
Scatter Chart
Column Chart
Pie Chart
Bubble Chart
Donut Chart
Bar Chart
Combo Chart
Org Chart
Treemap
Table
Name
Salary
Full Time
1 Marie $24,700
2 Albert $25,200
3 Enrico $25,700
4 Lise $26,600
Stepped Area Chart
Timeline
Gauge
Memory
92
CPU
74
Candlestick Chart
PROJECT ON IPL

AGENDA

- INTRODUCTION
- SYSTEM DESIGN
- IMPLEMENTATION
- CONCLUSION

INTRODUCTION

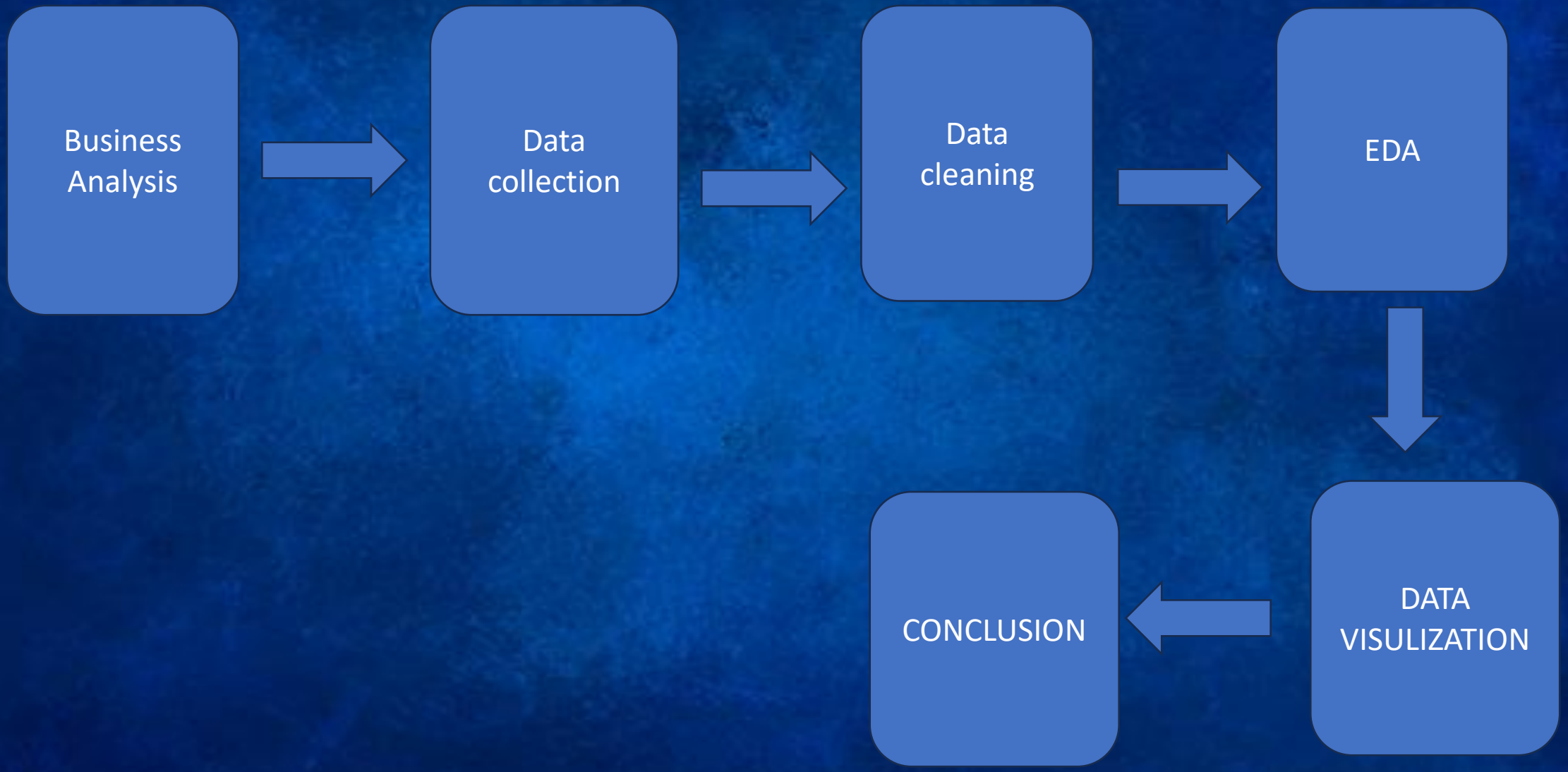
- IPL stands for Indian Premier League.
- It is a Twenty20 cricket tournament introduced by Board of Control of Cricket in India (BCCI), in 2008.
- It is the most watched Twenty20 tournament and the second-best-paying sporting league globally. IPL was established in 2008 and currently consists of ten teams in ten cities across India.

DATA SET:

Here my data set consist of 816 colunmns and 17 rows:

- ID
- City
- Date
- Player_of_match
- Venue
- Team1
- Team2
- Toss_winner
- Toss_decision
- Winner
- Result
- Result_margin
- Eliminatore
- Umpire1 and 2

SYSTEM DESIGN



IMPLEMENTATION

Import required library and required data set from external source:

```
In [125]:
1 import numpy as np
2 import pandas as pd
3 import seaborn as sns
4 import matplotlib.pyplot as plt
5 import warnings
6 import matplotlib.pyplot as plt
7 import geopandas as gpd
```

```
1 IPL=pd.read_csv("E:\IPL.csv")
```

```
1 IPL
```

	id	city	date	player_of_match	venue	neutral_venue	team1	team2	toss_winner	toss_decision	winner	result	result_
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	Royal Challengers Bangalore	Kolkata Knight Riders	Royal Challengers Bangalore	field	Kolkata Knight Riders	runs	
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	Kings XI Punjab	Chennai Super Kings	Chennai Super Kings	bat	Chennai Super Kings	runs	
2	335984	Delhi	2008-04-19	MF Maharoof	Feroz Shah Kotla	0	Delhi Daredevils	Rajasthan Royals	Rajasthan Royals	bat	Delhi Daredevils	wickets	
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0	Mumbai Indians	Royal Challengers Bangalore	Mumbai Indians	bat	Royal Challengers Bangalore	wickets	
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0	Kolkata Knight Riders	Deccan Chargers	Deccan Chargers	bat	Kolkata Knight Riders	wickets	
...	...	...	...	...	...	...	...	...	...	...	...	...	...

We use head() to fetch the first 5 rows of data set:

```
In [158]:  1 IPL.head()
```

Out[158]:

	id	city	date	player_of_match	venue	neutral_venue	team1	team2	toss_winner	toss_decision	winner	result	result_ma
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	Royal Challengers Bangalore	Kolkata Knight Riders	Royal Challengers Bangalore	field	Kolkata Knight Riders	runs	1
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	Kings XI Punjab	Chennai Super Kings	Chennai Super Kings	bat	Chennai Super Kings	runs	
2	335984	Delhi	2008-04-19	MF Maharoof	Feroz Shah Kotla	0	Delhi Daredevils	Rajasthan Royals	Rajasthan Royals	bat	Delhi Daredevils	wickets	
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0	Mumbai Indians	Royal Challengers Bangalore	Mumbai Indians	bat	Royal Challengers Bangalore	wickets	
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0	Kolkata Knight Riders	Deccan Chargers	Deccan Chargers	bat	Kolkata Knight Riders	wickets	

Shape() Is used to know how many rows and columns are present in data set:

```
In [18]:  1 print (IPL.shape)

(816, 17)
```

Info() function is used to get the information about the data set:

```
In [19]: 1 IPL.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 816 entries, 0 to 815
Data columns (total 17 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   id              816 non-null    int64  
 1   city            803 non-null    object 
 2   date            816 non-null    object 
 3   player_of_match 812 non-null    object 
 4   venue           816 non-null    object 
 5   neutral_venue   816 non-null    int64  
 6   team1           816 non-null    object 
 7   team2           816 non-null    object 
 8   toss_winner     816 non-null    object 
 9   toss_decision   816 non-null    object 
 10  winner          812 non-null    object 
 11  result          812 non-null    object 
 12  result_margin   799 non-null    float64
 13  eliminator      812 non-null    object 
 14  method          19 non-null     object 
 15  umpire1         816 non-null    object 
 16  umpire2         816 non-null    object 
dtypes: float64(1), int64(2), object(14)
memory usage: 108.5+ KB
```

To check null values we use null() or isnull() function:

```
In [21]: 1 IPL.isnull().sum()
```

```
Out[21]: id                    0
         city                 13
         date                  0
         player_of_match       4
         venue                 0
         neutral_venue         0
         team1                 0
         team2                 0
         toss_winner           0
         toss_decision         0
         winner                4
         result                4
         result_margin        17
         eliminator            4
         method              797
         umpire1               0
         umpire2               0
         dtype: int64
```

Here we observed that null values are present in some columns.

From the above in formation we can tell that null values are present in data set,columns like city , result _margin eliminator, method and winner

CLEANING THE DATA

```
In [61]: 1 IPL['city']=IPL['city'].fillna("unknown")

In [62]: 1 IPL['city'].unique()

Out[62]: array(['Bangalore', 'Chandigarh', 'Delhi', 'Mumbai', 'Kolkata', 'Jaipur',
                'Hyderabad', 'Chennai', 'Cape Town', 'Port Elizabeth', 'Durban',
                'Centurion', 'East London', 'Johannesburg', 'Kimberley',
                'Bloemfontein', 'Ahmedabad', 'Cuttack', 'Nagpur', 'Dharamsala',
                'Kochi', 'Indore', 'Visakhapatnam', 'Pune', 'Raipur', 'Ranchi',
                'Abu Dhabi', 'unknown', 'Rajkot', 'Kanpur', 'Bengaluru', 'Dubai',
                'Sharjah'], dtype=object)

In [63]: 1 IPL['result_margin']=IPL['result_margin'].fillna("unknown")

In [64]: 1 IPL['eliminator'].unique()

Out[64]: array(['N', 'Y', nan], dtype=object)

In [65]: 1 IPL['eliminator']=IPL['eliminator'].fillna("no result")

In [159]: 1 IPL.drop(['method'],axis=1,inplace=True)
          2 IPL
```

Out[159]:

	id	city	date	player_of_match	venue	neutral_venue	team1	team2	toss_winner	toss_decision	winner	result	result_
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	Royal Challengers Bangalore	Kolkata Knight Riders	Royal Challengers Bangalore	field	Kolkata Knight Riders	runs	
					Punjab								

Here we are cleaning the data,replacing the null values by respected values or any other values that are required for data set.
Here the colums which are having null values like city,result_margin,winner are replaced with respected value.
And I have dropped one column since it is not required.

Similarly we can see in city the bangalore name has give 2 time in different name hence using replace function we can resolve them

```
In [76]: 1 IPL['city'].value_counts()
```

Out[76]:

Mumbai	101
Kolkata	77
Delhi	74
Bangalore	65
Hyderabad	64
Chennai	57
Chandigarh	56
Jaipur	47
Pune	38
Abu Dhabi	29
Dubai	26
Bengaluru	15
Durban	15
unknown	13
Visakhapatnam	13
Ahmedabad	12
Sharjah	12
Centurion	12
Rajkot	10
Dharamsala	9
Indore	9
Johannesburg	8
Cape Town	7
Port Elizabeth	7
Cuttack	7
Ranchi	7
Raipur	6
Kochi	5
Kanpur	4
East London	3
Kimberley	3
Nagpur	3
Bloemfontein	2

```
In [77]: 1 IPL.replace(to_replace='Bengaluru',value='Bangalore',inplace=True)
         2 IPL['city'].value_counts()
```

Out[77]:

Mumbai	101
Bangalore	80
Kolkata	77
Delhi	74
Hyderabad	64
Chennai	57
Chandigarh	56
Jaipur	47
Pune	38
Abu Dhabi	29
Dubai	26
Durban	15
unknown	13
Visakhapatnam	13
Ahmedabad	12
Sharjah	12
Centurion	12
Rajkot	10
Dharamsala	9
Indore	9
Johannesburg	8
Cuttack	7
Cape Town	7
Ranchi	7
Port Elizabeth	7
Raipur	6
Kochi	5
Kanpur	4
East London	3
Kimberley	3
Nagpur	3
Bloemfontein	2

Similary in teams also we have same data In different names ,using replace function we are resolving them:

```
1 IPL['team1'].value_counts()
```

```
Royal Challengers Bangalore    108
Mumbai Indians                  97
Kolkata Knight Riders           95
Chennai Super Kings             94
Kings XI Punjab                 92
Delhi Daredevils                83
Rajasthan Royals                70
Sunrisers Hyderabad             59
Deccan Chargers                 39
Pune Warriors                   23
Delhi Capitals                  19
Gujarat Lions                   16
Kochi Tuskers Kerala             7
Rising Pune Supergiants          7
Rising Pune Supergiant           7
Name: team1, dtype: int64
```

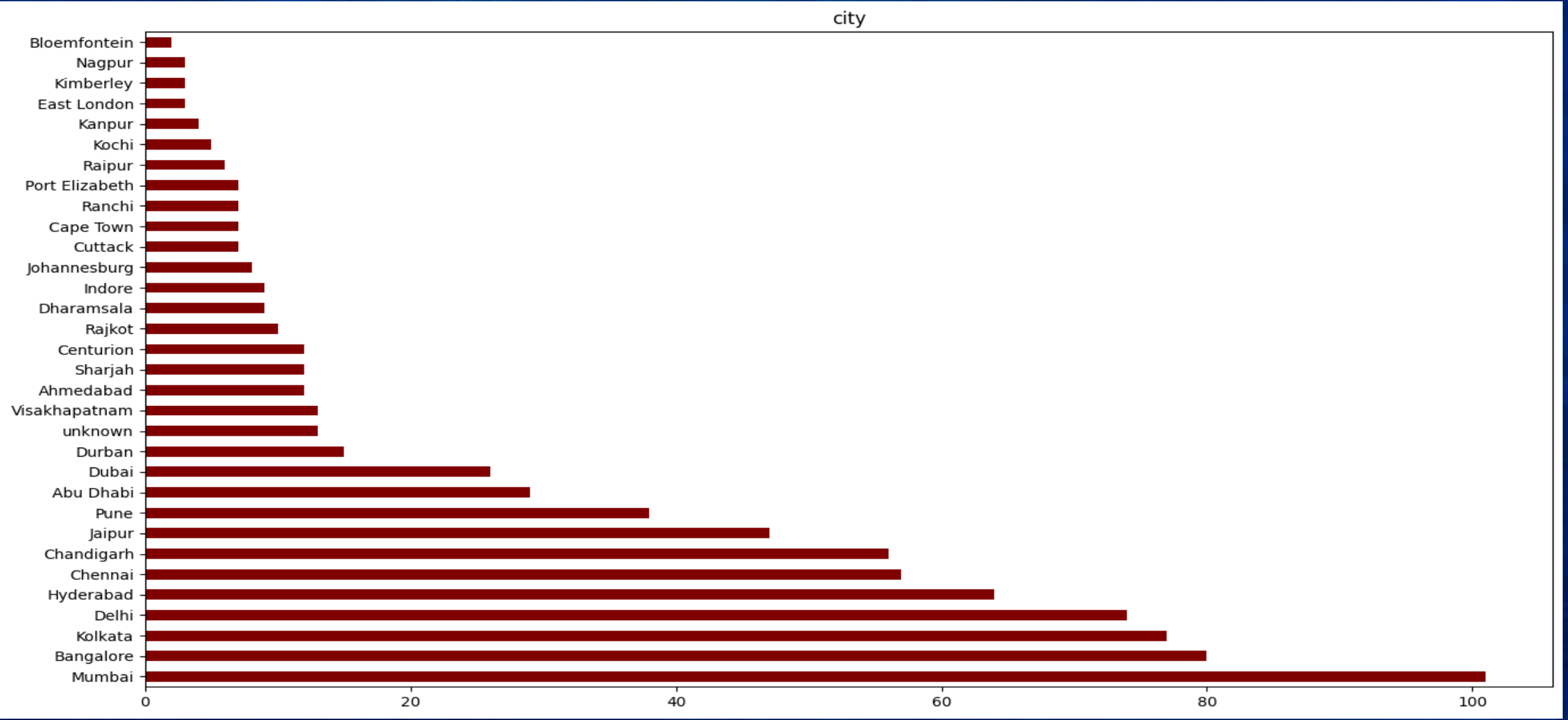
```
9]: 1 IPL.replace(to_replace='Delhi Daredevils',value='Delhi Capitals',inplace=True)
    2 IPL['team1'].value_counts()
```

```
9]: Royal Challengers Bangalore    108
    Delhi Capitals                 102
    Mumbai Indians                  97
    Kolkata Knight Riders           95
    Chennai Super Kings             94
    Kings XI Punjab                 92
    Rajasthan Royals                70
    Sunrisers Hyderabad             59
    Deccan Chargers                 39
    Pune Warriors                   23
    Gujarat Lions                   16
    Kochi Tuskers Kerala             7
    Rising Pune Supergiants          7
    Rising Pune Supergiant           7
    Name: team1, dtype: int64
```

Here I have replaced delhi dare devils to delhi capitals similarly even Rising pune super gaints also have replaced.

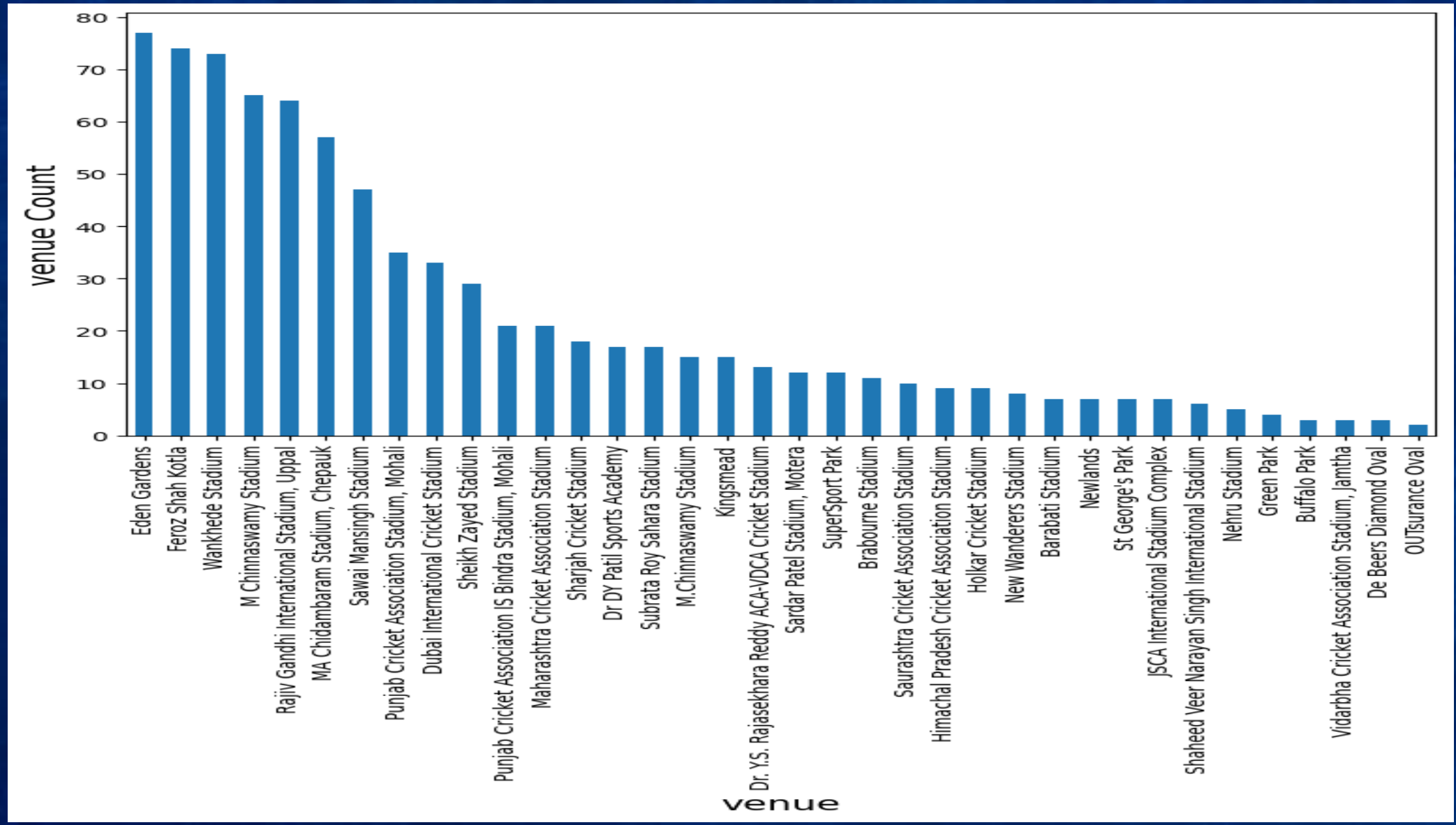
EDA AND VISUALIZATION

1.City where most of the matches played

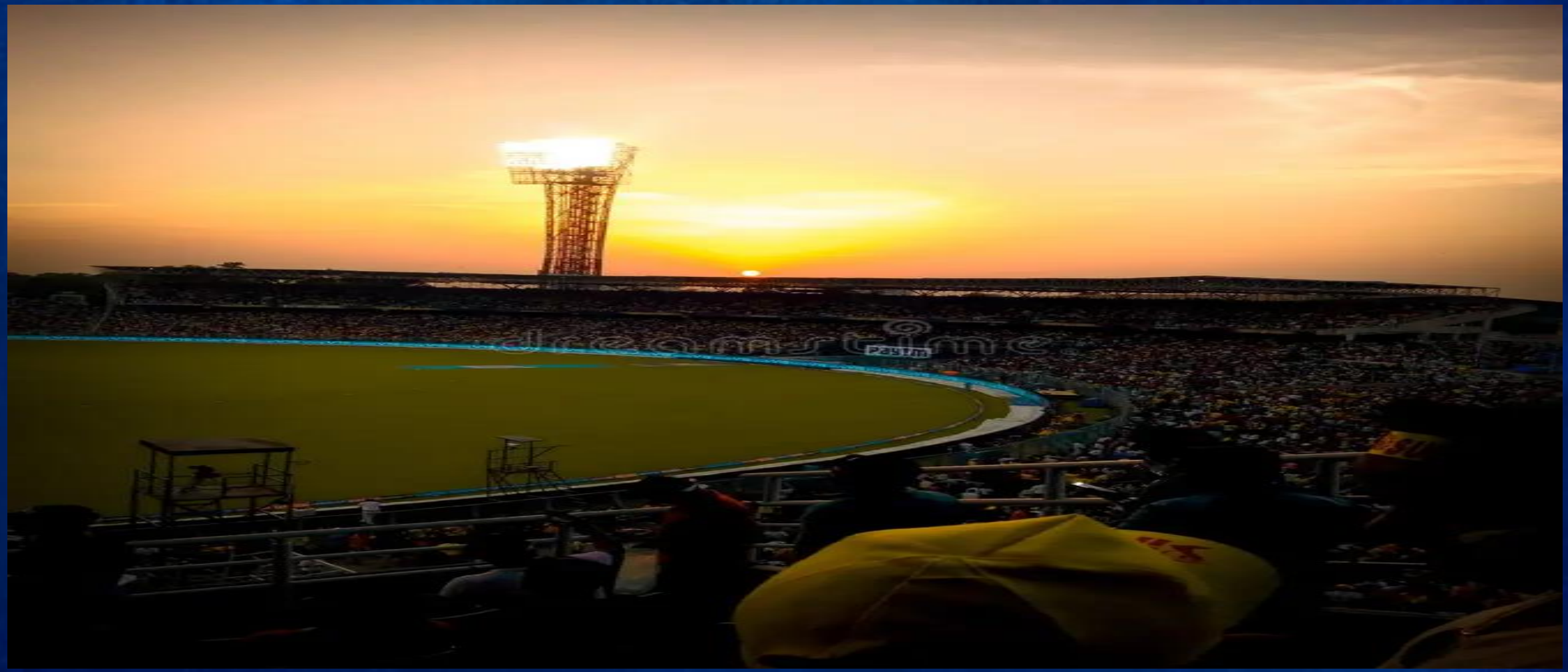


- The above plot shows the city where highest matches played .
- Mumbai is the city where highest matches played and followed by Banglore is second highest
- And least is played city is Bloenmfontien.

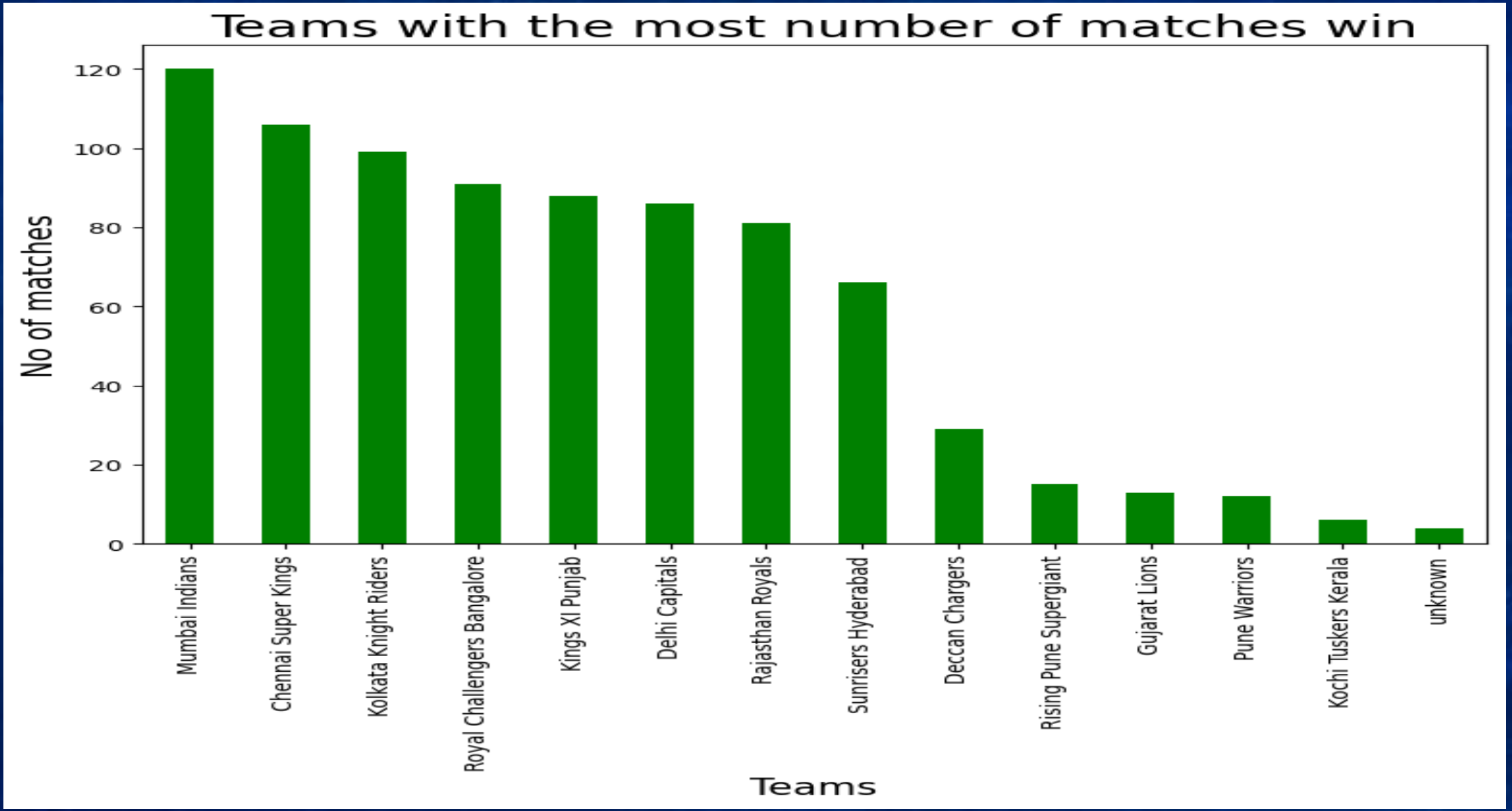
2.Venue where number of matches played.



- The above plot shows the venue where matches played.
- Here we can say that highest matches played venue is Eden garden .
- And least played venue is Outsurance oval.

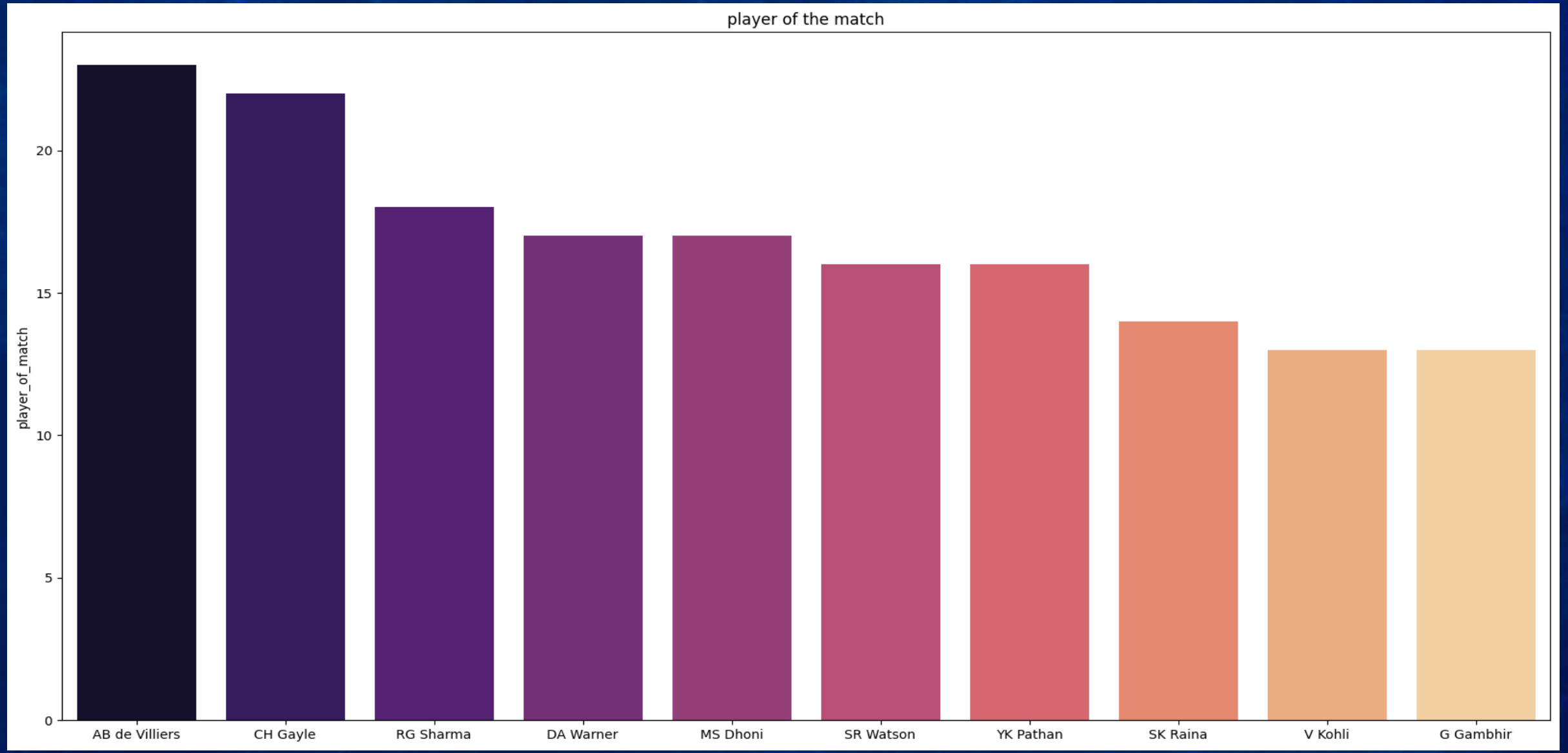


3.Which team have won maximum number of matches in IPL



- The above plot show which team have won many times’
- From the plot we can tell that Mumbai inadians are the team who won many times, second wining team is Chennai super kings.
- And the least won team is Kochi tuskers kerala ,Gujarat titanas and pune .

4.Player of the match



- The plot shows the player who have won man of the match many times.
- From graph we can say that AB De Villers is the player who have won many matches.
- Followed by CH Gayle, RG Sharma, DA warner and MS Dhoni these are 5 players who won man of the match many times.

AB De Villers

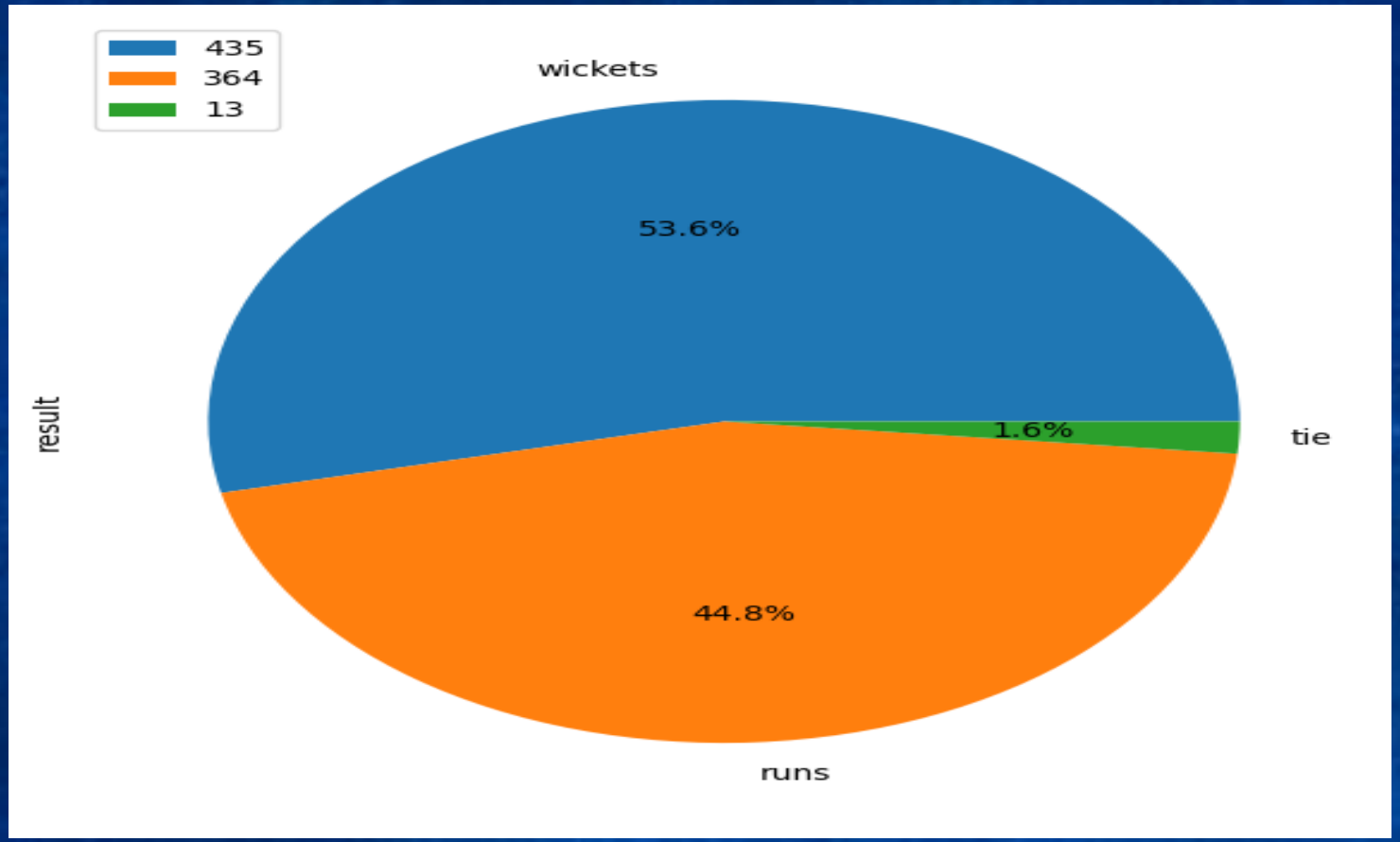
CH gayle

RG sharma

Da warner

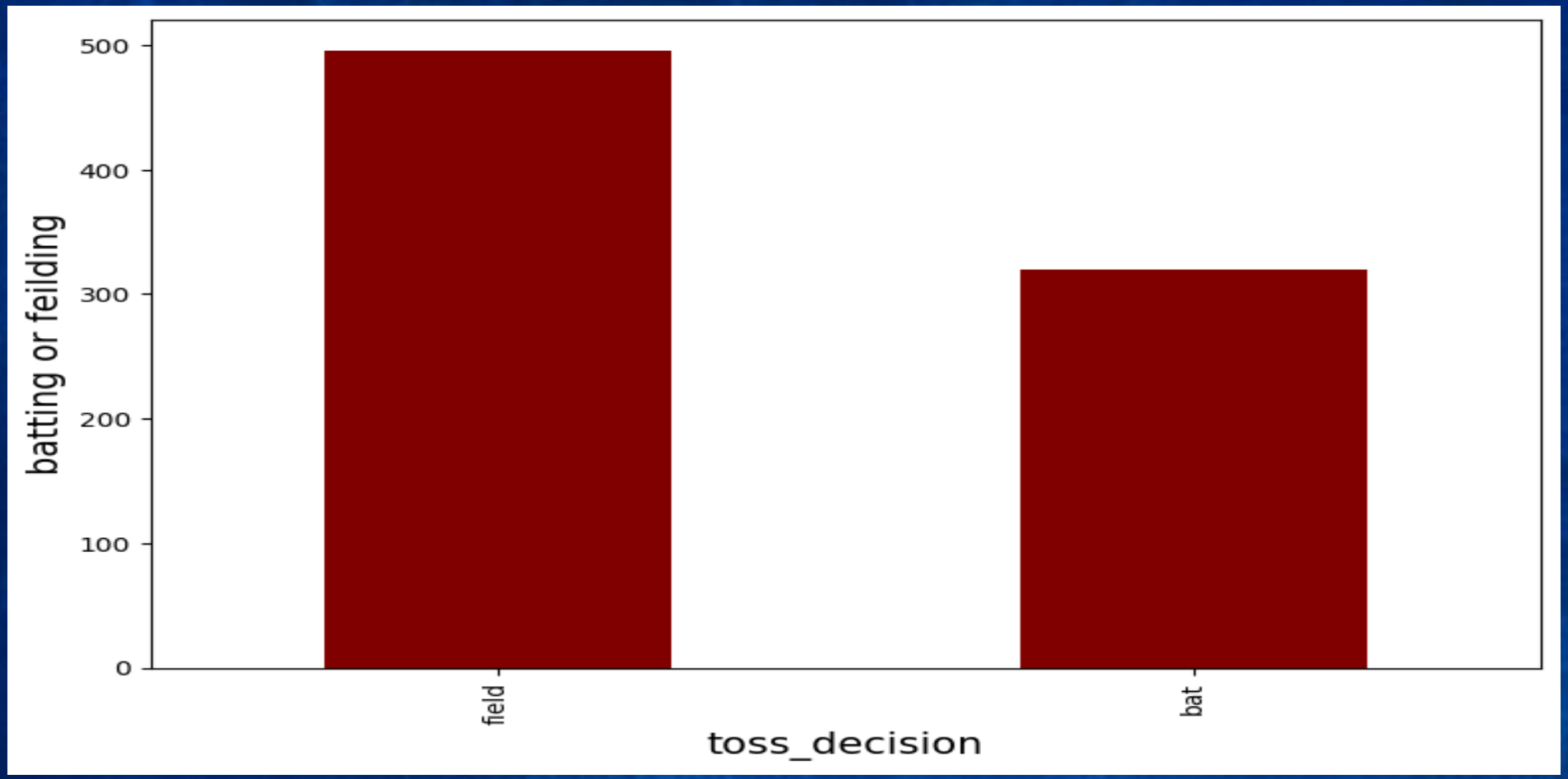
MS Dhoni

5.Percentage of results.



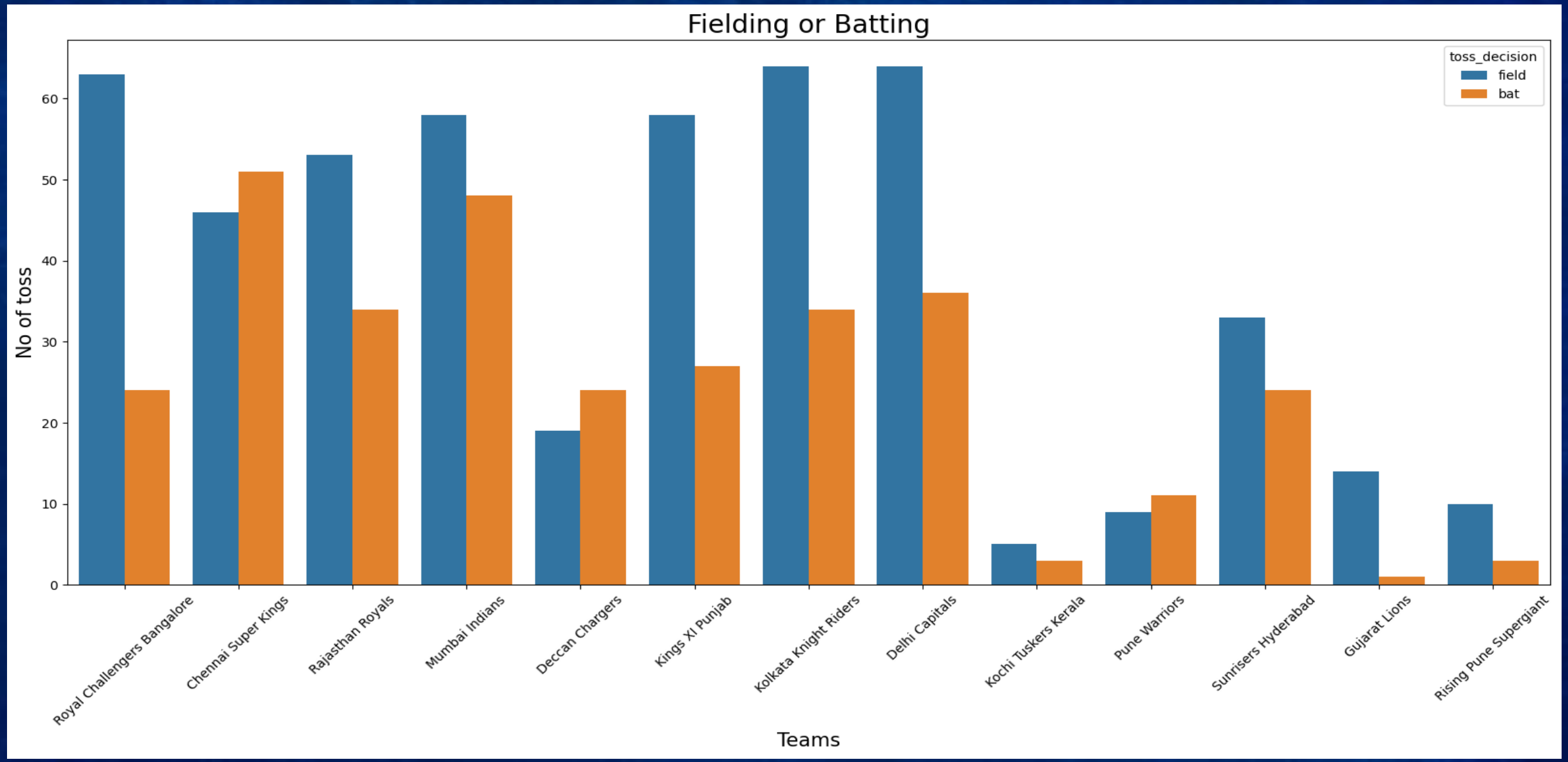
- The bar plot show the percentage of the results like the matches are won by wickets, runs or it is tie.
- So from plot we can see that winning by wickets are more followed by runs and tie matches are less.
- Almost 435(53%) matches are won by wickets and 364 (44%)matches are won by runs and remaining 13 (2.2%)are tie

6.Batting and Feilding

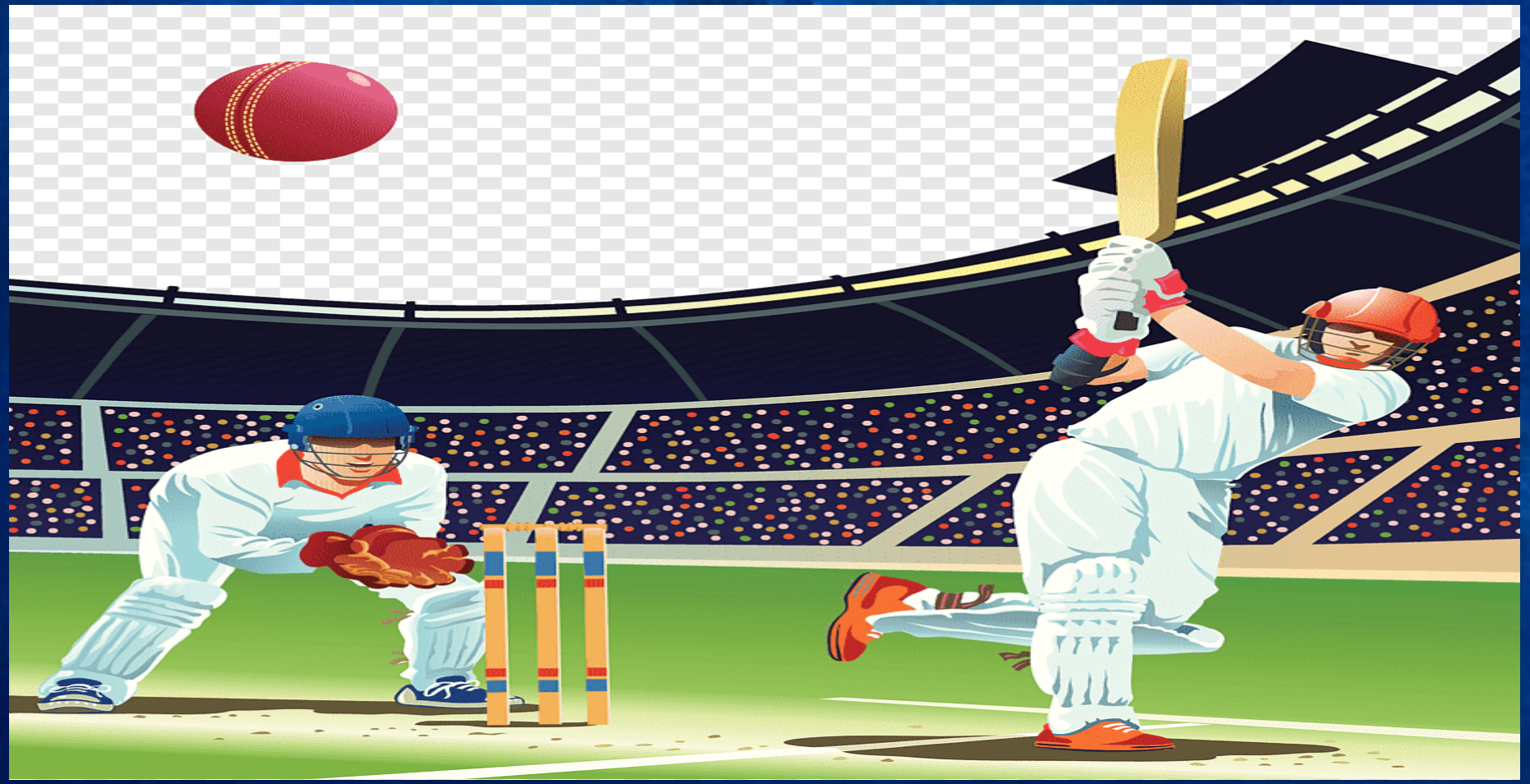


- The plot shows the value of feilding or batting.
- Clearly we can say that feilding are chosen more compared to batting.

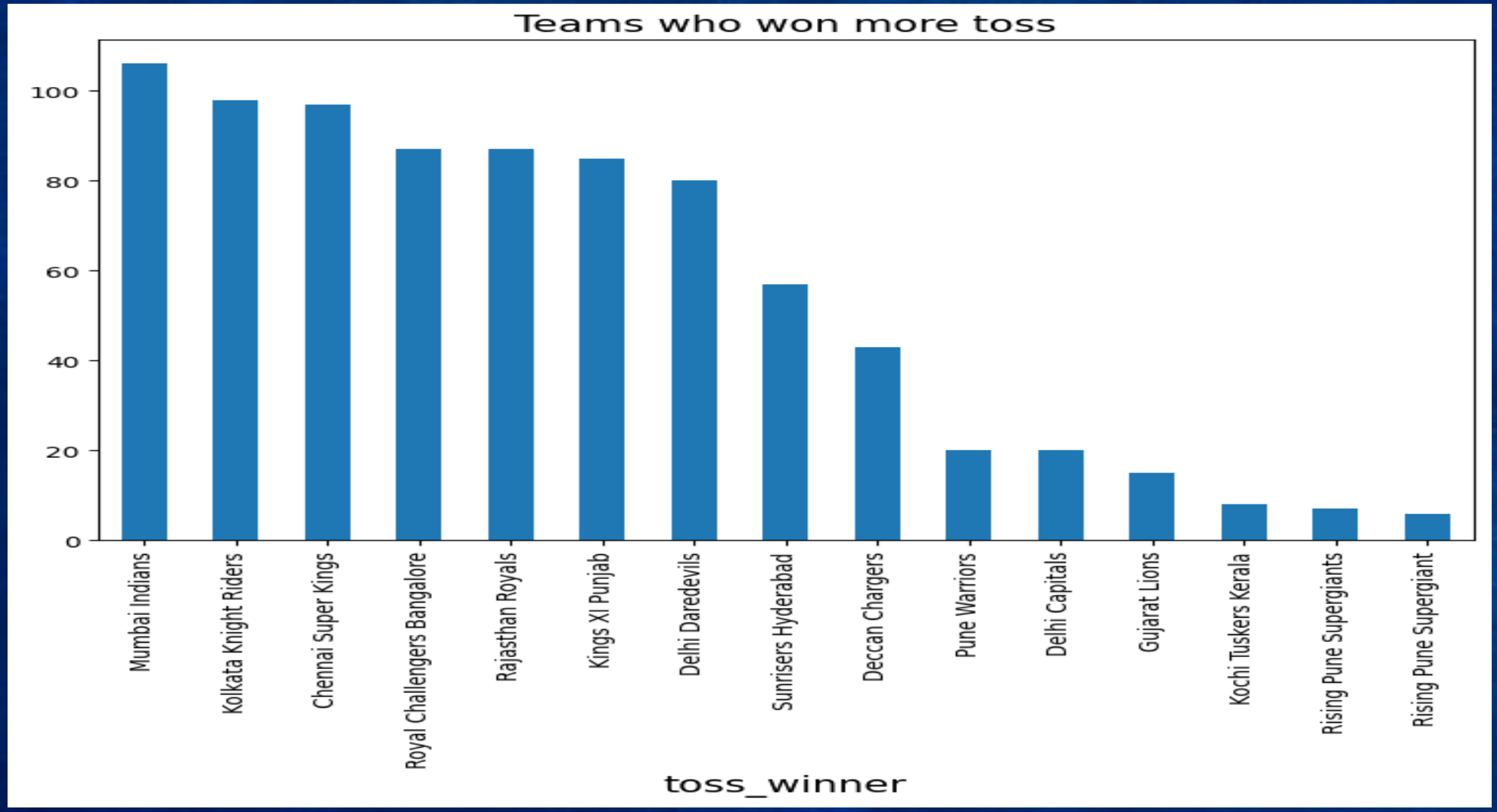
7. Teams who have choosen fielding and batting



- From the plot we can compare that which team have choosen batting and fielding more.
- Here in this Kolkata knight riders ,royal challengers and delhi captials have choose fielding many times compared to other teams.
- Similarly Chennai super kings have choosen Batting many times.



8.Team who won toss many times.



- The graph show the teams who won toss many times .
- From the graph we can tell that Mumbai indains have won toss many times and follwe by Chennai super kings and Kolkata knight riders have won toss .
- And least team is rising pune.

9.Result of particular Team

In [115]:
```
ipl1=IPL[['team1','team2','winner','result']]
ipl1_df=IPL.winner=='Royal Challengers Bangalore'
```

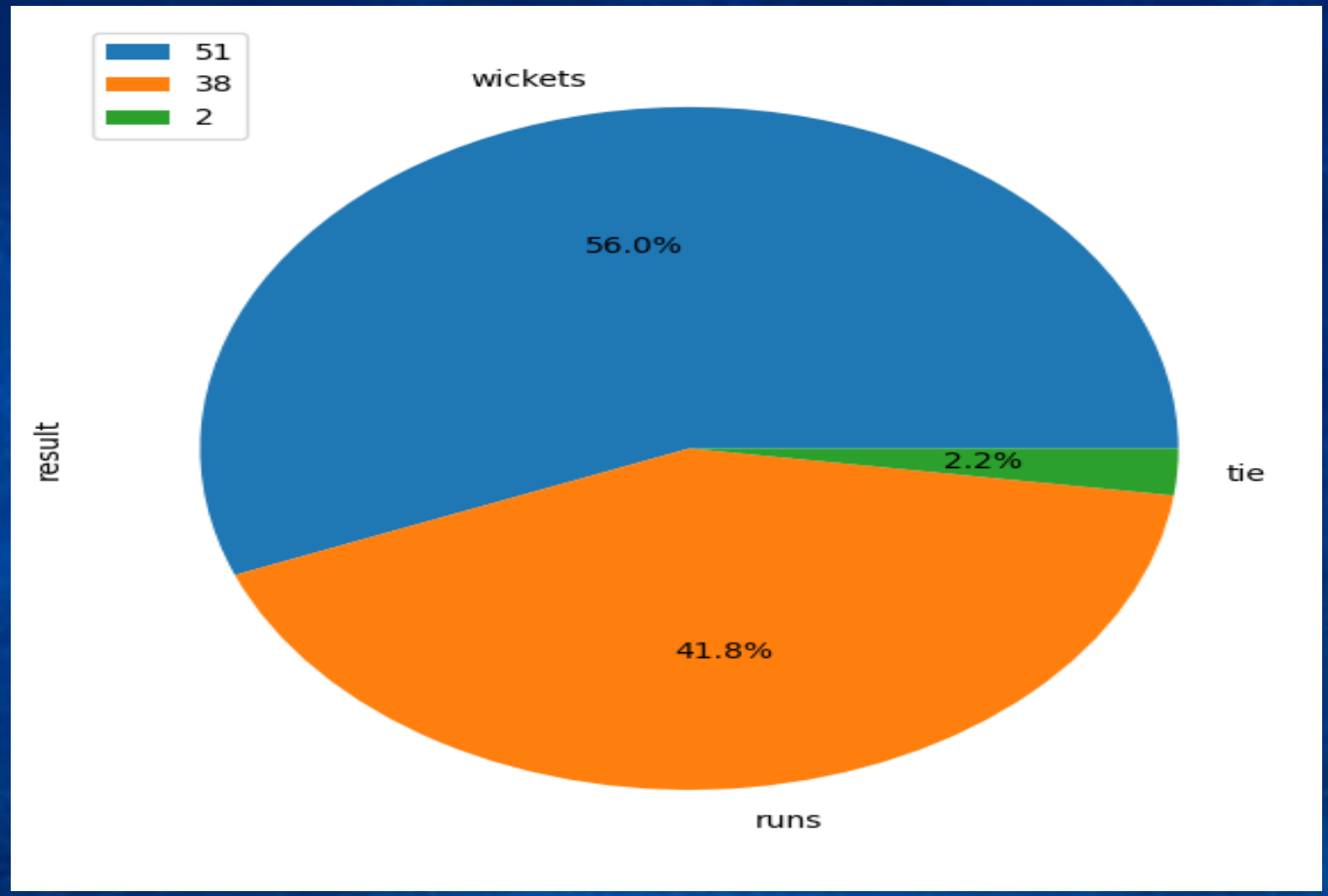
In [116]:
```
ipl2=ipl1[ipl1_df]
ipl2_df=ipl2.sort_values(by=['winner'],ascending=False)
ipl2_df
```

Out[116]:

	team1	team2	winner	result
3	Mumbai Indians	Royal Challengers Bangalore	Royal Challengers Bangalore	wickets
572	Delhi Capitals	Royal Challengers Bangalore	Royal Challengers Bangalore	wickets
564	Kolkata Knight Riders	Royal Challengers Bangalore	Royal Challengers Bangalore	wickets
560	Royal Challengers Bangalore	Gujarat Lions	Royal Challengers Bangalore	runs
555	Kings XI Punjab	Royal Challengers Bangalore	Royal Challengers Bangalore	runs
...	...	...	...	...
231	Royal Challengers Bangalore	Kolkata Knight Riders	Royal Challengers Bangalore	wickets
228	Rajasthan Royals	Royal Challengers Bangalore	Royal Challengers Bangalore	wickets
223	Royal Challengers Bangalore	Kochi Tuskers Kerala	Royal Challengers Bangalore	wickets
220	Royal Challengers Bangalore	Kings XI Punjab	Royal Challengers Bangalore	runs
811	Royal Challengers Bangalore	Mumbai Indians	Royal Challengers Bangalore	tie

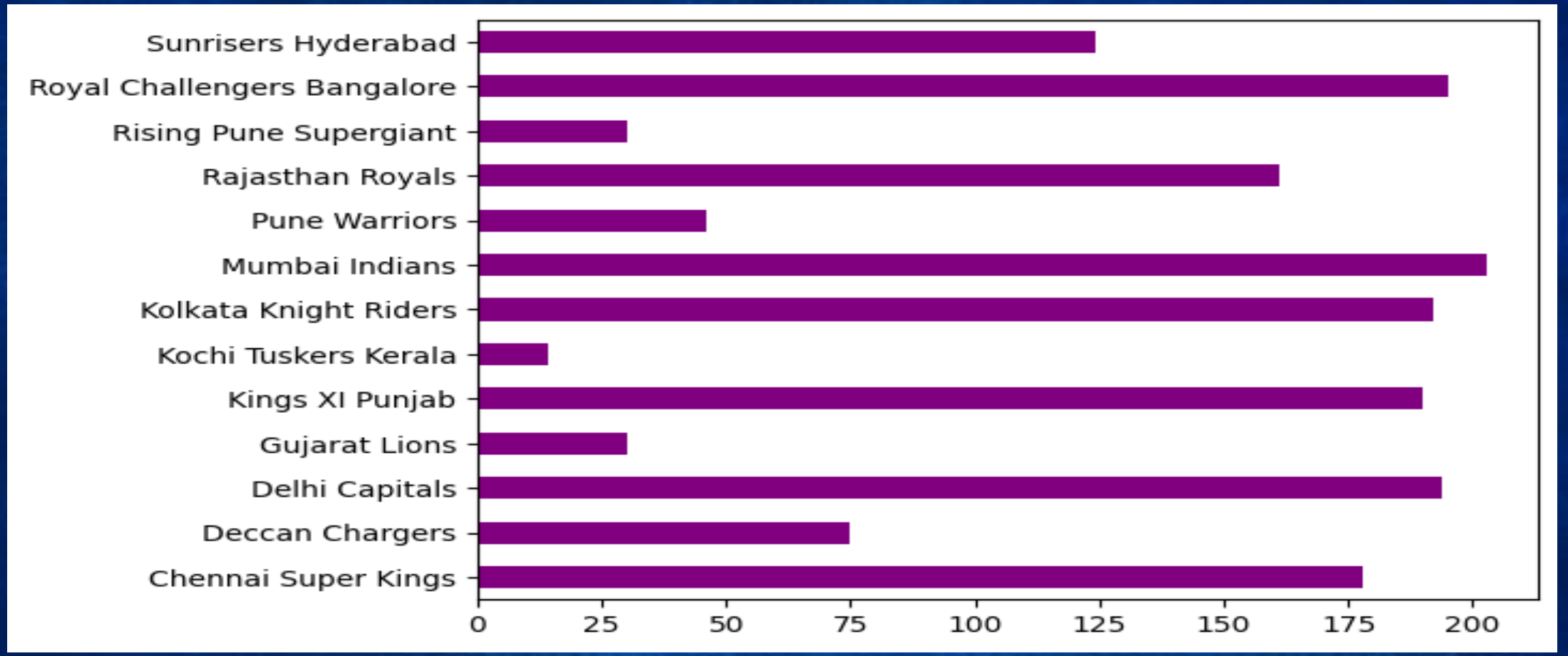
91 rows × 4 columns

- Here to understand and to fetch the data of particular team.
- Example here I have taken Royal challengers bangalore and I fetched the result data of that particular team.



- The above bar plot show the royal challemgers bangalore winning result .
- Here rcb have won by wickets more compared to other result like runs and tie.
- Almost 56% of winning is by wickets.

10.Match played by each team in IPL(2008_2020)



We count the value of each team playing in column one and add to the count of each team from team two to get the desired output. For example, if CSK played 90 times from team one and 85 times from team 2 then the total of175 matches are shown in the graph. You can see that Mumbai Indians played the highest number of matches in the IPL.

CONCLUSION

- Mumbai is the city where highest matches played and followed by Banglore is second highest
- Here we can say that highest matches played venue is Eden garden
- AB De Villers is the player who have won many matches.
- Followed by CH Gayle, RG Sharma, DA warner and MS Dhoni these are 5 players who won man of the match many times.
- Almost 435(53%) matches are won by wickets and 364 (44%)matches are won by runs and remaining 13 (2.2%)are tie
- we can say that feilding are chosen more compared to batting


THANK YOU