

IPLDATA

March 31, 2024

```
[7]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[9]: df=pd.read_csv("matches.csv")
```

```
[13]: df.head() #overview of data
```

```
[13]:
```

	id	season	city	date	team1 \
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad
1	2	2017	Pune	2017-04-06	Mumbai Indians
2	3	2017	Rajkot	2017-04-07	Gujarat Lions
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore

		team2	toss_winner	toss_decision \
0	Royal Challengers Bangalore	Royal Challengers Bangalore	field	
1	Rising Pune Supergiant	Rising Pune Supergiant	field	
2	Kolkata Knight Riders	Kolkata Knight Riders	field	
3	Kings XI Punjab	Kings XI Punjab	field	
4	Delhi Daredevils	Royal Challengers Bangalore	bat	

	result	dl_applied	winner	win_by_runs \
0	normal	0	Sunrisers Hyderabad	35
1	normal	0	Rising Pune Supergiant	0
2	normal	0	Kolkata Knight Riders	0
3	normal	0	Kings XI Punjab	0
4	normal	0	Royal Challengers Bangalore	15

	win_by_wickets	player_of_match	venue \
0	0	Yuvraj Singh	Rajiv Gandhi International Stadium, Uppal
1	7	SPD Smith	Maharashtra Cricket Association Stadium
2	10	CA Lynn	Saurashtra Cricket Association Stadium
3	6	GJ Maxwell	Holkar Cricket Stadium
4	0	KM Jadhav	M Chinnaswamy Stadium

	umpire1	umpire2	umpire3
0	AY Dandekar	NJ Llong	NaN
1	A Nand Kishore	S Ravi	NaN
2	Nitin Menon	CK Nandan	NaN
3	AK Chaudhary	C Shamshuddin	NaN
4	NaN	NaN	NaN

```
[15]: #getting freq of most man of the match awards
df["player_of_match"].value_counts()
```

```
[15]: player_of_match
CH Gayle          21
AB de Villiers    20
RG Sharma         17
MS Dhoni          17
DA Warner         17
..
PD Collingwood    1
NV Ojha           1
AC Voges          1
J Theron          1
S Hetmyer         1
Name: count, Length: 226, dtype: int64
```

```
[17]: #top ten players

df["player_of_match"].value_counts()[0:10]
```

```
[17]: player_of_match
CH Gayle          21
AB de Villiers    20
RG Sharma         17
MS Dhoni          17
DA Warner         17
YK Pathan         16
SR Watson         15
SK Raina          14
G Gambhir         13
MEK Hussey        12
Name: count, dtype: int64
```

```
[23]: #top 5 players

df["player_of_match"].value_counts()[0:5]
```

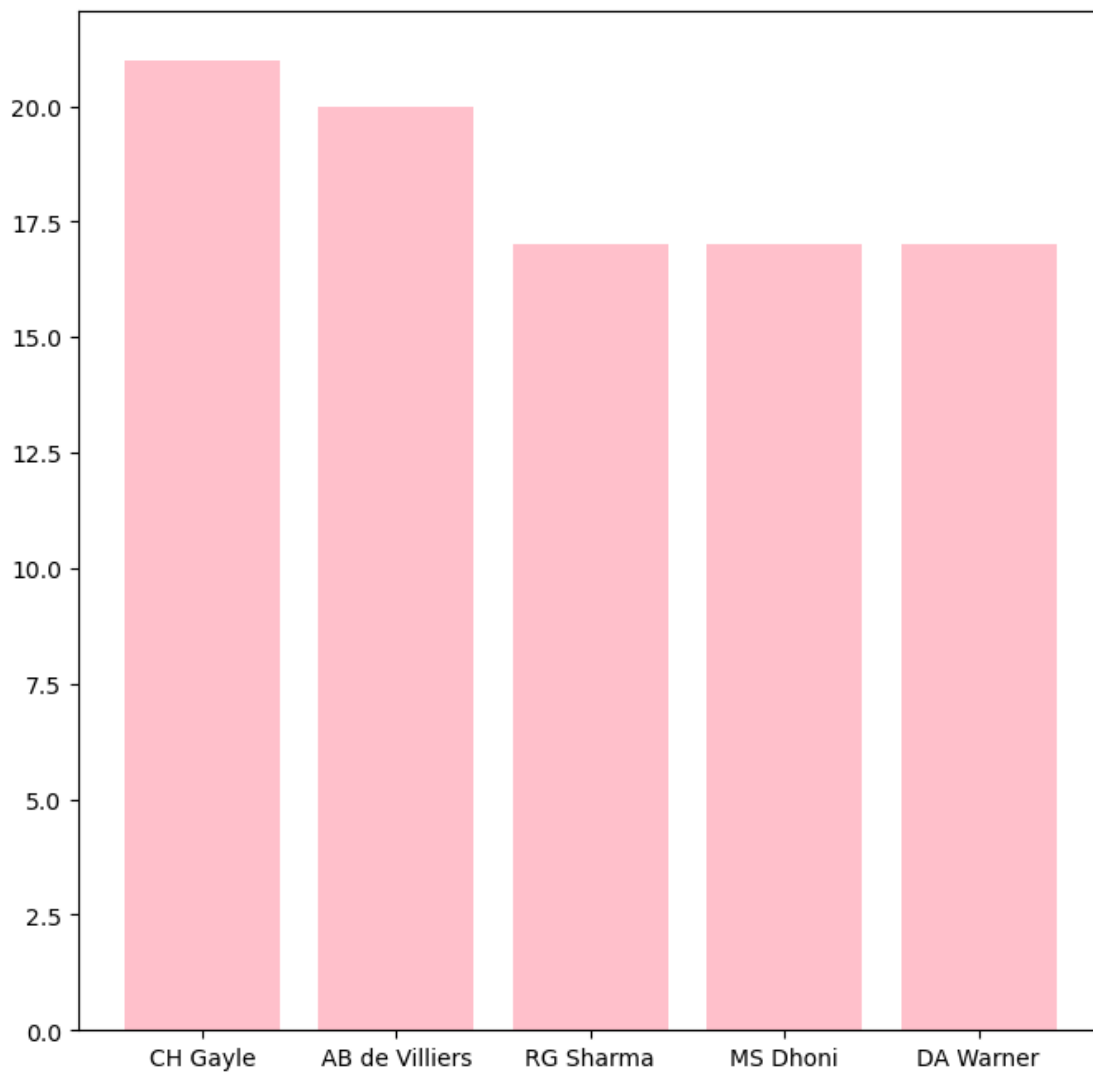
```
[23]: player_of_match
CH Gayle          21
AB de Villiers    20
```

```
RG Sharma          17
MS Dhoni            17
DA Warner           17
Name: count, dtype: int64
```

```
[29]: list(df["player_of_match"].value_counts()[0:5].keys())
```

```
[29]: ['CH Gayle', 'AB de Villiers', 'RG Sharma', 'MS Dhoni', 'DA Warner']
```

```
[45]: plt.figure(figsize=(8, 8))
plt.bar(list(df["player_of_match"].value_counts()[0:5].keys()),
        list(df["player_of_match"].value_counts()[0:5]), color="pink")
plt.show()
```



```
[39]: #getting frew of the results column
df["result"].value_counts()
```

```
[39]: result
normal      743
tie          9
no result    4
Name: count, dtype: int64
```

```
[41]: #finding out number of toss wins wrt. each team

df["toss_winner"].value_counts()
```

```
[41]: toss_winner
Mumbai Indians      98
Kolkata Knight Riders  92
Chennai Super Kings  89
Royal Challengers Bangalore  81
Kings XI Punjab      81
Delhi Daredevils     80
Rajasthan Royals     80
Sunrisers Hyderabad  46
Deccan Chargers      43
Pune Warriors        20
Gujarat Lions        15
Delhi Capitals        10
Kochi Tuskers Kerala   8
Rising Pune Supergiants  7
Rising Pune Supergiant  6
Name: count, dtype: int64
```

```
[62]: #records where a team won batting first
batting_first=df[df["win_by_runs"]!=0]
```

```
[64]: batting_first.head()
```

```
[64]:
```

	id	season	city	date	team1 \
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore
8	9	2017	Pune	2017-04-11	Delhi Daredevils
13	14	2017	Kolkata	2017-04-15	Kolkata Knight Riders
14	15	2017	Delhi	2017-04-15	Delhi Daredevils

		team2	toss_winner	toss_decision \
0	Royal Challengers Bangalore	Royal Challengers Bangalore	field	
4	Delhi Daredevils	Royal Challengers Bangalore	bat	
8	Rising Pune Supergiant	Rising Pune Supergiant	field	

13		Sunrisers Hyderabad	Sunrisers Hyderabad	field
14		Kings XI Punjab	Delhi Daredevils	bat

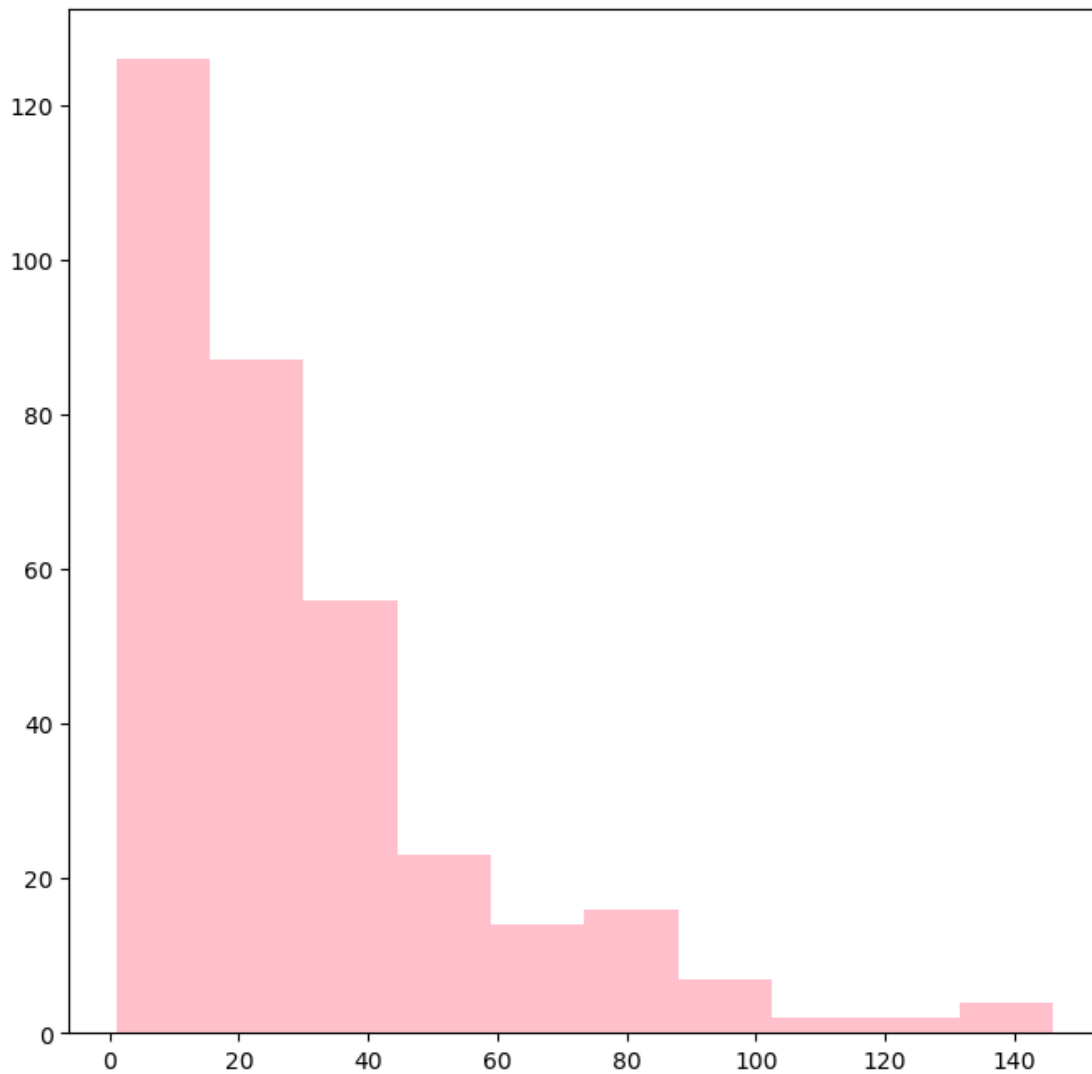
	result	dl_applied	winner	win_by_runs	\
0	normal	0	Sunrisers Hyderabad	35	
4	normal	0	Royal Challengers Bangalore	15	
8	normal	0	Delhi Daredevils	97	
13	normal	0	Kolkata Knight Riders	17	
14	normal	0	Delhi Daredevils	51	

	win_by_wickets	player_of_match	venue	\
0	0	Yuvraj Singh	Rajiv Gandhi International Stadium, Uppal	
4	0	KM Jadhav	M Chinnaswamy Stadium	
8	0	SV Samson	Maharashtra Cricket Association Stadium	
13	0	RV Uthappa	Eden Gardens	
14	0	CJ Anderson	Feroz Shah Kotla	

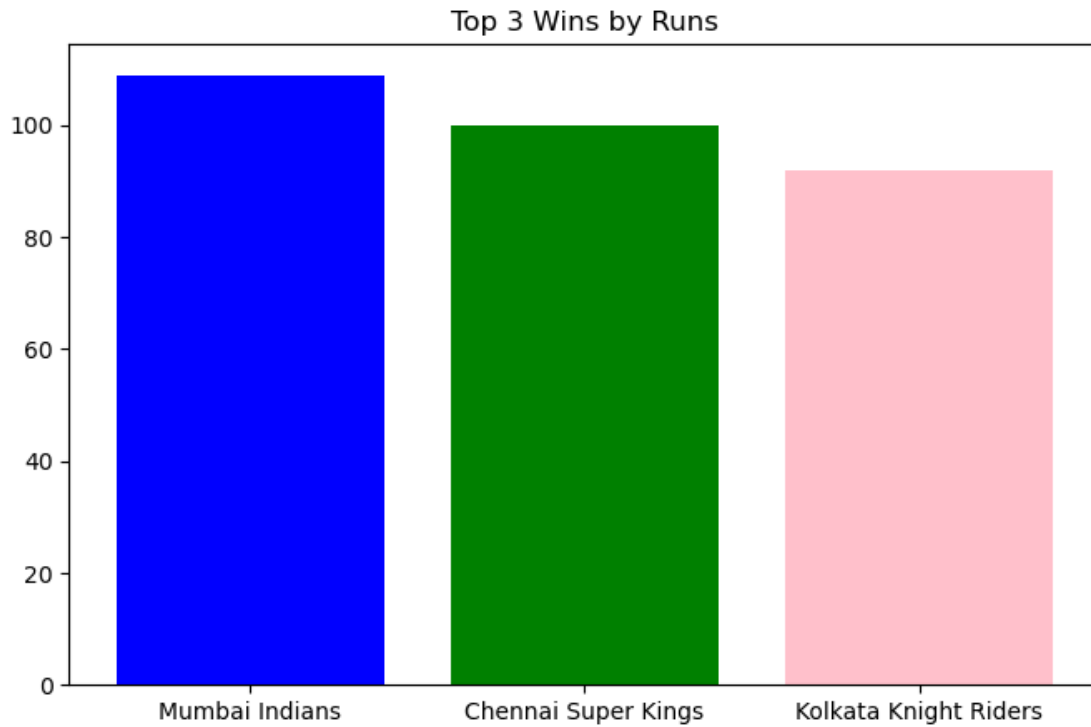
	umpire1	umpire2	umpire3
0	AY Dandekar	NJ Llong	NaN
4	NaN	NaN	NaN
8	AY Dandekar	S Ravi	NaN
13	AY Dandekar	NJ Llong	NaN
14	YC Barde	Nitin Menon	NaN

```
[82]: #histogram

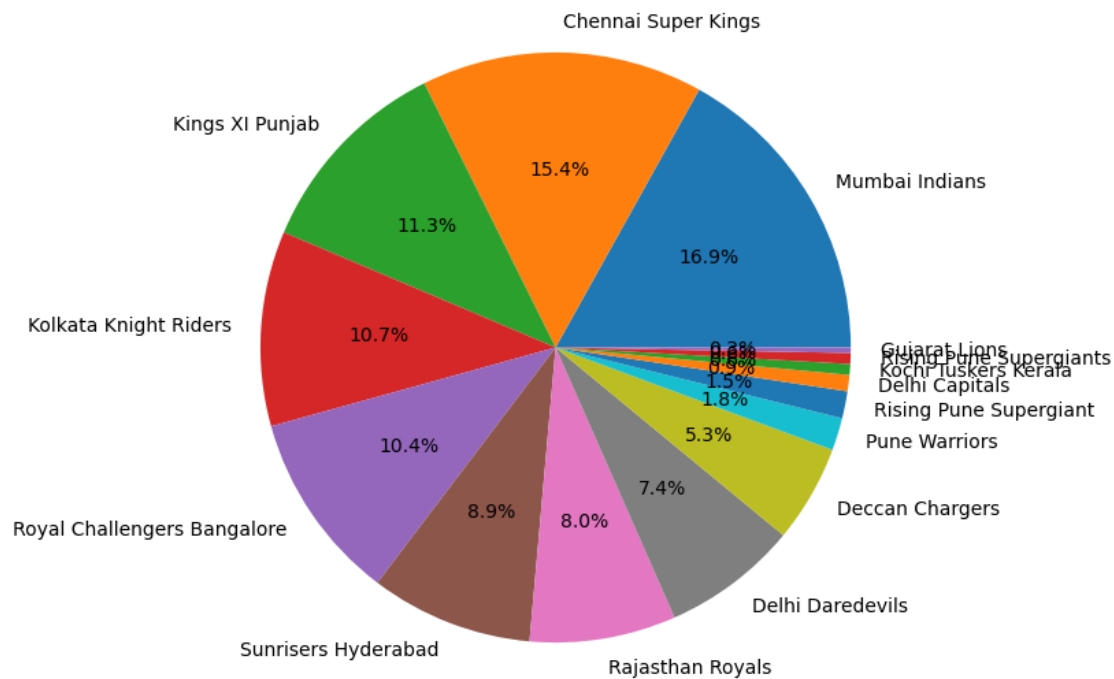
plt.figure(figsize=(8,8))
plt.hist(batting_first["win_by_runs"],color='pink')
plt.show()
```



```
[106]: # bar plot
plt.figure(figsize=(8, 5))
plt.bar(list(df["winner"].value_counts()[0:3].keys()), list(df["winner"].
    ↳value_counts()[0:3]), color=["blue", "green", "pink"])
plt.title('Top 3 Wins by Runs')
plt.show()
```



```
[115]: #Making a pie chart
plt.figure(figsize=(7,7))
plt.pie(list(batting_first['winner'].
    ↳value_counts()),labels=list(batting_first['winner'].value_counts().
    ↳keys()),autopct='%0.1f%%')
plt.show()
```



[123]: *#records after batting second*

```
batting_second=df[df["win_by_wickets"]!=0]
```

[125]: `batting_second.head()`

```
[125]:
```

	id	season	city	date	team1 \
1	2	2017	Pune	2017-04-06	Mumbai Indians
2	3	2017	Rajkot	2017-04-07	Gujarat Lions
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant
5	6	2017	Hyderabad	2017-04-09	Gujarat Lions
6	7	2017	Mumbai	2017-04-09	Kolkata Knight Riders

	team2	toss_winner	toss_decision	result \
1	Rising Pune Supergiant	Rising Pune Supergiant	field	normal
2	Kolkata Knight Riders	Kolkata Knight Riders	field	normal
3	Kings XI Punjab	Kings XI Punjab	field	normal
5	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal
6	Mumbai Indians	Mumbai Indians	field	normal

	dl_applied	winner	win_by_runs	win_by_wickets \
1	0	Rising Pune Supergiant	0	7
2	0	Kolkata Knight Riders	0	10

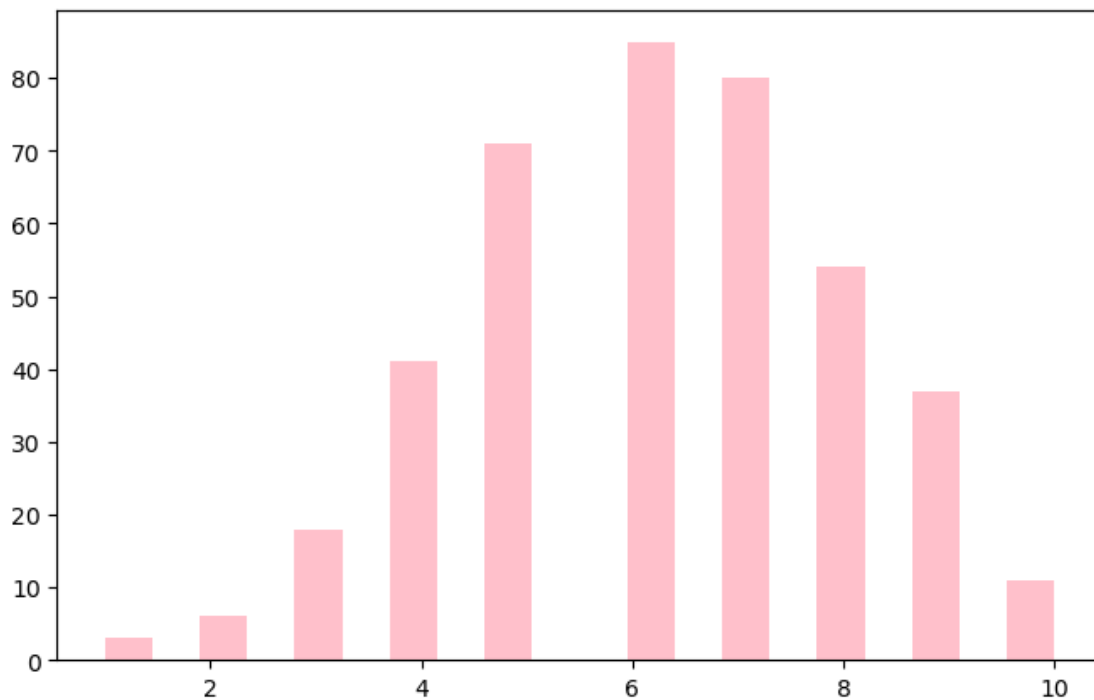
3	0	Kings XI Punjab	0	6
5	0	Sunrisers Hyderabad	0	9
6	0	Mumbai Indians	0	4

	player_of_match	venue	umpire1 \
1	SPD Smith	Maharashtra Cricket Association Stadium	A Nand Kishore
2	CA Lynn	Saurashtra Cricket Association Stadium	Nitin Menon
3	GJ Maxwell	Holkar Cricket Stadium	AK Chaudhary
5	Rashid Khan	Rajiv Gandhi International Stadium, Uppal	A Deshmukh
6	N Rana	Wankhede Stadium	Nitin Menon

	umpire2	umpire3
1	S Ravi	NaN
2	CK Nandan	NaN
3	C Shamshuddin	NaN
5	NJ Llong	NaN
6	CK Nandan	NaN

[133]: *#histogram*

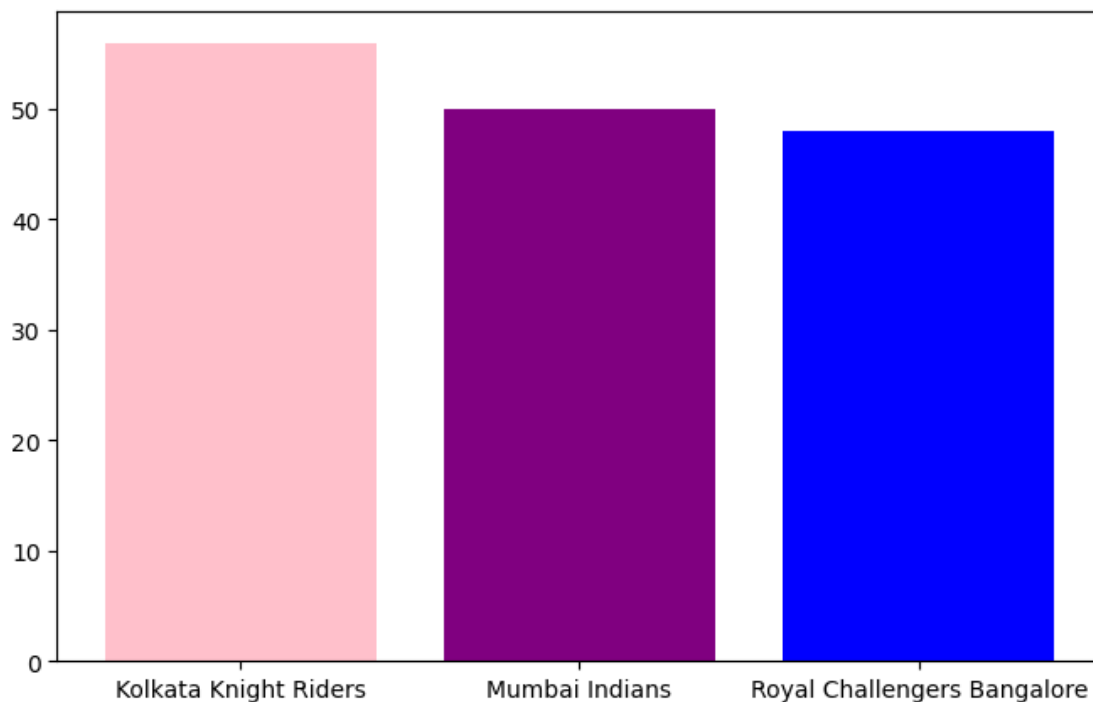
```
plt.figure(figsize=(8,5))
plt.hist(batting_second["win_by_wickets"],bins=20,color="pink")
plt.show()
```



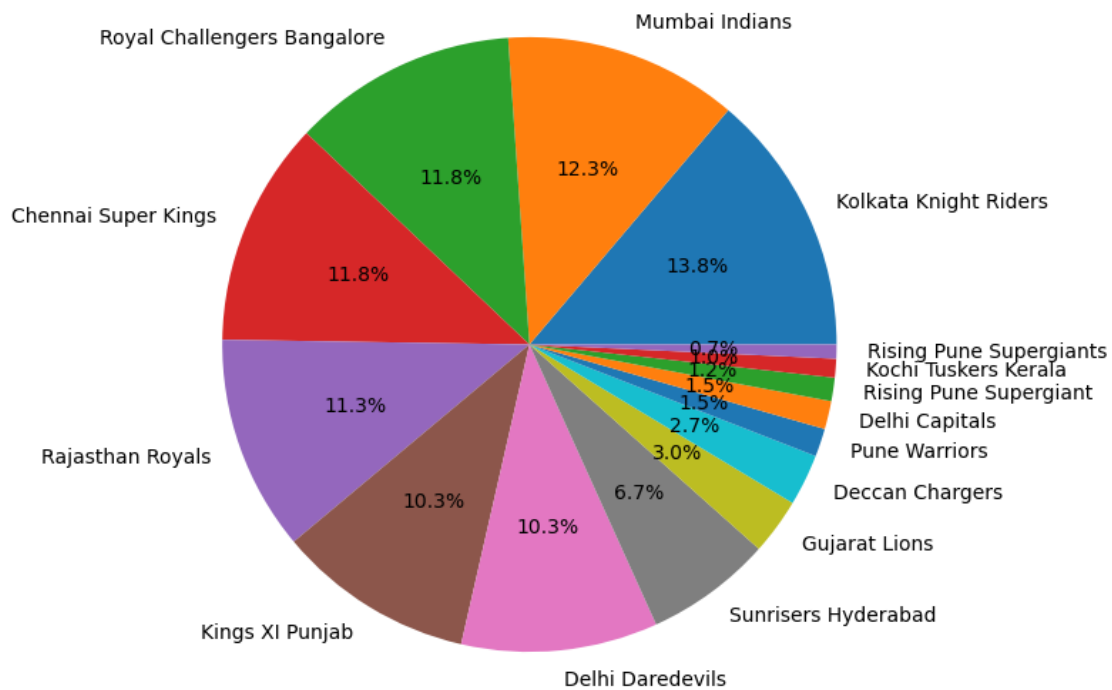
```
[135]: batting_second["winner"].value_counts()
```

```
[135]: winner
Kolkata Knight Riders      56
Mumbai Indians             50
Royal Challengers Bangalore 48
Chennai Super Kings        48
Rajasthan Royals           46
Kings XI Punjab            42
Delhi Daredevils           42
Sunrisers Hyderabad        27
Gujarat Lions              12
Deccan Chargers            11
Pune Warriors              6
Delhi Capitals              6
Rising Pune Supergiant      5
Kochi Tuskers Kerala        4
Rising Pune Supergiants     3
Name: count, dtype: int64
```

```
[141]: # Bar chart
plt.figure(figsize=(8,5))
plt.bar(list(batting_second["winner"].value_counts()[0:3].keys()),
        ↪list(batting_second["winner"].value_counts()[0:3]),
        ↪color=["pink","purple","blue"])
plt.show()
```



```
[149]: # Pie chart
plt.figure(figsize=(7,7))
plt.pie(list(batting_second["winner"].value_counts()),
        labels=list(batting_second["winner"].value_counts().keys()), autopct='%0.1f%%')
plt.show()
```



```
[152]: #Looking at the number of matches played in each city
df["city"].value_counts()
```

```
[152]: city
Mumbai      101
Kolkata      77
Delhi        74
Bangalore    66
Hyderabad    64
Chennai      57
Jaipur       47
```

Chandigarh	46
Pune	38
Durban	15
Bengaluru	14
Visakhapatnam	13
Centurion	12
Ahmedabad	12
Rajkot	10
Mohali	10
Indore	9
Dharamsala	9
Johannesburg	8
Cuttack	7
Ranchi	7
Port Elizabeth	7
Cape Town	7
Abu Dhabi	7
Sharjah	6
Raipur	6
Kochi	5
Kanpur	4
Nagpur	3
Kimberley	3
East London	3
Bloemfontein	2

Name: count, dtype: int64

[154]: *#Looking at the number of matches played in each season*

```
df["season"].value_counts()
```

[154]: season

2013	76
2012	74
2011	73
2010	60
2014	60
2016	60
2018	60
2019	60
2017	59
2015	59
2008	58
2009	57

Name: count, dtype: int64

```
[160]: #finding out how many times a team has won the match after winning the toss  
import numpy as np  
np.sum(df["toss_winner"]==df["winner"])
```

```
[160]: 393
```

```
[166]: (393/636)*100
```

```
[166]: 61.79245283018868
```

```
[ ]:
```