IPL Data Analysis

In [1]:

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

In [3]:

```
match_data = pd.read_csv()
ball_data = pd.read_csv()
```

In [4]:

```
match_data.head()
```

Out[4]:

	id	city	date	player_of_match	venue	neutral_venue	team1	
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	Royal Challengers Bangalore	
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	Kings XI Punjab	
2	335984	Delhi	2008-04-19	MF Maharoof	Feroz Shah Kotla	0	Delhi Daredevils	R
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0	Mumbai Indians	Cha B
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0	Kolkata Knight Riders	(

In [5]:

```
ball_data.head()
```

Out[5]:

	id	inning	over	ball	batsman	non_striker	bowler	batsman_runs	extra_runs	tota
0	335982	1	6	5	RT Ponting	BB McCullum	AA Noffke	1	0	
1	335982	1	6	6	BB McCullum	RT Ponting	AA Noffke	1	0	
2	335982	1	7	1	BB McCullum	RT Ponting	Z Khan	0	0	
3	335982	1	7	2	BB McCullum	RT Ponting	Z Khan	1	0	
4	335982	1	7	3	RT Ponting	BB McCullum	Z Khan	1	0	

In [6]:

```
match_data.isnull().sum()
```

Out[6]:

```
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
```

In [7]:

```
ball_data.isnull().sum()
```

Out[7]:

```
id                      0
inning                  0
over                    0
ball                    0
batsman                 0
non_striker             0
bowler                  0
batsman_runs            0
extra_runs              0
total_runs              0
non_boundary            0
is_wicket               0
dismissal_kind     183973
player_dismissed   183973
fielder            186684
extras_type        183235
batting_team            0
bowling_team          191
dtype: int64
```

In [9]:

```
ball_data.shape
```

Out[9]:

```
(193468, 18)
```

In [10]:

```
match_data.columns
```

Out[10]:

```
Index(['id', 'city', 'date', 'player_of_match', 'venue', 'neutral_venue',
       'team1', 'team2', 'toss_winner', 'toss_decision', 'winner', 'resul
t',
       'result_margin', 'eliminator', 'method', 'umpire1', 'umpire2'],
      dtype='object')
```

In [13]:

```
print('Matches played so far:', match_data.shape[0])
print('\n Cities played at:', match_data['city'].unique())
print('\n Teams participated:', match_data['team1'].unique())
```

```
Matches played so far: 816

 Cities played at: ['Bangalore' 'Chandigarh' 'Delhi' 'Mumbai' 'Kolkata' 'J
aipur' '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' nan 'Rajkot' 'Kanpur' 'Bengaluru' 'Dubai' 'Sharjah']

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

In [14]:

```
match_data['Season'] = pd.DatetimeIndex(match_data['date']).year
match_data.head()
```

Out[14]:

	id	city	date	player_of_match	venue	neutral_venue	team1	
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	Royal Challengers Bangalore	
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	Kings XI Punjab	
2	335984	Delhi	2008-04-19	MF Maharoof	Feroz Shah Kotla	0	Delhi Daredevils	R
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0	Mumbai Indians	Cha B
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0	Kolkata Knight Riders	(

In [16]:

```
match_per_season = match_data.groupby(['Season'])['id'].count().reset_index().rename(co
lumns={'id':'matches'})
match_per_season
```

Out[16]:

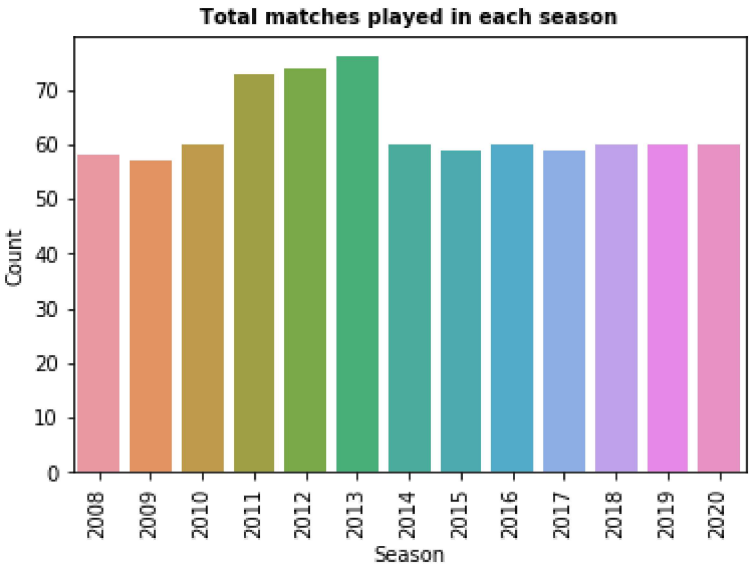
	Season	matches
0	2008	58
1	2009	57
2	2010	60
3	2011	73
4	2012	74
5	2013	76
6	2014	60
7	2015	59
8	2016	60
9	2017	59
10	2018	60
11	2019	60
12	2020	60

In [20]:

```
sns.countplot(match_data['Season'])
plt.xticks(rotation=90, fontsize=10)
plt.yticks(fontsize=10)
plt.xlabel('Season', fontsize=10)
plt.ylabel('Count', fontsize=10)
plt.title('Total matches played in each season', fontsize = 10, fontweight = "bold")
```

Out[20]:

```
Text(0.5, 1.0, 'Total matches played in each season')
```



In [21]:

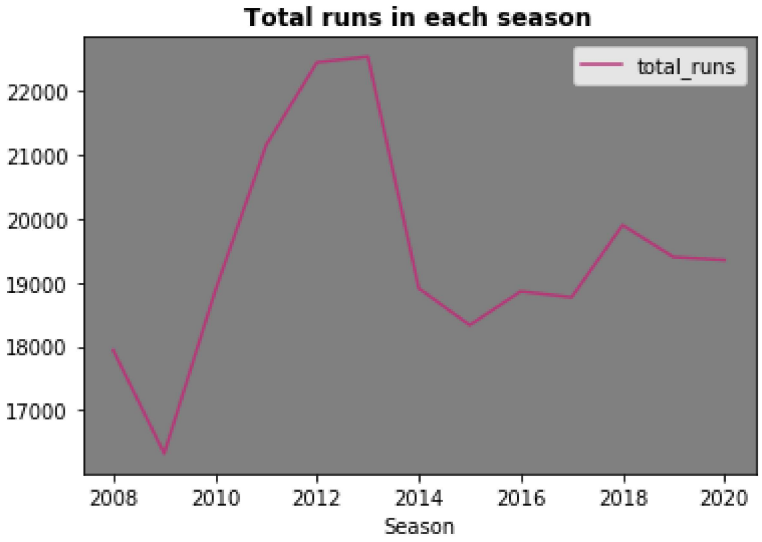
```
season_data=match_data[['id','Season']].merge(ball_data, left_on = 'id', right_on = 'id', how = 'left').drop('id', axis = 1)
season_data.head()
```

Out[21]:

	Season	inning	over	ball	batsman	non_striker	bowler	batsman_runs	extra_runs	tot
0	2008	1	6	5	RT Ponting	BB McCullum	AA Noffke	1	0	
1	2008	1	6	6	BB McCullum	RT Ponting	AA Noffke	1	0	
2	2008	1	7	1	BB McCullum	RT Ponting	Z Khan	0	0	
3	2008	1	7	2	BB McCullum	RT Ponting	Z Khan	1	0	
4	2008	1	7	3	RT Ponting	BB McCullum	Z Khan	1	0	

In [26]:

```
season=season_data.groupby(['Season'])['total_runs'].sum().reset_index()
p=season.set_index('Season')
ax = plt.axes()
ax.set(facecolor = "grey")
sns.lineplot(data=p,palette="magma")
plt.title('Total runs in each season',fontsize=12,fontweight="bold")
plt.show()
```



In [27]:

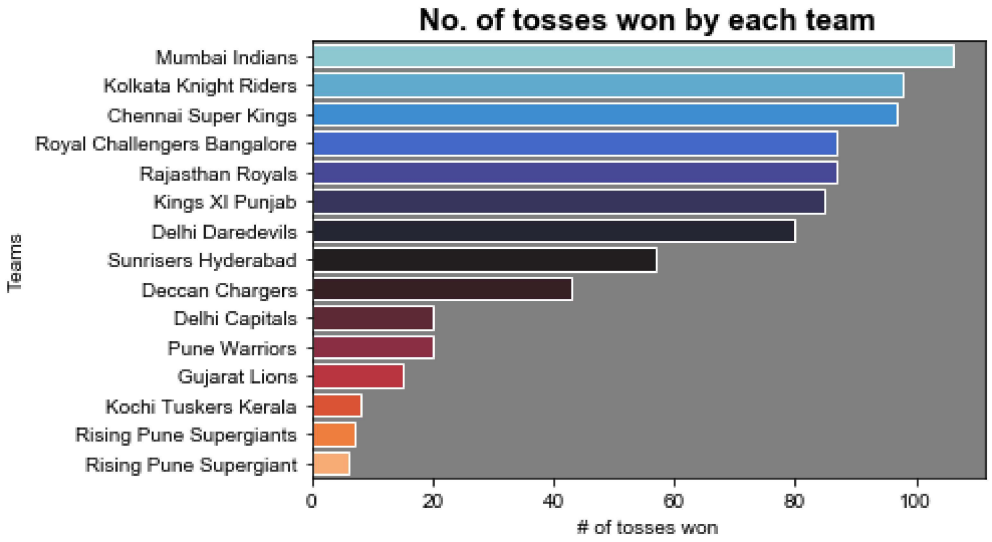
```
runs_per_season=pd.concat([match_per_season,season.iloc[:,1]],axis=1)
runs_per_season['Runs scored per match']=runs_per_season['total_runs']/runs_per_season[
'matches']
runs_per_season.set_index('Season',inplace=True)
runs_per_season
```

Out[27]:

Season	matches	total_runs	Runs scored per match
2008	58	17937	309.258621
2009	57	16320	286.315789
2010	60	18864	314.400000
2011	73	21154	289.780822
2012	74	22453	303.418919
2013	76	22541	296.592105
2014	60	18909	315.150000
2015	59	18332	310.711864
2016	60	18862	314.366667
2017	59	18769	318.118644
2018	60	19901	331.683333
2019	60	19400	323.333333
2020	60	19352	322.533333

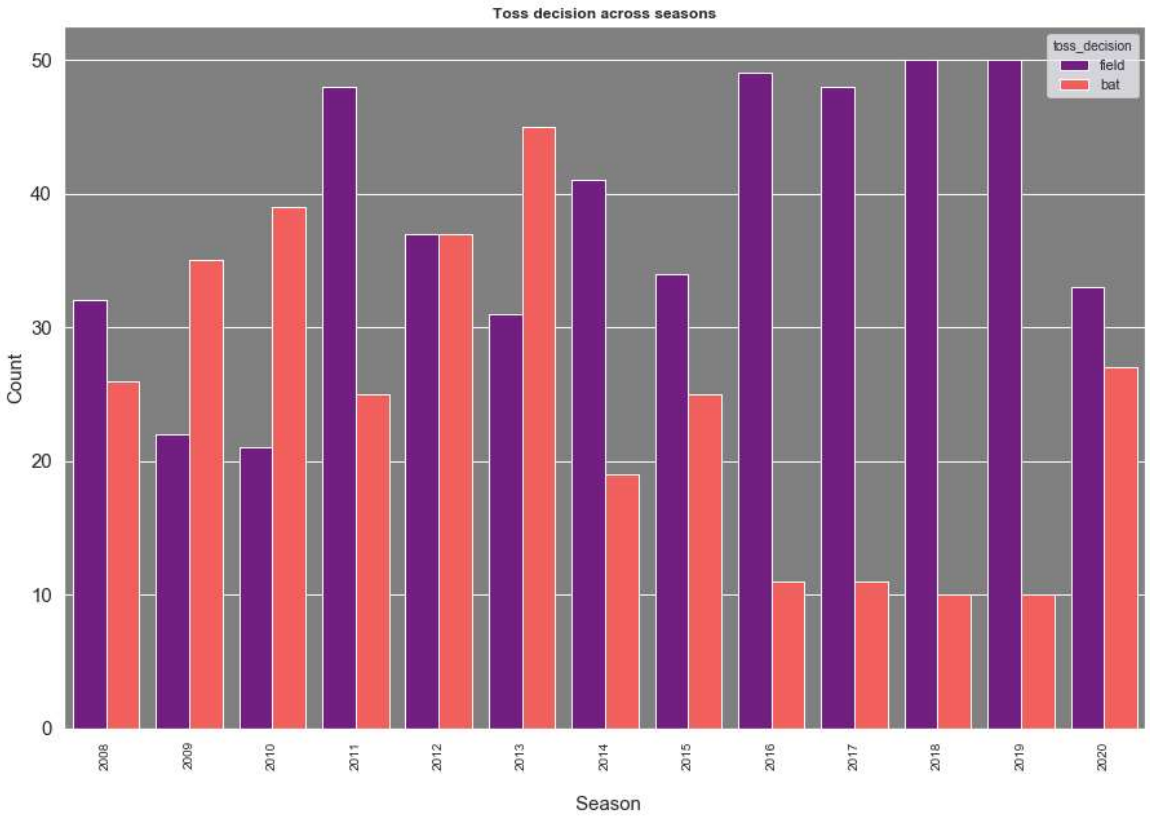
In [28]:

```
toss=match_data['toss_winner'].value_counts()
ax = plt.axes()
ax.set(facecolor = "grey")
sns.set(rc={'figure.figsize':(15,10)},style='darkgrid')
ax.set_title('No. of tosses won by each team',fontsize=15,fontweight="bold")
sns.barplot(y=toss.index, x=toss, orient='h',palette="icefire",saturation=1)
plt.xlabel('# of tosses won')
plt.ylabel('Teams')
plt.show()
```



In [29]:

```
ax = plt.axes()
ax.set(facecolor = "grey")
sns.countplot(x='Season', hue='toss_decision', data=match_data,palette="magma",saturati
on=1)
plt.xticks(rotation=90,fontsize=10)
plt.yticks(fontsize=15)
plt.xlabel('\n Season',fontsize=15)
plt.ylabel('Count',fontsize=15)
plt.title('Toss decision across seasons',fontsize=12,fontweight="bold")
plt.show()
```



In [31]:

```
match_data['result'].value_counts()
```

Out[31]:

```
wickets    435
runs       364
tie         13
Name: result, dtype: int64
```

In [32]:

```
match_data.venue[match_data.result!='runs'].mode()
```

Out[32]:

```
0    Eden Gardens
dtype: object
```

In [33]:

```
match_data.venue[match_data.result!='wickets'].mode()
```

Out[33]:

```
0    Feroz Shah Kotla
dtype: object
```

In [36]:

```
match_data.venue[match_data.toss_winner=='Kolkata Knight Riders'][match_data.winner=='K
olkata Knight Riders'].mode()
```

Out[36]:

```
0    Eden Gardens
dtype: object
```

In [37]:

```
match_data.winner[match_data.result!='runs'].mode()
```

Out[37]:

```
0    Kolkata Knight Riders
1           Mumbai Indians
dtype: object
```

In [38]:

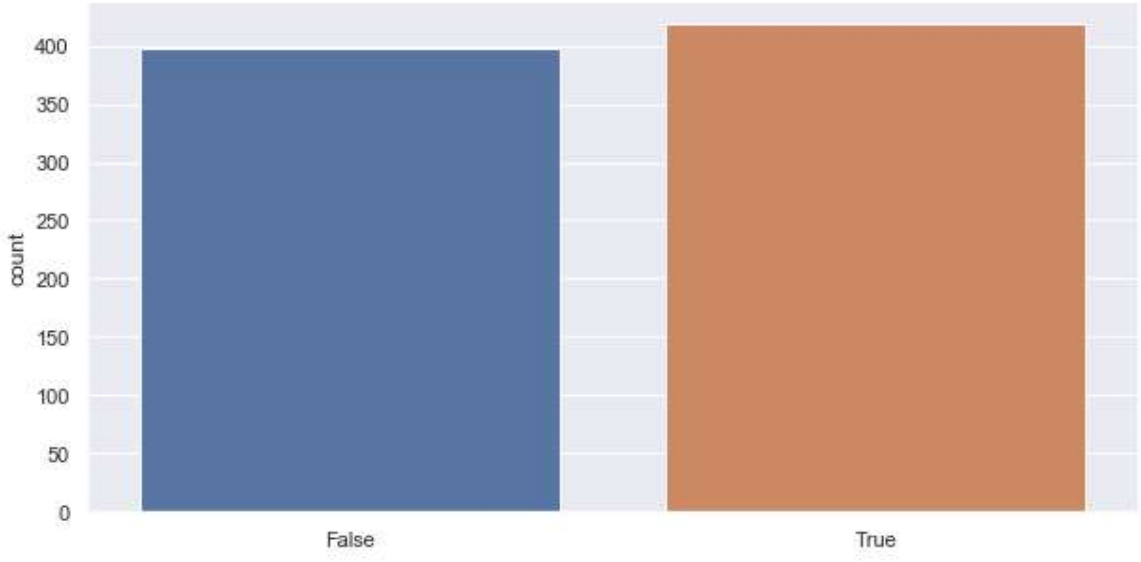
```
match_data.winner[match_data.result!='wickets'].mode()
```

Out[38]:

```
0    Mumbai Indians
dtype: object
```

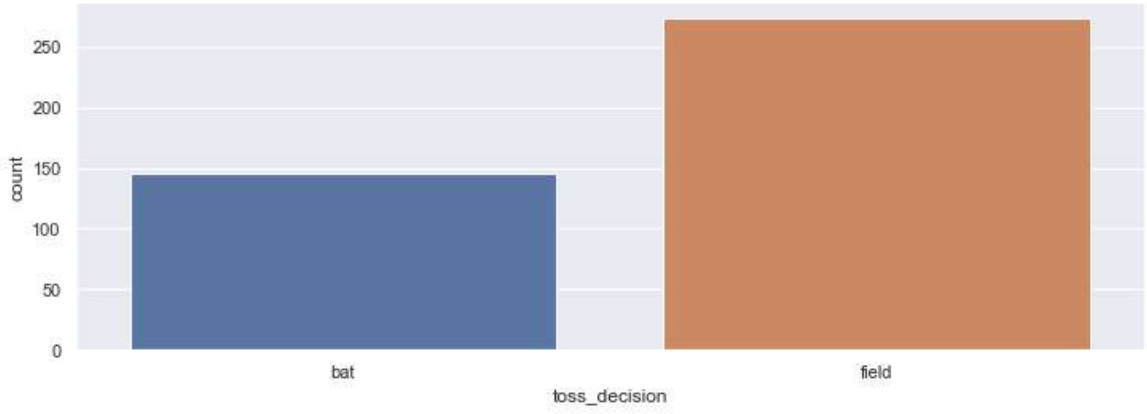
In [39]:

```
toss = match_data['toss_winner'] == match_data['winner']
plt.figure(figsize=(10,5))
sns.countplot(toss)
plt.show()
```



In [40]:

```
plt.figure(figsize=(12,4))
sns.countplot(match_data.toss_decision[match_data.toss_winner == match_data.winner])
plt.show()
```



In [43]:

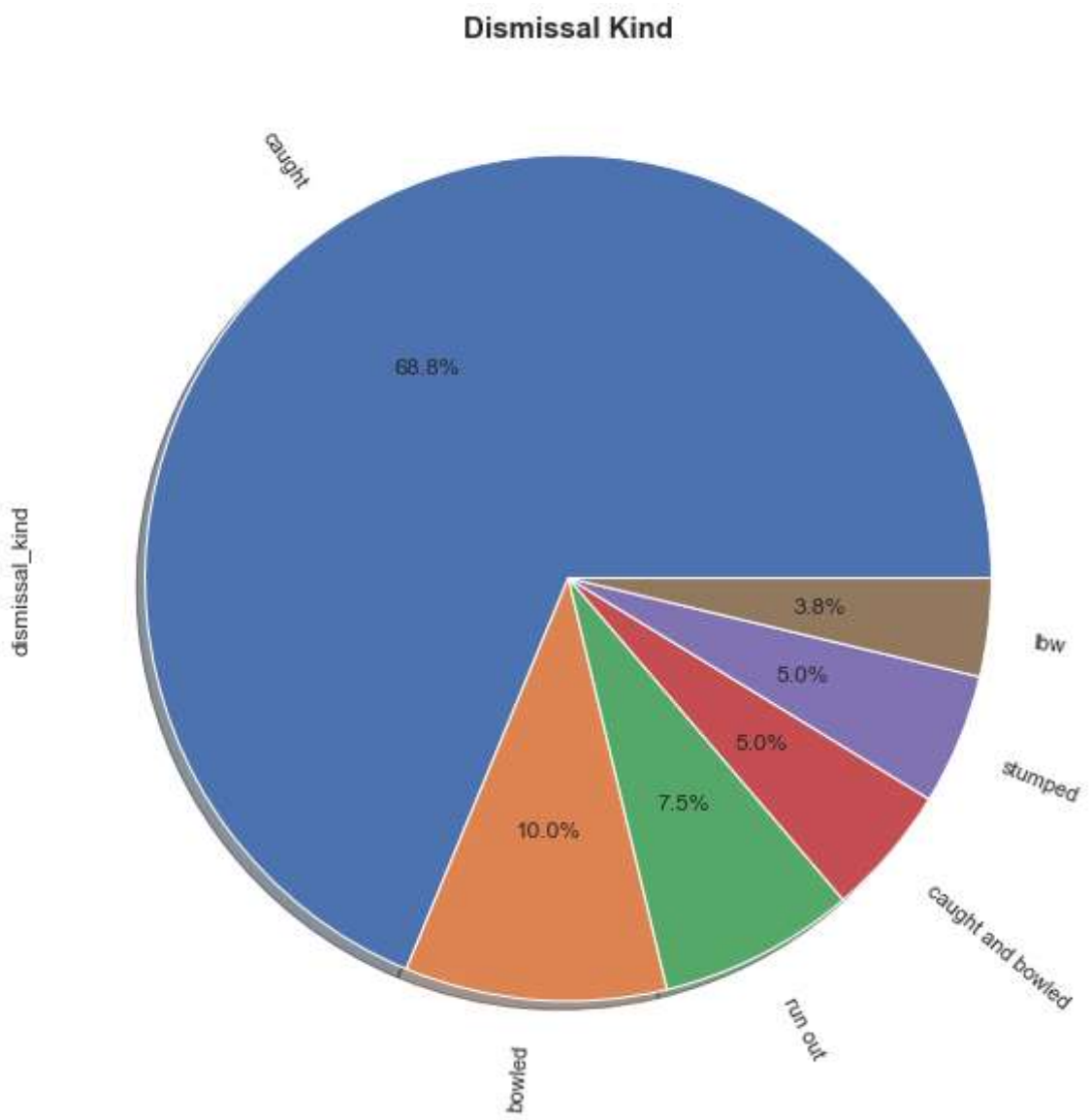
```
player = (ball_data['batsman']=='SK Raina')
df_raina=ball_data[player]
df_raina.head()
```

Out[43]:

	id	inning	over	ball	batsman	non_striker	bowler	batsman_runs	extra_runs	to
246	335983	1	10	3	SK Raina	MEK Hussey	PP Chawla	2	0	
247	335983	1	10	4	SK Raina	MEK Hussey	PP Chawla	0	0	
248	335983	1	10	5	SK Raina	MEK Hussey	PP Chawla	6	0	
249	335983	1	10	6	SK Raina	MEK Hussey	PP Chawla	4	0	
253	335983	1	11	4	SK Raina	MEK Hussey	K Goel	6	0	

In [44]:

```
df_raina['dismissal_kind'].value_counts().plot.pie(autopct='%1.1f%%',shadow=True,rotate
labels=True)
plt.title("Dismissal Kind",fontweight="bold",fontsize=15)
plt.show()
```



In [45]:

```
def count(df_raina,runs):
    return len(df_raina[df_raina['batsman_runs']==runs])*runs
```

In [46]:

```
print("Runs scored from 1's :",count(df_raina,1))
print("Runs scored from 2's :",count(df_raina,2))
print("Runs scored from 3's :",count(df_raina,3))
print("Runs scored from 4's :",count(df_raina,4))
print("Runs scored from 6's :",count(df_raina,6))
```

```
Runs scored from 1's : 1666
Runs scored from 2's : 528
Runs scored from 3's : 33
Runs scored from 4's : 1972
Runs scored from 6's : 1164
```

In [47]:

```
match_data[match_data['result_margin']==match_data['result_margin'].max()]
```

Out[47]:

	id	city	date	player_of_match	venue	neutral_venue	team1	team2	toss
620	1082635	Delhi	2017-05-06	LMP Simmons	Feroz Shah Kotla	0	Delhi Daredevils	Mumbai Indians	Da

In [49]:

```
runs = ball_data.groupby(['batsman'])['batsman_runs'].sum().reset_index()
runs.columns = ['Batsman', 'runs']
y = runs.sort_values(by='runs', ascending = False).head(10).reset_index().drop('index',
axis=1)
y
```

Out[49]:

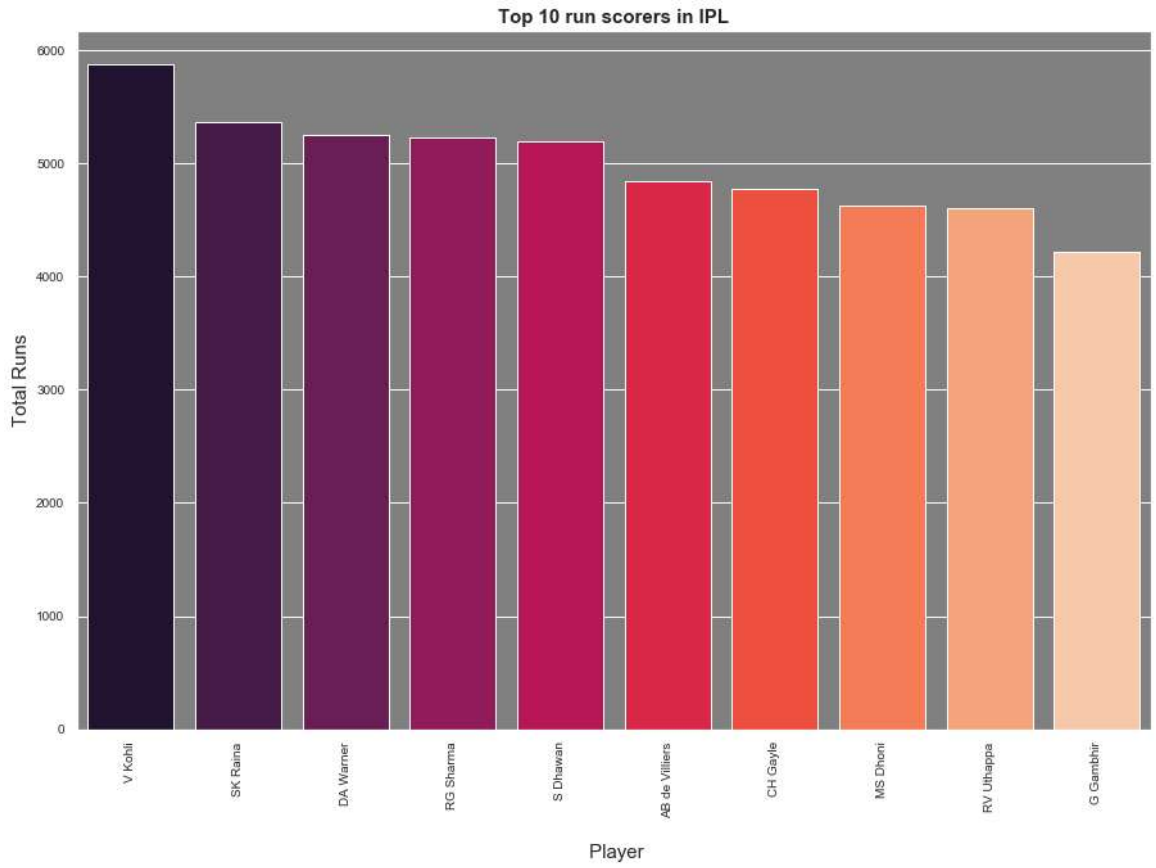
	Batsman	runs
0	V Kohli	5878
1	SK Raina	5368
2	DA Warner	5254
3	RG Sharma	5230
4	S Dhawan	5197
5	AB de Villiers	4849
6	CH Gayle	4772
7	MS Dhoni	4632
8	RV Uthappa	4607
9	G Gambhir	4217

In [50]:

```
ax = plt.axes()
ax.set(facecolor = "grey")
sns.barplot(x=y['Batsman'],y=y['runs'],palette='rocket',saturation=1)
plt.xticks(rotation=90,fontsize=10)
plt.yticks(fontsize=10)
plt.xlabel('\n Player',fontsize=15)
plt.ylabel('Total Runs',fontsize=15)
plt.title('Top 10 run scorers in IPL',fontsize=15,fontweight="bold")
```

Out[50]:

```
Text(0.5, 1.0, 'Top 10 run scorers in IPL')
```

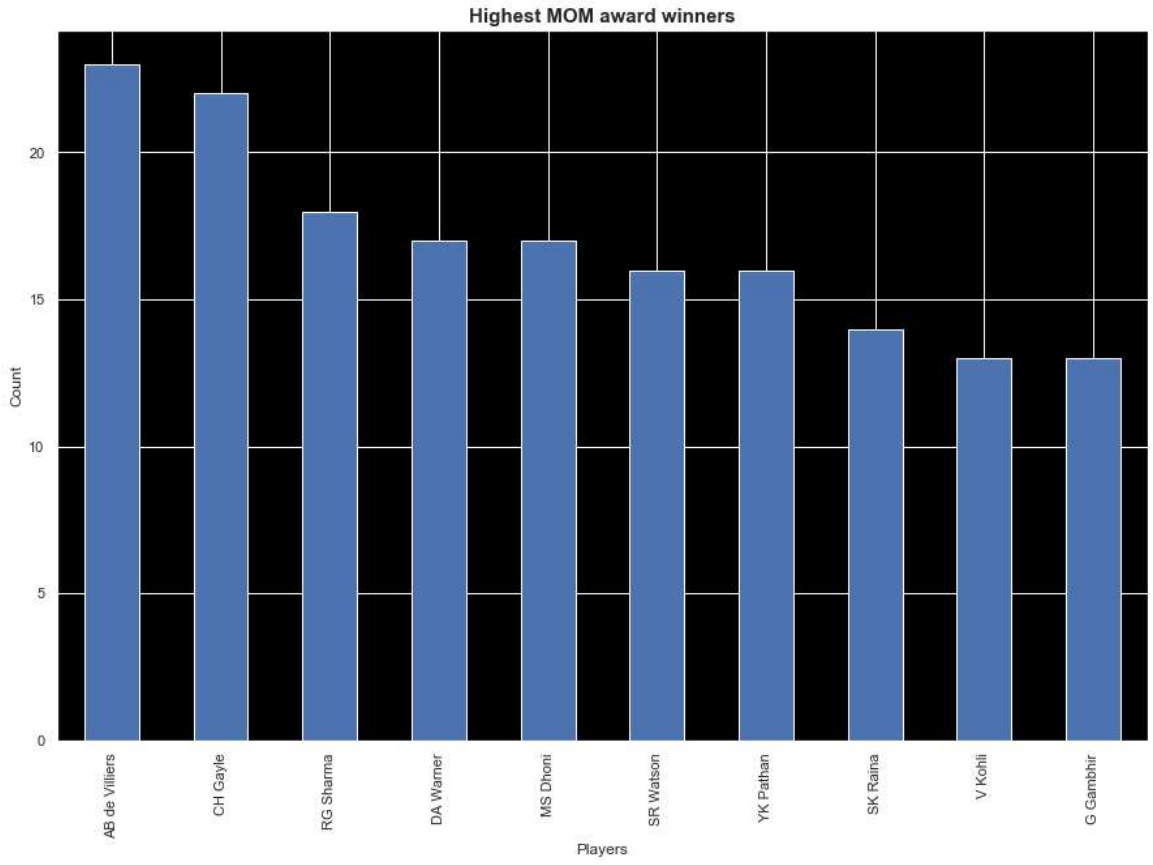


In [51]:

```
ax = plt.axes()
ax.set(facecolor = "black")
match_data.player_of_match.value_counts()[:10].plot(kind='bar')
plt.xlabel('Players')
plt.ylabel("Count")
plt.title("Highest MOM award winners",fontsize=15,fontweight="bold")
```

Out[51]:

```
Text(0.5, 1.0, 'Highest MOM award winners')
```



In [ ]: