Name: - Muhammad Huzaifa Waseem (2303-KHI-DEG-021)

Pair Partner 1: - Muhammad Faizan Rafique (2303.005.KHI.DEG)

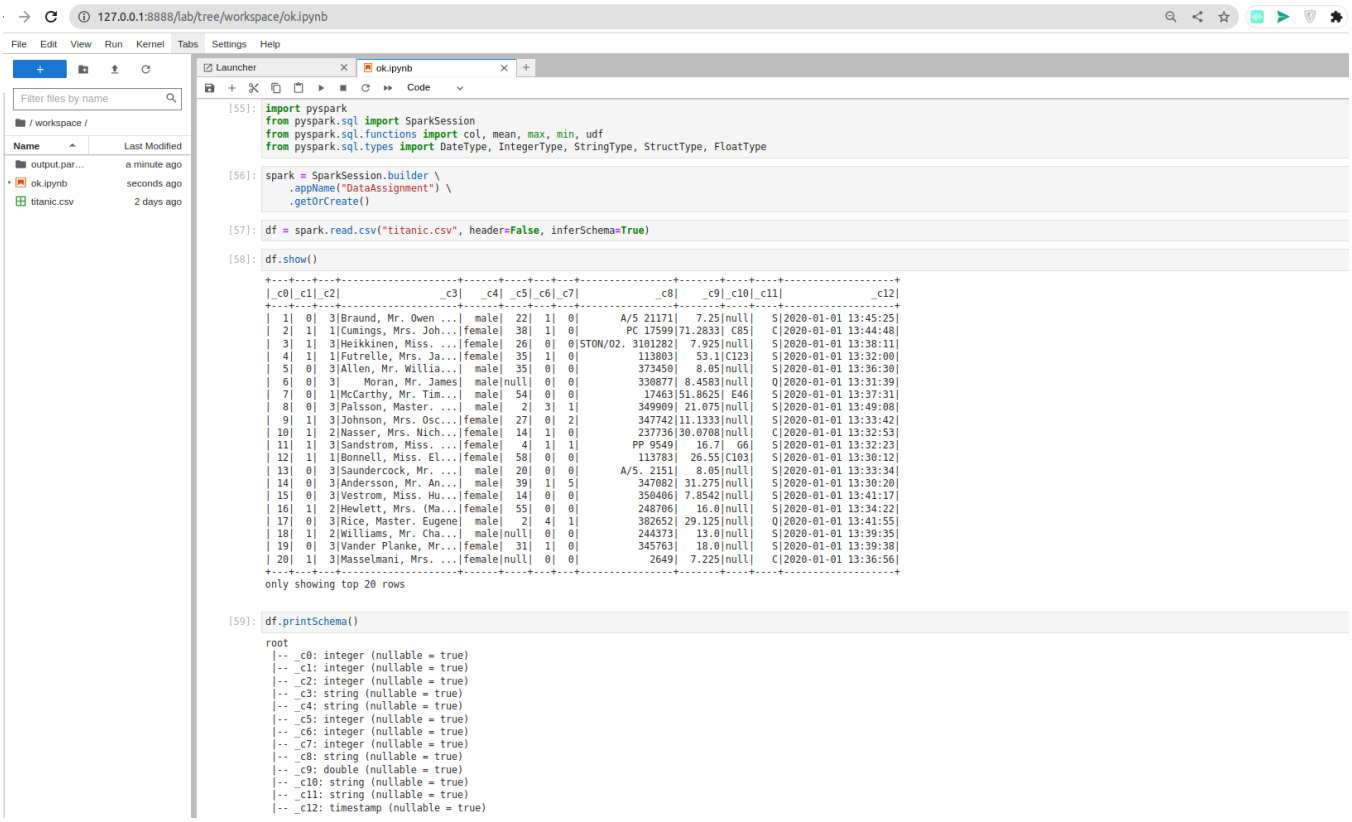
Pair Partner 2: - Syed Muhammad Hammad Irshad(2303.KHI.DEG.032)

UNIT 5.3:

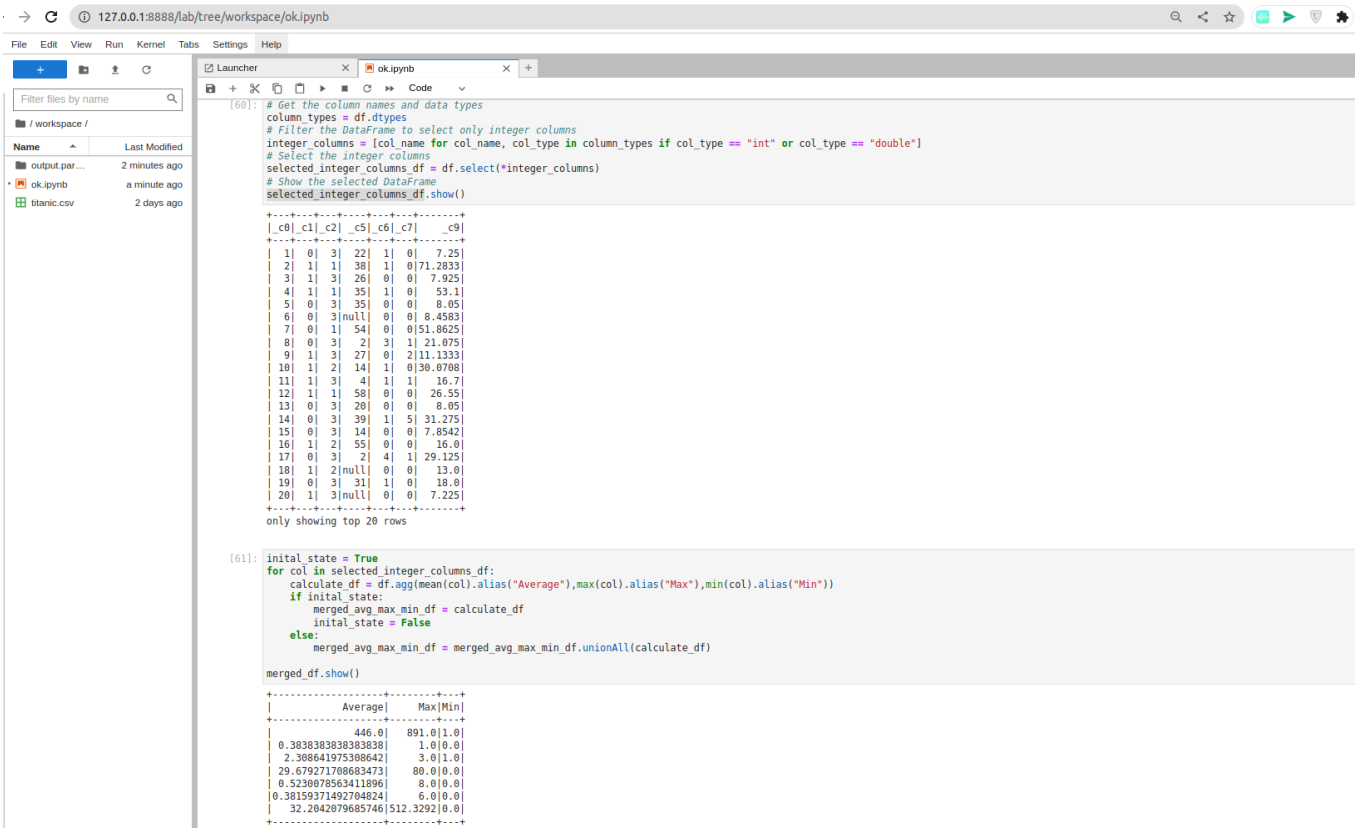
Assignment

Read data from source to DataFrame in local Spark setup and display DataFrame schema.

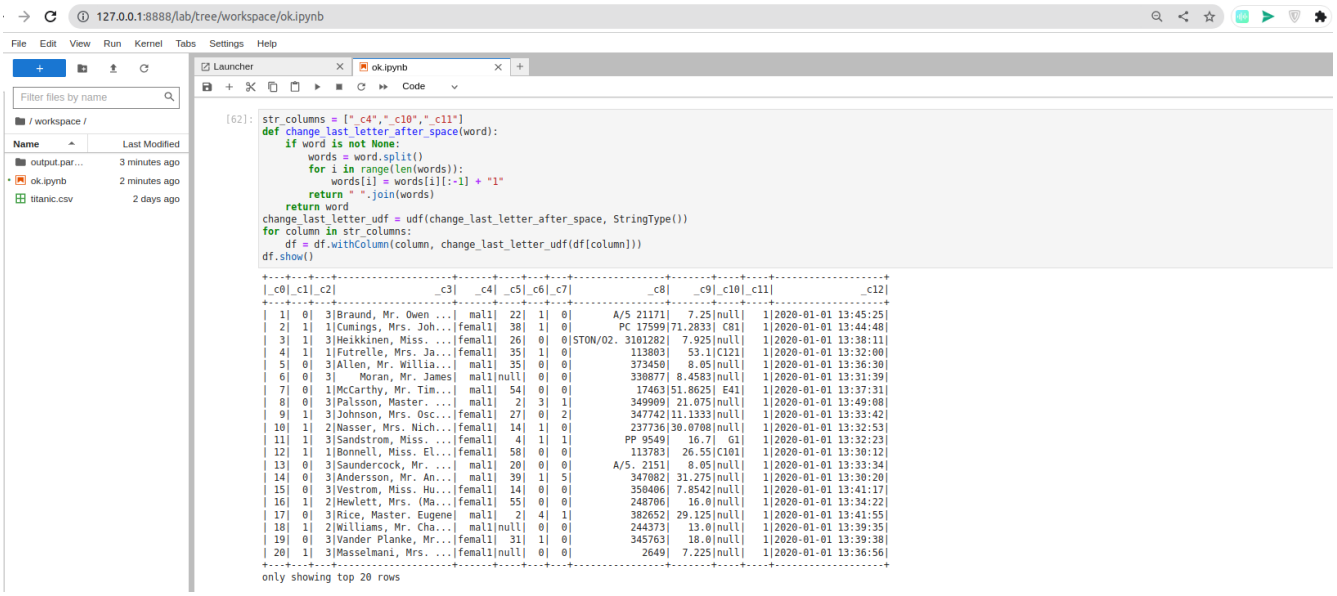
`tasks/4_data_pipelines/day_3_spark/data_assignment`



For numerical columns, calculate minimum, maximum and average values.



For categorical columns, create and apply UDF that will change the last letter of every word to “1”.



Sort DataFrame by the first column and save the results to the Parquet file.