Exploratory Data Analysis - Sports

NOTE:- This is only a reference file. you do not need to copy this file exactly you can analyze your data in your own way

In this task, we will be performing exploratory data analysis on the dataset "Sport" and try to find out the best team in IPL. Also, we will find the best bolwer and various insights.

The steps demonstrated in this notebook are:

1. Loading the data
2. Familiarizing with data
3. Visualizing the data
4. Data Analysis
5. Conclusion

1. Loading Data:

In [1]:
```
#importing required libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [3]:
```
#Loading data into dataframe
```

2. Familiarizing with Data:

Analysing Deliveries Dataset

In [9]:
```
#Shape of dataframe
```

In [8]:
```
#Listing the features of the dataset
```

In [7]:
```
#Information about the dataset
```

In [6]:
```
#checking for null value
```

In [5]:
```
# describtion of dataset
```

In [4]:
```
#unique value in dataset
```

Analysing Matches Dataset

In [10]:
```
#Shape of dataframe
```

In [11]:
```
#Listing the features of the dataset
```

In [12]:
```
#Information about the dataset
```

In [13]:
```
#checking for null value
```

In [14]:
```
# describtion of dataset
```

In [15]:
```
#unique value in dataset
```

In [17]:
```
#code
```

```
City in which most matches have been won        : Mumbai
Team that has won most matches                  : Mumbai Indians
Player who has been man of the match most times : CH Gayle
Most frequent Umpire 1                          : HDPK Dharmasena
Most frequent Umpire 2                          : C Shamshuddin
```

In [18]:
```
#fill null values
```

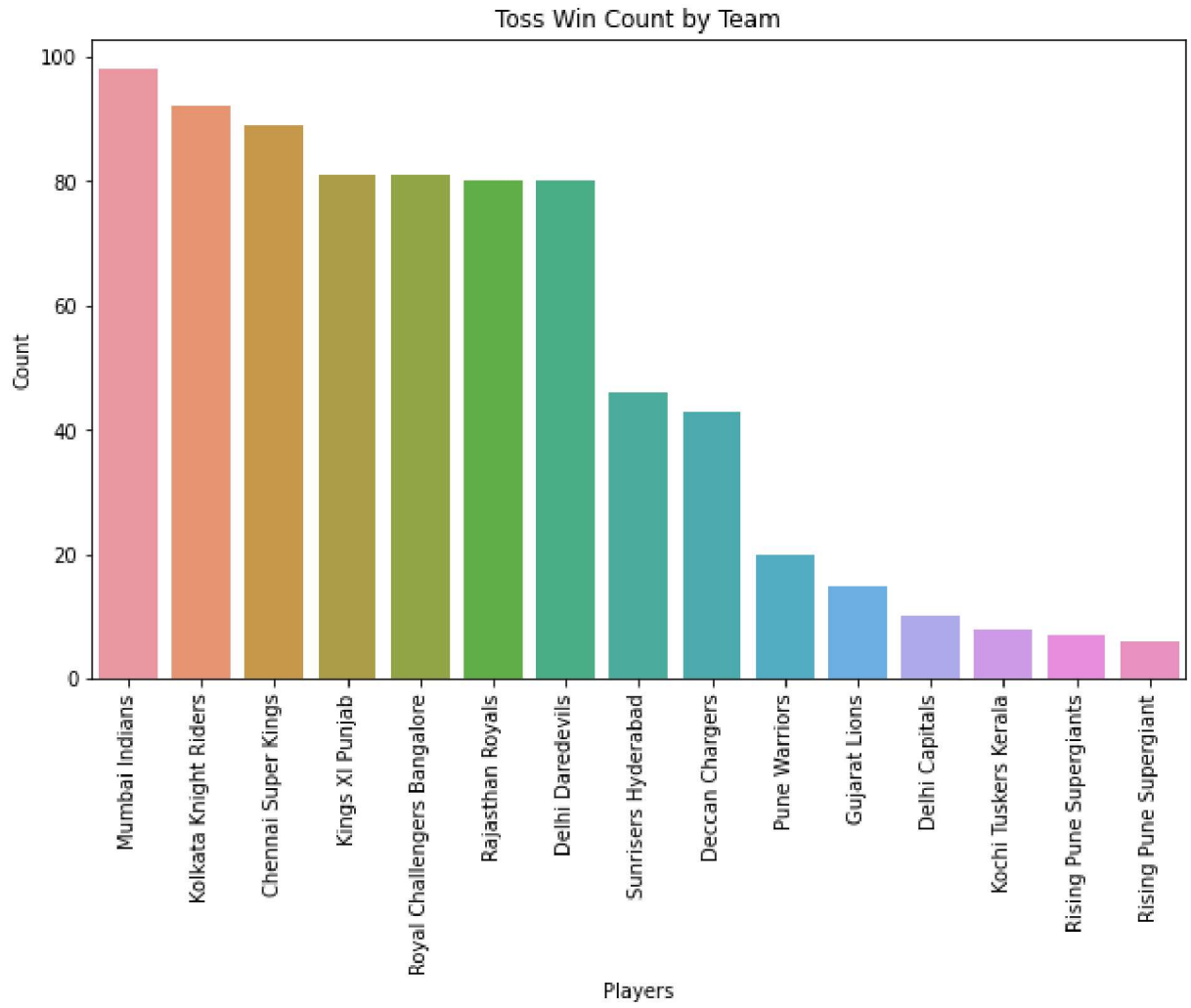
In [16]:
```
#check duplicate values
```

3. Visualizing the data:

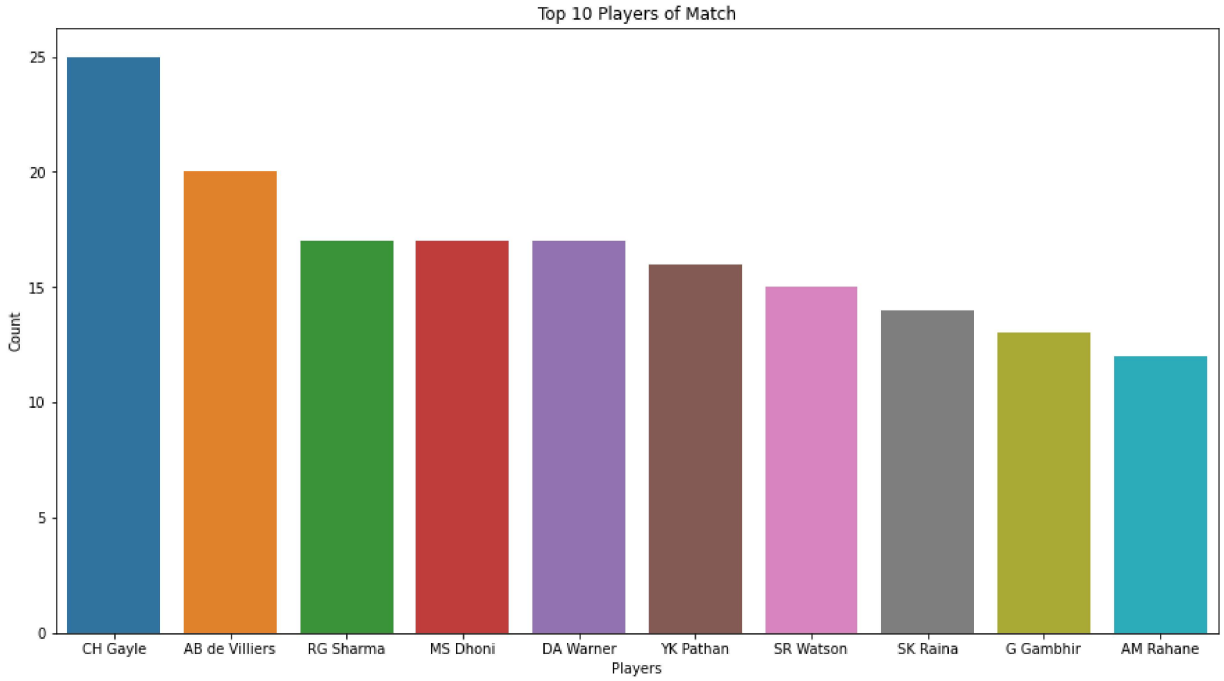
Few plots and graphs are displayed to find how the data is distributed and the how features are related to each other.

Finding Top Teams and Players

In [21]:
```
#code
```



In [22]:
```
#code
```



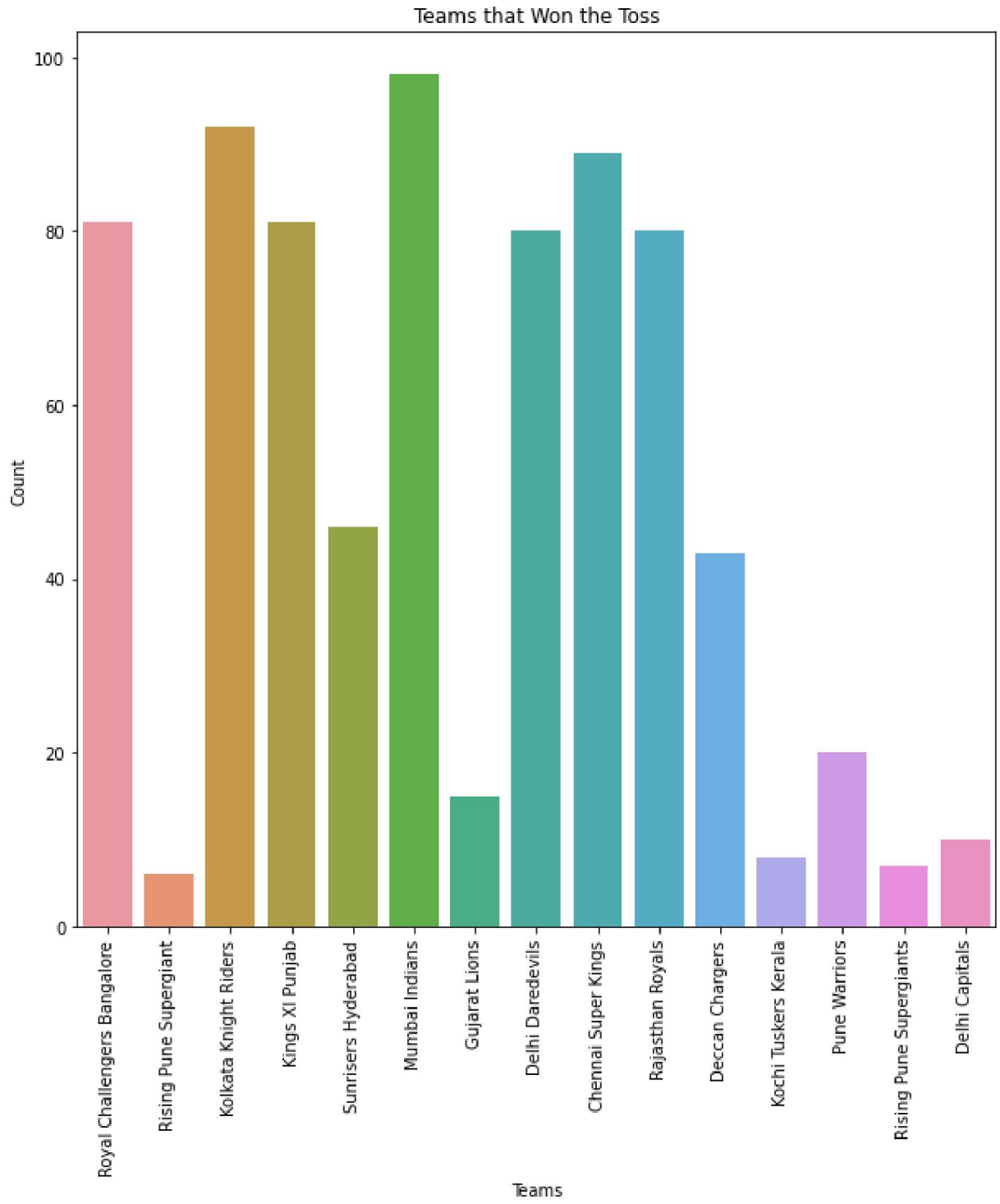
Finding the Factors Affecting the Victory

In [23]:
```
#code
```

Out[23]:

	id	season	dl_applied	Bat_1	Ball_1
id	1.000000	0.668304	0.011658	0.039403	0.012239
season	0.668304	1.000000	0.001116	0.037529	0.009379
dl_applied	0.011658	0.001116	1.000000	0.016349	0.011631
Bat_1	0.039403	0.037529	0.016349	1.000000	0.560420
Ball_1	0.012239	0.009379	0.011631	0.560420	1.000000

In [26]:
```
#code
```



In [27]:
```
#code
```

```
Team that won most matches by Batting First:  Mumbai Indians
```

4. Data Analysis:

4.1. Merging the two Datasets into a new Dataset and Reading it (join on match-id)

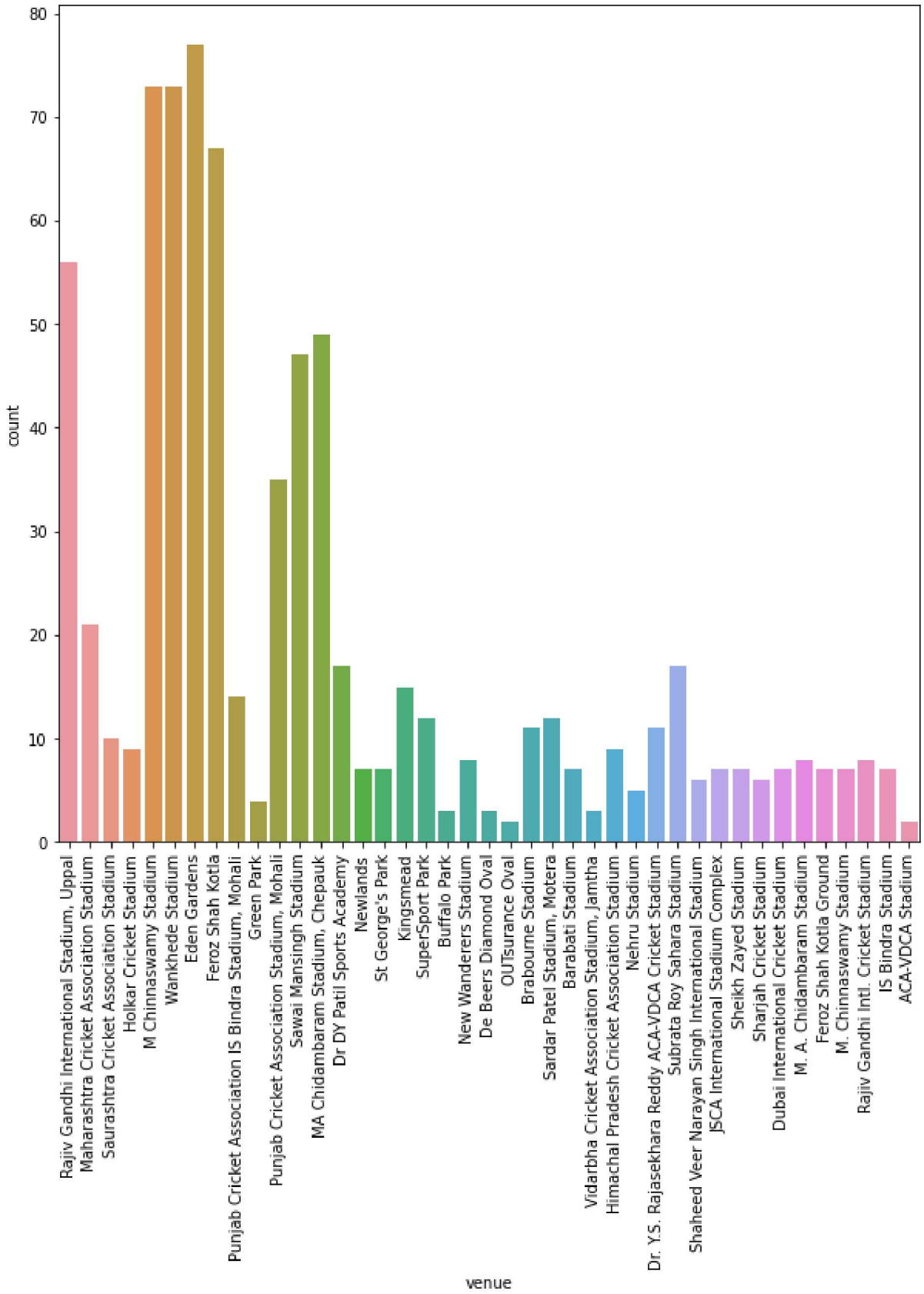
In [17]:
```
#code
```

In [18]:
```
#check shape
```

In [19]:
```
#check duplicate and if any then drop duplicate
```

Number of Matches Played in Each Stadium

In [36]:
```
#code
```


count
venue
0
10
20
30
40
50
60
70
80
Rajiv Gandhi International Stadium, Uppal
Maharashtra Cricket Association Stadium
Saurashtra Cricket Association Stadium
Holkar Cricket Stadium
M Chinnaswamy Stadium
Wankhede Stadium
Eden Gardens
Feroz Shah Kotla
Punjab Cricket Association IS Bindra Stadium, Mohali
Green Park
Punjab Cricket Association Stadium, Mohali
Sawai Mansingh Stadium
MA Chidambaram Stadium, Chepauk
Dr DY Patil Sports Academy
Newlands
St George's Park
Kingsmead
SuperSport Park
Buffalo Park
New Wanderers Stadium
De Beers Diamond Oval
OUTsurance Oval
Brabourne Stadium
Sardar Patel Stadium, Motera
Barabati Stadium
Vidarbha Cricket Association Stadium, Jamtha
Himachal Pradesh Cricket Association Stadium
Nehru Stadium
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium
Subrata Roy Sahara Stadium
Shaheed Veer Narayan Singh International Stadium
JSCA International Stadium Complex
Sheikh Zayed Stadium
Sharjah Cricket Stadium
Dubai International Cricket Stadium
M. A. Chidambaram Stadium
Feroz Shah Kotla Ground
M. Chinnaswamy Stadium
Rajiv Gandhi Intl. Cricket Stadium
IS Bindra Stadium
ACA-VDCA Stadium

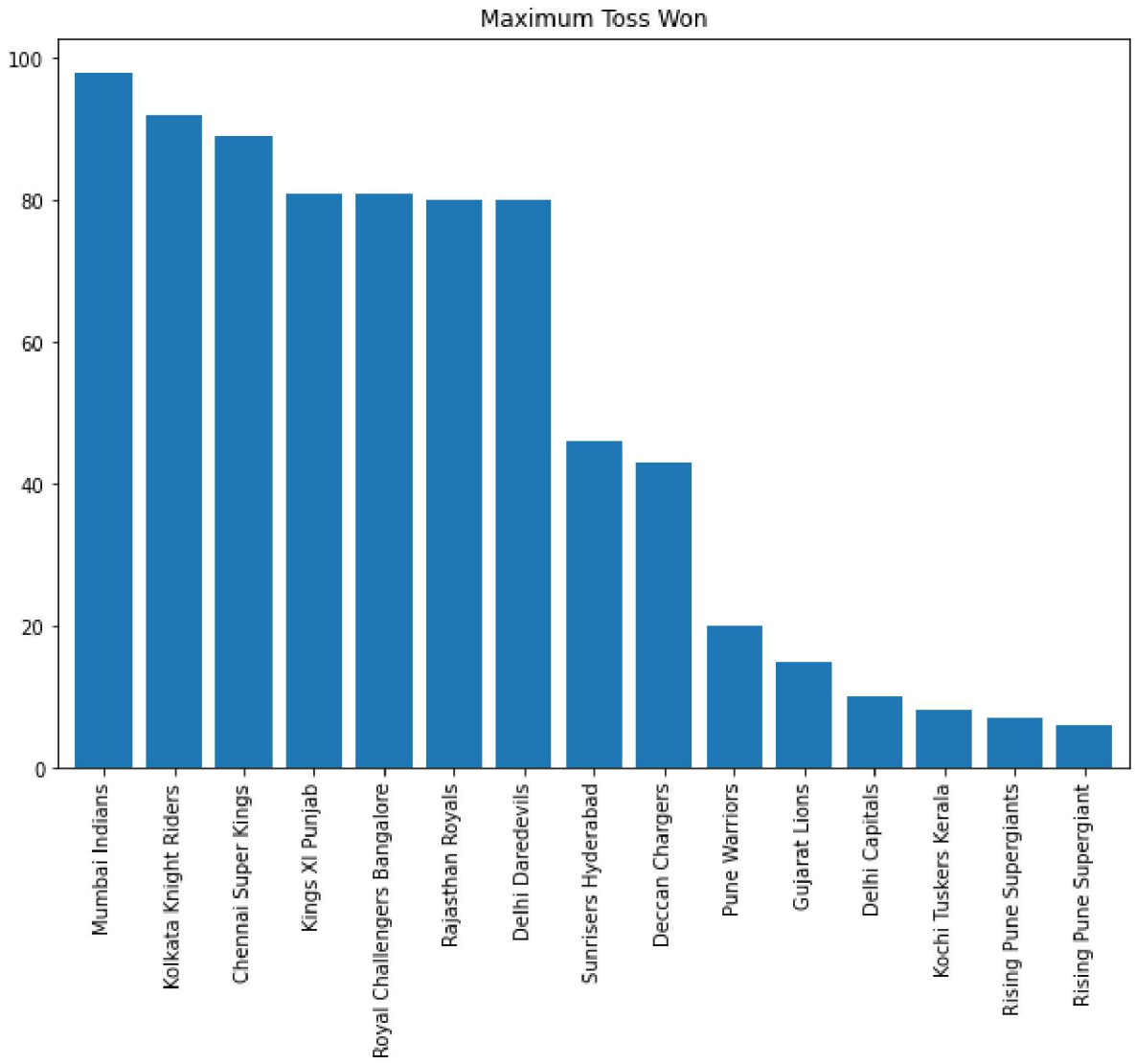
In [37]:
```
#team Stats
```

Out[37]:

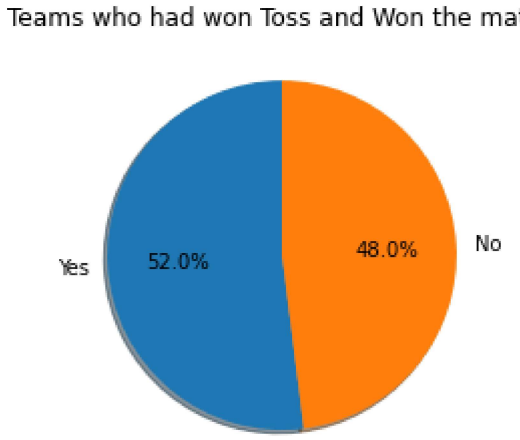
	Teams	Total Matches played	Total won	Toss won	Total lost	Winloss Ratio
0	Chennai Super Kings	164	100	89	64	0.61
1	Deccan Chargers	75	29	43	46	0.39
2	Delhi Capitals	16	10	10	6	0.62
3	Delhi Daredevils	161	67	80	94	0.42
4	Gujarat Lions	30	13	15	17	0.43
5	Kings XI Punjab	176	82	81	94	0.47
6	Kochi Tuskers Kerala	14	6	8	8	0.43
7	Kolkata Knight Riders	178	92	92	86	0.52
8	Mumbai Indians	187	113	98	74	0.60
9	Pune Warriors	46	12	20	34	0.26
10	Rajasthan Royals	147	75	80	72	0.51
11	Rising Pune Supergiant	16	10	6	6	0.62
12	Rising Pune Supergiants	14	5	7	9	0.36
13	Royal Challengers Bangalore	180	84	81	96	0.47
14	Sunrisers Hyderabad	108	58	46	50	0.54

Maximum Toss Won

In [38]:
```
#code
```

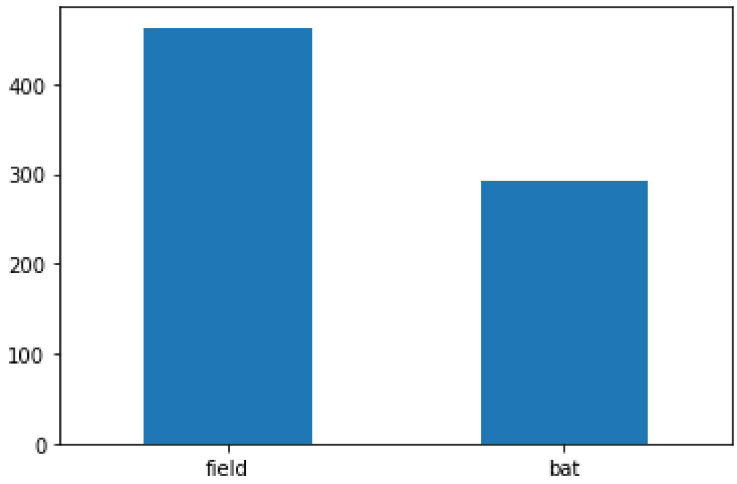

Maximum Toss Won
100
80
60
40
20
0
Mumbai Indians
Kolkata Knight Riders
Chennai Super Kings
Kings XI Punjab
Royal Challengers Bangalore
Rajasthan Royals
Delhi Daredevils
Sunrisers Hyderabad
Deccan Chargers
Pune Warriors
Gujarat Lions
Delhi Capitals
Kochi Tuskers Kerala
Rising Pune Supergiants
Rising Pune Supergiant

In [39]:
```
#code
```



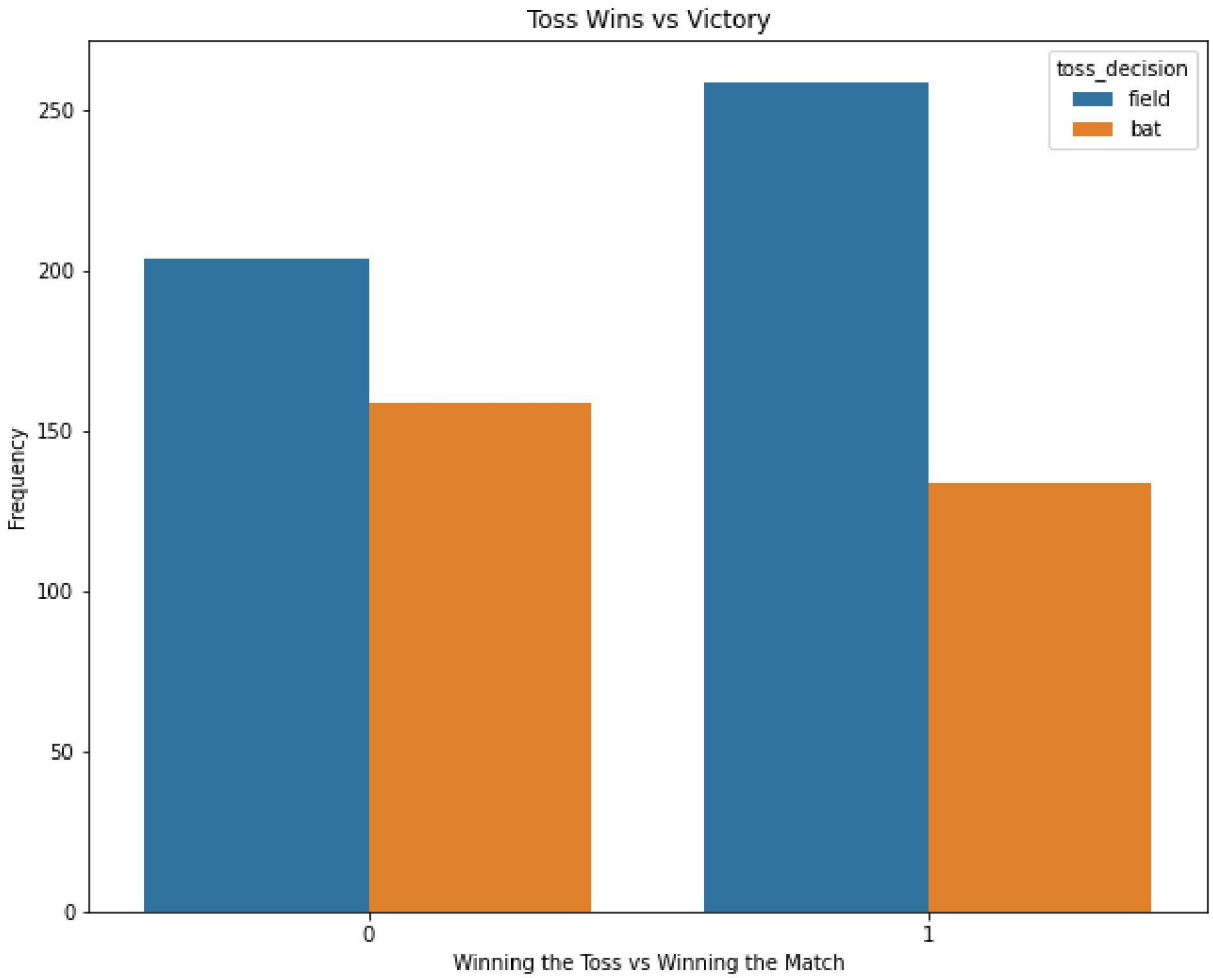
Deciding Whether to Bat or Field After Winning the Toss

In [40]:
```
#code
```



Relation between Winning toss and victory

In [41]:
```
#code
```



Batsmen overview

In [42]:
```
#code
```

Out[42]:

batsman	balls	runs	batting_strike_rate
V Kohli	4211	5434	129.04
SK Raina	4044	5415	133.90
RG Sharma	3816	4914	128.77
S Dhawan	3776	4632	122.67
G Gambhir	3524	4223	119.84
RV Uthappa	3492	4446	127.32
DA Warner	3398	4741	139.52
MS Dhoni	3318	4477	134.93
AM Rahane	3215	3850	119.75
CH Gayle	3131	4560	145.64

In [44]:
```
#code
```

Out[44]:

batsman	balls	runs	batting_strike_rate
V Kohli	4211	5434	129.04
SK Raina	4044	5415	133.90
RG Sharma	3816	4914	128.77
S Dhawan	3776	4632	122.67
G Gambhir	3524	4223	119.84
RV Uthappa	3492	4446	127.32
DA Warner	3398	4741	139.52
MS Dhoni	3318	4477	134.93
AM Rahane	3215	3850	119.75
CH Gayle	3131	4560	145.64
AB de Villiers	2977	4428	148.74
KD Karthik	2890	3688	127.61
AT Rayudu	2681	3326	124.06
SR Watson	2639	3614	136.95
PA Patel	2444	2874	117.59
MK Pandey	2425	2872	118.43
YK Pathan	2334	3241	138.86
JH Kallis	2291	2427	105.94
BB McCullum	2272	2893	127.33
Yuvraj Singh	2207	2765	125.28

In [45]:
```
#code
```

Out[45]:

batsman	balls	runs	batting_strike_rate	0s	4s	1s	6s	3s	2s	5s	7s
V Kohli	4211	5434	129.04	1493	482	1741	191	11	293	0	0
SK Raina	4044	5415	133.90	1381	495	1695	195	11	266	1	0
RG Sharma	3816	4914	128.77	1390	431	1589	194	5	205	1	1
S Dhawan	3776	4632	122.67	1455	526	1473	96	18	205	3	0
G Gambhir	3524	4223	119.84	1351	492	1358	59	15	249	0	0
RV Uthappa	3492	4446	127.32	1382	436	1295	156	13	206	4	0
DA Warner	3398	4741	139.52	1254	459	1213	181	18	271	2	0
MS Dhoni	3318	4477	134.93	1111	297	1383	207	14	304	0	2
AM Rahane	3215	3850	119.75	1198	405	1308	74	15	214	1	0
CH Gayle	3131	4560	145.64	1423	376	919	327	3	83	0	0

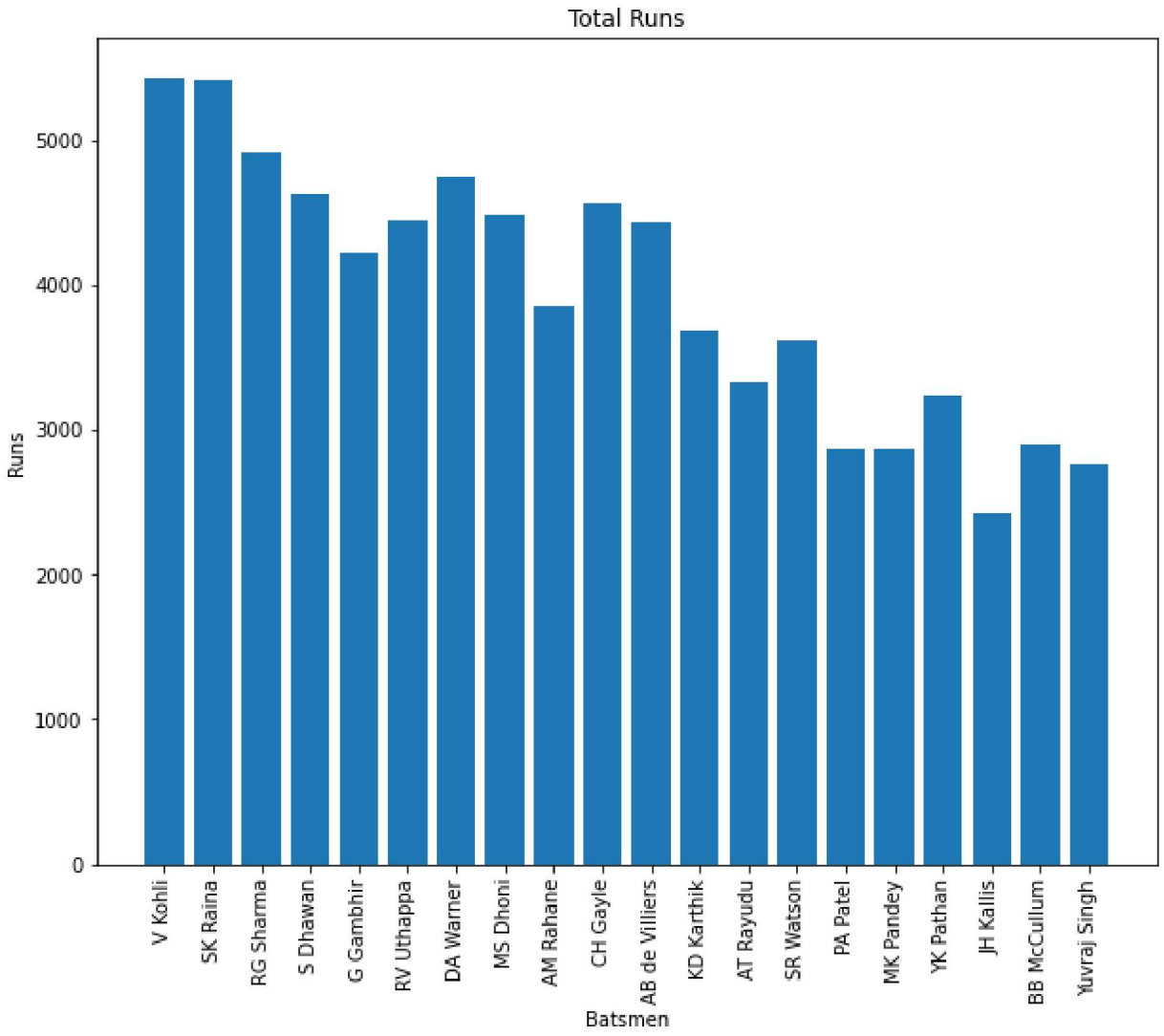
In [47]:
```
#code
```

Out[47]:

batsman	balls	runs	batting_strike_rate	0s	4s	1s	6s	3s	2s	5s	7s	matches_played
V Kohli	4211	5434	129.04	1493	482	1741	191	11	293	0	0	169
SK Raina	4044	5415	133.90	1381	495	1695	195	11	266	1	0	189
RG Sharma	3816	4914	128.77	1390	431	1589	194	5	205	1	1	182
S Dhawan	3776	4632	122.67	1455	526	1473	96	18	205	3	0	158
G Gambhir	3524	4223	119.84	1351	492	1358	59	15	249	0	0	151

Total runs by each batsmen

In [48]:
```
#code
```

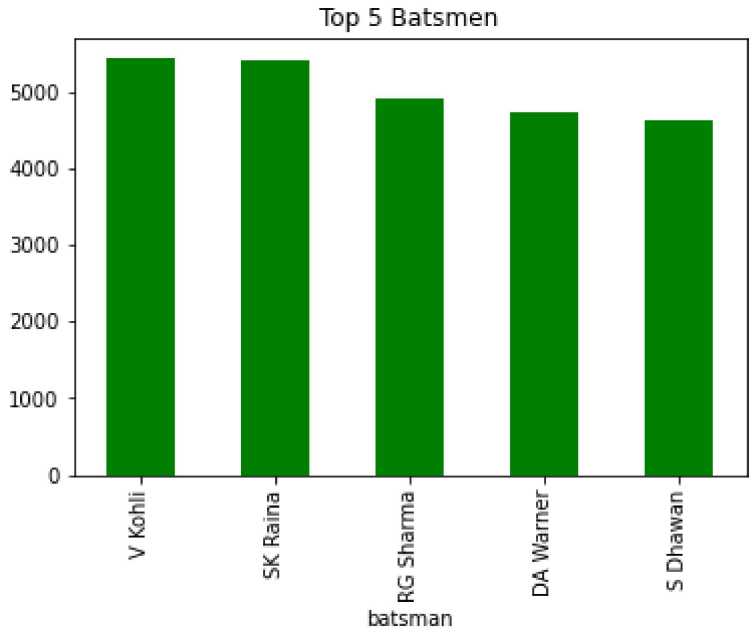


Each batsmen strike rate

In [20]:
```
#code
```

Top 5 Batsmen

In [50]:
```
#code
```



Bowler information

In [51]:
```
#code
```

Out[51]:

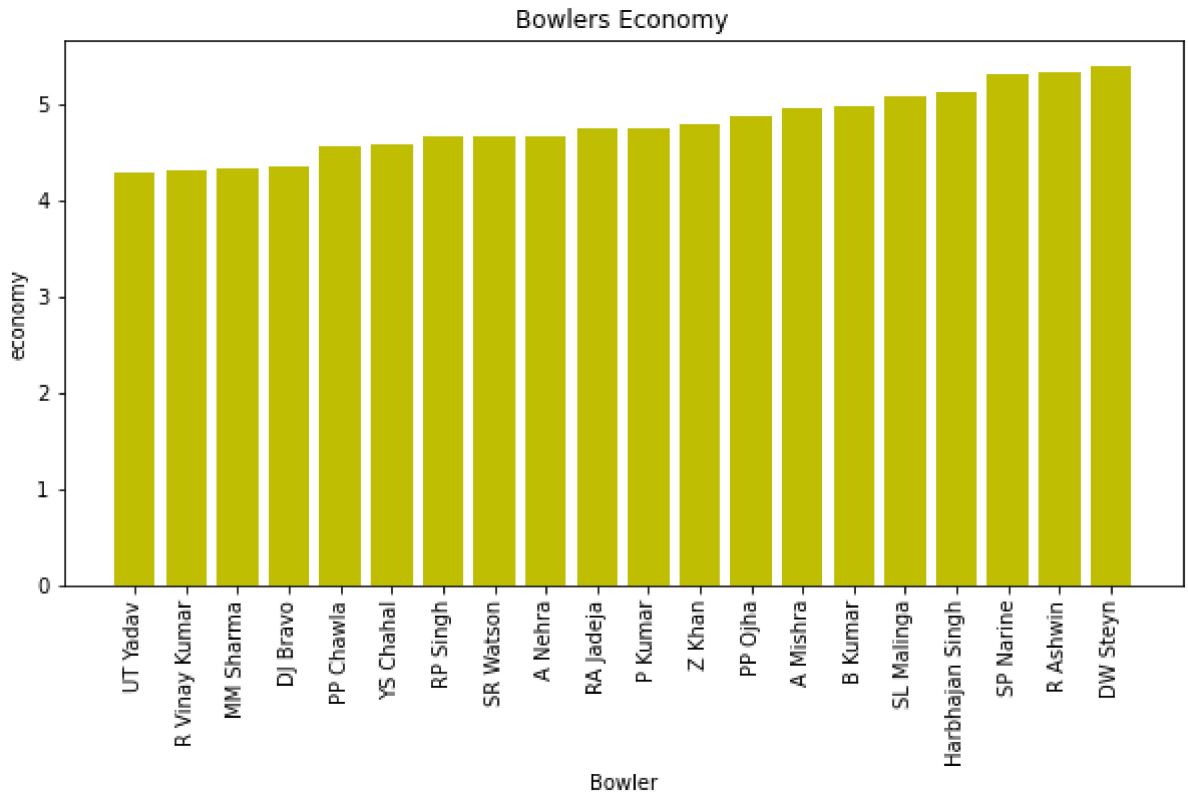
bowler	runs	balls	wickets
SL Malinga	2974	3511	188
DJ Bravo	2711	3733	168
A Mishra	3172	3850	165
Harbhajan Singh	3451	4050	161
PP Chawla	3157	4153	156
B Kumar	2707	3264	141
R Ashwin	3016	3391	138
SP Narine	2600	2939	137
UT Yadav	2605	3640	136
R Vinay Kumar	2186	3043	127
A Nehra	1974	2537	121
Z Khan	2276	2860	119
RA Jadeja	2541	3221	116
SR Watson	2137	2751	107
DW Steyn	2207	2454	104
YS Chahal	1841	2416	102
P Kumar	2637	3342	102
RP Singh	1874	2417	100
PP Ojha	1945	2399	99
MM Sharma	1770	2457	99

In [52]:
```
#code
```

Out[52]:

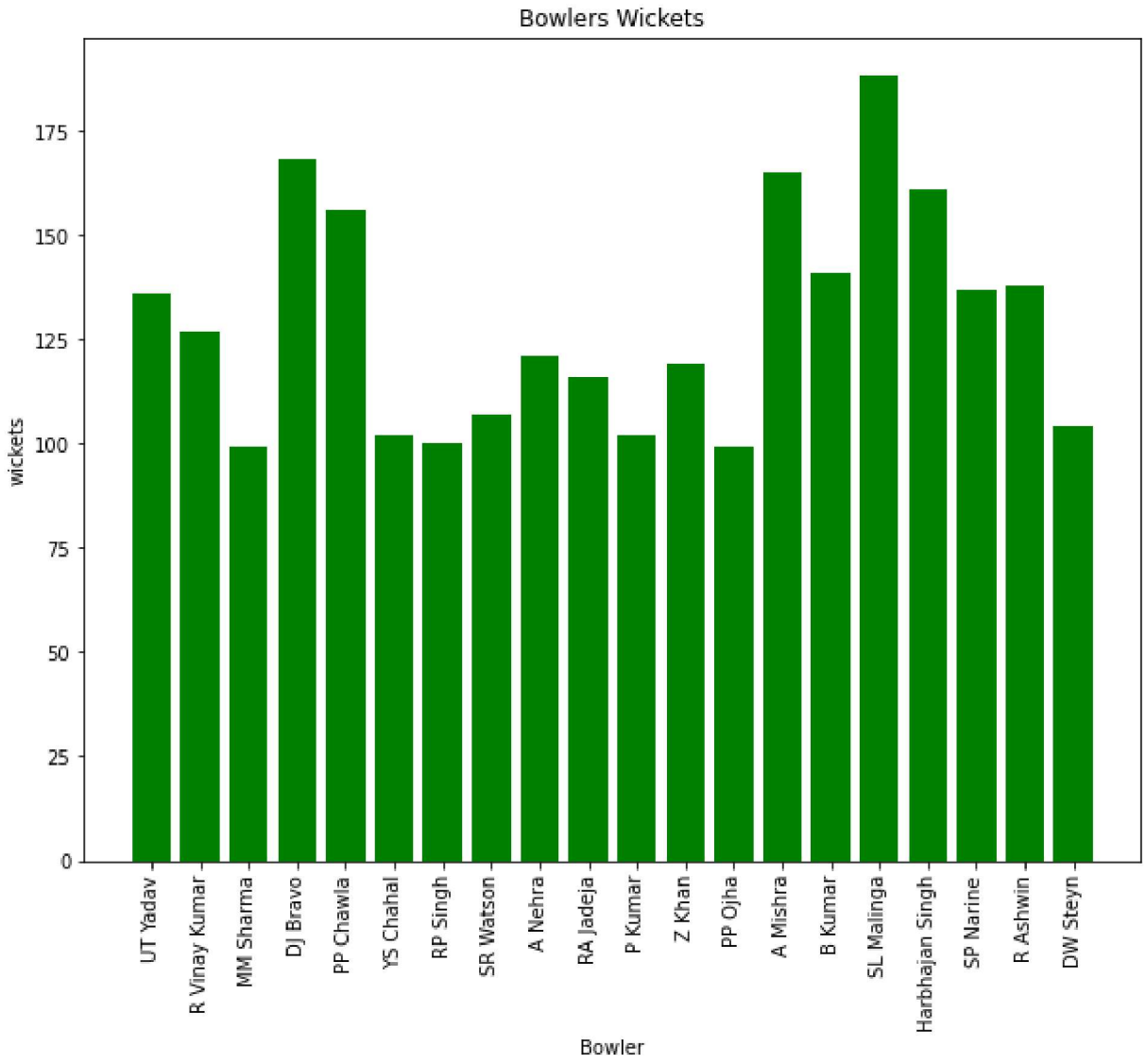
bowler	runs	balls	wickets	economy
UT Yadav	2605	3640	136	4.293956
R Vinay Kumar	2186	3043	127	4.310220
MM Sharma	1770	2457	99	4.322344
DJ Bravo	2711	3733	168	4.357353
PP Chawla	3157	4153	156	4.561040
YS Chahal	1841	2416	102	4.572020
RP Singh	1874	2417	100	4.652048
SR Watson	2137	2751	107	4.660851
A Nehra	1974	2537	121	4.668506
RA Jadeja	2541	3221	116	4.733313
P Kumar	2637	3342	102	4.734291
Z Khan	2276	2860	119	4.774825
PP Ojha	1945	2399	99	4.864527
A Mishra	3172	3850	165	4.943377
B Kumar	2707	3264	141	4.976103
SL Malinga	2974	3511	188	5.082313
Harbhajan Singh	3451	4050	161	5.112593
SP Narine	2600	2939	137	5.307928
R Ashwin	3016	3391	138	5.336479
DW Steyn	2207	2454	104	5.396088

In [53]:
```
#code
```



Wickets taken by a bowler

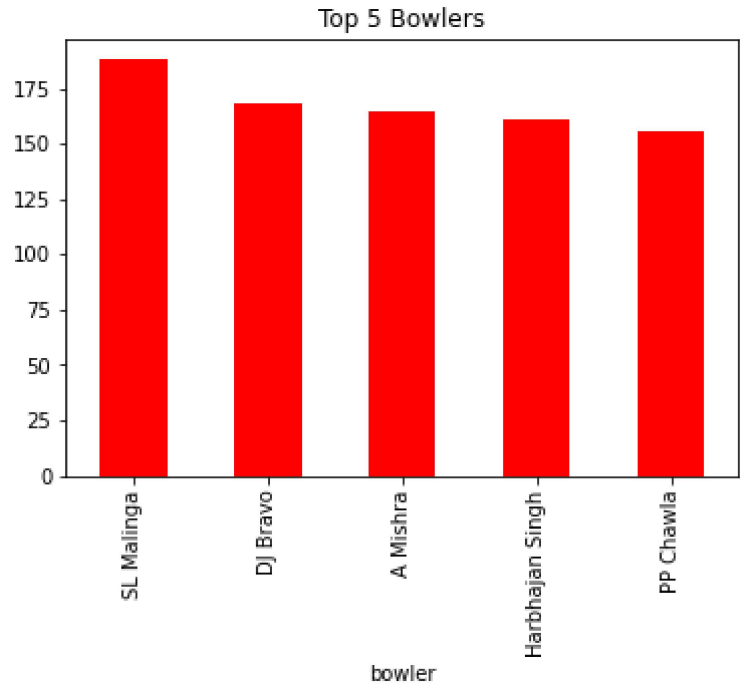
In [54]:
```
#code
```



Top 5 Bowlers

In [55]:
```
#code
```

Out[55]: Text(0.5, 1.0, 'Top 5 Bowlers')



5. Conclusion:

write conclusion here

In [ ]:
```

```