

Capstone Project

EDA Of Indian Premier League (IPL)

Team Members

Navin Kodam
Jyoti Chiluka

Content:

- ❑ Introduction
- ❑ Loading Libraries (Numpy,Pandas,Matplotlib,Seaborn)
- ❑ Reading input dataset by `df.info()` , `df.describe()` ,`df.columns`
- ❑ Cleaning the Data (by replacing NaN values with Zeros and renaming the column names which we get connect easily)
- ❑ Analyzing the data and plotting the different charts from the given Dataset
- ❑ Conclusion

Intro → Quick Facts about IPL

- ❑ IPL is the professional Twenty20 league in India contested during April and May of every year by teams representing Indian Cities.
- ❑ Founded by BCCI and it is now the most attended cricket league in the world and ranks 6th among all sports league.
- ❑ In 2010, the IPL became the first sporting event in the world to be broadcasted live on YouTube

Loading Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

- Numpy is python library used for working with arrays
- Pandas allows importing data from various file formats such as CSV, Excel, Json, SQL
- Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy
- Seaborn is a data visualization library built on top of matplotlib and closely integrated with pandas data structures in Python.

Reading the information of Dataset by .info and .columns

▶ matches_df.info()

▶ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 756 entries, 0 to 755
Data columns (total 18 columns):
Column Non-Null Count Dtype
--- --
0 id 756 non-null int64
1 Season 756 non-null object
2 city 749 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 752 non-null object
11 win_by_runs 756 non-null int64
12 win_by_wickets 756 non-null int64
13 player_of_match 752 non-null object
14 venue 756 non-null object
15 umpire1 754 non-null object
16 umpire2 754 non-null object
17 umpire3 119 non-null object
dtypes: int64(4), object(14)
memory usage: 106.4+ KB

▶ matches_df.columns

▶ Index(['id', 'Season', 'city', 'date', 'team1', 'team2', 'toss_winner',
'toss_decision', 'result', 'dl_applied', 'winner', 'win_by_runs',
'win_by_wickets', 'player_of_match', 'venue', 'umpire1', 'umpire2',
'umpire3'],
dtype='object')

Most important step is to cleaning the data by replacing the NaN values with zeros by .fillna(0)

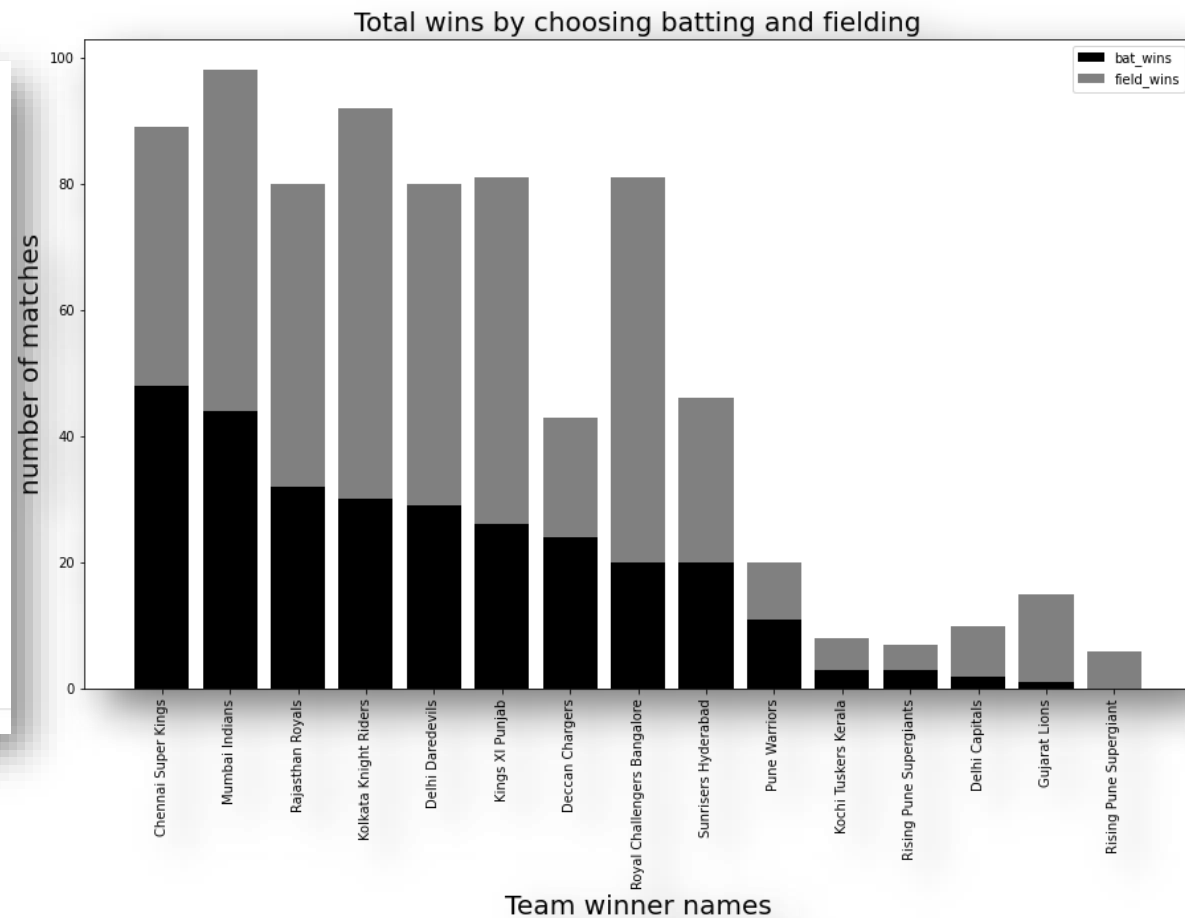
umpire1	umpire2	umpire3
AY Dandekar	NJ Llong	NaN
A Nand Kishore	S Ravi	NaN
Nitin Menon	CK Nandan	NaN
AK Chaudhary	C Shamshuddin	NaN
NaN	NaN	NaN
...

umpire1	umpire2	umpire3
AY Dandekar	NJ Llong	0
A Nand Kishore	S Ravi	0
Nitin Menon	CK Nandan	0
AK haudhary	C Shamshuddin	0
0	0	0
...

Here we did the replacing of NaN values of umpire(1,2,3) columns

1. How many times the team wins (if they choose field or bat)

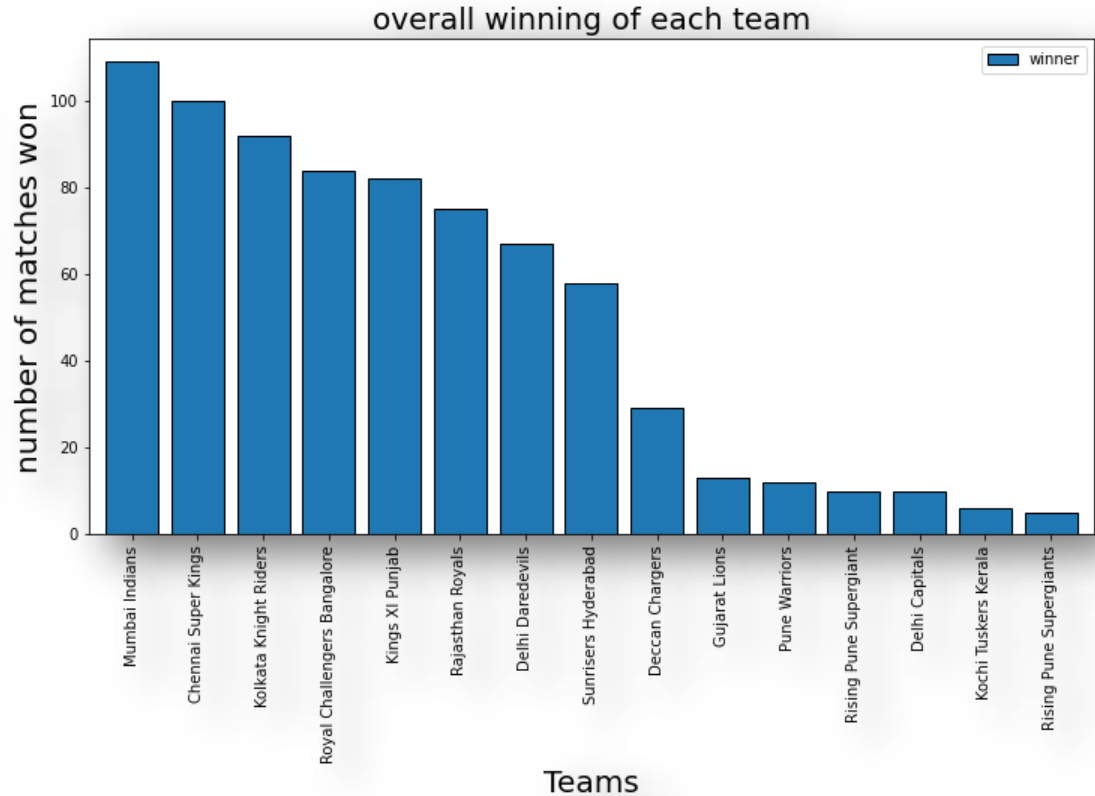
	Toss_winner	bat_wins	field_wins
8	Chennai Super Kings	48	41
5	Mumbai Indians	44	54
9	Rajasthan Royals	32	48
2	Kolkata Knight Riders	30	62
7	Delhi Daredevils	29	51
3	Kings XI Punjab	26	55
10	Deccan Chargers	24	19
0	Royal Challengers Bangalore	20	61
4	Sunrisers Hyderabad	20	26
12	Pune Warriors	11	9
11	Kochi Tuskers Kerala	3	5
13	Rising Pune Supergiants	3	4
14	Delhi Capitals	2	8
6	Gujarat Lions	1	14
1	Rising Pune Supergiant	0	6



2. what is the overall winning of each team

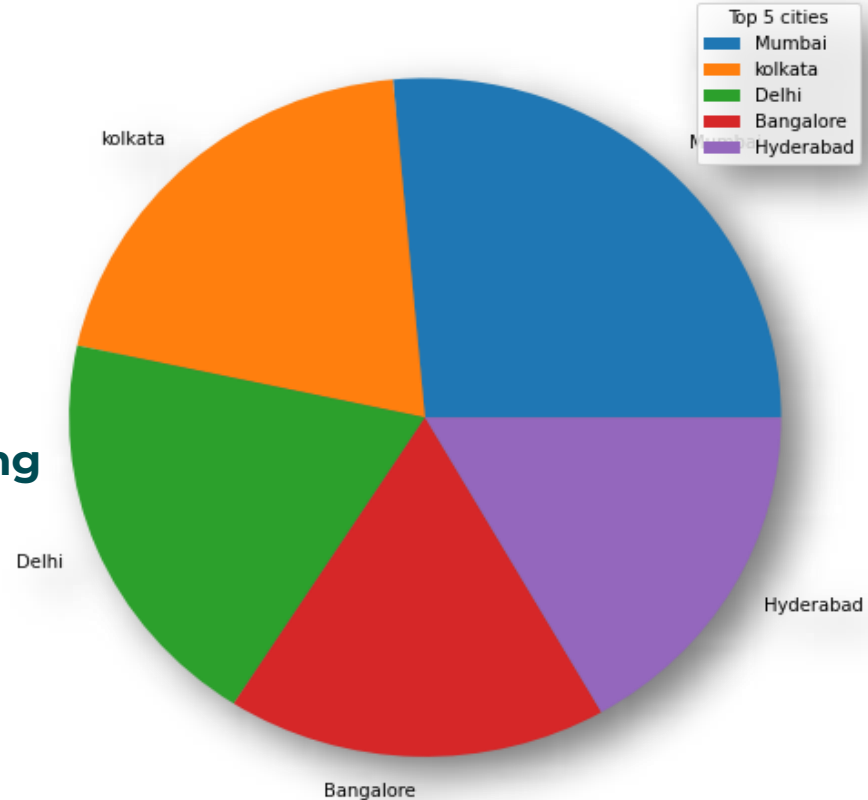
```
wins_df = new_matches_df['winner'].value_counts()  
wins_df
```

```
Mumbai Indians      109  
Chennai Super Kings  100  
Kolkata Knight Riders 92  
Royal Challengers Bangalore 84  
Kings XI Punjab      82  
Rajasthan Royals     75  
Delhi Daredevils     67  
Sunrisers Hyderabad  58  
Deccan Chargers      29  
Gujarat Lions        13  
Pune Warriors        12  
Delhi Capitals       10  
Rising Pune Supergiant 10  
Kochi Tuskers Kerala  6  
Rising Pune Supergiants 5  
0                     4  
Name: winner, dtype: int64
```



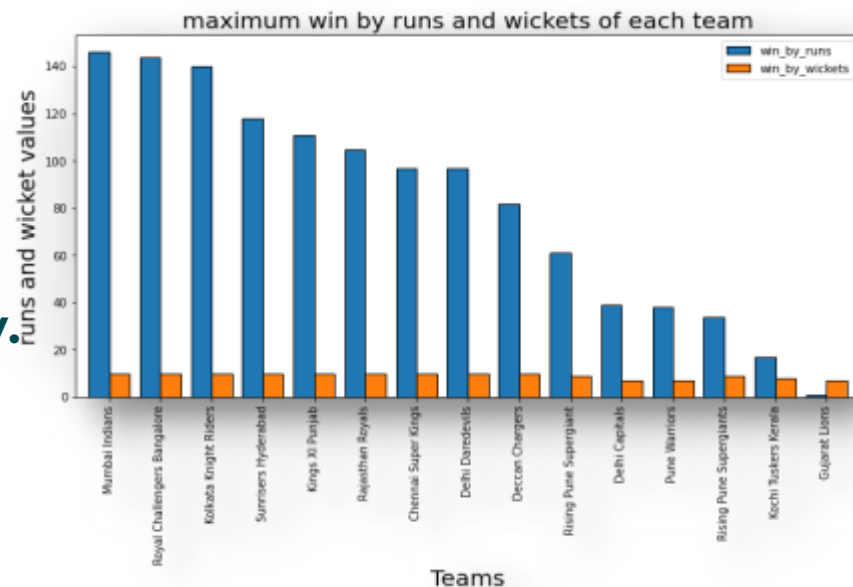
3. Find the Top 5 cities which held the total number of matches

Here we analyzed the total number of matches held in cities and we sorted in descending manner and we are interested in top 5 cities



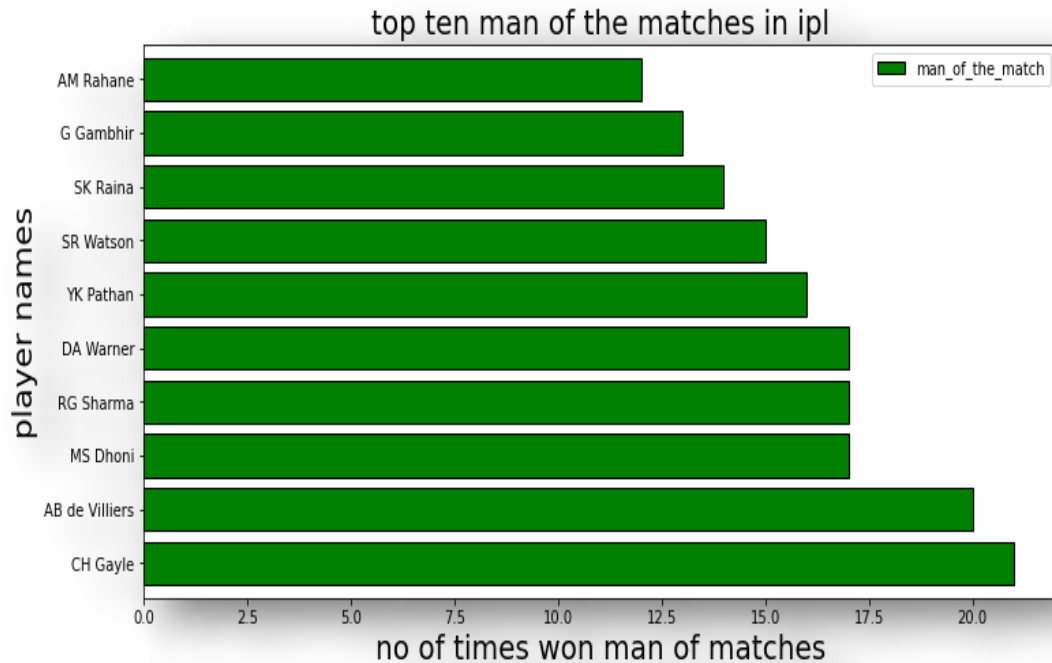
4.What is the Maximum winning of each team (by runs and by wickets)

Mumbai Indians secured the maximum win by runs ,win by wickets among all the teams (146 runs, 10 wickets) respectively. The least one is Gujarat lions (1 runs , 7 wickets)



5. The first top ten man of the match players in IPL

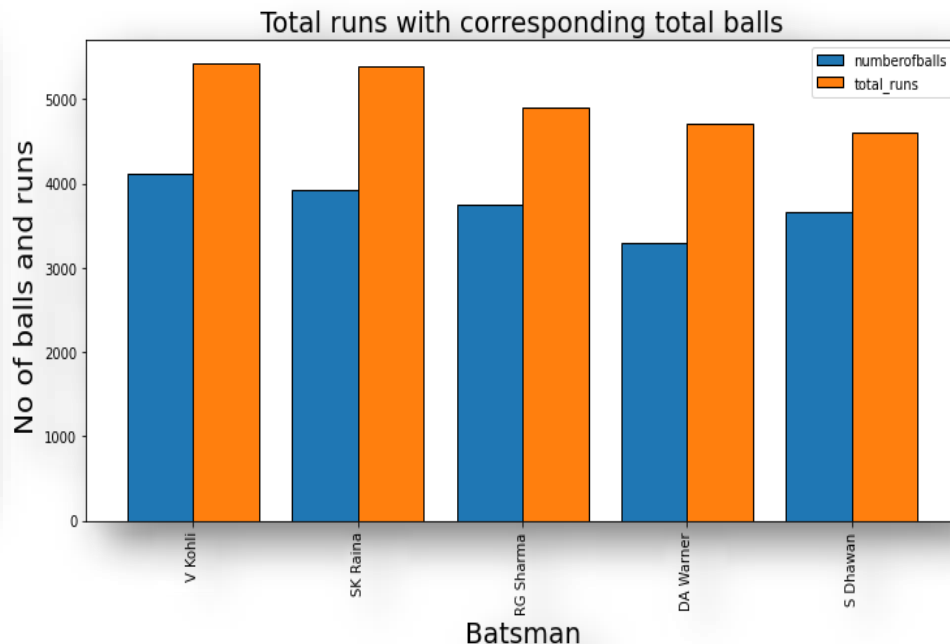
The most of the IPL viewers are always curious about “Man of the match”. Here we analyzed the top Ten Most number of Man of the matches secured



6. The top five batsman of total runs with the corresponding deliveries

	batsman	total_runs	out	numberofballs	average	strikerate
0	V Kohli	5426	152	4111	35.697368	131.987351
1	SK Raina	5386	160	3916	33.662500	137.538304
2	RG Sharma	4902	161	3742	30.447205	130.999466
3	DA Warner	4717	114	3292	41.377193	143.286756
4	S Dhawan	4601	137	3665	33.583942	125.538881

By sorting the total runs in descending order with the corresponding deliveries we got this statistics.
The most runs got from virat kohli

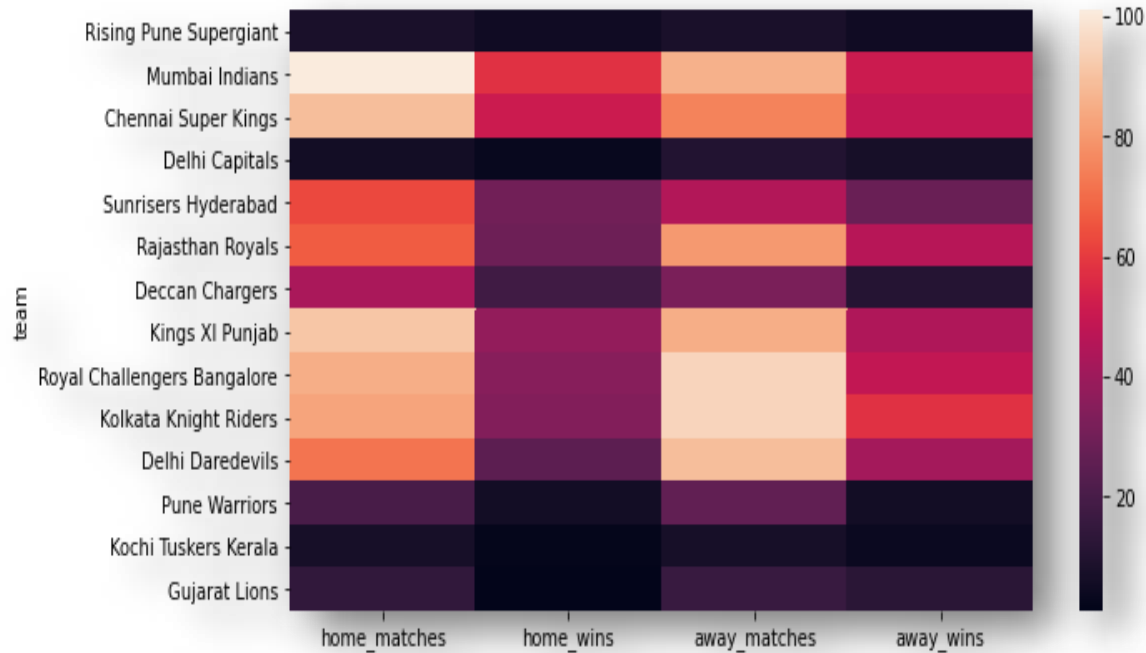


7. Dataset of team wise wins and matches

Here we tried some sought of team-wise analysis of home ground wins and away ground wins with the Corresponding matches..

team	home_wins	away_wins	home_matches	away_matches	home_win_percentage	away_win_percentage
Rising Pune Supergiant	5	5	8	8	62.500000	62.500000
Mumbai Indians	58	51	101	86	57.425743	59.302326
Chennai Super Kings	51	49	89	75	57.303371	65.333333
Delhi Capitals	3	7	6	10	50.000000	70.000000
Sunrisers Hyderabad	30	28	63	45	47.619048	62.222222
Rajasthan Royals	29	46	67	80	43.283582	57.500000
Deccan Chargers	18	11	43	32	41.860465	34.375000
Kings XI Punjab	38	44	91	85	41.758242	51.764706
Royal Challengers Bangalore	35	49	85	95	41.176471	51.578947
Kolkata Knight Riders	34	58	83	95	40.963855	61.052632
Delhi Daredevils	25	42	72	89	34.722222	47.191011
Pune Warriors	6	6	20	26	30.000000	23.076923
Kochi Tuskers Kerala	2	4	7	7	28.571429	57.142857
Gujarat Lions	1	12	14	16	7.142857	75.000000

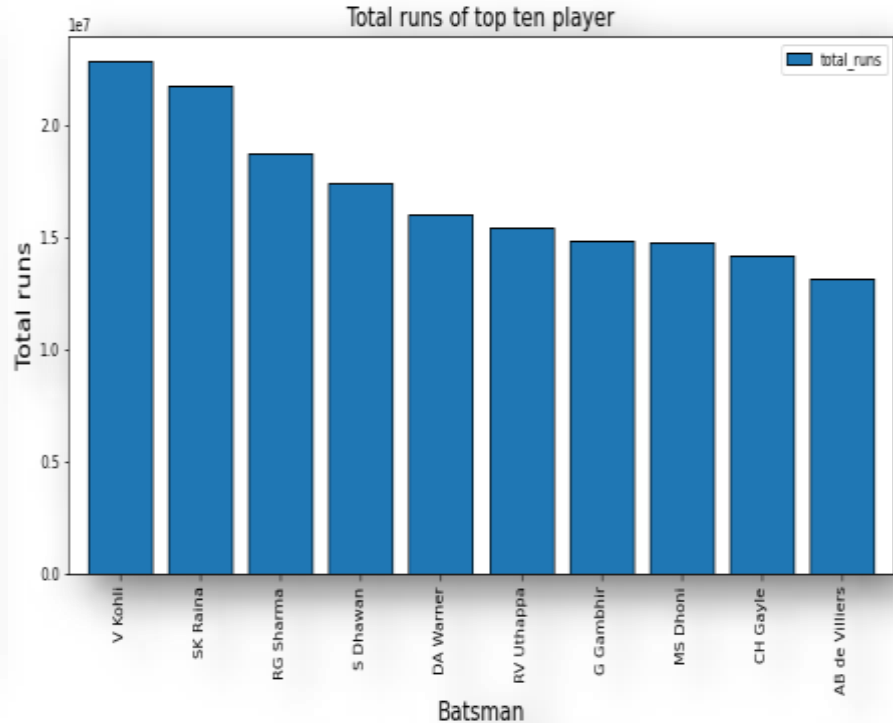
Contd..



8. The Top Ten total number of runs scored by Batsman

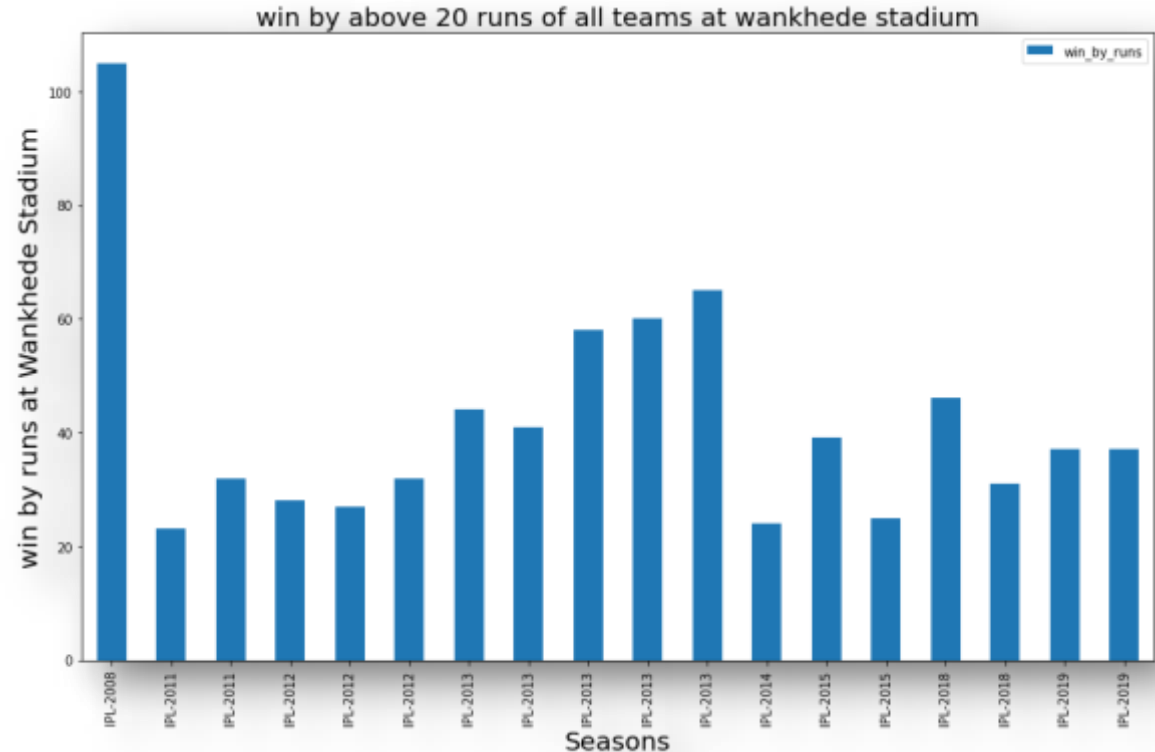
Virat kohli secured the top position in total runs In the whole ipl season

batsman	total_runs
V Kohli	22848886
SK Raina	21780984
RG Sharma	18706032
S Dhawan	17373376
DA Warner	16028366
RV Uthappa	15434640
G Gambhir	14867756
MS Dhoni	14765100
CH Gayle	14167775
AB de Villiers	13140478



9.The winning of all teams above 20 runs in wankhede stadium in all seasons

By analyzing these question we get to see is the maximum no of win by runs at wankhede stadium in Year of 2008



Conclusion:

- ✓ Chennai super kings has most number of wins if the toss decision takes bat → 48 times & Kolkata knight riders has most number of wins if the toss decision takes field → 62 times
- ✓ Mumbai Indians wins most number of matches → 109 wins
- ✓ Chris Gayle won the 21 man of the matches
- ✓ In the overall IPL virat kohli is the top batsman with 22848886 runs
- ✓ 101 home matches and 58 home wins of Mumbai Indians and 86 away matches and 51 away wins respectively

Q & A