

Práctica 2 - Spark - Dataframes - SQL

De los problemas que se enuncian a continuación debe entregar el código fuente o notebook de jupyter con la resolución.

Fuente: <https://www.kaggle.com/marshallproject/crime-rates>

1. Cargue en un dataframe el archivo report.csv. Imprima en pantalla el esquema (columnas y tipos de datos asociados).
2. Muestre en pantalla 3 registros del dataframe.
3. ¿Qué datos contiene el archivo? ¿Qué información útil se podría obtener?
4. Crear una “vista” llamada “crimes”.
5. Sobre la vista creada en el inciso anterior realice las siguientes consultas utilizando exclusivamente SQL:
 - a. Obtener la cantidad de crímenes violentos por año en Detroit, Michigan. Mostrar los resultados de los años recientes primero.
 - b. Mostrar la cantidad de homicidios en la ciudad de Las Vegas por año. Mostrar los resultados por año de forma creciente.
6. Realizar las consultas del ejercicio 5 utilizando los métodos del dataframe (select(), filter(), where(), etc).
7. Guardar el resultado de la consulta 5.a en un único archivo JSON.