Pyspark Assignment

Q.1 For Superheroes dataset:

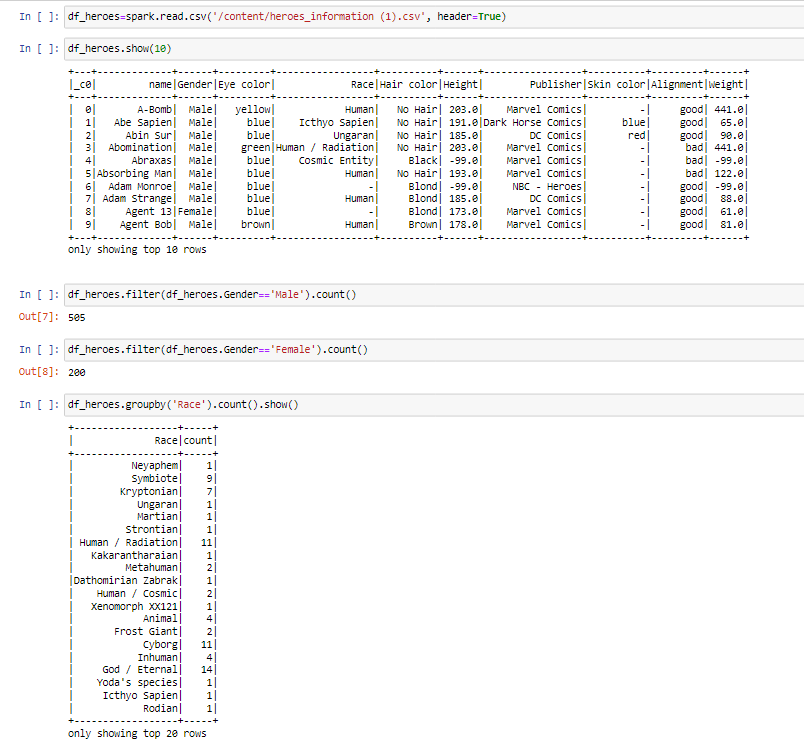
1. Reading Data from file

2. Show only top 10 records.

3. Find the count of male heroes.

4. Find the count of Female heroes.

5. Find the count of the particular race.



Q.2. For World cup players dataset:

1. Reading Data from file

2. Schema of Dataframe : printSchema

3. count of the number of Rows and Columns of Dataframe.

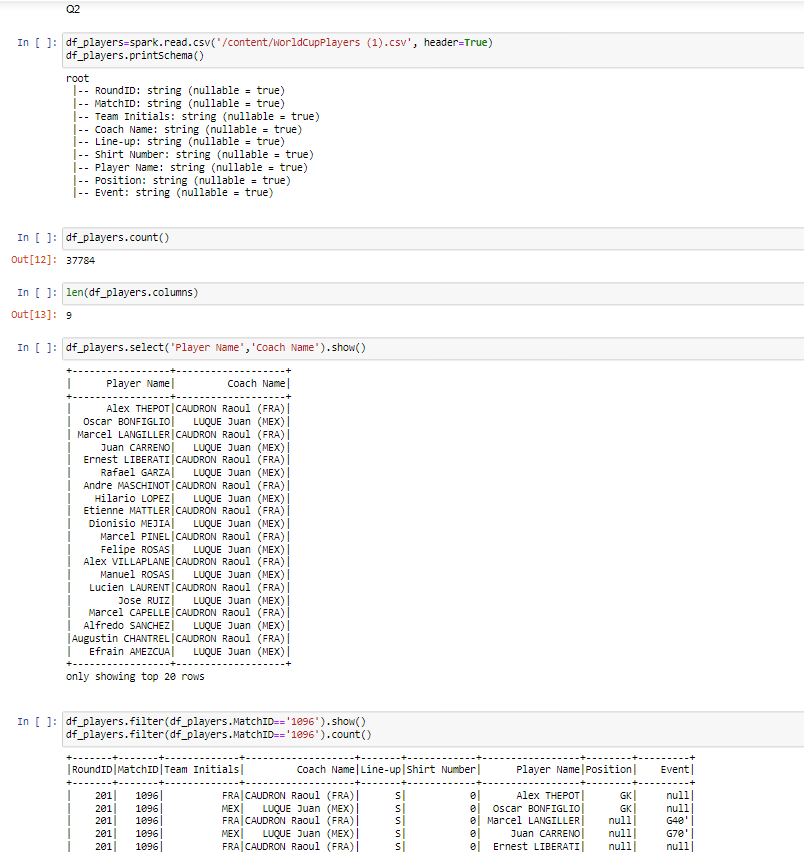
4. Show the player name and coach name

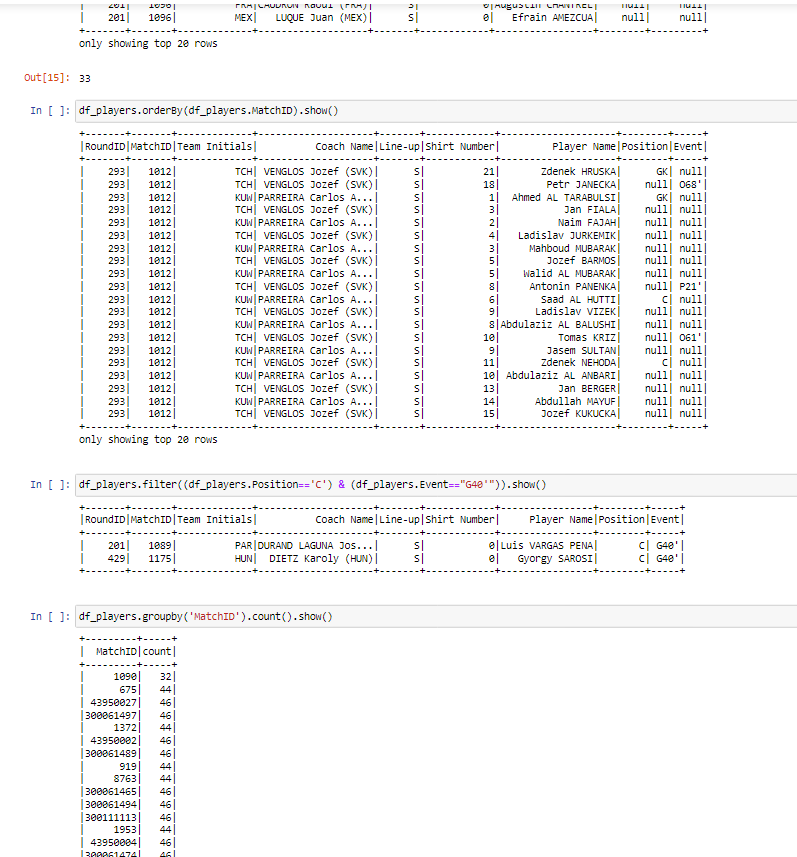
5. Show the records where Match ID must be equal to 1096 and calculate how many records/rows are there in the filtered output.

6. Sort the data by match id

7. Show the records of position C and event is G40.

8. Show the unique records with respect to match id.





Q.3 use olympics.csv file

1. Load the file

2. Filter the records of India

3. Show the final result as country and the total number of medals won by each country in all the Olympic games.

4. Display the Athlete name and his country

5. Count the number of records in q.5

6. Show the count of each country in a particular sport.

