

PS4

Anais Ouedraogo

February 21, 2023

0.1 Data sources in my interest

I would be interested in scrapping tweets that have the "Police violence" hashtag. I would like to analyze data on that issue. I will also be interested in using IMDb API to analyze the different movies and data about them.

0.2 Answers to spark questions

The two data frames are different types. df1 is a data frame while df is a spark type with sql and tbl. The columns' names are different across the two. df has dots in column names while df1 has underscores in column names. This can be due to the fact that spark does not allow dots in column names.