Capstone Project - 1

IPL T20 Analysis

Exploration and Analysis of IPL T20 Cricket:

1. Removal and Handling of null or NaN values in the Dataset.
2. Exploratory Data Analysis.
3. Data Correlation and Trends in the data.

What is IPL T20?

The **Indian Premier League** (**IPL**) is a professional Twenty20 cricket league, contested by eight teams based out of eight different Indian cities. The league was founded by the Board of Control for Cricket in India (BCCI) in 2007. It is usually held between March and May of every year and has an exclusive window in the ICC Future Tours Programme.

The IPL is the most-attended cricket league in the world and in 2014 was ranked sixth by average attendance among all sports leagues. In 2010, the IPL became the first sporting event in the world to be broadcast live on YouTube. The brand value of the IPL in 2019 was ₹475 billion (US$6.7 billion), according to Duff & Phelps. According to BCCI, the 2015 IPL season contributed ₹11.5 billion (US$160 million) to the GDP of the Indian economy.IPL 2020 set a Massive Viewership Record With 31.57 Million Average Impressions and with an overall consumption increase of 23 percent from 2019 season.

There have been thirteen seasons of the IPL tournament. The current IPL title holders are the Mumbai Indians, who won the 2020 season. The venue for the 2020 season was moved due to the COVID-19 pandemic and games were played in the United Arab Emirates.

Data Processing

- **Data Preprocessing**: Deletion of NaN values and replacing it with the respective values to process the data machine readable for ML and DL purposes.

- **EDA**: Exploratory Data Analysis is done on the dataset to get inference from the data and to see the visible trends.

Data Preprocessing

As the **First step**, the datasets are made free from NaN values by removing the NaN values and replacing it with other values:

1. **Deliveries df** had ~510000 null values and those are mainly in the player dismissed column and those are removed through filling the na values with zeros as it is more intuitive.
2. **Matches df** had ~1000 entries as null so that they are filled with one hot encoding, and replacing the null values by web surfing since there is no open source API available at this moment.
3. **Players df** had ~350 null values and those are replaced with mean age, and the country column is filled using the same distribution from the players df.
4. **Players Stats df** had ~30 null values which are replaced by calculating the average of the players.
5. **Teams df** had no null values.
6. **Teams home and away wins df** had no null values.

Exploratory Data Analysis

And the second step is the **Exploratory Data Analysis(EDA)** part, where the data is correlated and the trends in the data are discussed. The statistics obtained are as follows:

- **By Teams:**
  - Highest Runs
  - Lowest runs
  - Most Wins
  - Least Wins
  - Most Winning Percentage
  - Least Winning Percentage

- **By Players:**
  - Most runs
  - Most Sixes
  - Most Fours
  - Highest Strike Rate
  - Highest Average

Data Summary

Deliveries Dataset:

#	Column	Non-Null Count	Dtype
0	match_id	179078 non-null	int64
1	inning	179078 non-null	int64
2	batting_team	179078 non-null	object
3	bowling_team	179078 non-null	object
4	over	179078 non-null	int64
5	ball	179078 non-null	int64
6	batsman	179078 non-null	object
7	non_striker	179078 non-null	object
8	bowler	179078 non-null	object
9	is_super_over	179078 non-null	int64
10	wide_runs	179078 non-null	int64
11	bye_runs	179078 non-null	int64
12	legbye_runs	179078 non-null	int64
13	noball_runs	179078 non-null	int64
14	penalty_runs	179078 non-null	int64
15	batsman_runs	179078 non-null	int64
16	extra_runs	179078 non-null	int64
17	total_runs	179078 non-null	int64
18	player_dismissed	179078 non-null	object
19	dismissal_kind	179078 non-null	object
20	fielder	179078 non-null	object

Data Summary

Matches Dataset:

#	Column	Non-Null Count	Dtype
0	id	756 non-null	int64
1	Season	756 non-null	object
2	city	756 non-null	object
3	date	756 non-null	object
4	team1	756 non-null	object
5	team2	756 non-null	object
6	toss_winner	756 non-null	object
7	toss_decision	756 non-null	object
8	result	756 non-null	object
9	dl_applied	756 non-null	int64
10	winner	756 non-null	object
11	win_by_runs	756 non-null	int64
12	win_by_wickets	756 non-null	int64
13	player_of_match	756 non-null	object
14	venue	756 non-null	object
15	umpire1	756 non-null	object
16	umpire2	756 non-null	object

Data Summary

Players Dataset:

#	Column	Non-Null Count	Dtype
0	Player_Name	566 non-null	object
1	DOB	566 non-null	object
2	Batting_Hand	566 non-null	object
3	Bowling_Skill	566 non-null	object
4	Country	566 non-null	object

Data Summary

Players Stats Dataset:

#	Column	Non-Null Count	Dtype
0	batsman	516 non-null	object
1	total_runs	516 non-null	int64
2	out	516 non-null	int64
3	numberofballs	516 non-null	int64
4	average	516 non-null	float64
5	strikerate	516 non-null	float64

Data Summary

Teams Dataset:

#	Column	Non-Null Count	Dtype
0	team1	15 non-null	object

Teams home and away Dataset:

#	Column	Non-Null Count	Dtype
0	team	14 non-null	object
1	home_wins	14 non-null	int64
2	away_wins	14 non-null	int64
3	home_matches	14 non-null	int64
4	away_matches	14 non-null	int64
5	home_win_percentage	14 non-null	float64
6	away_win_percentage	14 non-null	float64

EDA:

By Teams:

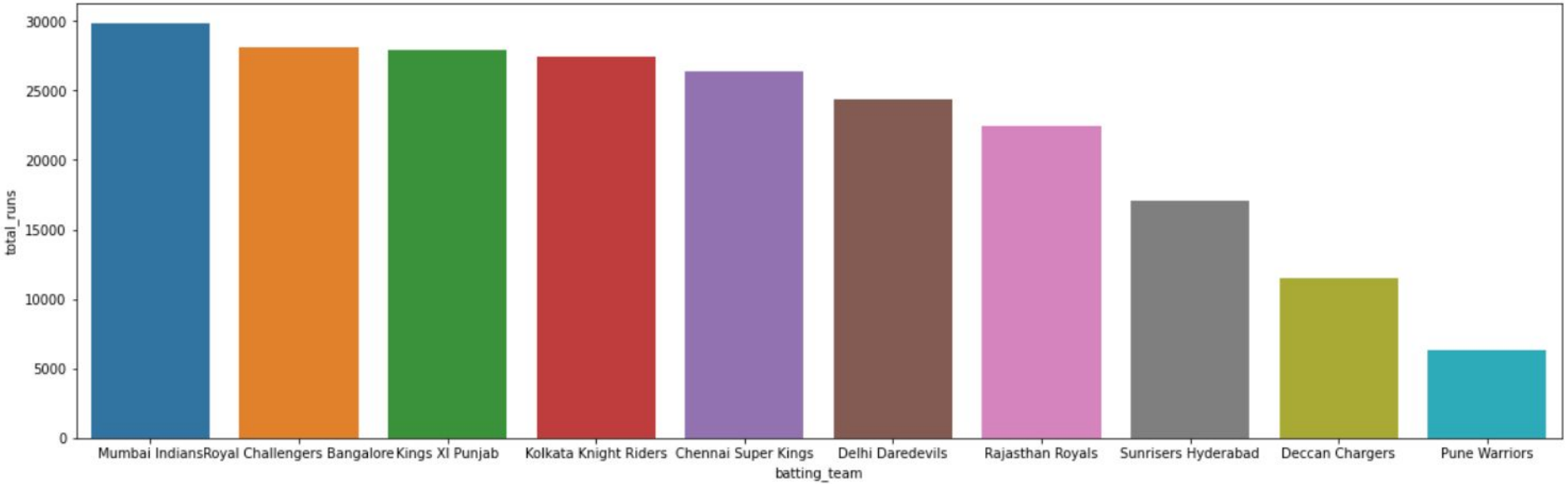
1. Highest Runs
2. Lowest Runs
3. Most Wins
4. Least Wins
5. Most Winning Percentage
6. Least Winning Percentage

By Players:

1. Most Runs
2. Most Sixes
3. Most Fours
4. Highest Strike Rate
5. Highest Average

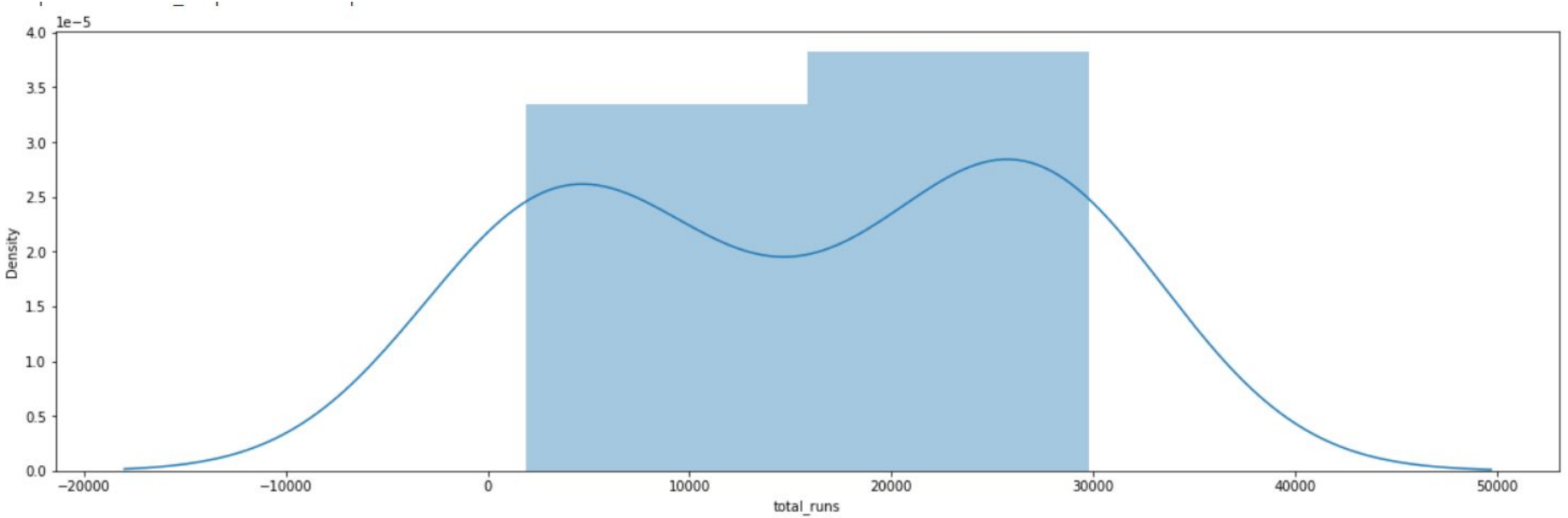
EDA(By Teams)

Highest Runs:



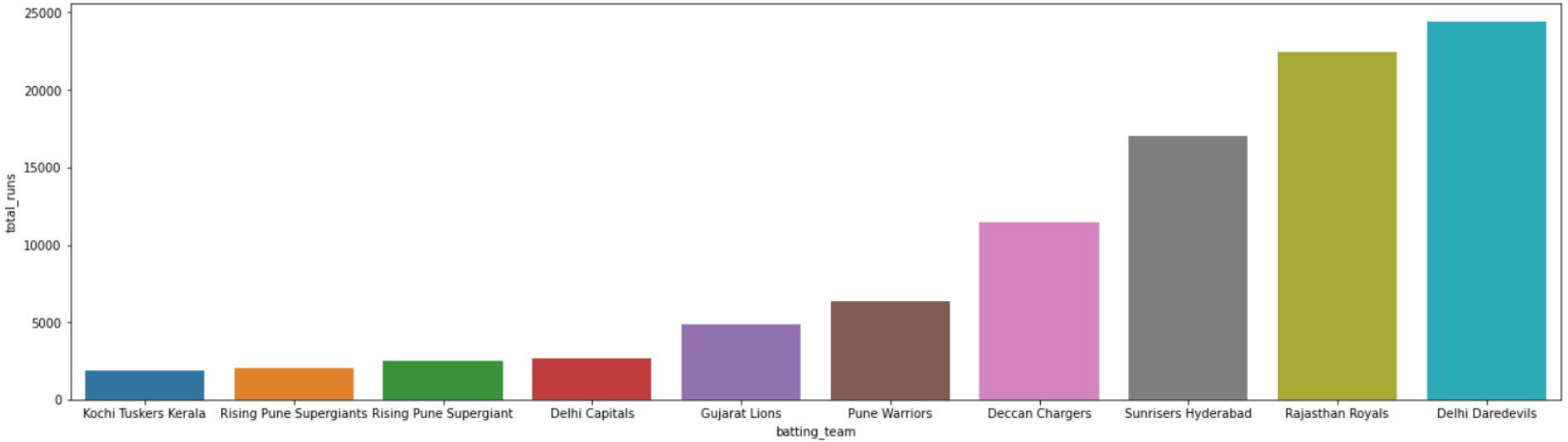
EDA(By Teams)

Highest Runs:



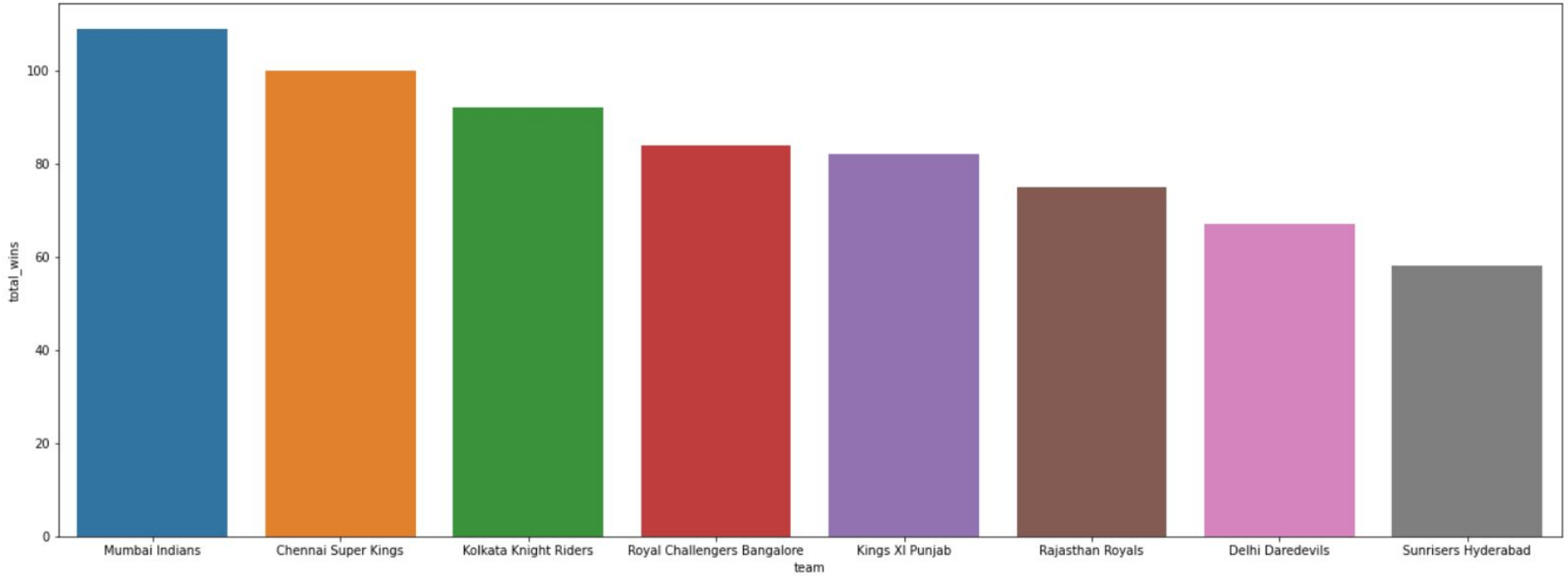
EDA(By Teams)

Lowest Runs:



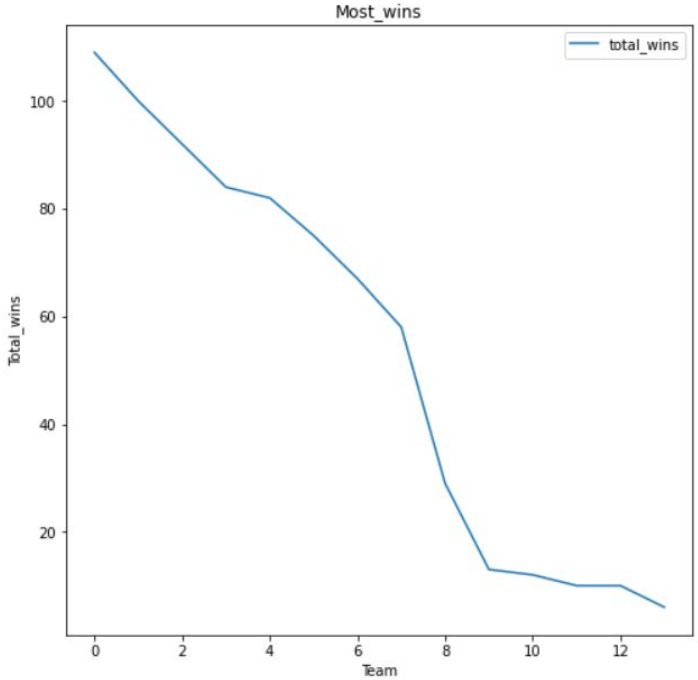
EDA(By Teams)

Most Wins:



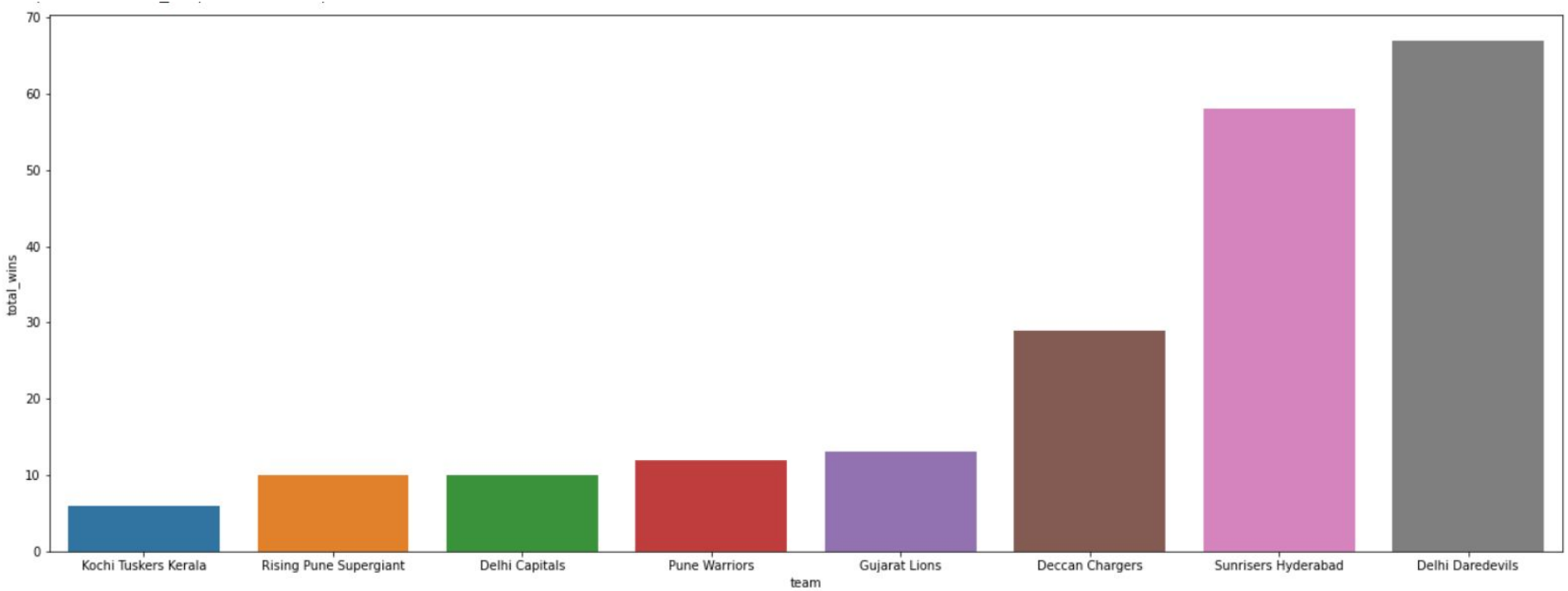
EDA(By Teams)

Most Wins:



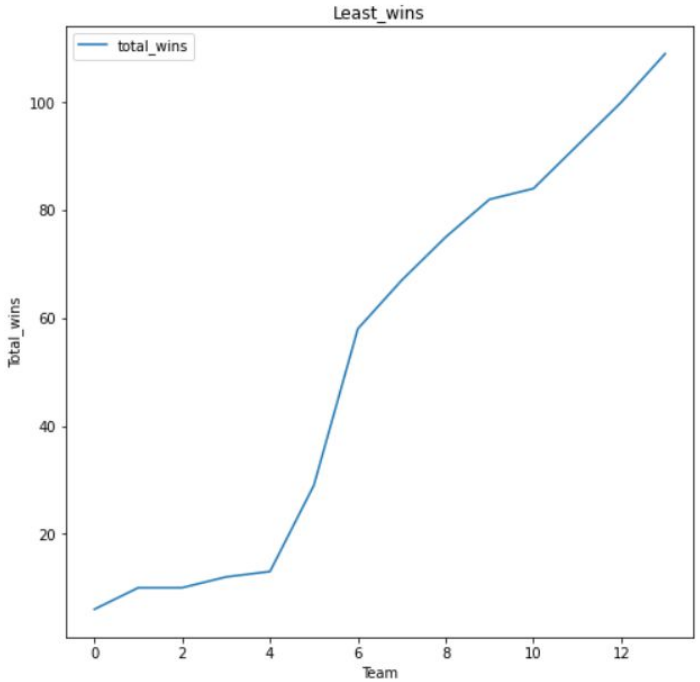
EDA(By Teams)

Least Wins:



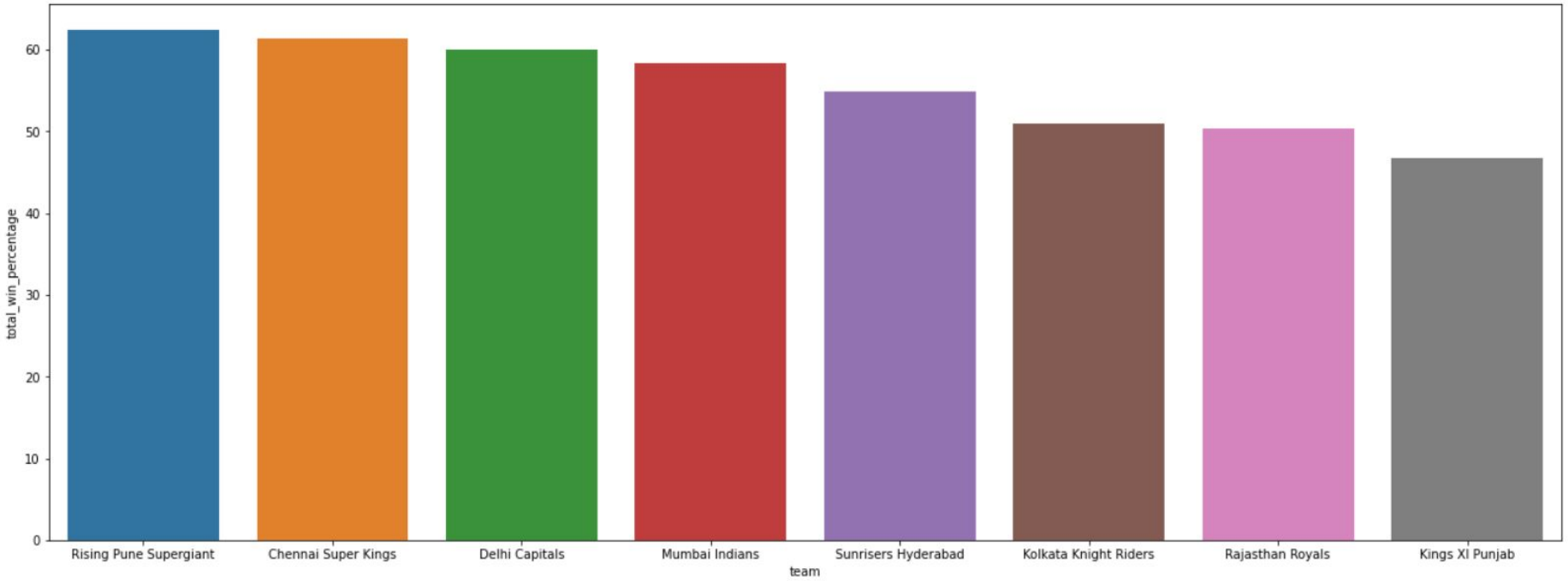
EDA(By Teams)

Least Wins:



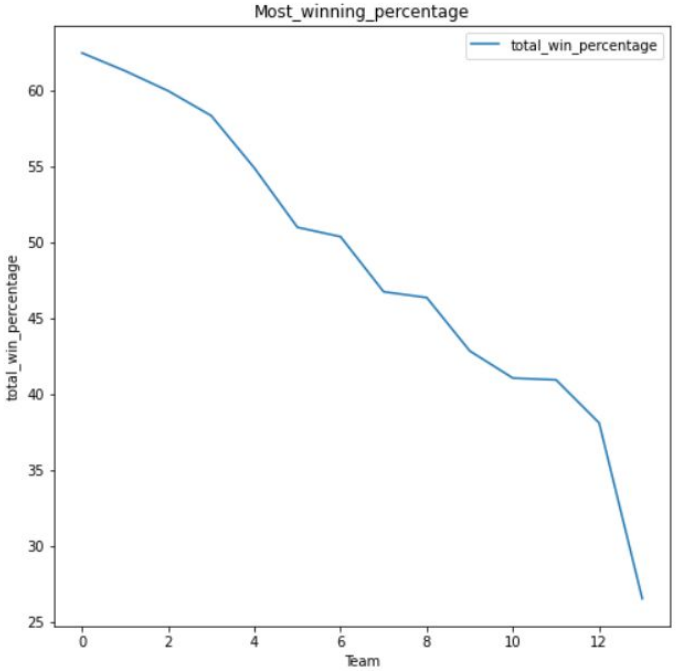
EDA(By Teams)

Most Winning Percentage:



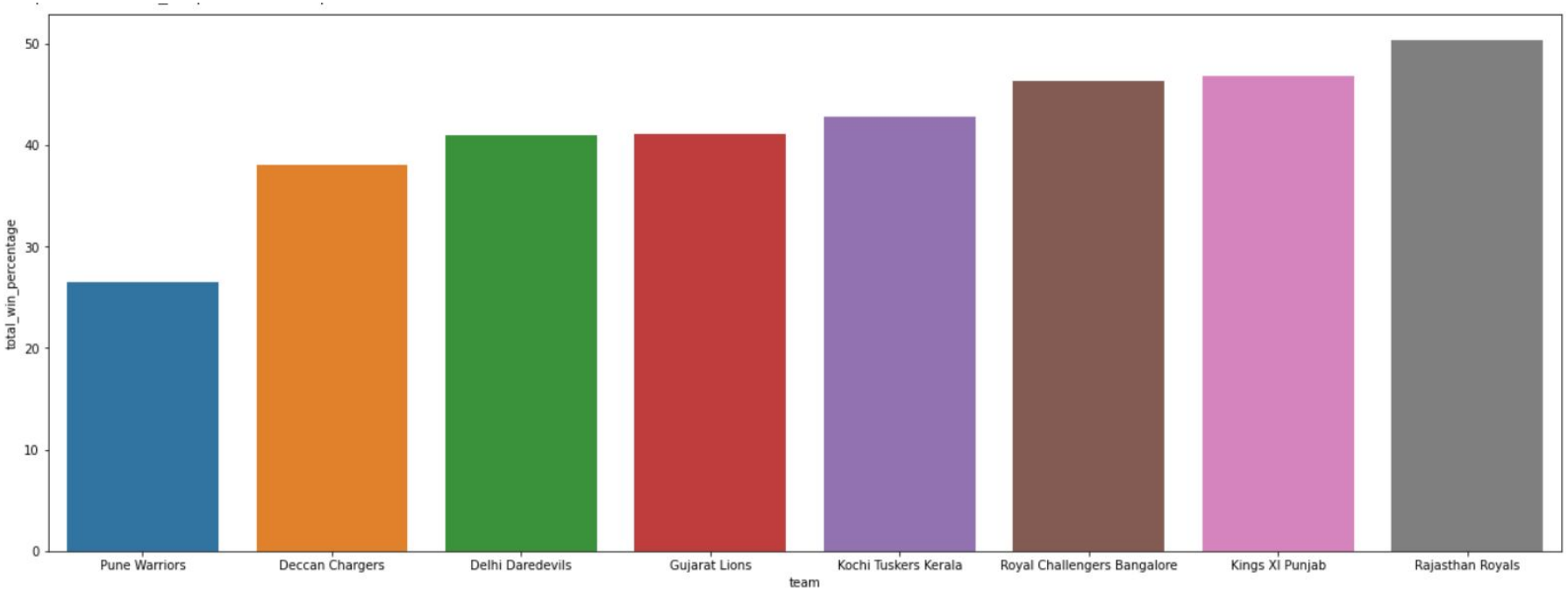
EDA(By Teams)

Most Winning Percentage:



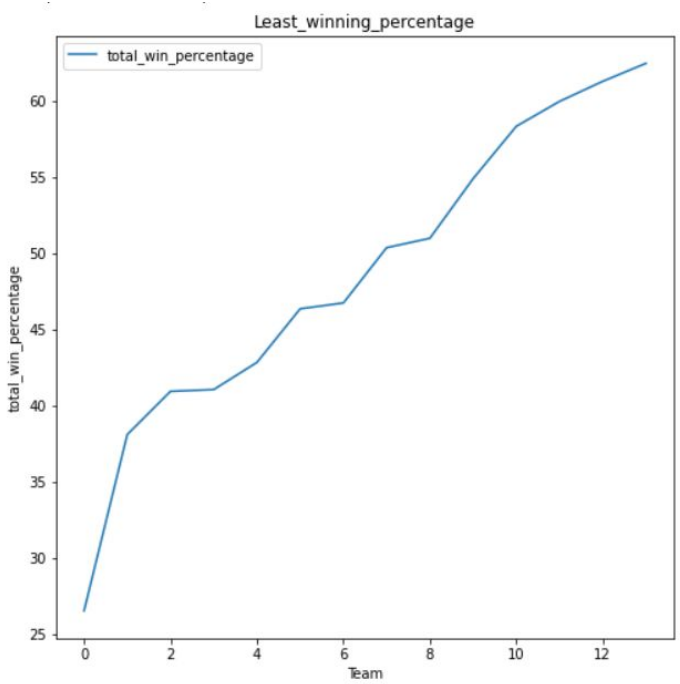
EDA(By Teams)

Least Winning Percentage:



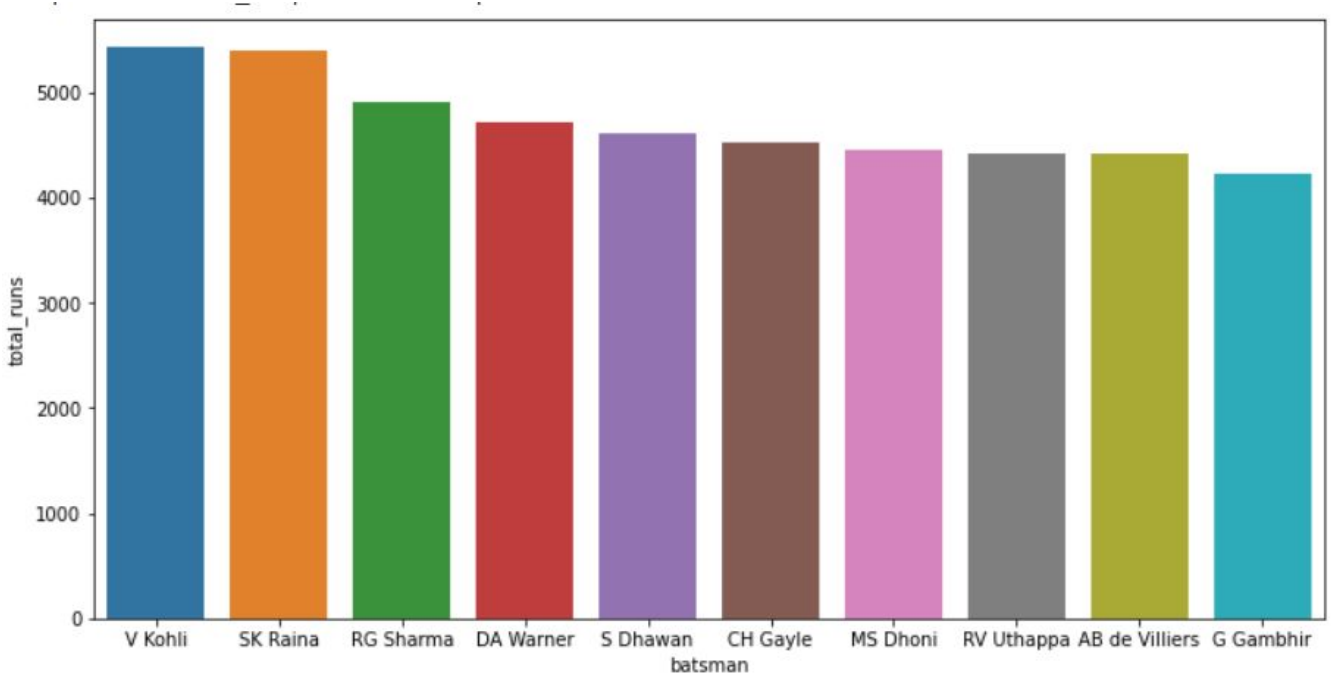
EDA(By Teams)

Least Winning Percentage:



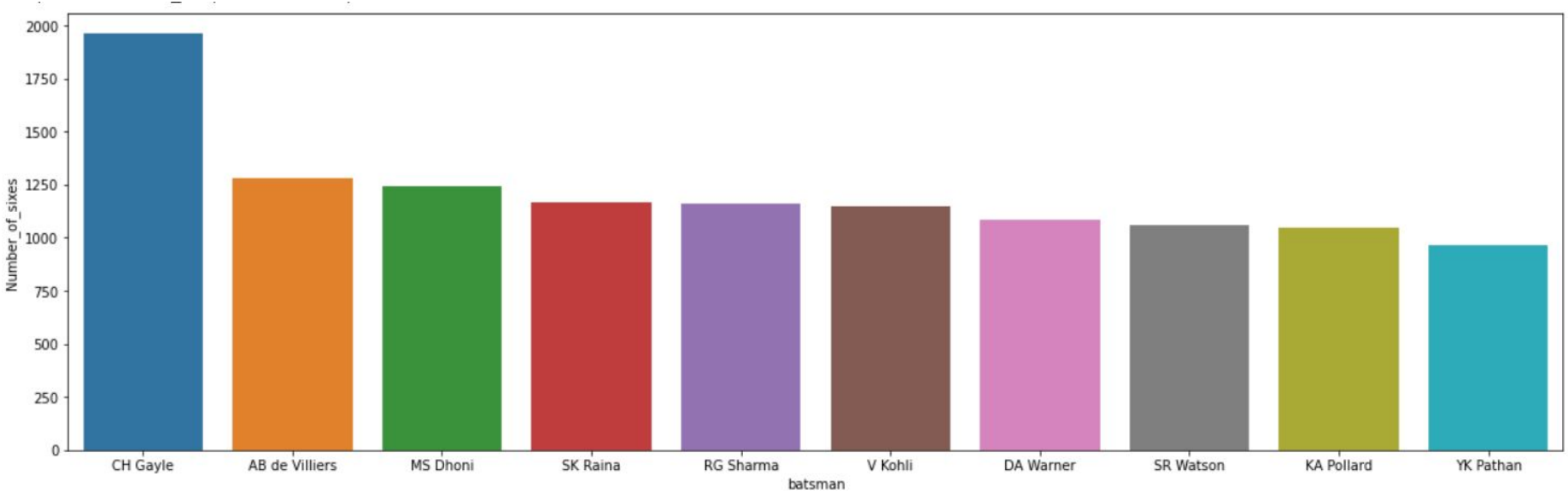
EDA(By Players)

Most Runs:



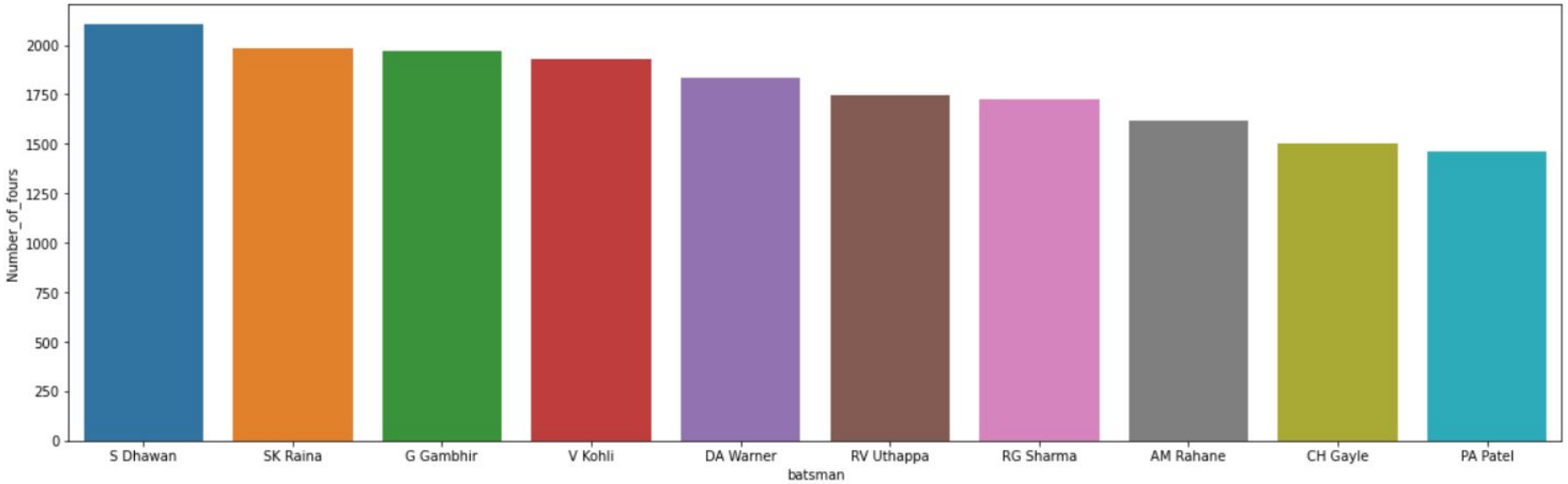
EDA(By Players)

Most Sixes:



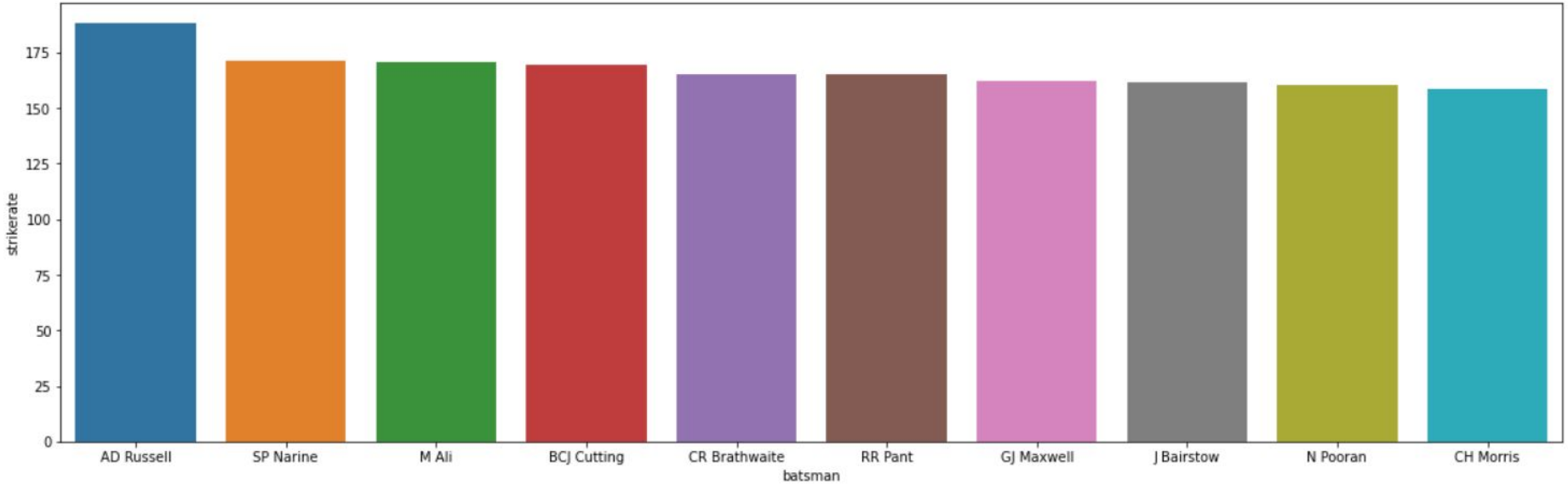
EDA(By Players)

Most Fours:



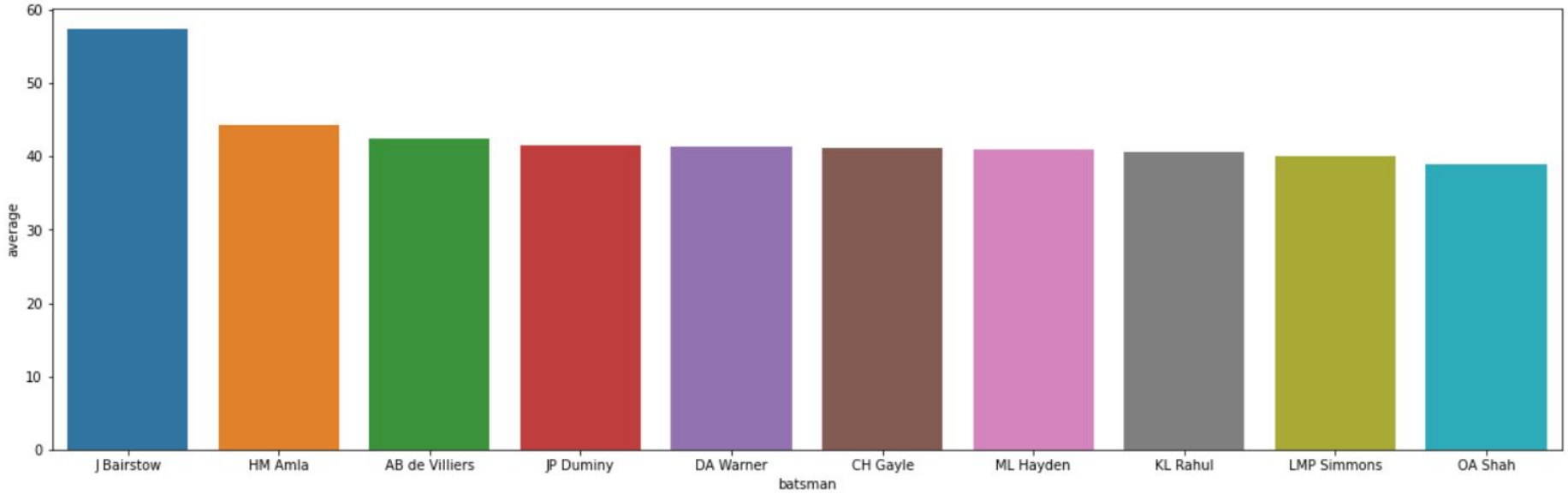
EDA(By Players)

Highest Strikerate:



EDA(By Players)

Most Sixes:



The End