Alternative Project – Applied Statistics

Domain: Sports

Context: La Liga is the men's top professional football division of the Spanish football league system. The dataset contains information on all the teams that have 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.

Data Description: Laliga.csv - The data set contains information on all the teams so far participated in all the past tournaments

Attribute Information:

Pos - Position in among the list of all teams

Team Seasons - how many seasons team has played so far

Points - total number of points of the team

GamesPlayed - total number of games played so far

GamesWon - total number of games won so far

GamesDrawn - total number of games drawn so far

GamesLost - total number of games lost so far

GoalsFor - total number of goals by the team

GoalsAgainst - total number of goals against the team

Champion - total number of times it team is a champion

Runner-up - total number of times it team is a runner-up

Third / Fourth/ Fifth/ Sixth - total number of times it team came in a third/fourth.fifth/sixth position

Debut - debut year

BestPosition - best position of the team

Objective:

We want to use statistical techniques to come up with metrics with which can be used to gauge the winning team in the upcoming La Liga cup (Football tournament)

Steps and tasks:

1. Read the data set and replace dashes with 0 to make sure you can perform arithmetic operations on the data. (10 points)
2. Print all the teams which have started playing between 1930-1980. Use “Debut” column (Include year 1930 only) (10 points)
3. Print the list of teams which came Top 5 in terms of points (5 points)
4. Write a function with the name “Goal_diff_count” which should return all the teams with their Goal Differences. (5 points)
5. Using the same function, find the team which has the maximum and minimum goal difference. (5 points)

   Hint: Goal_diff_count = GoalsFor - GoalsAgainst
6. Create a new column with the name “Winning Percent” and append it to the data set (5 points)

   Hint: Percentage of Winning = (GamesWon / GamesPlayed)*100. If there are any numerical error, replace it with 0%
7. Print the top 5 teams which have the highest Winning percentage (5 points)
8. Group teams based on their “Best position” and print the sum of their points for all positions (15 points)

   Eg:

Best Position	Points
1	25000
2	7000

Learning Outcomes:

- Exploratory Data Analysis
- Practicing statistics using Python

References:

- Applications of Data science in sports domain