

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

In [2]: df=pd.read_csv("data/matches.csv")
df=pd.read_csv("data/deliveries.csv")
```

```
In [3]: df.head(5)
```

	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
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal		0	Sunrisers Hyderabad	35	0	Yuvraj Singh	Rajiv Gandhi International Stadium, Uppal	AY Dandekar	NJ
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal		0	Rising Pune Supergiant	0	7	SPD Smith	Maharashtra Cricket Association Stadium	A Nand Kishore	S
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal		0	Kolkata Knight Riders	0	10	CA Lynn	Saurashtra Cricket Association Stadium	Nitin Menon	CK N
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal		0	Kings XI Punjab	0	6	GJ Maxwell	Holkar Cricket Stadium	AK Chaudhary	Shamsh
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal		0	Royal Challengers Bangalore	15	0	KM Jadhav	M Chinnaswamy Stadium	NaN	

```
In [4]: df1.head(5)
```

	match_id	inning	bating_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs	legbye_runs	noball_runs	penalty_runs	batsman_runs	extra_runs	total_runs
0	1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0	0	0
1	1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0	0	0
2	1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0	4	0
3	1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0	0	0
4	1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0	2	2

5 rows × 21 columns

```
In [5]: print('Total Matches Played:',df.shape[0])
print('\n Venues Played At:',df['city'].unique())
print('\n Teams :',df['team1'].unique())

Total Matches Played: 636

Venues Played At: ['Hyderabad' 'Pune' 'Rajkot' 'Indore' 'Bangalore' 'Mumbai' 'Kolkata'
'Delhi' 'Chandigarh' 'Kanpur' 'Jaipur' 'Chennai' 'Cape Town'
'Port Elizabeth' 'Durban' 'Centurion' 'East London' 'Johannesburg'
'Kimberley' 'Bloemfontein' 'Ahmedabad' 'Cuttack' 'Nagpur' 'Dharamsala'
'Kochi' 'Visakhapatnam' 'Raipur' 'Ranchi' 'Abu Dhabi' 'Sharjah' 'nan']

Teams: ['Sunrisers Hyderabad' 'Mumbai Indians' 'Gujarat Lions'
'Rising Pune Supergiant' 'Royal Challengers Bangalore'
'Kolkata Knight Riders' 'Delhi Daredevils' 'Kings XI Punjab'
'Chennai Super Kings' 'Rajasthan Royals' 'Deccan Chargers'
'Kochi Tuskers Kerala' 'Pune Warriors' 'Rising Pune Supergiants']
```

```
In [6]: print('Total venues played at:',df['city'].nunique())
print('\nTotal umpires ',df['umpire1'].nunique())

Total venues played at: 30

Total umpires 44
```

```
In [7]: print((df['player_of_match'].value_counts()).idxmax(),': has most man of the match awards')
print(((df['winner']).value_counts()).idxmax(),': has the highest number of match wins')

CH Gayle : has most man of the match awards
Mumbai Indians : has the highest number of match wins
```

KOLKATA KNIGHT RIDERS WON WITH THE MOST WICKETS

```

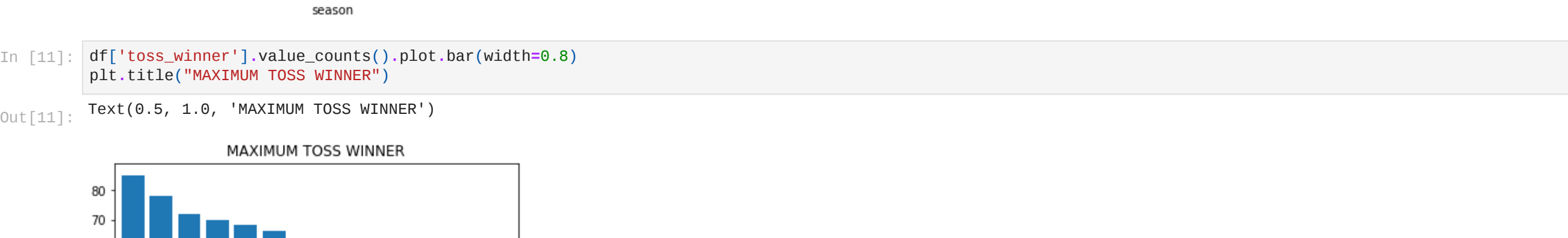
sns.countplot(data=df, x='season', hue='toss_decision')
plt.title("TOSS DECISIONS ACROSS SEASONS")
Text(0.5, 1.0, 'TOSS DECISIONS ACROSS SEASONS')

```

Season	bat	field
44	45	0
45	45	0
46	45	0
47	45	0
48	45	0

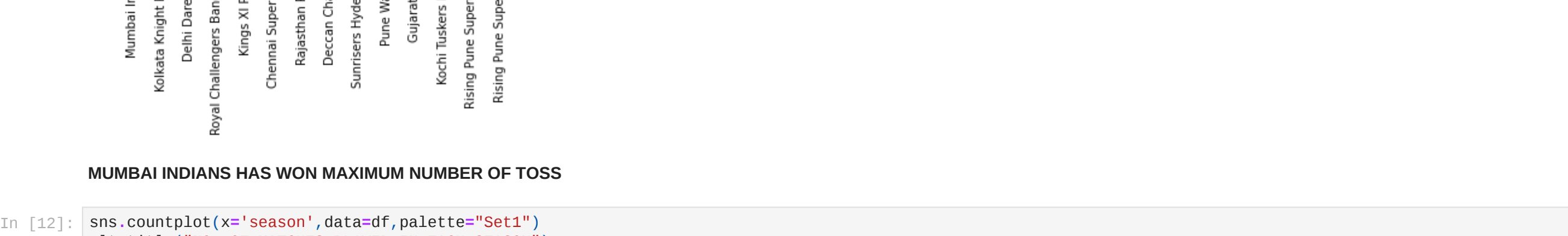
Team	bet (won by runs)	blue (won by wickets)
Gujarat Lions	27	33
Kolkata Knight Riders	36	22
Mumbai Indians	38	38
Chennai Super Kings	39	31
Royal Challengers Bangalore	26	42

```
In [10]: sns.countplot(data=df,x='season',hue='toss_decision')
plt.title('TOSS DECISIONS ACROSS SEASONS')
```



```
In [11]: df['toss_winner'].value_counts().plot.bar(width=0.8)
plt.title('MAXIMUM TOSS WINNER')
```

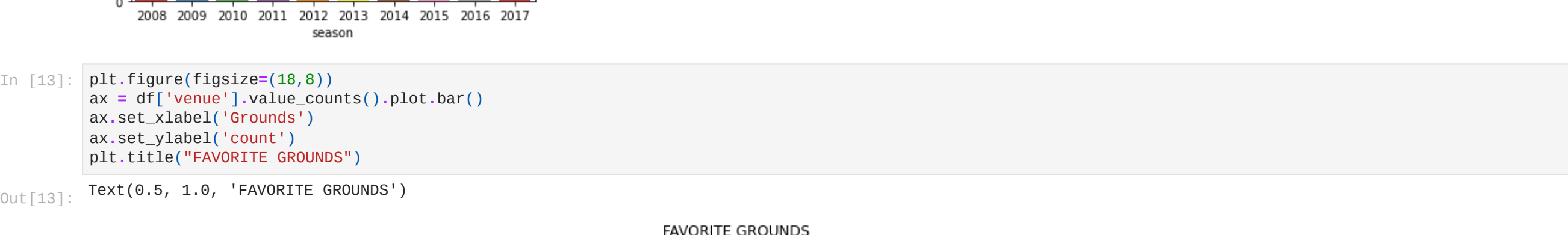
Text(0.5, 1.0, 'MAXIMUM TOSS WINNER')



MUMBAI INDIANS HAS WON MAXIMUM NUMBER OF TOSS

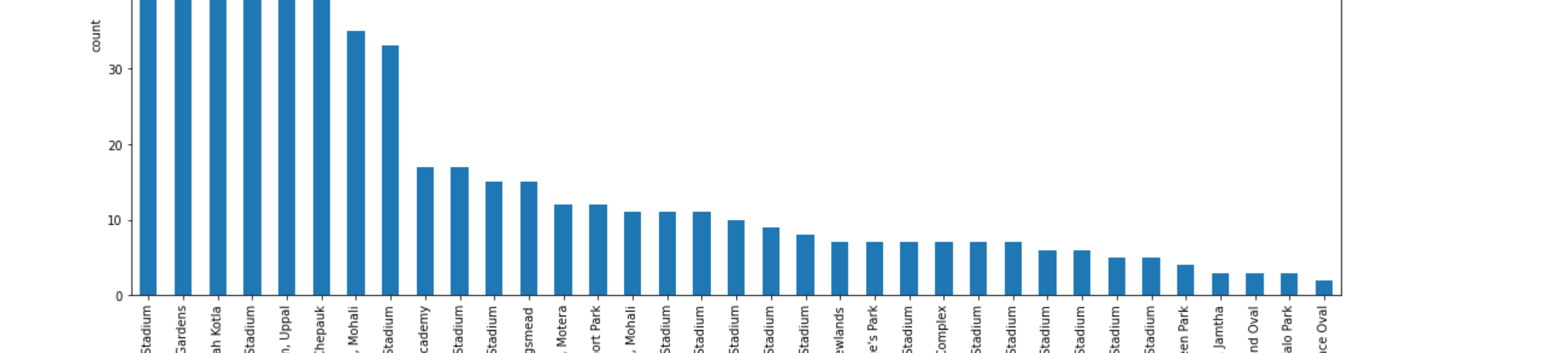
```
In [12]: sns.countplot(x='season',data=df,palette="Set1")
plt.title('NO. OF MATCHES PLAYED IN EACH SEASON')
```

Text(0.5, 1.0, 'NO. OF MATCHES PLAYED IN EACH SEASON')



```
In [13]: plt.figure(figsize=(18,8))
ax = df['venue'].value_counts().plot.bar()
ax.set_xlabel('grounds')
ax.set_ylabel('count')
```

Text(0.5, 1.0, 'FAVORITE GROUNDS')



MAXIMUM NUMBER OF MATCHES HAVE BEEN PLAYED IN M CHINNASWAMY STATDIUM

```
In [14]: a = df['player_of_match'].value_counts().head(10).plot.bar(color='r')
a.set_xlabel('player_of_match')
a.set_ylabel('count')
```

Text(0.5, 1.0, 'MAXIMUM MAN OF THE MATCH AWARD')



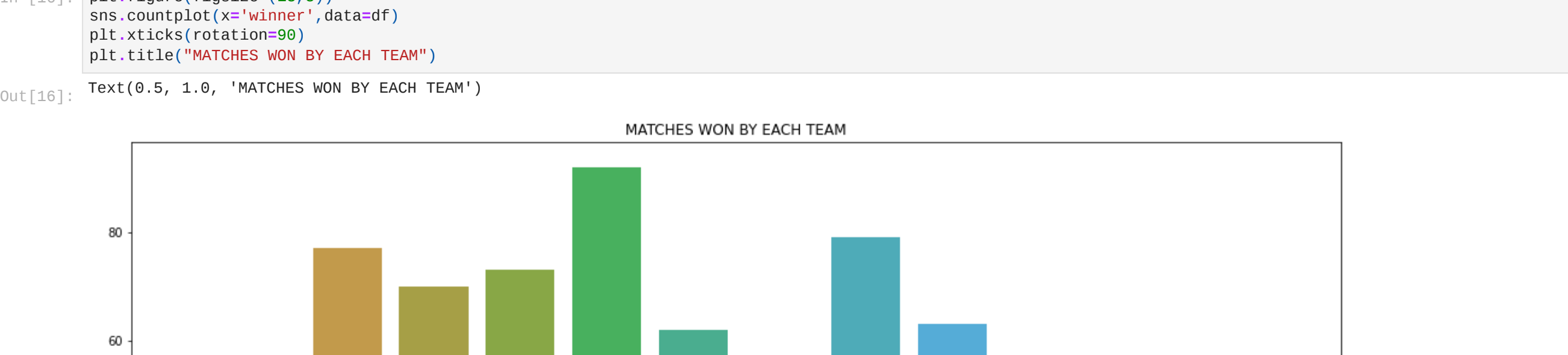
CHRIS GAYLE HAS RECEIVED MAXIMUM MOM AWARDS

```
In [15]: print(' Total Matches with Super Overs:',df1[df['is_super_over']==1].match_id.nunique())

Total Matches with Super Overs: 7
```

```
In [16]: plt.figure(figsize=(18,8))
sns.countplot(x='winner',data=df)
plt.xticks(rotation=90)
plt.title('MATCHES WON BY EACH TEAM')
```

Text(0.5, 1.0, 'MATCHES WON BY EACH TEAM')



MUMBAI INDIANS HAVE WON THE MOST NUMBER OF MATCHES

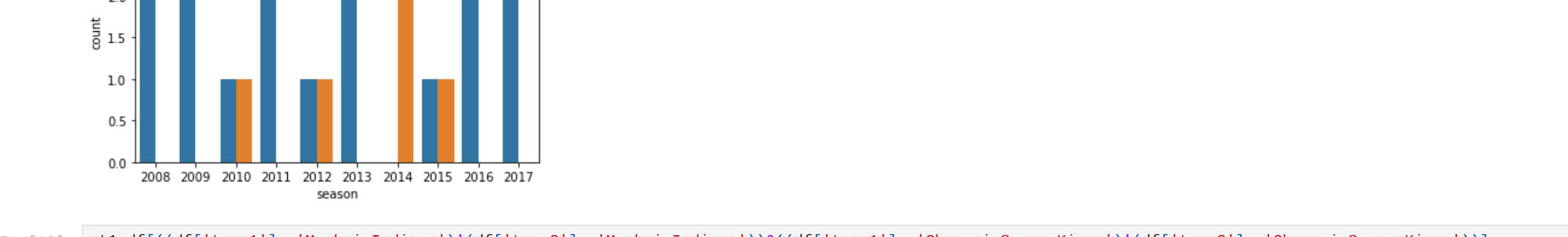
```
In [17]: mti=df[(df['team1']!='Mumbai Indians')](df['team2']=='Mumbai Indians')&((df['team1']=='Kolkata Knight Riders')|(df['team2']=='Kolkata Knight Riders'))
sns.countplot(x='season', hue='winner', data=mti)
plt.title('MUMBAI INDIANS vs KOLKATA KNIGHT RIDERS')
```

Text(0.5, 1.0, 'MUMBAI INDIANS vs KOLKATA KNIGHT RIDERS')



```
In [18]: mti=df[(df['team1']!='Mumbai Indians')](df['team2']=='Mumbai Indians')&((df['team1']=='Chennai Super Kings')|(df['team2']=='Chennai Super Kings'))
sns.countplot(x='season', hue='winner', data=mti)
plt.title('MUMBAI INDIANS vs CHENNAI SUPER KINGS')
```

Text(0.5, 1.0, 'MUMBAI INDIANS vs CHENNAI SUPER KINGS')



TEAMS WHO SCORED MORE THAN 200 RUNS IN AN INNINGS

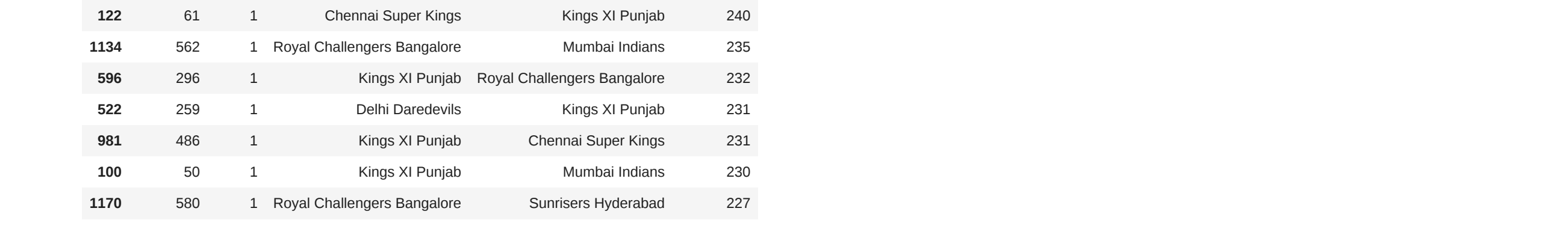
```
In [19]: high_scores=df1.groupby(['match_id','inning','batting_team','bowling_team'])['total_runs'].sum().reset_index()
a=high_scores.nlargest(10,'total_runs')
```

```
Out [19]:
```

	match_id	inning	batting_team	bowling_team	total_runs
829	411	1	Royal Challengers Bangalore	Pune Warriors	263
1250	620	1	Royal Challengers Bangalore	Gujarat Lions	248
416	206	1	Chennai Super Kings	Rajasthan Royals	246
122	61	1	Chennai Super Kings	Kings XI Punjab	240
1134	562	1	Royal Challengers Bangalore	Mumbai Indians	235
596	296	1	Kings XI Punjab	Royal Challengers Bangalore	232
522	259	1	Delhi Daredevils	Kings XI Punjab	231
981	486	1	Kings XI Punjab	Chennai Super Kings	231
100	50	1	Kings XI Punjab	Mumbai Indians	230
1170	580	1	Royal Challengers Bangalore	Sunrisers Hyderabad	227

```
In [20]: sns.countplot(x='batting_team',data=a,palette="Set1")
plt.xticks(rotation=90)
plt.title('No. of times 200+ Scored')
```

Text(0.5, 1.0, 'No. of times 200+ Scored')



ROYAL CHALLENGERS BANGALORE HAVE SCORED 200+ 4 TIMES

```
In [21]: max_runs=df1.groupby(['batting_team','batsman_runs']).sum()
max_runs.sort_values(ascending=False)[:5].plot.bar(width=0.8,color='r')
```

Text(0.5, 1.0, 'Top 5 batsman with most runs')



SK RAINA HAS THE MOST RUNS IN TOTAL

```
In [22]: df1[df1['batsman_runs']==6].batsman.value_counts()[0:5].plot.bar(width=0.8,color='g')
```

Text(0.5, 1.0, 'MAXIMUM SIXEX BY TEAM')



CHRIS GAYLE HAS HIT THE MOST NUMBER OF SIXEX

```
In [23]: df1[df1['batsman_runs']==6].batting_team.value_counts().plot.bar(width=0.8,color='g')
```

Text(0.5, 1.0, 'MAXIMUM SIXEX BY TEAM')



RCB HAS HIT THE MOST NUMBER OF SIXEX

```
In [24]: top_scores = df1.groupby(['match_id','batsman','batting_team'])['batsman_runs'].sum().reset_index()
top_scores.sort_values('batsman_runs', ascending=0).head(10)
```

```
Out [24]:
```

	match_id	batsman	batting_team	batsman_runs
6200	411	CH Gayle	Royal Challengers Bangalore	175
900	60	BB McCullum	Kolkata Knight Riders	158
8426	562	AB de Villiers	Royal Challengers Bangalore	153
9257	620	AB de Villiers	Royal Challengers Bangalore	129
5585	372	CH Gayle	Royal Challengers Bangalore	128
3129	206	M Vijay	Chennai Super Kings	127
529	36	DA Warner	Sunrisers Hyderabad	126
7752	516	V Selhway	Kings XI Punjab	122
3686	242	PC Valthaty	Kings XI Punjab	120
4223	279	V Selhway	Delhi Daredevils	119

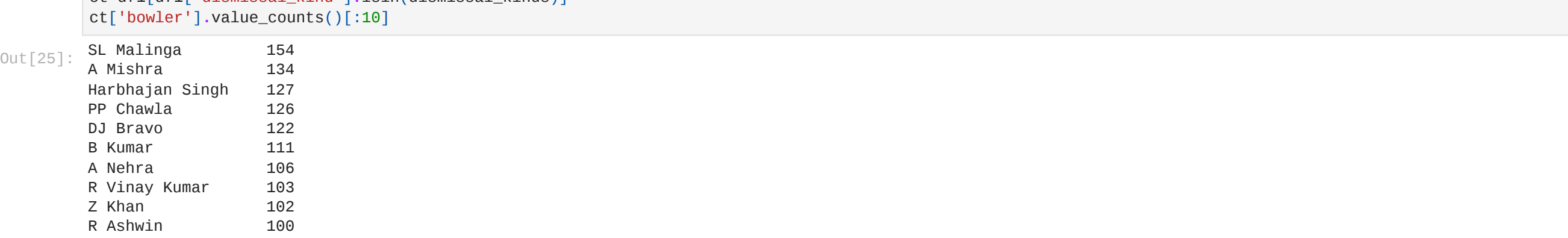
```
In [25]: dismissal_kinds = ["bowled", "caught", "lbw", "stumped", "caught and bowled", "hit wicket"]
ct=df1[df1["dismissal_kind"].isin(dismissal_kinds)]
ct['bowler'].value_counts()[0:10]
```

Out [25]:

```
SL Malinga      154
A Mishra       134
Harbhajan Singh 127
PP Chawla      126
DJ Bravo       122
B Kumar        111
A Nehra        106
R Vinay Kumar  103
Z Khan         102
P Ashwin       100
Name: bowler, dtype: int64
```

```
In [26]: dismissal_kinds = ["bowled", "caught", "lbw", "stumped", "caught and bowled", "hit wicket"]
ct=df1[df1["dismissal_kind"].isin(dismissal_kinds)]
ct['bowler'].value_counts()[0:10].plot.bar(width=0.8,color='b')
```

Text(0.5, 1.0, 'HIGHEST WICKET TAKER')



MALINGA HAS TAKEN THE MOST NUMBER OF WICKETS

```
In [ ]:
```