Dear Participant,

Please find below the Project for Statistical Learning course. This is an individual assignment. Kindly submit it before the deadline.

Please find below the dataset to be used for the project

Github Link:

Think of yourself, currently working as a Business analyst in one of the top sports companies. The senior management team has asked you to come up with metrics with which they can gauge which team will win the upcoming La Liga cup (Football tournament).

The data set contains information on all the teams so far participated in all the past tournaments. It has data about how many goals each team scored, conceded, how many times they came within the first 6 positions, how many seasons they have qualified, their best position in the past, etc.

You are required to do the following:

1. Read the data set and replace dashes with 0 to make sure you can perform arithmetic operations on the data. (5 points)
2. Print all the teams which have started playing between 1930-1980. Use “Debut” column (5 points)
3. Print the list of teams which came Top 5 in terms of points (2.5 points)
4. Write a function with name “Goal_diff_count” which should return all the teams with their Goal Differences. Using the same function, find the team which has maximum and minimum goal difference. (5 points)

Goal_diff_count = GoalsFor - GoalsAgainst

5. Create a new column with the name “Winning Percent” and append it to the data set (5 points)

Percentage of Winning = (GamesWon / GamesPlayed)*100
If there are any numerical error, replace it with 0%
Print the top 5 teams which have the highest Winning percentage

6. Group teams based on their “Best position” and print the sum of their points for all positions (7.5 points)

Eg:

Best Position	Points
1	25000
2	7000

Please Note:

- The points for each question is mentioned above.
- Total marks allotted for this assignment is 30.
- Assignment final submission should be a URL to your Github repository.

Regards

Program Office