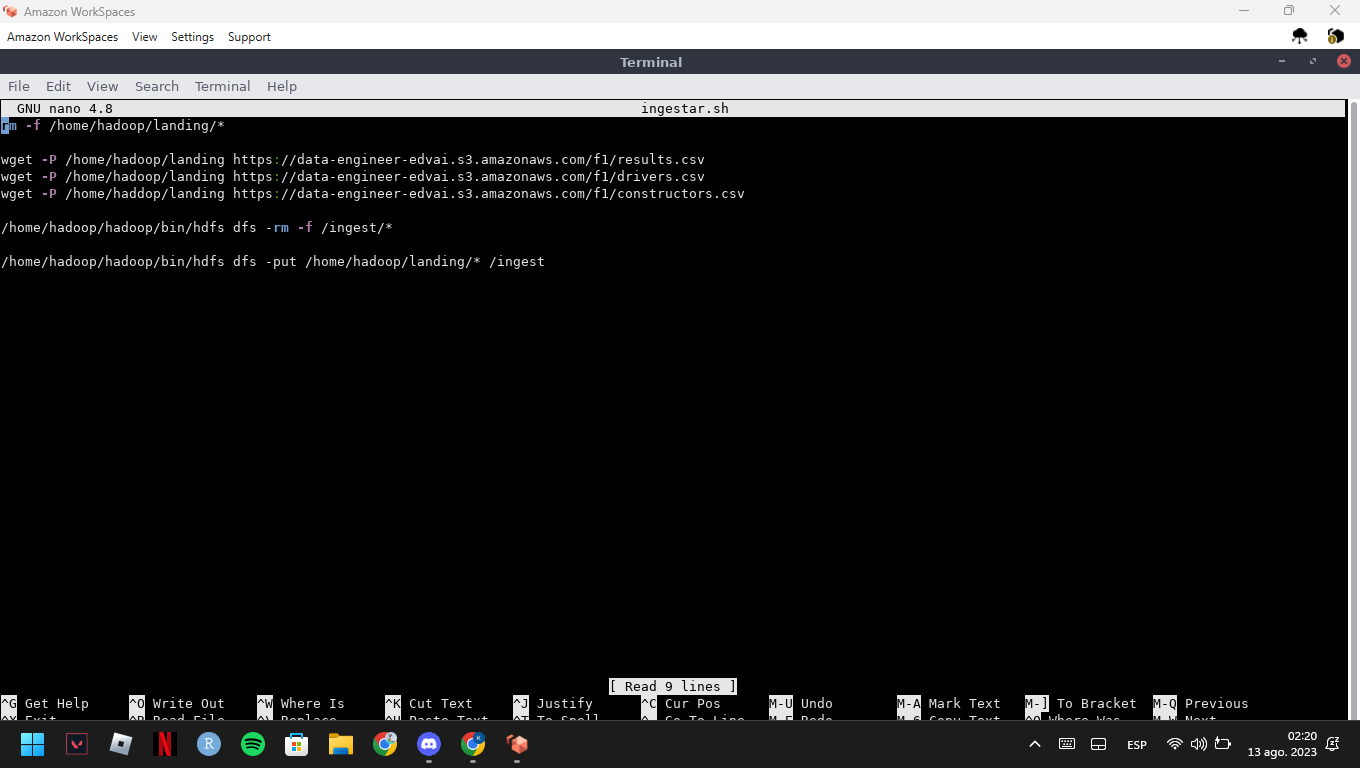
1. Crear la siguientes tablas externas en la base de datos f1 en hive: a. driver\_results (driver\_forename, driver\_surname, driver\_nationality, points) b. constructor\_results (constructorRef, cons\_name, cons\_nationality, url, points)

2. En Hive, mostrar el esquema de driver\_results y constructor\_results.

NOTA: aquí hubo un error al crear la tabla constructor, puse a la columna constructorRef como INT, asi que hice un drop table y una nueva creación con el tipo de dato STRING. Me di cuenta del error al intentar hacer el insert en la tabla constructor\_results.

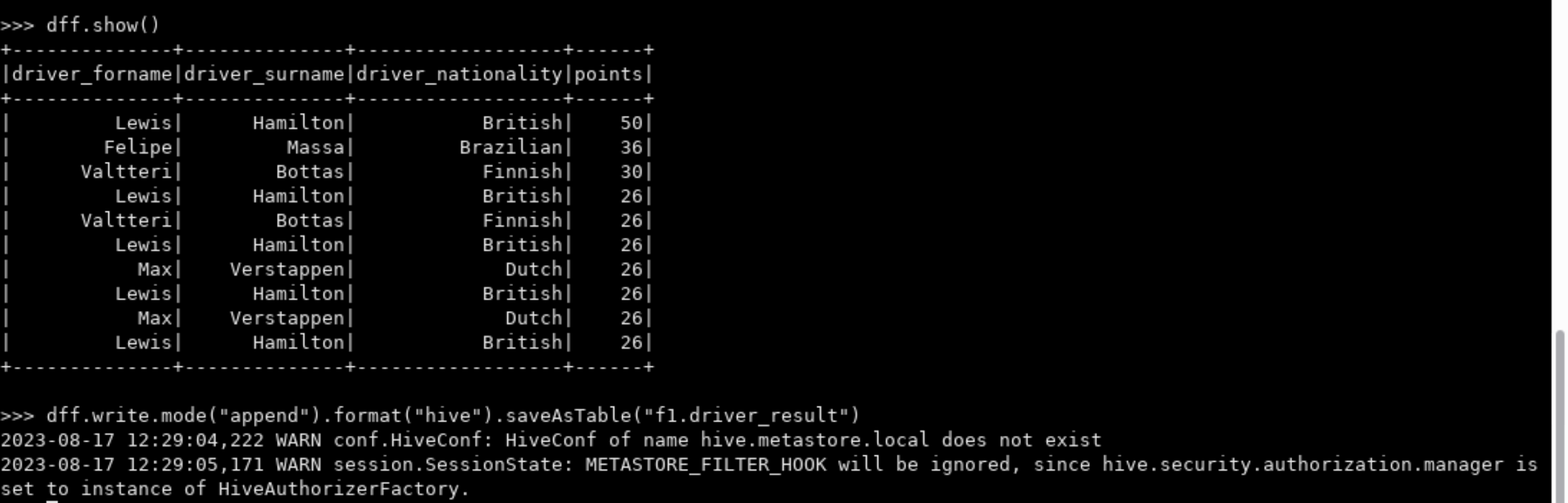
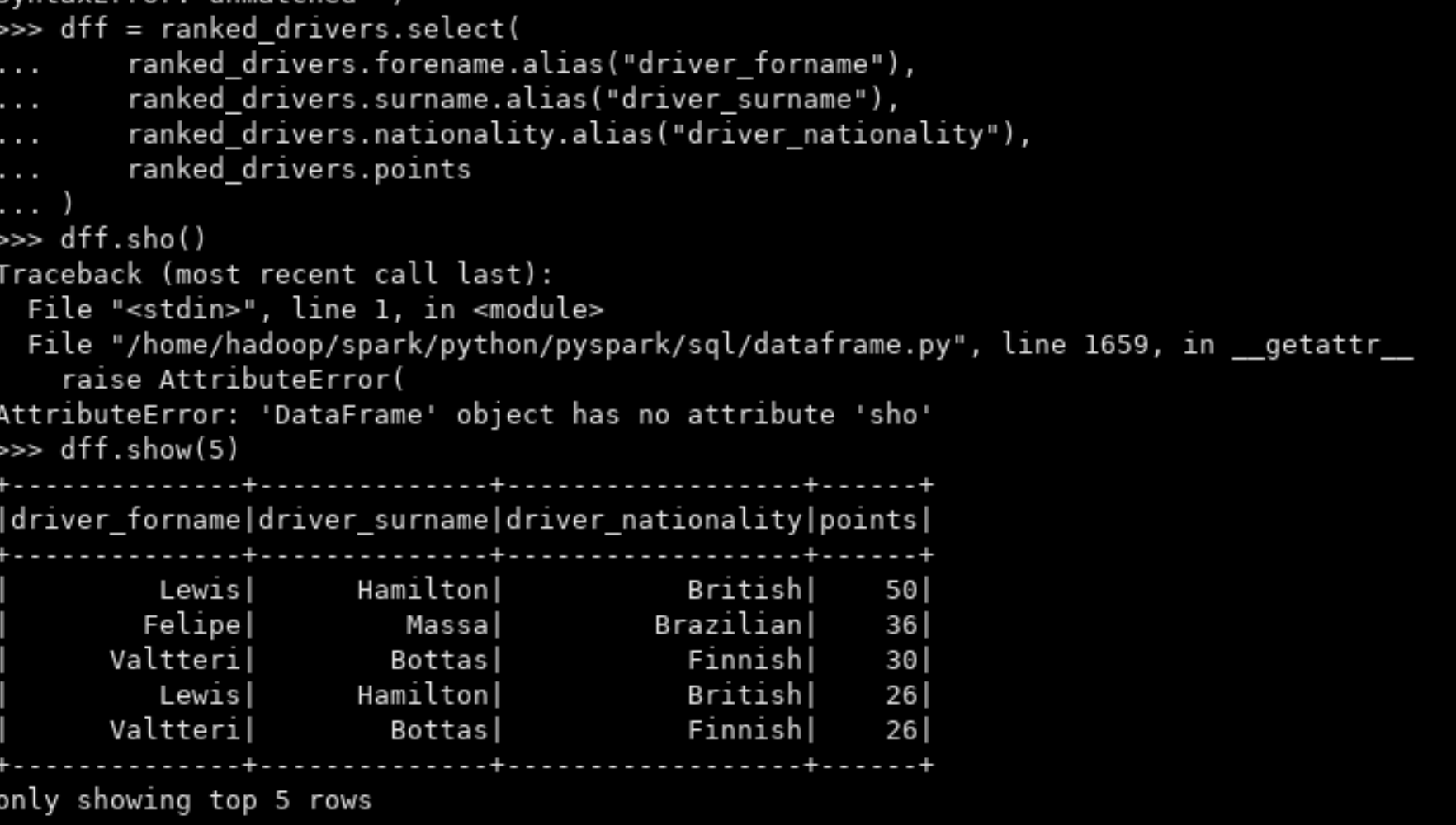
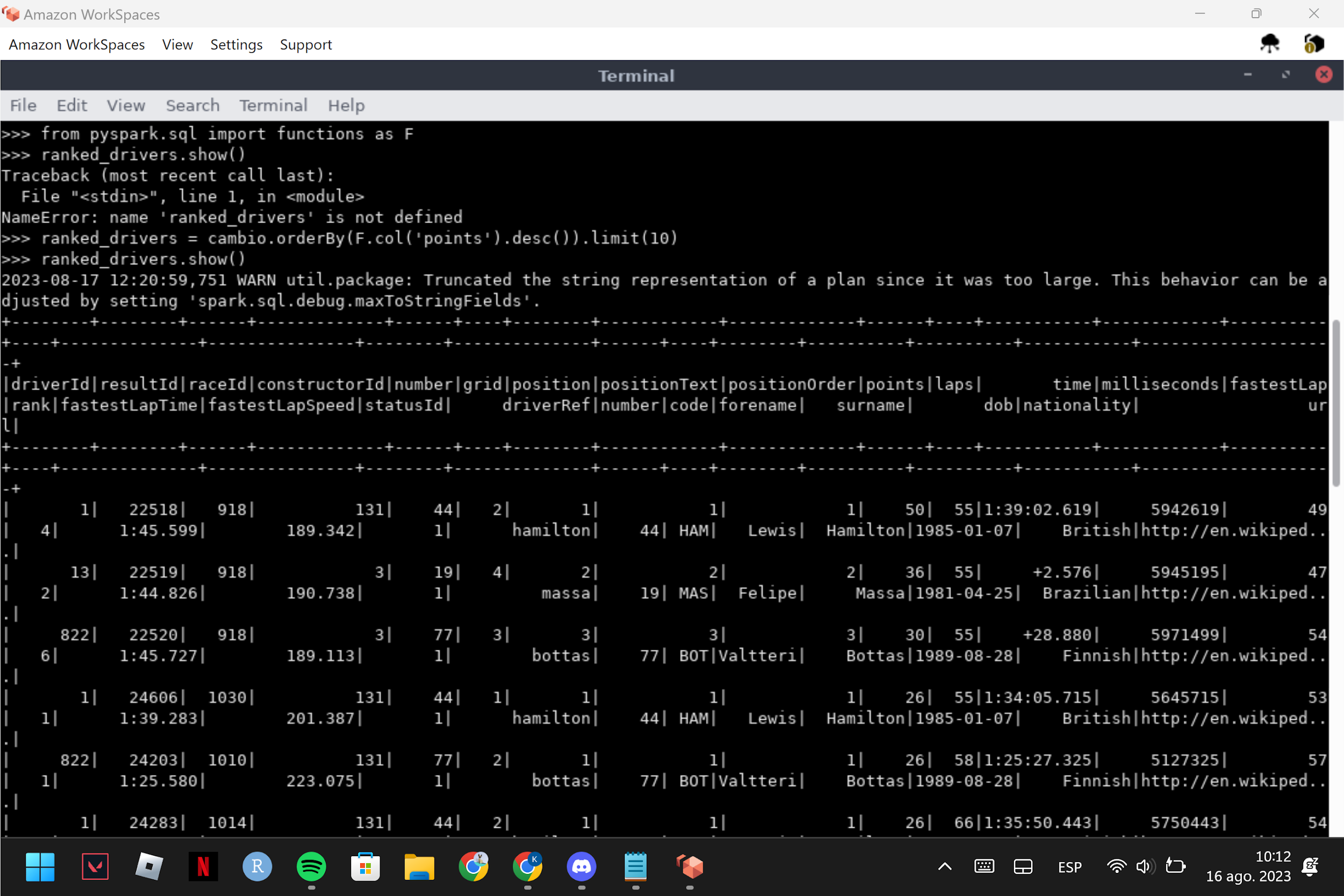
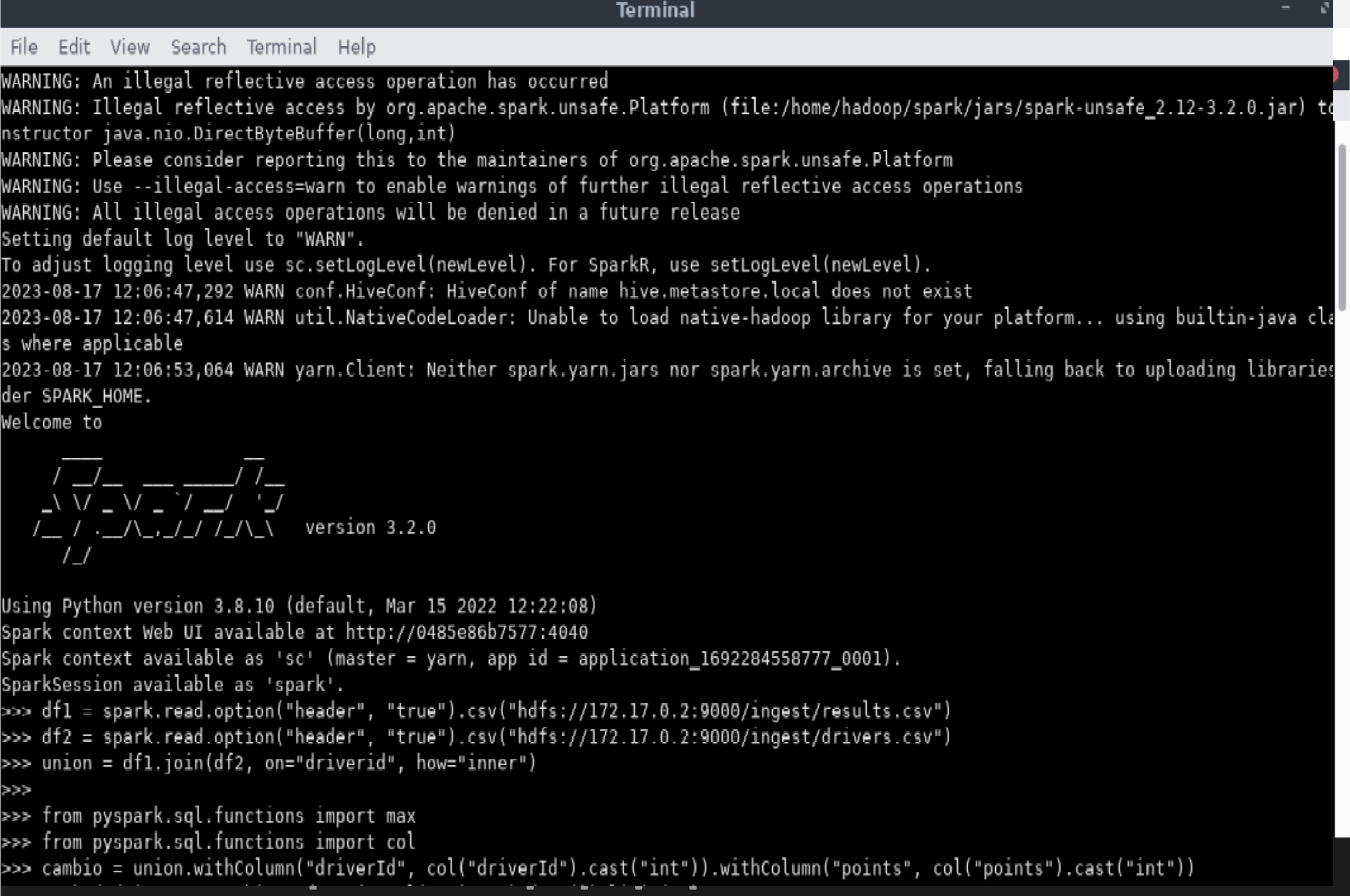


3. Crear un archivo .bash que permita descargar los archivos mencionados abajo e ingestarlos en HDFS: results.csv https://data-engineer-edvai.s3.amazonaws.com/f1/results.csv drivers.csv https://data-engineer-edvai.s3.amazonaws.com/f1/drivers.csv constructors.csv <https://data-engineer-edvai.s3.amazonaws.com/f1/constructors.csv>

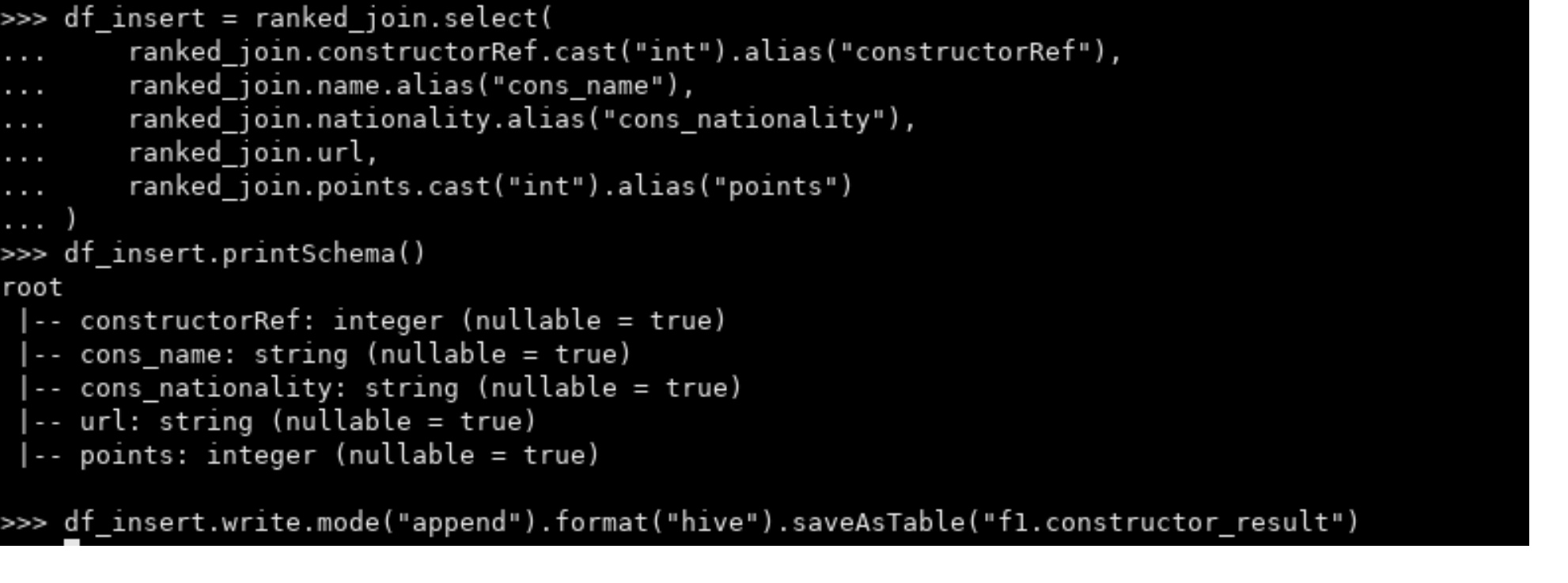
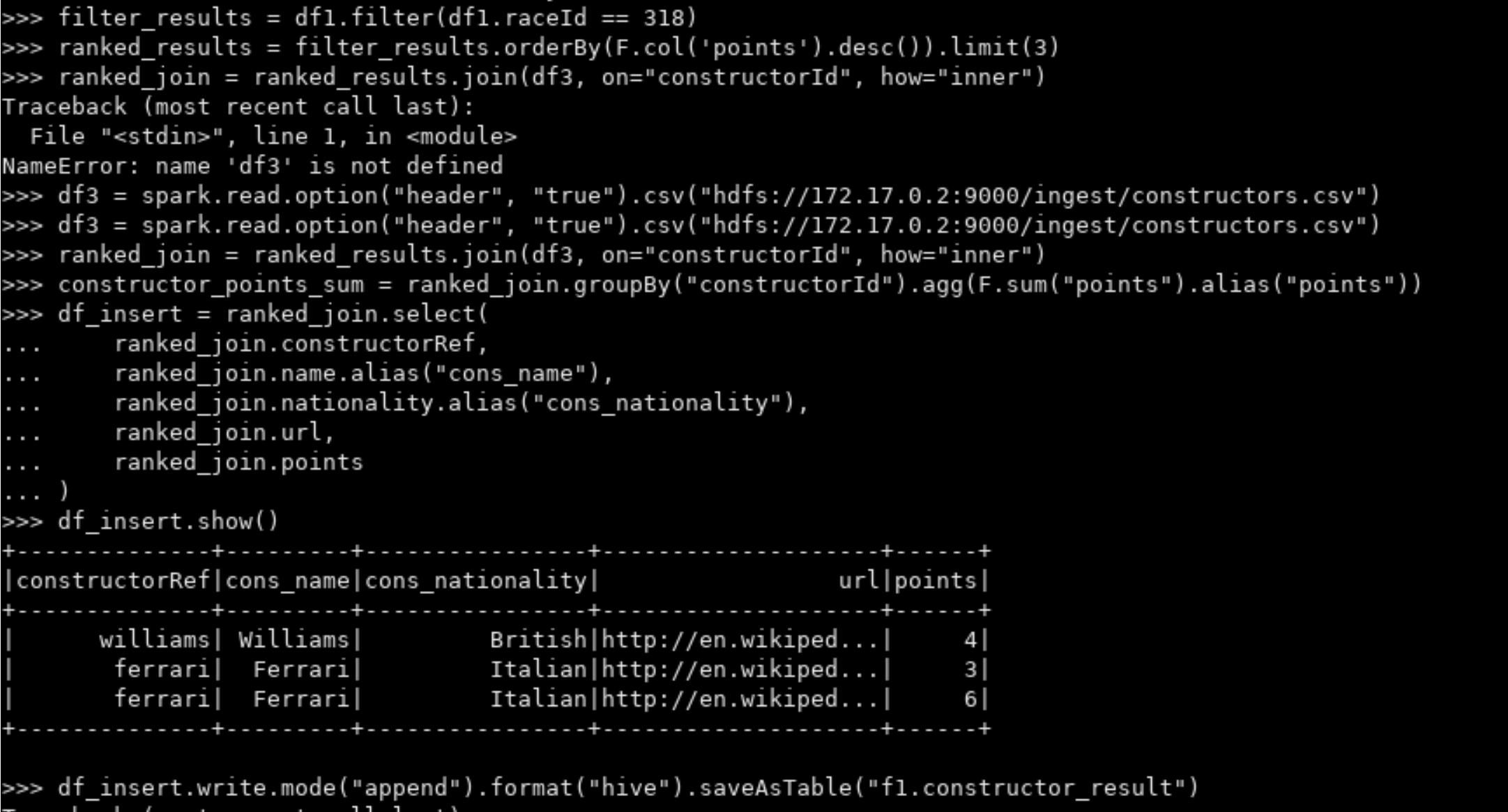


4. Generar un archivo .py que permita, mediante Spark:

a. insertar en la tabla driver\_results los corredores con mayor cantidad de puntos en la historia.



b. insertar en la tabla constructor\_result quienes obtuvieron más puntos en el Spanish Grand Prix en el año 1991.



5. Realizar un proceso automático en Airflow que orqueste los archivos creados en los puntos 3 y 4. Correrlo y mostrar una captura de pantalla (del DAG y del resultado en la base de datos)

