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

In [2]:
```
match_data= pd.read_csv("C:\\Users\\AJAY\\Desktop\\Project\\IPL Matches 2008-2020.csv")
match_data
```

Out[2]:

	id	city	date	player_of_match	venue	neutral_venue	team1	team2	toss_winner	toss_decision	winn
0	335982	Bangalore	18-04-2008	BB McCullum	M Chinnaswamy Stadium	0	Royal Challengers Bangalore	Kolkata Knight Riders	Royal Challengers Bangalore	field	Kolka Knig Ride
1	335983	Chandigarh	19-04-2008	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	Kings XI Punjab	Chennai Super Kings	Chennai Super Kings	bat	Chenn Sup Kin
2	335984	Delhi	19-04-2008	MF Maharoof	Feroz Shah Kotla	0	Delhi Daredevils	Rajasthan Royals	Rajasthan Royals	bat	De Daredev
3	335985	Mumbai	20-04-2008	MV Boucher	Wankhede Stadium	0	Mumbai Indians	Royal Challengers Bangalore	Mumbai Indians	bat	Roy Challenge Bangalo
4	335986	Kolkata	20-04-2008	DJ Hussey	Eden Gardens	0	Kolkata Knight Riders	Deccan Chargers	Deccan Chargers	bat	Kolka Knig Ride
...	...	...	...	...	...	...	...	...	...	...	
811	1216547	Dubai	28-09-2020	AB de Villiers	Dubai International Cricket Stadium	0	Royal Challengers Bangalore	Mumbai Indians	Mumbai Indians	field	Roy Challenge Bangalo
812	1237177	Dubai	05-11-2020	JJ Bumrah	Dubai International Cricket Stadium	0	Mumbai Indians	Delhi Capitals	Delhi Capitals	field	Mumb India
813	1237178	Abu Dhabi	06-11-2020	KS Williamson	Sheikh Zayed Stadium	0	Royal Challengers Bangalore	Sunrisers Hyderabad	Sunrisers Hyderabad	field	Sunrise Hyderab
814	1237180	Abu Dhabi	08-11-2020	MP Stoinis	Sheikh Zayed Stadium	0	Delhi Capitals	Sunrisers Hyderabad	Delhi Capitals	bat	De Capit
815	1237181	Dubai	10-11-2020	TA Boult	Dubai International Cricket Stadium	0	Delhi Capitals	Mumbai Indians	Delhi Capitals	bat	Mumb India

816 rows × 17 columns

In [3]:
```
#head function gives top 5 rows
match_data.head()
```

Out[3]:

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

In [4]:
```
ball_data=pd.read_csv("C:\\Users\\AJAY\\Desktop\\Project\\IPL Ball-by-Ball 2008-2020.csv")
```

In [5]:
```
ball_data
```

Out[5]:

	id	inning	over	ball	batsman	non_striker	bowler	batsman_runs	extra_runs	total_runs	non_boundary	is_wicket	dismiss
0	335982	1	6	5	RT Ponting	BB McCullum	AA Noffke	1	0	1	0	0	
1	335982	1	6	6	BB McCullum	RT Ponting	AA Noffke	1	0	1	0	0	
2	335982	1	7	1	BB McCullum	RT Ponting	Z Khan	0	0	0	0	0	
3	335982	1	7	2	BB McCullum	RT Ponting	Z Khan	1	0	1	0	0	
4	335982	1	7	3	RT Ponting	BB McCullum	Z Khan	1	0	1	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
193463	1237181	1	12	5	RR Pant	SS Iyer	NM Coulter-Nile	0	0	0	0	0	
193464	1237181	1	12	6	RR Pant	SS Iyer	NM Coulter-Nile	1	0	1	0	0	
193465	1237181	1	13	1	RR Pant	SS Iyer	KH Pandya	0	1	1	0	0	
193466	1237181	1	13	2	RR Pant	SS Iyer	KH Pandya	1	0	1	0	0	
193467	1237181	1	13	3	SS Iyer	RR Pant	KH Pandya	1	0	1	0	0	

193468 rows × 18 columns

In [6]:
```
ball_data.head()
```

Out[6]:

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

In [7]:
```
#counts number of null values in each column
match_data.isnull().sum()
```

Out[7]:
```
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 [8]:
```
ball_data.isnull().sum()
```

Out[8]:
```
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]:
```
match_data.shape
```

Out[9]:
```
(816, 17)
```

In [10]:
```
ball_data.shape
```

Out[10]:
```
(193468, 18)
```

In [11]:
```
match_data.columns
```

Out[11]:
```
Index(['id', 'city', 'date', 'player_of_match', 'venue', 'neutral_venue',
       'team1', 'team2', 'toss_winner', 'toss_decision', 'winner', 'result',
       'result_margin', 'eliminator', 'method', 'umpire1', 'umpire2'],
      dtype='object')
```

In [12]:
```
print("Matches played so  far:",  match_data.shape[0])
print("Venues played at:", len(match_data['city'].unique()))
print('Teams participated thusfar:', match_data['team1'].unique())
```

```
Matches played so far: 816
Venues played at: 33
Teams participated thusfar: ['Royal Challengers Bangalore' 'Kings XI Punjab' 'Delhi 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]:
```
#Add a new column containing season

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

Out[14]:

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

In [ ]:

```
#Finding out the number of matches played per season

matches_per_season= match_data.groupby(['Season'])['id'].count().reset_index().rename(columns={'id':'matches'})
matches_per_season
```

barplot using seaborn

sns.countplot(match_data['Season']) plt.xticks(rotation=45, fontsize=10) plt.yticks(fontsize=10) plt.xlabel('Season', fontsize=10) plt.ylabel('Matches', fontsize=10) plt.title('Total Matches played', fontsize=20, fontweight='bold')

In [15]:

```
#Merge two tables using left join

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

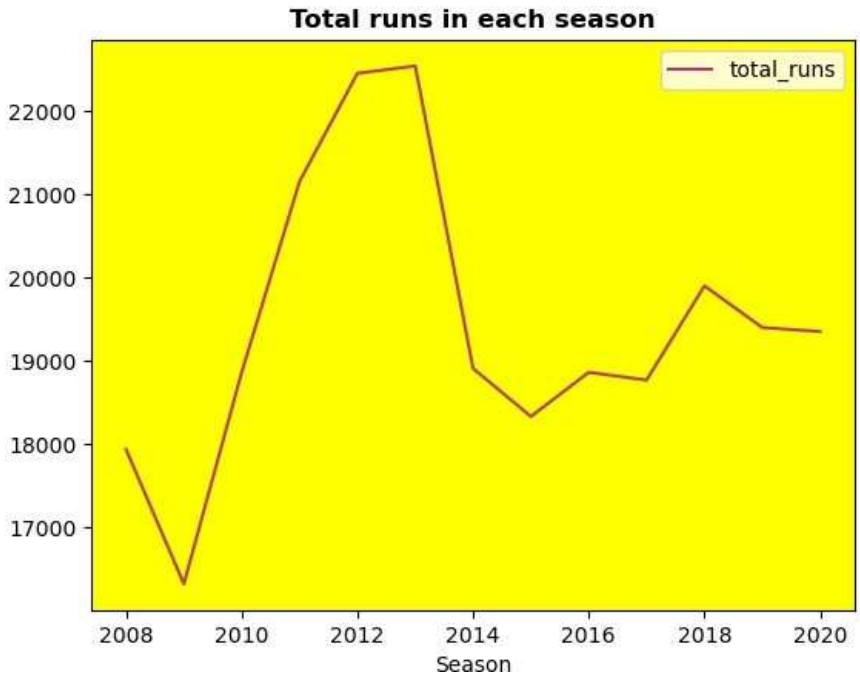
Out[15]:

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

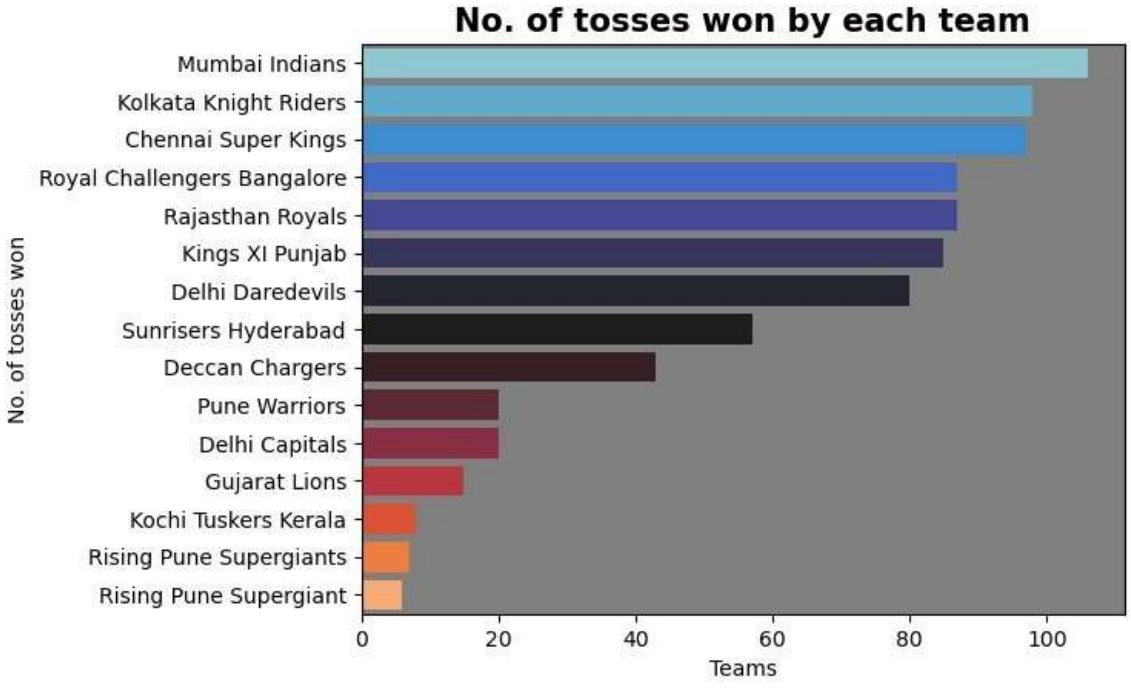
In [16]:

```
#Finding total runs scored in each season

season= season_data.groupby(['Season'])['total_runs'].sum().reset_index()
p=season.set_index('Season')

ax=plt.axes()
ax.set(facecolor= 'yellow')
```

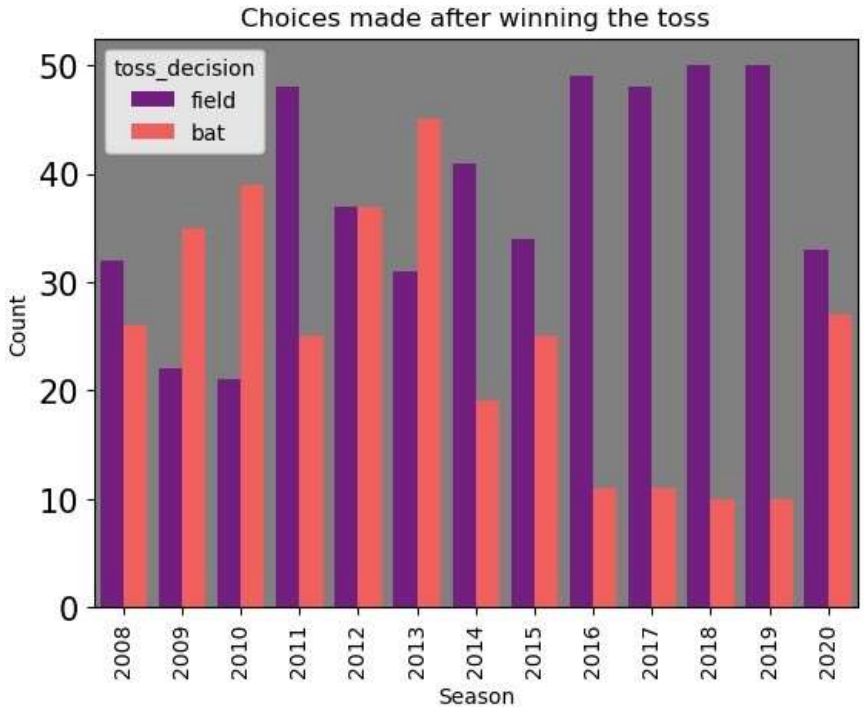
```
sns.lineplot(data=p, palette='magma')
plt.title('Total runs in each season', fontsize=12, fontweight='bold')
plt.show()
```



In [17]:
```
 toss=match_data['toss_winner'].value_counts()
ax=plt.axes()
ax.set(facecolor= 'grey')
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.ylabel('No. of tosses won')
plt.xlabel('Teams')
plt.show()
```



In [18]:
```
ax= plt.axes()
ax.set(facecolor= 'grey')
sns.countplot(x='Season', hue='toss_decision', data=match_data, palette='magma', saturation= 1)
plt.xticks(rotation=90, fontsize=10)
plt.yticks(fontsize=15)
plt.ylabel('Count')
plt.xlabel('Season')
plt.title('Choices made after winning the toss')
plt.show()
```



In [19]:
```
match_data['result'].value_counts()
```

Out[19]:
```
wickets    435
runs       364
tie         13
Name: result, dtype: int64
```

In [20]:
```
match_data.venue[match_data.result!='runs'].mode()
```

Out[20]:
```
0    Eden Gardens
Name: venue, dtype: object
```

In [21]:
```
match_data.venue[match_data.result!='wickets'].mode()
```

Out[21]:
```
0    Feroz Shah Kotla
Name: venue, dtype: object
```

In [22]:
```
match_data.venue[match_data.toss_winner=='Mumbai Indians'][match_data.winner=='Mumbai Indians'].mode()
```

Out[22]:
```
0    Wankhede Stadium
Name: venue, dtype: object
```

In [23]:
```
match_data.winner[match_data.result!='runs'].mode()
```

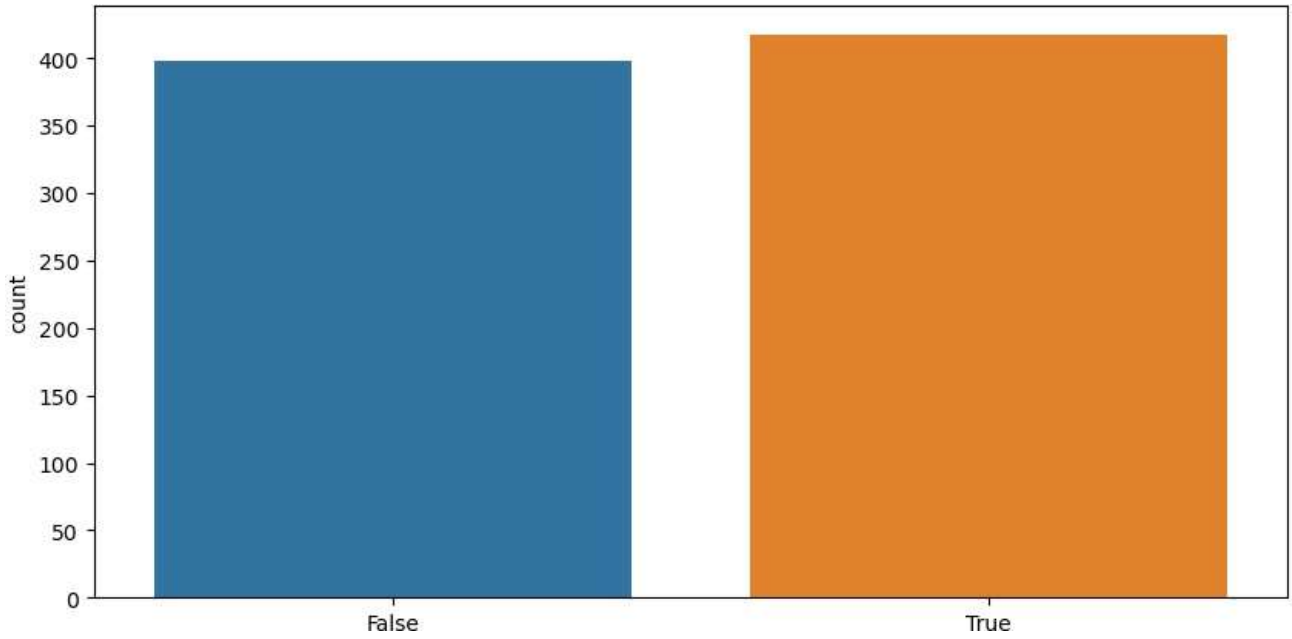
Out[23]:
```
0    Kolkata Knight Riders
1           Mumbai Indians
Name: winner, dtype: object
```

In [24]:
```
match_data.winner[match_data.result!='wickets'].mode()
```

Out[24]:
```
0    Mumbai Indians
Name: winner, dtype: object
```

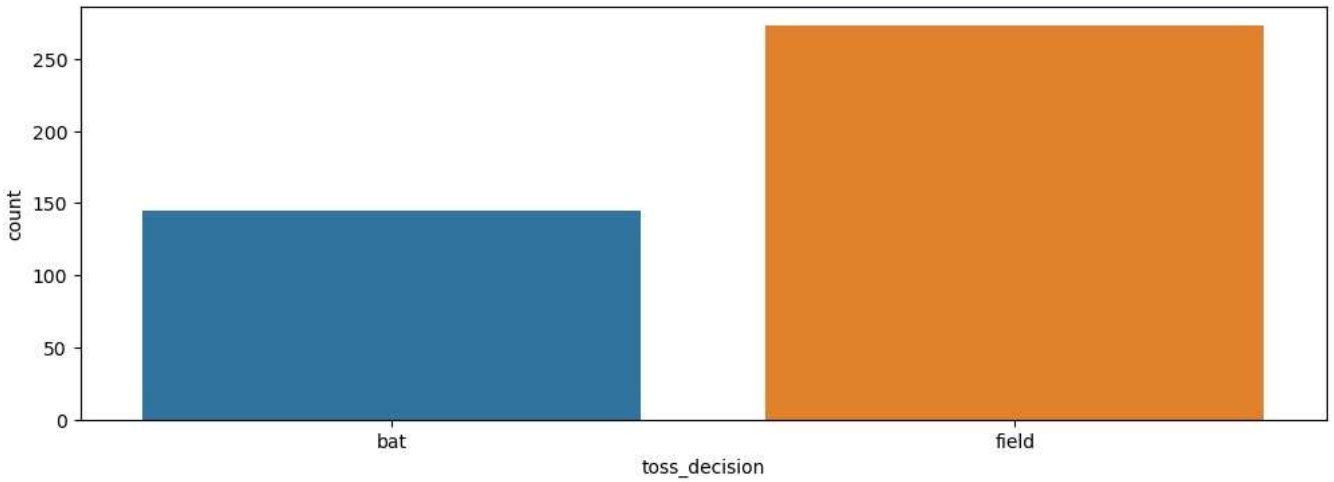
In [25]:
```
toss= match_data['toss_winner']== match_data['winner']
plt.figure(figsize=(10,5))
sns.countplot(toss)
plt.show()
```

```
C:\Users\AJAY\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```



In [26]:
```
plt.figure(figsize=(12,4))
sns.countplot(match_data.toss_decision[match_data.toss_winner== match_data.winner])
plt.show()
```

```
C:\Users\AJAY\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable
as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other ar
guments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```



In [27]:
```
player= (ball_data['batsman']=='SK Raina')
df_raina= ball_data[player]
df_raina
```

Out[27]:

	id	inning	over	ball	batsman	non_striker	bowler	batsman_runs	extra_runs	total_runs	non_boundary	is_wicket	dis
246	335983	1	10	3	SK Raina	MEK Hussey	PP Chawla	2	0	2	0	0	
247	335983	1	10	4	SK Raina	MEK Hussey	PP Chawla	0	0	0	0	0	
248	335983	1	10	5	SK Raina	MEK Hussey	PP Chawla	6	0	6	0	0	
249	335983	1	10	6	SK Raina	MEK Hussey	PP Chawla	4	0	4	0	0	
253	335983	1	11	4	SK Raina	MEK Hussey	K Goel	6	0	6	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
178795	1181768	2	7	5	SK Raina	SR Watson	RD Chahar	0	0	0	0	0	
178796	1181768	2	7	6	SK Raina	SR Watson	RD Chahar	1	0	1	0	0	
178797	1181768	2	8	1	SK Raina	SR Watson	MJ McClenaghan	1	0	1	0	0	
178886	1181768	2	9	1	SK Raina	SR Watson	RD Chahar	0	0	0	0	0	
178887	1181768	2	9	2	SK Raina	SR Watson	RD Chahar	0	0	0	0	1	

4041 rows × 18 columns

In [28]:
```
df_raina['dismissal_kind'].value_counts()
```

Out[28]:
```
caught                110
bowled                 16
run out                12
caught and bowled       8
stumped                 8
lbw                     6
Name: dismissal_kind, dtype: int64
```

In [33]:
```
def count(df_raina, runs):
    return len(df_raina[df_raina['batsman_runs']== runs]) * runs
```

In [32]:
```
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 [34]:
```
match_data[match_data['result_margin']== match_data['result_margin'].max()]
```

Out[34]:

	id	city	date	player_of_match	venue	neutral_venue	team1	team2	toss_winner	toss_decision	winner	result	result_ma
620	1082635	Delhi	06-05-2017	LMP Simmons	Feroz Shah Kotla	0	Delhi Daredevils	Mumbai Indians	Delhi Daredevils	field	Mumbai Indians	runs	1

In [35]:
```
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[35]:

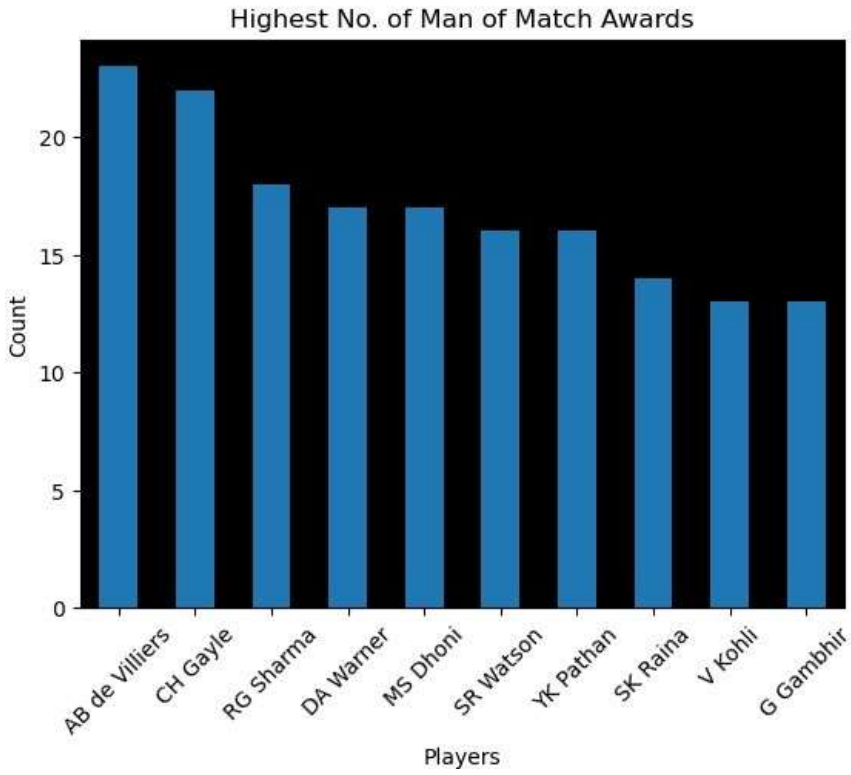
	Batsman	runs
0	V Kohli	5878
1	SK Raina	5368
2	DA Warner	5254
3	RG Sharma	5230
4	S Dhawan	5197
5	AB de Villiers	4849
6	CH Gayle	4772
7	MS Dhoni	4632
8	RV Uthappa	4607
9	G Gambhir	4217

In [36]:

```
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.xticks(rotation= 45)
plt.title("Highest No. of Man of Match Awards")
```

Out[36]:

```
Text(0.5, 1.0, 'Highest No. of Man of Match Awards')
```



In [ ]: