In [2]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [7]:
```
matches = pd.read_csv("desktop/IPL Matches.csv")
scores = pd.read_csv("desktop/IPL.csv")
matches.head()
```

Out[7]:

	id	city	date	player_of_match	venue	neutral_venue	team1	team2
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	Royal Challengers Bangalore	Kolkat Knigh Rider
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	Kings XI Punjab	Chenna Supe King
2	335984	Delhi	2008-04-19	MF Maharoof	Feroz Shah Kotla	0	Delhi Daredevils	Rajasthan Royal
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0	Mumbai Indians	Roya Challenger Bangalor
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0	Kolkata Knight Riders	Decca Charger

In [8]:
```
scores.head()
```

Out[8]:

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

In [9]:
```
print(matches.info())
print(scores.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
None
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 193468 entries, 0 to 193467
Data columns (total 18 columns):
 #   Column            Non-Null Count   Dtype 
---  ------            --------------   ----- 
 0   id                193468 non-null  int64 
 1   inning            193468 non-null  int64 
 2   over              193468 non-null  int64 
 3   ball              193468 non-null  int64 
 4   batsman           193468 non-null  object
 5   non_striker       193468 non-null  object
 6   bowler            193468 non-null  object
 7   batsman_runs      193468 non-null  int64 
 8   extra_runs        193468 non-null  int64 
 9   total_runs        193468 non-null  int64 
 10  non_boundary      193468 non-null  int64 
 11  is_wicket         193468 non-null  int64 
 12  dismissal_kind    9495 non-null    object
 13  player_dismissed  9495 non-null    object
 14  fielder           6784 non-null    object
 15  extras_type       10233 non-null   object
 16  batting_team      193468 non-null  object
 17  bowling_team      193277 non-null  object
dtypes: int64(9), object(9)
memory usage: 26.6+ MB
None
```

In [14]:
```
matches.columns
```

Out[14]:
```
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 [15]:
```
scores.columns
```

Out[15]:
```
Index(['id', 'inning', 'over', 'ball', 'batsman', 'non_striker', 'bowler',
       'batsman_runs', 'extra_runs', 'total_runs', 'non_boundary', 'is_wicket',
       'dismissal_kind', 'player_dismissed', 'fielder', 'extras_type',
       'batting_team', 'bowling_team'],
      dtype='object')
```

In [16]:
```
matches.shape
```

Out[16]:
```
(816, 17)
```

In [17]:
```
scores.shape
```

Out[17]:
```
(193468, 18)
```

In [32]:
```
scores.tail()
```

Out[32]:

	id	inning	over	ball	batsman	non_striker	bowler	batsman_runs	extra_runs	tota
193463	1237181	1	12	5	RR Pant	SS Iyer	NM Coulter-Nile	0	0	
193464	1237181	1	12	6	RR Pant	SS Iyer	NM Coulter-Nile	1	0	
193465	1237181	1	13	1	RR Pant	SS Iyer	KH Pandya	0	1	
193466	1237181	1	13	2	RR Pant	SS Iyer	KH Pandya	1	0	
193467	1237181	1	13	3	SS Iyer	RR Pant	KH Pandya	1	0	

In [21]:
```
scores.head()
```

Out[21]:

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

In [22]:
```
scores["fielder"].isnull().sum()
```

Out[22]:
```
186684
```

In [23]:
```
matches["winner"][3]
```

Out[23]:
```
'Royal Challengers Bangalore'
```

In [25]:
```
scores.describe()
```

Out[25]:

	id	inning	over	ball	batsman_runs	extra_runs
count	1.934680e+05	193468.000000	193468.000000	193468.000000	193468.000000	193468.000000
mean	7.567688e+05	1.482131	9.177027	3.615967	1.240231	0.066414
std	3.060971e+05	0.499682	5.676848	1.807128	1.610867	0.339991
min	3.359820e+05	1.000000	0.000000	1.000000	0.000000	0.000000
25%	5.012270e+05	1.000000	4.000000	2.000000	0.000000	0.000000
50%	7.292970e+05	1.000000	9.000000	4.000000	1.000000	0.000000
75%	1.082628e+06	2.000000	14.000000	5.000000	1.000000	0.000000
max	1.237181e+06	2.000000	19.000000	9.000000	6.000000	7.000000

In [26]:
```
matches['id'].max()
```

Out[26]:
```
1237181
```

```
In [30]: matches['city'].unique()
```

```
Out[30]: 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', nan, 'Rajkot', 'Kanpur', 'Bengaluru', 'Dubai',
                'Sharjah'], dtype=object)
```

```
In [ ]:
```

```
In [35]: scores.iloc[scores['total_runs'].idxmax()]
```

```
Out[35]: id                             335988
         inning                              1
         over                               19
         ball                                6
         batsman                     WPUJC Vaas
         non_striker                   RP Singh
         bowler                       VY Mahesh
         batsman_runs                        6
         extra_runs                          1
         total_runs                          7
         non_boundary                        0
         is_wicket                           0
         dismissal_kind                    NaN
         player_dismissed                  NaN
         fielder                           NaN
         extras_type                   noballs
         batting_team           Deccan Chargers
         bowling_team          Delhi Daredevils
         Name: 1497, dtype: object
```

team won by maximum and minimum result

```
In [39]: matches.iloc[matches['result_margin'].idxmax()]['winner']
```

```
Out[39]: 'Mumbai Indians'
```

```
In [42]: matches.iloc[matches['result_margin'].idxmin()]['winner']
```

```
Out[42]: 'Kings XI Punjab'
```

```
In [44]: matches.head(1)
```

Out[44]:

	id	city	date	player_of_match	venue	neutral_venue	team1	team2	to
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	Royal Challengers Bangalore	Kolkata Knight Riders	C

In [45]:

```
scores.head(1)
```

Out[45]:

	id	inning	over	ball	batsman	non_striker	bowler	batsman_runs	extra_runs	total_runs
0	335982	1	6	5	RT Ponting	BB McCullum	AA Noffke	1	0	1

In [ ]:

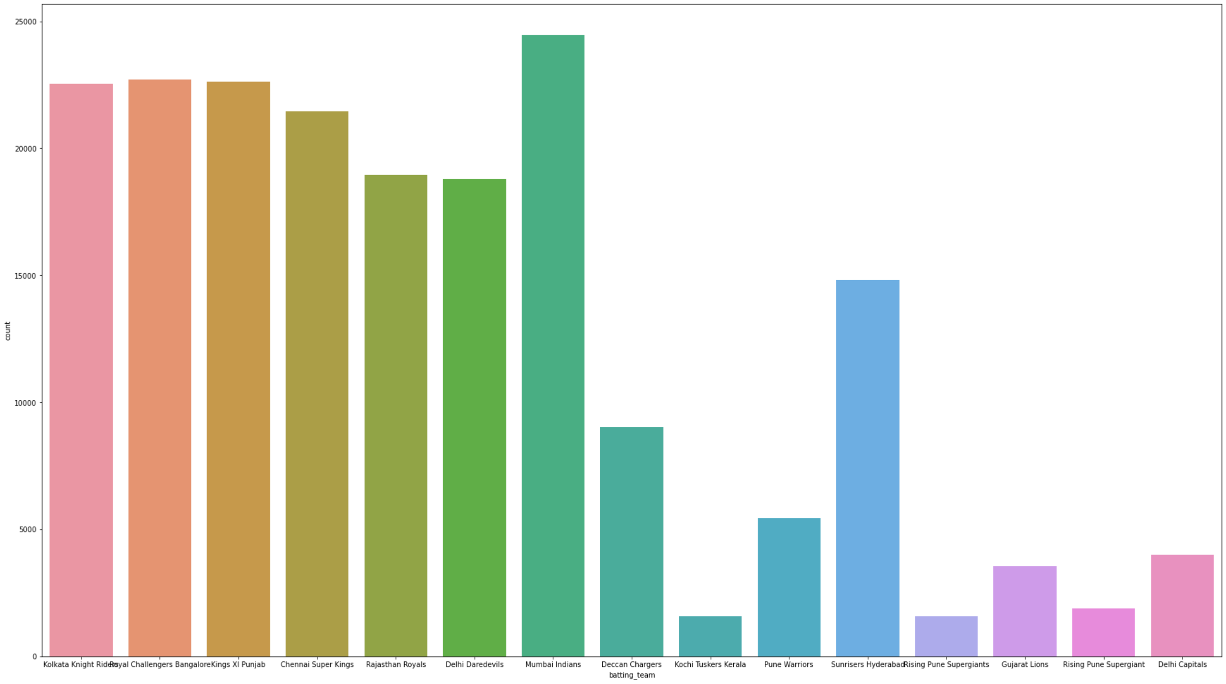
In [47]:

```
scores.iloc[scores[scores['is_wicket'].ge(1)].is_wicket.idxmax()]
```

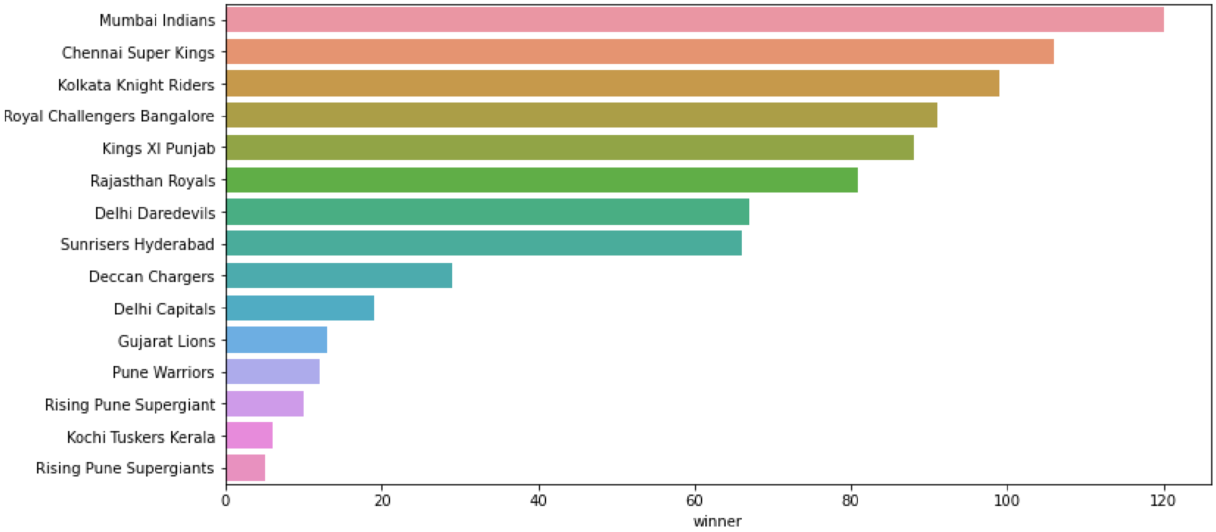
Out[47]:

```
id                                       335982
inning                                        1
over                                         12
ball                                          1
batsman                              RT Ponting
non_striker                         BB McCullum
bowler                                JH Kallis
batsman_runs                                  0
extra_runs                                    0
total_runs                                    0
non_boundary                                  0
is_wicket                                     1
dismissal_kind                           caught
player_dismissed                     RT Ponting
fielder                                 P Kumar
extras_type                                 NaN
batting_team              Kolkata Knight Riders
bowling_team        Royal Challengers Bangalore
Name: 32, dtype: object
```

In [61]:
```
plt.figure(figsize=(30,17))
sns.countplot(x='batting_team', data=scores)
plt.show()
```

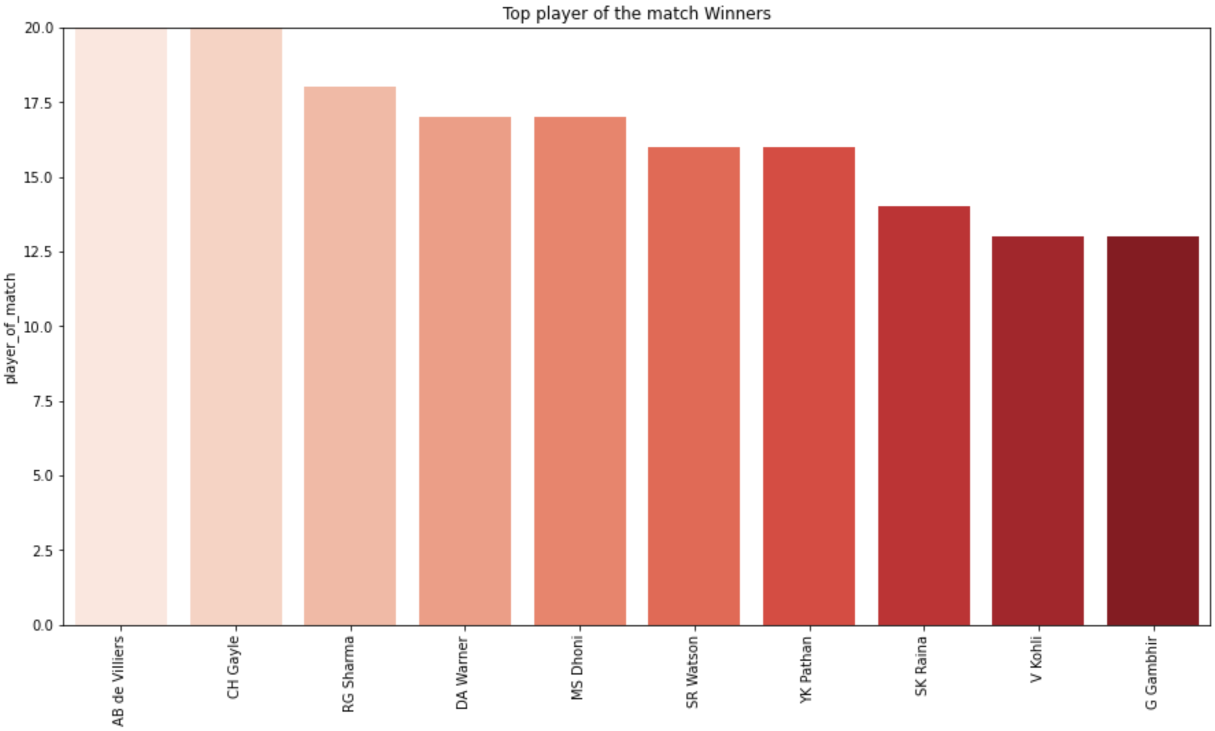


In [62]:
```
plt.figure(figsize=(12,6))
data = matches.winner.value_counts()
sns.barplot(y = data.index, x = data, orient='h')
plt.show()
```

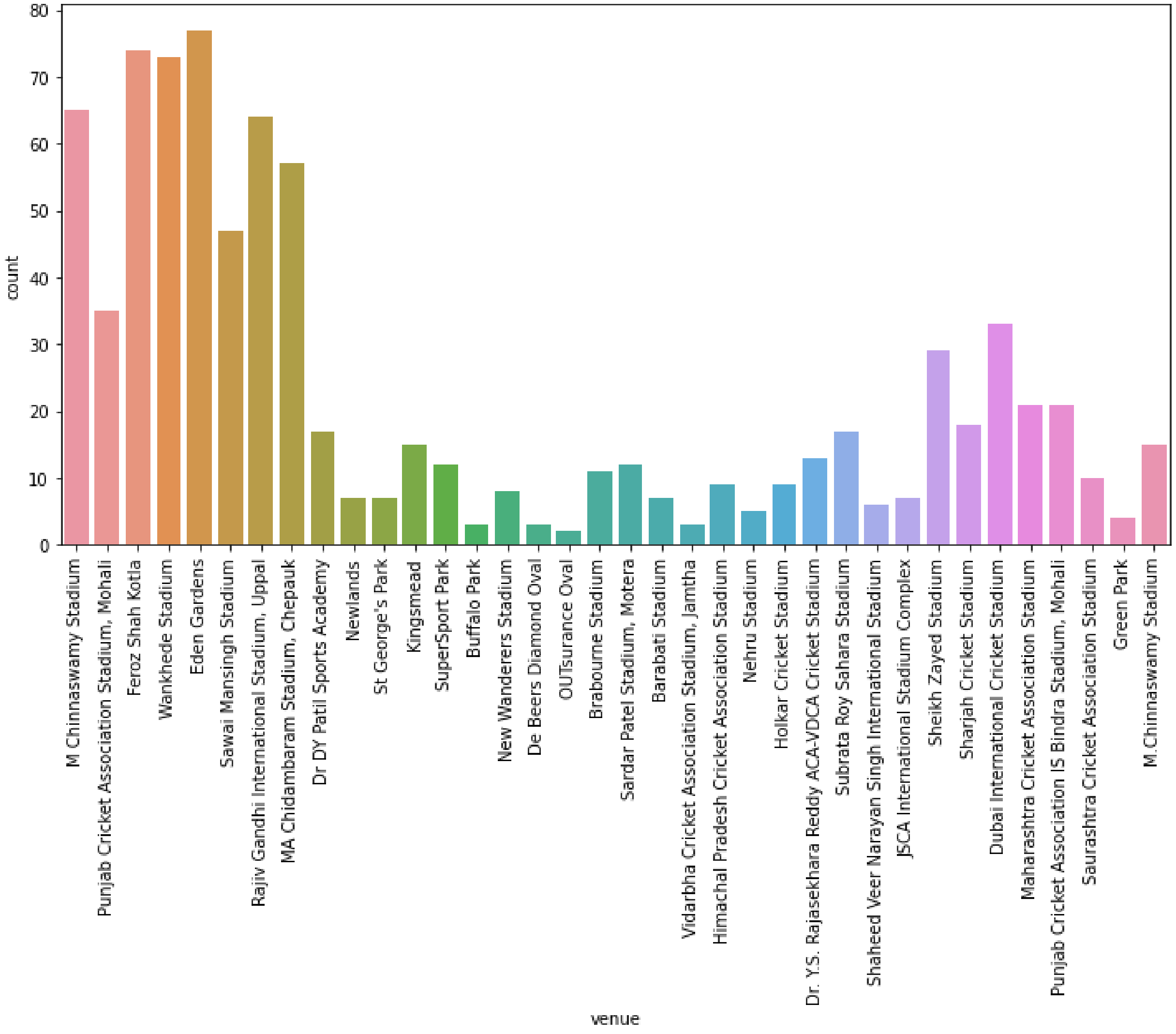


In [68]:

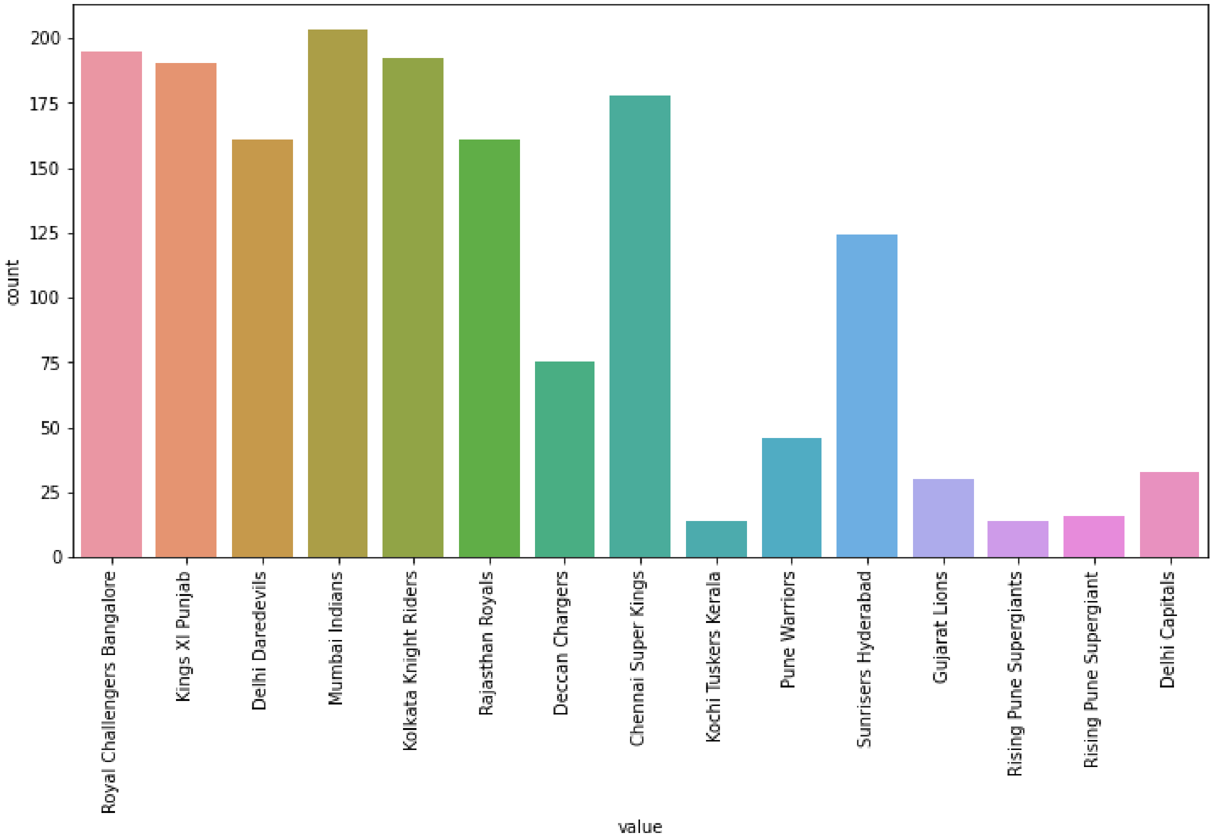
```
top_players = matches.player_of_match.value_counts()[:10]
#sns.barplot(x="day", y="total_bill", data=df)
fig, ax = plt.subplots(figsize=(15,8))
ax.set_ylim([0,20])
ax.set_ylabel("Count")
ax.set_title("Top player of the match Winners")
top_players.plot.bar()
sns.barplot(x = top_players.index, y = top_players, orient='v', palette="Reds");
plt.show()
```



In [69]:
```
plt.figure(figsize=(12,6))
sns.countplot(x='venue', data=matches)
plt.xticks(rotation='vertical')
plt.show()
```

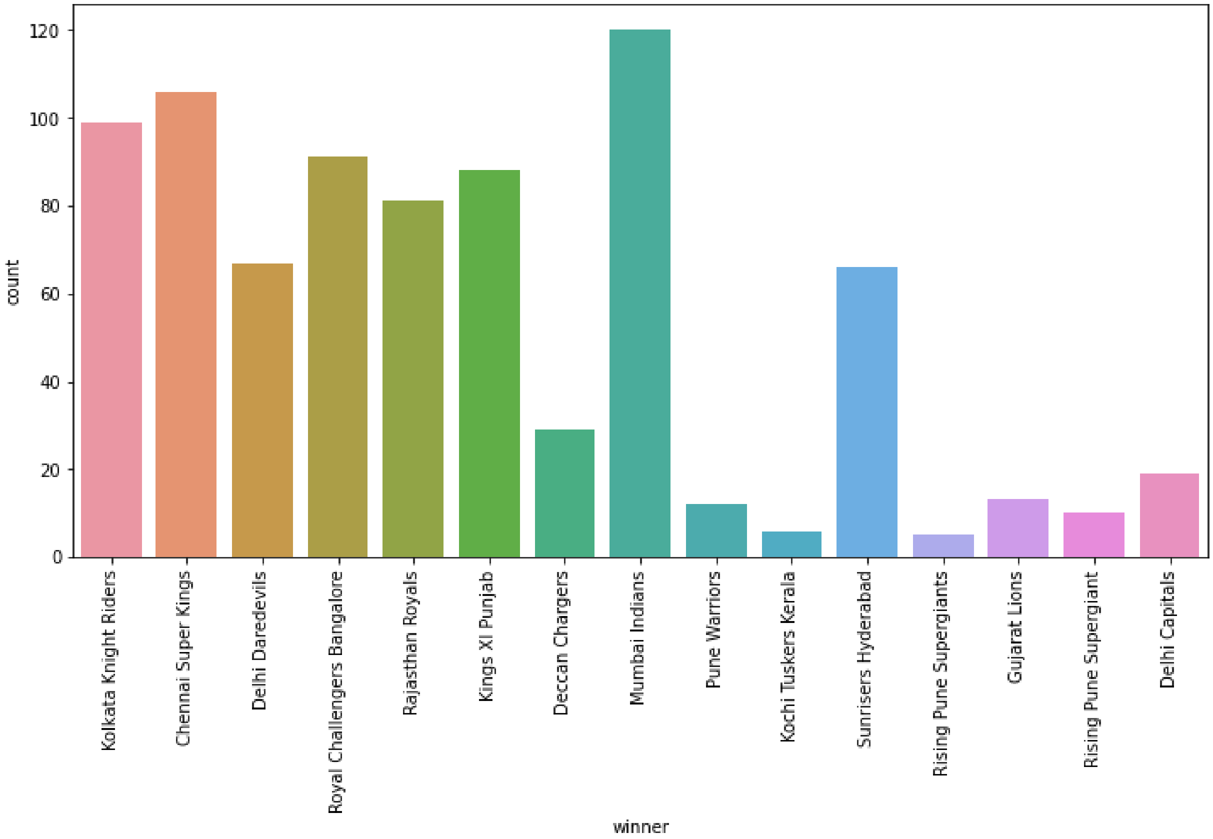


In [71]:
```
temp_df = pd.melt(matches, id_vars=['id','date'], value_vars=['team1', 'team2'])

plt.figure(figsize=(12,6))
sns.countplot(x='value', data=temp_df)
plt.xticks(rotation='vertical')
plt.show()
```



In [73]:

```
plt.figure(figsize=(12,6))
sns.countplot(x='winner', data=matches)
plt.xticks(rotation=90)
plt.show()
```



In [75]:
```
temp_df = matches.drop_duplicates(subset=['date'], keep='last')[['date', 'winner'
temp_df
```

Out[75]:

	date	winner
0	2008-04-18	Kolkata Knight Riders
1	2008-04-19	Delhi Daredevils
2	2008-04-20	Kolkata Knight Riders
3	2008-04-21	Rajasthan Royals
4	2008-04-22	Delhi Daredevils
...	...	...
591	2020-09-28	Royal Challengers Bangalore
592	2020-11-05	Mumbai Indians
593	2020-11-06	Sunrisers Hyderabad
594	2020-11-08	Delhi Capitals
595	2020-11-10	Mumbai Indians

596 rows × 2 columns

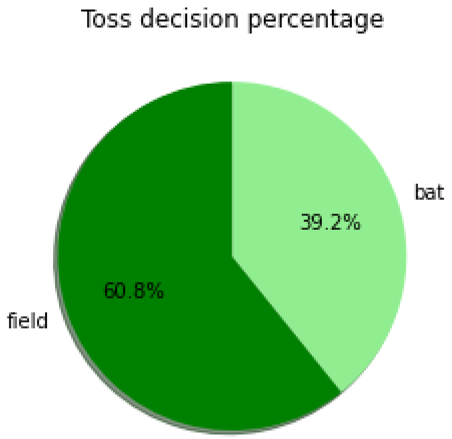
In [77]:
```
temp_df = matches.drop_duplicates(subset=['city'], keep='last')[['city', 'winner'
temp_df
```

Out[77]:

	city	winner
0	Cape Town	Kings XI Punjab
1	East London	Delhi Daredevils
2	Kimberley	Deccan Chargers
3	Port Elizabeth	Chennai Super Kings
4	Bloemfontein	Delhi Daredevils
5	Durban	Chennai Super Kings
6	Centurion	Deccan Chargers
7	Johannesburg	Deccan Chargers
8	Nagpur	Deccan Chargers
9	Kochi	Kochi Tuskers Kerala
10	Dharamsala	Kings XI Punjab
11	NaN	Sunrisers Hyderabad
12	Cuttack	Kolkata Knight Riders
13	Ahmedabad	Royal Challengers Bangalore
14	Ranchi	Chennai Super Kings
15	Raipur	Royal Challengers Bangalore
16	Rajkot	Mumbai Indians
17	Kanpur	Sunrisers Hyderabad
18	Bangalore	Mumbai Indians
19	Indore	Royal Challengers Bangalore
20	Pune	Chennai Super Kings
21	Jaipur	Rajasthan Royals
22	Kolkata	Kolkata Knight Riders
23	Delhi	Delhi Capitals
24	Bengaluru	Royal Challengers Bangalore
25	Chandigarh	Kings XI Punjab
26	Mumbai	Mumbai Indians
27	Chennai	Mumbai Indians
28	Visakhapatnam	Chennai Super Kings
29	Hyderabad	Mumbai Indians
30	Sharjah	Royal Challengers Bangalore
31	Abu Dhabi	Delhi Capitals
32	Dubai	Mumbai Indians

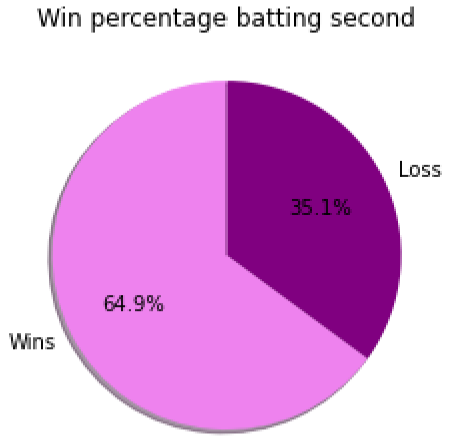
In [81]:

```
temp_series = matches.toss_decision.value_counts()
labels = (np.array(temp_series.index))
sizes = (np.array((temp_series / temp_series.sum())*100))
colors = ['green', 'lightgreen']
plt.pie(sizes, labels=labels, colors=colors,
        autopct='%1.1f%%', shadow=True, startangle=90)
plt.title("Toss decision percentage")
plt.show()
```



In [90]:

```
num_of_wins = (scores.total_runs>0).sum()
num_of_loss = (scores.total_runs==0).sum()
labels = ["Wins", "Loss"]
total = float(num_of_wins + num_of_loss)
sizes = [(num_of_wins/total)*100, (num_of_loss/total)*100]
colors = ['violet', 'purple']
plt.pie(sizes, labels=labels, colors=colors,
        autopct='%1.1f%%', shadow=True, startangle=90)
plt.title("Win percentage batting second")
plt.show()
```



In [86]:
```
matches.head(1)
```

Out[86]:

	id	city	date	player_of_match	venue	neutral_venue	team1	team2	tos
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	Royal Challengers Bangalore	Kolkata Knight Riders	C

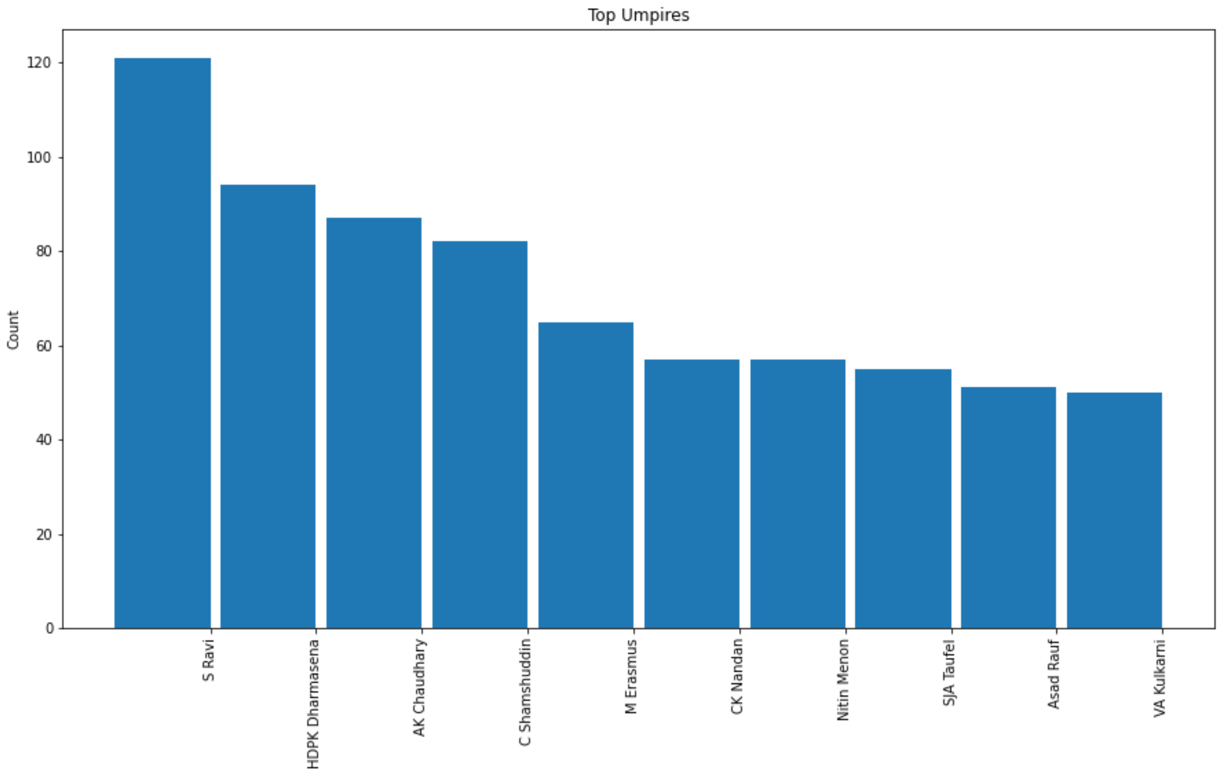
In [87]:
```
scores.head(1)
```

Out[87]:

	id	inning	over	ball	batsman	non_striker	bowler	batsman_runs	extra_runs	total_runs
0	335982	1	6	5	RT Ponting	BB McCullum	AA Noffke	1	0	1

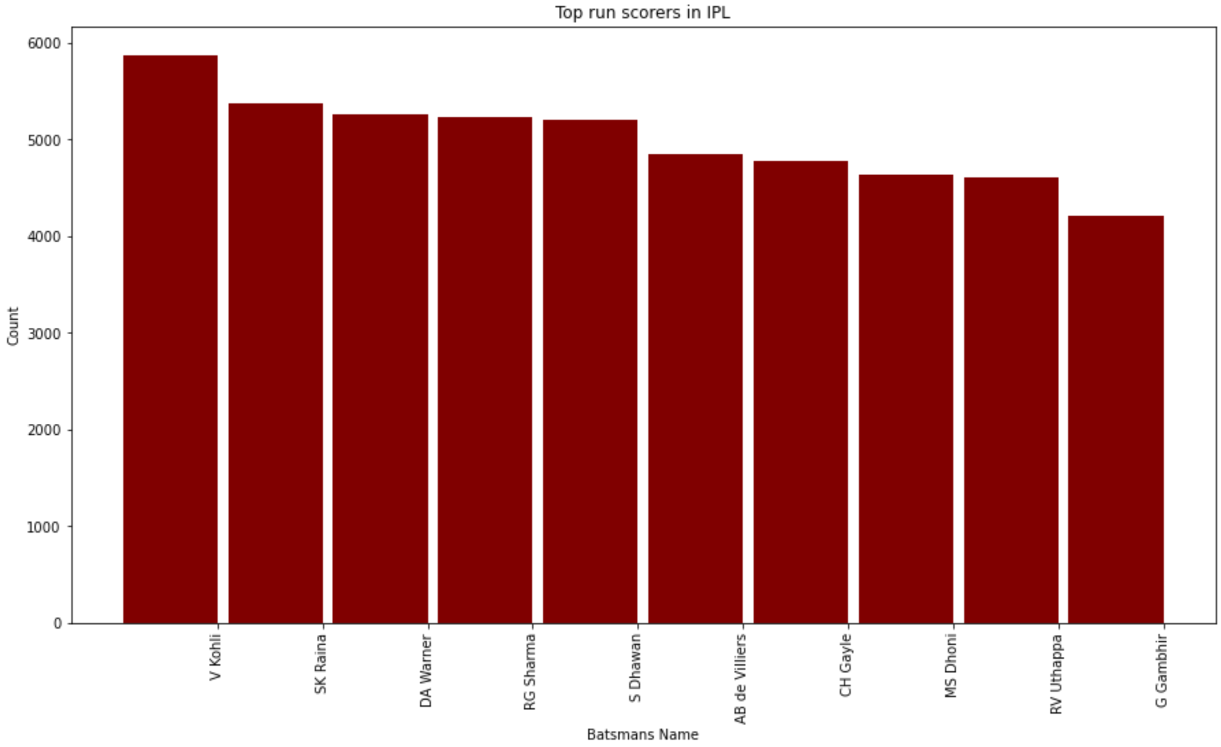
top umpires

In [94]:
```
temp_df = pd.melt(matches, id_vars=['id'], value_vars=['umpire1', 'umpire2'])
temp_series = temp_df.value.value_counts()[:10]
labels = np.array(temp_series.index)
ind = np.arange(len(labels))
width = 0.9
fig, ax = plt.subplots(figsize=(15,8))
rects = ax.bar(ind, np.array(temp_series), width=width,)
ax.set_xticks(ind+((width)/2.))
ax.set_xticklabels(labels, rotation='vertical')
ax.set_ylabel("Count")
ax.set_title("Top Umpires")
plt.show()
```

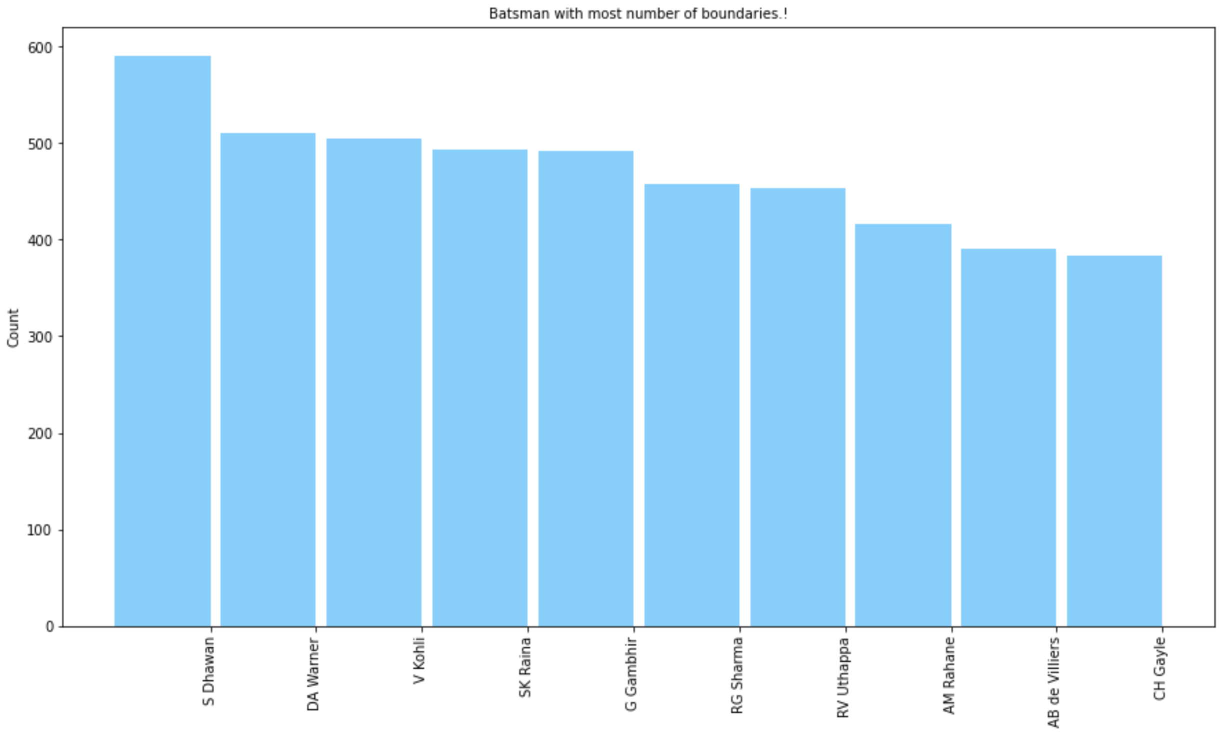


batsman analysis

In [101]:
```
temp_df = scores.groupby('batsman')['batsman_runs'].agg('sum').reset_index().sort
temp_df = temp_df.iloc[:10,:]

labels = np.array(temp_df['batsman'])
ind = np.arange(len(labels))
width = 0.9
fig, ax = plt.subplots(figsize=(15,8))
rects = ax.bar(ind, np.array(temp_df['batsman_runs']), width=width, color='maroon
ax.set_xticks(ind+((width)/2.))
ax.set_xticklabels(labels, rotation='vertical')
ax.set_ylabel("Count")
ax.set_title("Top run scorers in IPL")
ax.set_xlabel('Batsmans Name')
plt.show()
```

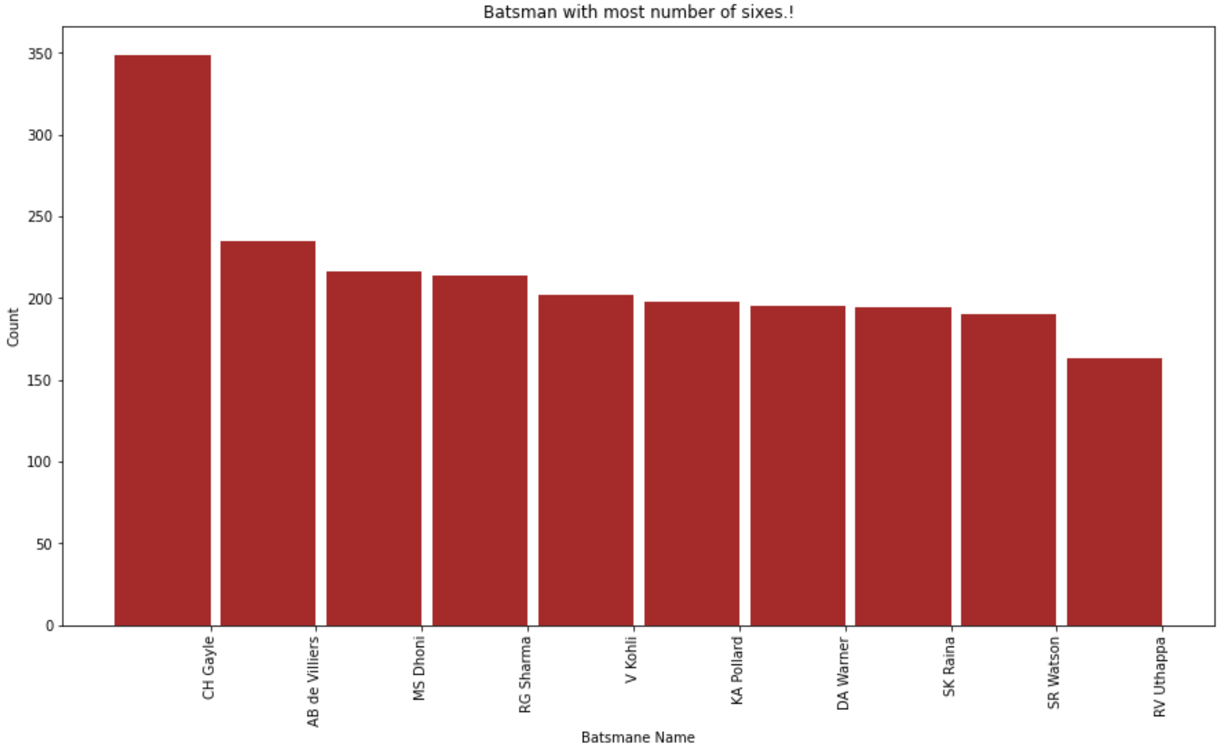


In [102]:
```
temp_df = scores.groupby('batsman')['batsman_runs'].agg(lambda x: (x==4).sum()).r
temp_df = temp_df.iloc[:10,:]

labels = np.array(temp_df['batsman'])
ind = np.arange(len(labels))
width = 0.9
fig, ax = plt.subplots(figsize=(15,8))
rects = ax.bar(ind, np.array(temp_df['batsman_runs']), width=width, color='lights
ax.set_xticks(ind+((width)/2.))
ax.set_xticklabels(labels, rotation='vertical')
ax.set_ylabel("Count")
ax.set_title("Batsman with most number of boundaries.!",fontsize = 10)
plt.show()
```



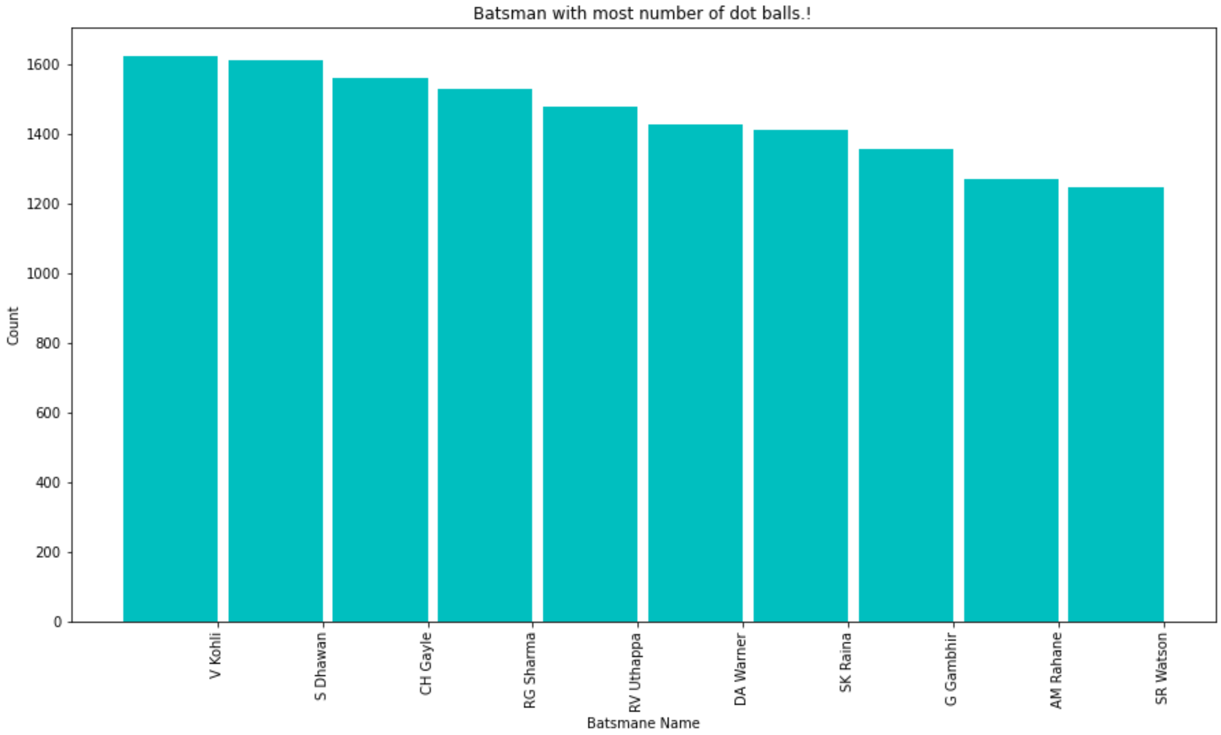
In [106]:

```
temp_df = scores.groupby('batsman')['batsman_runs'].agg(lambda x: (x==6).sum()).r
temp_df = temp_df.iloc[:10,:]

labels = np.array(temp_df['batsman'])
ind = np.arange(len(labels))
width = 0.9
fig, ax = plt.subplots(figsize=(15,8))
rects = ax.bar(ind, np.array(temp_df['batsman_runs']), width=width, color='brown'
ax.set_xticks(ind+((width)/2.))
ax.set_xticklabels(labels, rotation=90)
ax.set_ylabel("Count")
ax.set_title("Batsman with most number of sixes.!")
ax.set_xlabel('Batsmane Name')
plt.show()
```



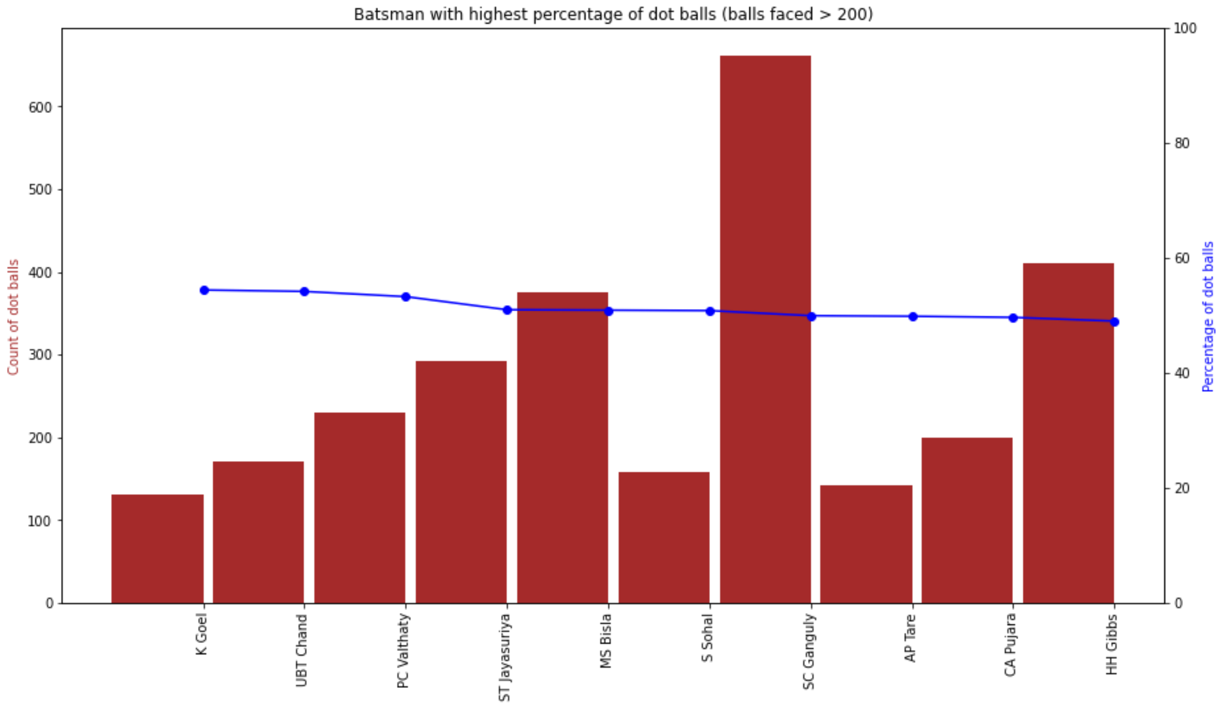
In [108]:

```
temp_df = scores.groupby('batsman')['batsman_runs'].agg(lambda x: (x==0).sum()).r
temp_df = temp_df.iloc[:10,:]

labels = np.array(temp_df['batsman'])
ind = np.arange(len(labels))
width = 0.9
fig, ax = plt.subplots(figsize=(15,8))
rects = ax.bar(ind, np.array(temp_df['batsman_runs']), width=width, color='c')
ax.set_xticks(ind+((width)/2.))
ax.set_xticklabels(labels, rotation='vertical')
ax.set_ylabel("Count")
ax.set_title("Batsman with most number of dot balls.!")
ax.set_xlabel('Batsmane Name')
plt.show()
```



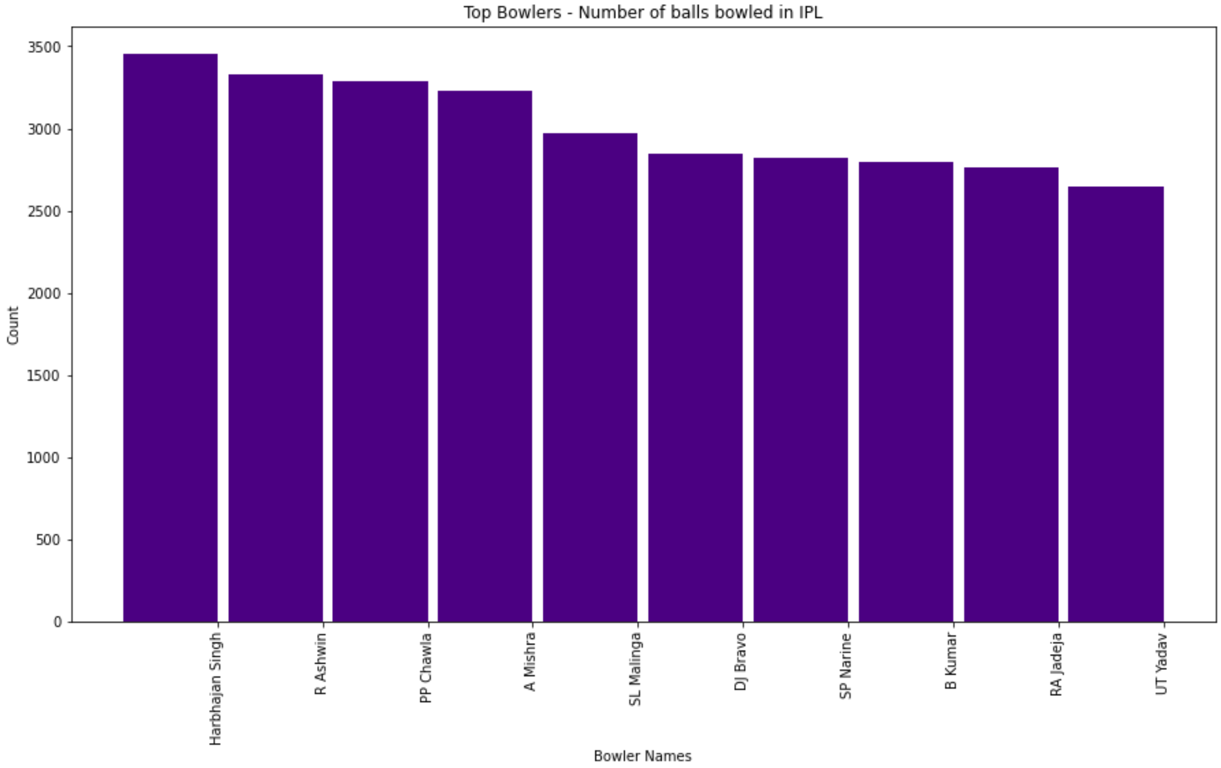
In [109]:
```
def balls_faced(x):
    return len(x)

def dot_balls(x):
    return (x==0).sum()

temp_df = scores.groupby('batsman')['batsman_runs'].agg([balls_faced, dot_balls])
temp_df = temp_df.loc[temp_df.balls_faced>200,:]
temp_df['percentage_of_dot_balls'] = (temp_df['dot_balls'] / temp_df['balls_faced
temp_df = temp_df.sort_values(by='percentage_of_dot_balls', ascending=False).rese
temp_df = temp_df.iloc[:10,:]
fig, ax1 = plt.subplots(figsize=(15,8))
ax2 = ax1.twinx()
labels = np.array(temp_df['batsman'])
ind = np.arange(len(labels))
width = 0.9
rects = ax1.bar(ind, np.array(temp_df['dot_balls']), width=width, color='brown')
ax1.set_xticks(ind+((width)/2.))
ax1.set_xticklabels(labels, rotation='vertical')
ax1.set_ylabel("Count of dot balls", color='brown')
ax1.set_title("Batsman with highest percentage of dot balls (balls faced > 200)")
ax2.plot(ind+0.45, np.array(temp_df['percentage_of_dot_balls']), color='b', marke
ax2.set_ylabel("Percentage of dot balls", color='b')
ax2.set_ylim([0,100])
ax2.grid(b=False)
plt.show()
```

```
C:\Users\Dell\AppData\Local\Temp\ipykernel_10760\1879917764.py:25: MatplotlibDe
precationWarning: The 'b' parameter of grid() has been renamed 'visible' since
Matplotlib 3.5; support for the old name will be dropped two minor releases lat
er.
  ax2.grid(b=False)
```



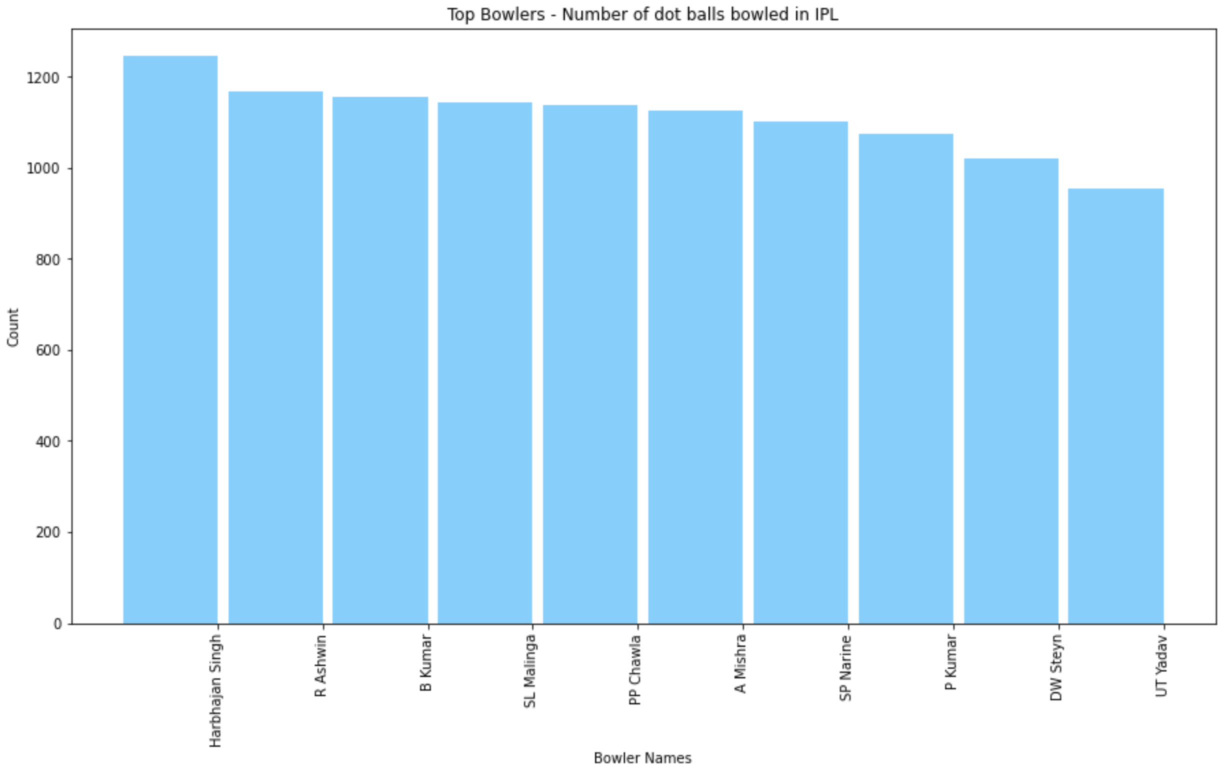
bowler analysis

In [111]:
```
temp_df = scores.groupby('bowler')['ball'].agg('count').reset_index().sort_values
temp_df = temp_df.iloc[:10,:]

labels = np.array(temp_df['bowler'])
ind = np.arange(len(labels))
width = 0.9
fig, ax = plt.subplots(figsize=(15,8))
rects = ax.bar(ind, np.array(temp_df['ball']), width=width, color='indigo')
ax.set_xticks(ind+((width)/2.))
ax.set_xticklabels(labels, rotation='vertical')
ax.set_ylabel("Count")
ax.set_title("Top Bowlers - Number of balls bowled in IPL")
ax.set_xlabel('Bowler Names')
plt.show()
```



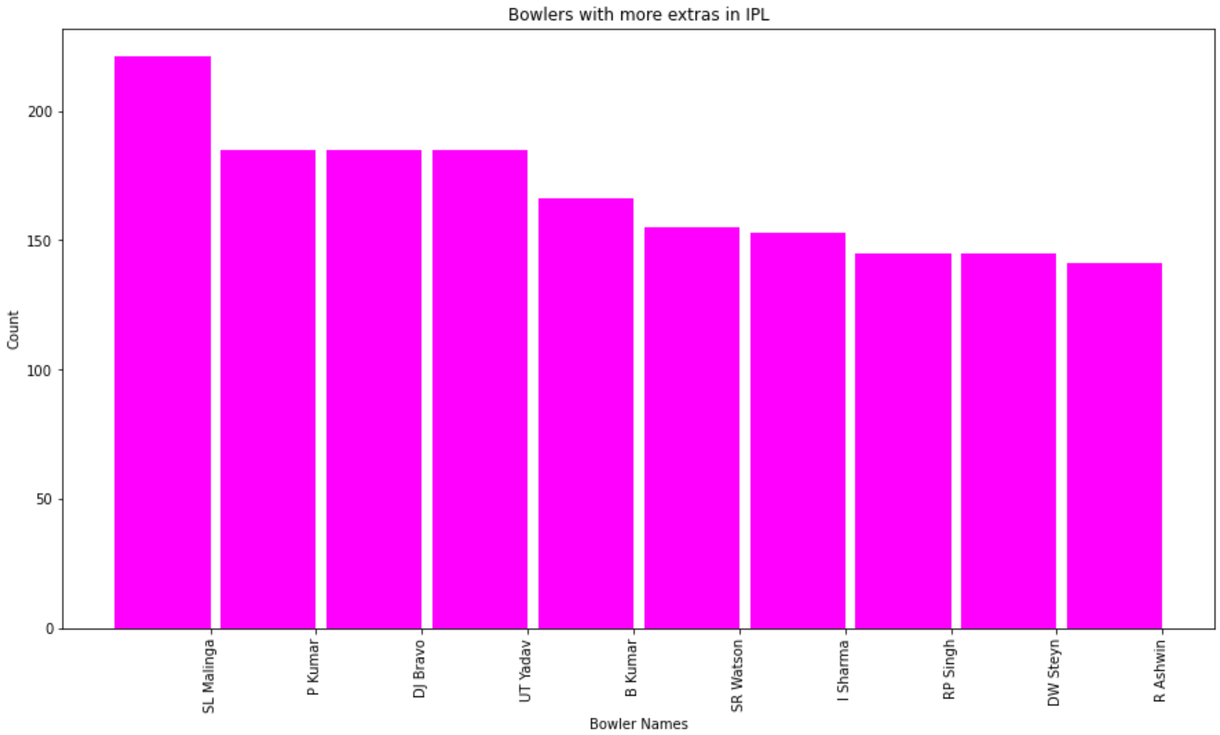
In [112]:

```
temp_df = scores.groupby('bowler')['total_runs'].agg(lambda x: (x==0).sum()).rese
temp_df = temp_df.iloc[:10,:]

labels = np.array(temp_df['bowler'])
ind = np.arange(len(labels))
width = 0.9
fig, ax = plt.subplots(figsize=(15,8))
rects = ax.bar(ind, np.array(temp_df['total_runs']), width=width, color='lightsky
ax.set_xticks(ind+((width)/2.))
ax.set_xticklabels(labels, rotation='vertical')
ax.set_ylabel("Count")
ax.set_title("Top Bowlers - Number of dot balls bowled in IPL")
ax.set_xlabel('Bowler Names')
plt.show()
```



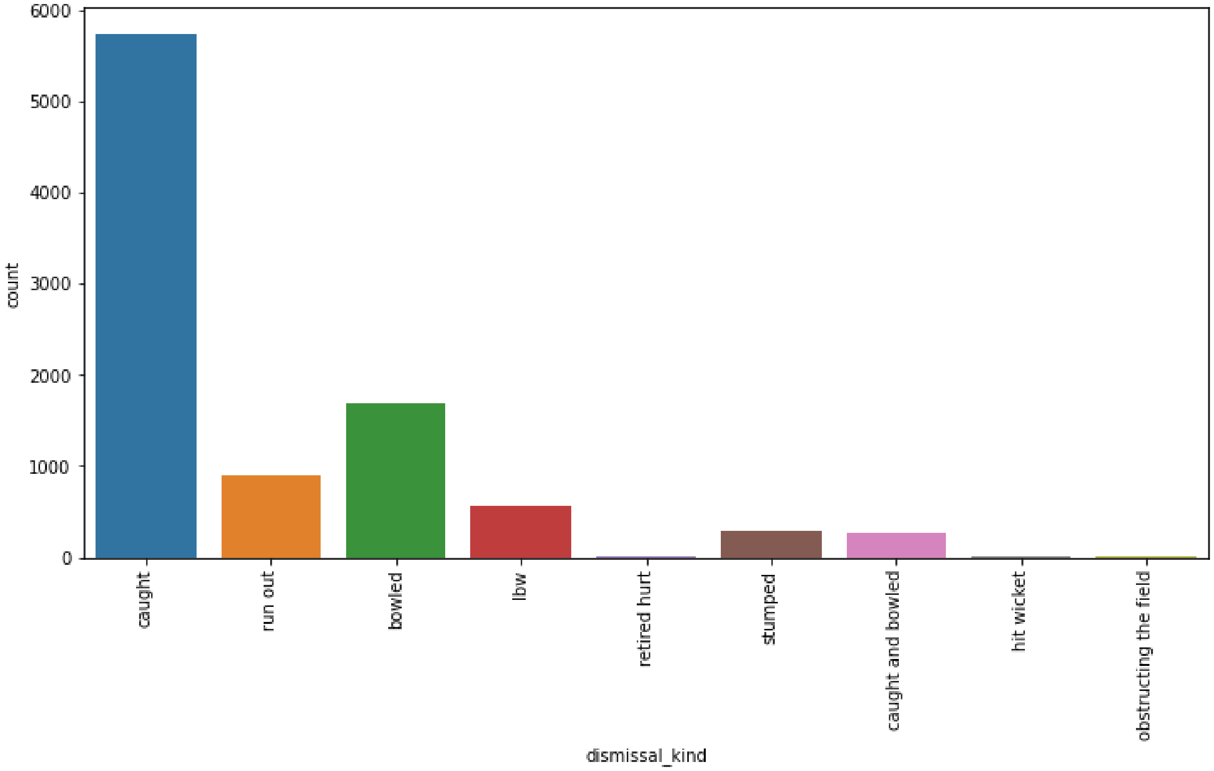
In [113]:

```
temp_df = scores.groupby('bowler')['extra_runs'].agg(lambda x: (x>0).sum()).reset
temp_df = temp_df.iloc[:10,:]

labels = np.array(temp_df['bowler'])
ind = np.arange(len(labels))
width = 0.9
fig, ax = plt.subplots(figsize=(15,8))
rects = ax.bar(ind, np.array(temp_df['extra_runs']), width=width, color='magenta'
ax.set_xticks(ind+((width)/2.))
ax.set_xticklabels(labels, rotation='vertical')
ax.set_ylabel("Count")
ax.set_title("Bowlers with more extras in IPL")
ax.set_xlabel('Bowler Names')
plt.show()
```



In [114]:

```
plt.figure(figsize=(12,6))
sns.countplot(x='dismissal_kind', data=scores)
plt.xticks(rotation='vertical')
plt.show()
```


count
caught
run out
bowled
lbw
retired hurt
stumped
caught and bowled
hit wicket
obstructing the field
dismissal_kind