Import all libraries

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

read ipl.csv file and print first 5 records

```
In [2]: df=pd.read_csv('ipl data.csv')
        df.head()
```

Out[2]:

	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 Llong	NaN
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 Ravi	NaN
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 Nandan	NaN
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	C Shamshuddin	NaN
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	NaN	NaN

check total number of columns,entries note down your findings

```
In [3]: #check total number of columns,entries note down your findings
        print(df.columns)
        #Findings:-
        df.shape
```

```
Index(['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'],
      dtype='object')
```

Out[3]: (636, 18)

In [4]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 636 entries, 0 to 635
Data columns (total 18 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   id              636 non-null    int64
 1   season          636 non-null    int64
 2   city            629 non-null    object
 3   date            636 non-null    object
 4   team1           636 non-null    object
 5   team2           636 non-null    object
 6   toss_winner     636 non-null    object
 7   toss_decision   636 non-null    object
 8   result          636 non-null    object
 9   dl_applied      636 non-null    int64
 10  winner          633 non-null    object
 11  win_by_runs     636 non-null    int64
 12  win_by_wickets  636 non-null    int64
 13  player_of_match 633 non-null    object
 14  venue           636 non-null    object
 15  umpire1         635 non-null    object
 16  umpire2         635 non-null    object
 17  umpire3         0 non-null      float64
dtypes: float64(1), int64(5), object(12)
memory usage: 89.6+ KB
```

find null values

In [5]: `df.isnull().sum()`

Out[5]:
```
id                 0
season             0
city               7
date               0
team1              0
team2              0
toss_winner        0
toss_decision      0
result             0
dl_applied         0
winner             3
win_by_runs        0
win_by_wickets     0
player_of_match    3
venue              0
umpire1            1
umpire2            1
umpire3          636
dtype: int64
```

drop umpire3 column as it contains more than 75% of null values

In [6]:
```python
df.drop(["umpire3"],axis=1, inplace=True)
df.head()
```

Out[6]:

	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
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 Llong
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 Ravi
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 Nandan
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	C Shamshuddin
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	NaN

drop null values of city , winner ,player_of_match,umpire1,umpire2 column

In [7]:
```python
df.dropna(axis=0, inplace=True)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 625 entries, 0 to 635
Data columns (total 17 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   id               625 non-null    int64
 1   season           625 non-null    int64
 2   city             625 non-null    object
 3   date             625 non-null    object
 4   team1            625 non-null    object
 5   team2            625 non-null    object
 6   toss_winner      625 non-null    object
 7   toss_decision    625 non-null    object
 8   result           625 non-null    object
 9   dl_applied       625 non-null    int64
 10  winner           625 non-null    object
 11  win_by_runs      625 non-null    int64
 12  win_by_wickets   625 non-null    int64
 13  player_of_match  625 non-null    object
 14  venue            625 non-null    object
 15  umpire1          625 non-null    object
 16  umpire2          625 non-null    object
dtypes: int64(5), object(12)
memory usage: 87.9+ KB
```
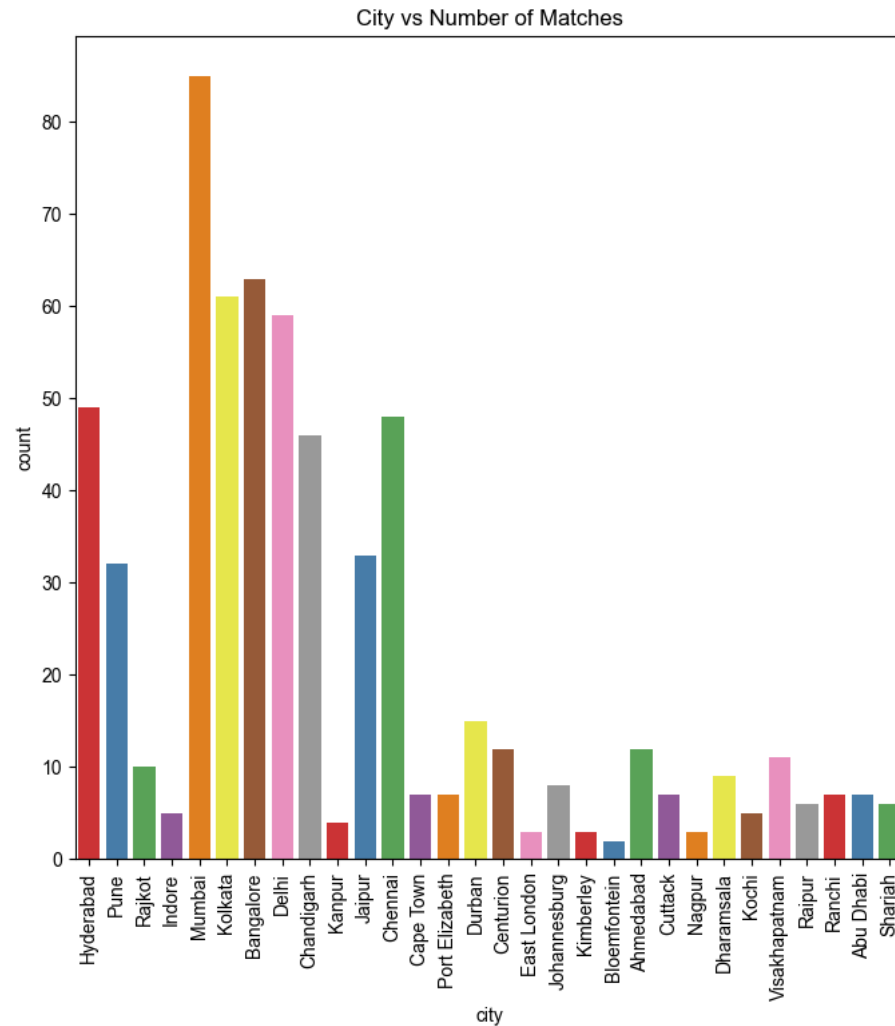
which city hosted most number of matches?

In [8]:
```python
df['city'].value_counts().head(1)
```

Out[8]:
```
Mumbai    85
Name: city, dtype: int64
```

draw bar plot and write down your insights

In [9]:
```python
zz=df.city.value_counts()
```

In [10]:
```python
plt.figure(figsize=(8,8))
plt.title('City vs Number of Matches')
plt.xticks(rotation=90)
sns.countplot(x='city',data=df,palette='Set1')
sns.set_style("darkgrid")
plt.show()
```


From graph we can see that,

```
Maximum Matches are played in Mummbai
Minimum matches are played in Bloemfontein
More than 50 Matches are played in-Mumbai,Bangalore ,Kolkata ,Delhi
Less than 10 Matches are Played in- Dharamsala ,Johannesburg, Cape Town,Abu Dhabi,Ranchi,Port Elizabeth ,Cuttack,Raipur,Sharjah,Kochi,Indore,Kanpur,Nagpur,Kimberley,East
London,Bloemfontein
```
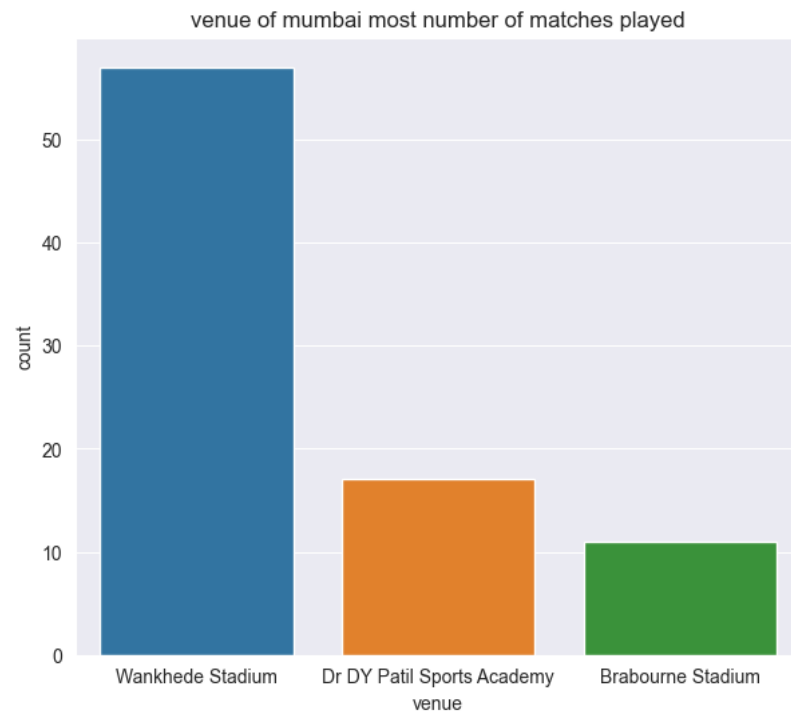
find all venue of mumbai city

In [11]: ```python
df[df['city']=="Mumbai"]['venue'].value_counts()
```

Out[11]:
```
Wankhede Stadium   57
Dr DY Patil Sports Academy 17
Brabourne Stadium   11
Name: venue, dtype: int64
```

## now compare in which venue of mumbai most number of matches played (draw bar plot and write down insights)

In [12]: ```python
plt.figure(figsize=(7,6))
mum=df[df['city']=="Mumbai"]
sns.countplot(data=mum,x="venue")
plt.title('venue of mumbai most number of matches played')
plt.show()
```


venue of mumbai most number of matches played

```
In Mumbai city Most of the IPL matches are played in Wankhede Stadium then followed by Dr DY Patil Stadium and Brabourn Stadium
Wankhede Stadium             57
Dr DY Patil Sports Academy   17
Brabourne Stadium            11
```

what is the preferred choice after winning a toss in mumbai

In [13]: `df[df['city']=='Mumbai']['toss_decision'].value_counts().idxmax()`

Out[13]: `'field'`

graphical representation of above question

In [14]:
```python
wm=df[df['city']=='Mumbai']
plt.title("MUMBAI vs Toss Decision")
sns.countplot(data=wm,x="city",hue='toss_decision',width=0.5)
plt.show()
```


which team won most number of toss :

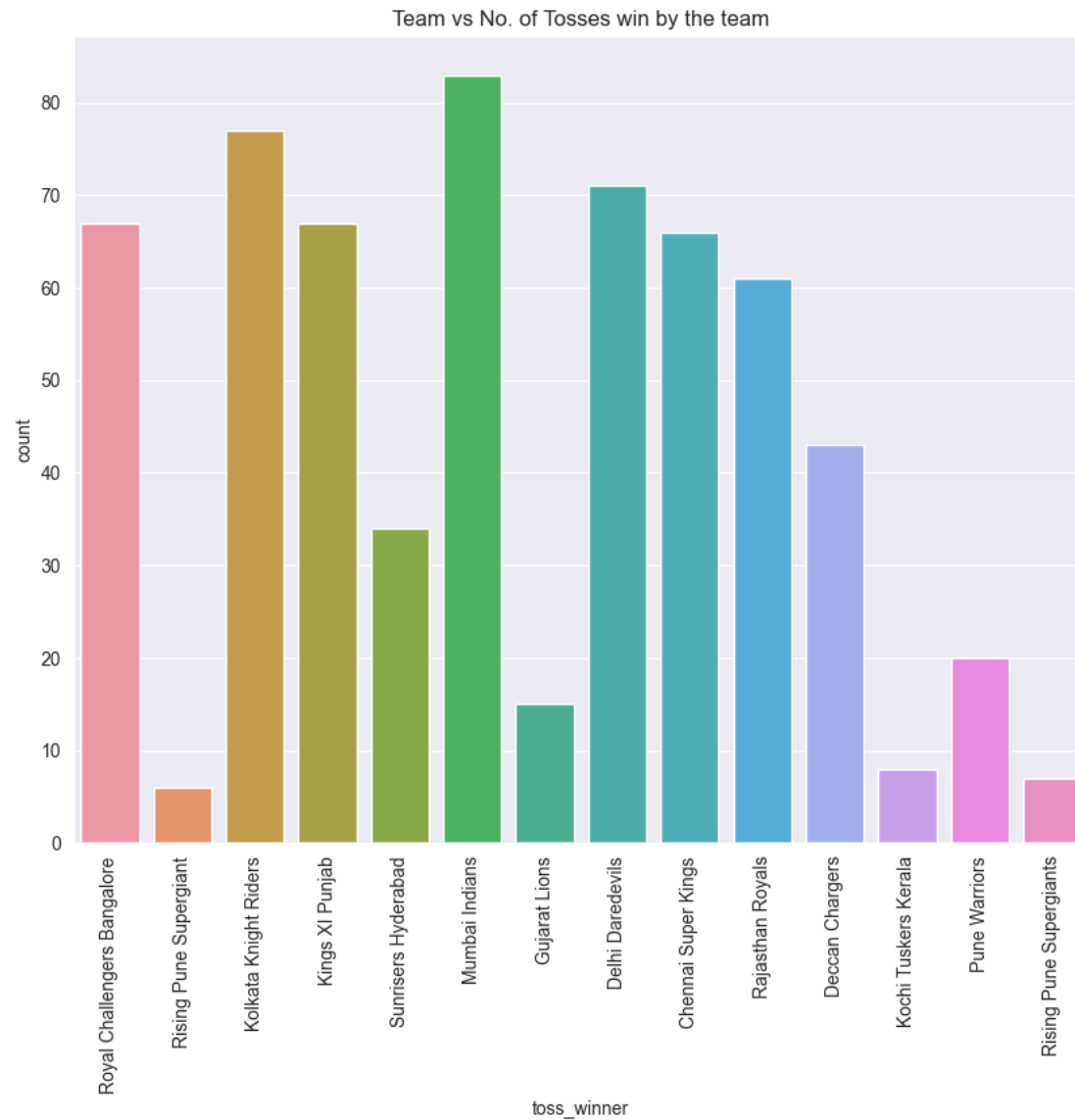In [15]: `df['toss_winner'].value_counts().idxmax()`

Out[15]: `'Mumbai Indians'`

In [16]:
```python
tw= df.value_counts('toss_winner')
tw
```

Out[16]:
```
toss_winner
Mumbai Indians                  83
Kolkata Knight Riders           77
Delhi Daredevils                71
Kings XI Punjab                 67
Royal Challengers Bangalore     67
Chennai Super Kings             66
Rajasthan Royals                61
Deccan Chargers                 43
Sunrisers Hyderabad             34
Pune Warriors                   20
Gujarat Lions                   15
Kochi Tuskers Kerala             8
Rising Pune Supergiants          7
Rising Pune Supergiant           6
dtype: int64
```

In [17]:
```python
#show graphical representation of above question

plt.figure(figsize=(10,8))
plt.title("Team vs No. of Tosses win by the team")
sns.countplot(x='toss_winner' ,data=df)#.sort_values('toss_winner'))
plt.xticks(rotation=90)

plt.show()
```


Team vs No. of Tosses win by the team

In [18]:
```python
df.value_counts('toss_winner')
```

Out[18]:
```
toss_winner
Mumbai Indians                83
Kolkata Knight Riders         77
Delhi Daredevils              71
Kings XI Punjab               67
Royal Challengers Bangalore   67
Chennai Super Kings           66
Rajasthan Royals              61
Deccan Chargers               43
Sunrisers Hyderabad           34
Pune Warriors                 20
Gujarat Lions                 15
Kochi Tuskers Kerala           8
Rising Pune Supergiants        7
Rising Pune Supergiant         6
dtype: int64
```

find what mumbai indians preferred after winning a toss?

In [19]:
```python
df[df['toss_winner']=='Mumbai Indians']['toss_decision'].value_counts().idxmax()
```

Out[19]: 'field'

head to head winning count of Mumbai Indians vs Chennai Super Kings

In [20]:
```python
df[((df['team1']=='Chennai Super Kings') & (df['team2']=='Mumbai Indians')) | ((df['team1']=='Mumbai Indians') & (df['team2']=='Chennai Super Kings'))]['winner'].value_counts()
```

Out[20]:
```
Mumbai Indians        12
Chennai Super Kings    9
Name: winner, dtype: int64
```

Which team won most of the matches in mumbai?

In [21]:
```python
df[df['city']=='Mumbai']['winner'].value_counts().idxmax()
```

Out[21]: 'Mumbai Indians'

how many times each team won the toss and won the match in mumbai

In [22]:
```python
df[(df['city']=='Mumbai') & (df['toss_winner']==df['winner'])]['winner'].value_counts()
```

Out[22]:
```
Mumbai Indians               26
Chennai Super Kings           5
Deccan Chargers               3
Rajasthan Royals              2
Delhi Daredevils              2
Royal Challengers Bangalore   2
Kochi Tuskers Kerala          1
Kolkata Knight Riders         1
Gujarat Lions                 1
Name: winner, dtype: int64
```

which venue hosted most number of matches

In [23]:
```python
df['venue'].value_counts().head(1)
```

Out[23]:
```
M Chinnaswamy Stadium    63
Name: venue, dtype: int64
```

find how many matches chennai super kings played at M Chinnaswamy stadium?

In [24]:
```python
df[((df['team1']=='Chennai Super Kings') | (df['team2']=='Chennai Super Kings')) & (df['venue']=='M Chinnaswamy Stadium')]['id'].count()
```

Out[24]: 7

who won most matches at M Chinnaswamy stadium?--Royal Challengers Bangalore

In [25]:
```python
df[df['venue']=='M Chinnaswamy Stadium']['winner'].value_counts().idxmax()
```

Out[25]: 'Royal Challengers Bangalore'

matches played in each year

In [26]:
```python
df.value_counts("season")
```

Out[26]:
```
season
2013    76
2012    74
2011    72
2010    60
2016    60
2008    58
2017    58
2009    57
2015    57
2014    53
dtype: int64
```

which city hosted most number of matches in 2013

In [27]:
```python
df[df["season"]==2013]['city'].value_counts()
```

Out[27]:
```
Kolkata       8
Bangalore     8
Hyderabad     8
Delhi         8
Chennai       8
Pune          8
Jaipur        8
Mumbai        8
Chandigarh    6
Dharamsala    2
Raipur        2
Ranchi        2
Name: city, dtype: int64
```

In [28]:
```python
#lets analyse ipl season held in 2013
#extract all the details of 2013 season

df13=df[df['season']==2013]
df13.head()
```

Out[28]:

	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
381	382	2013	Kolkata	2013-04-03	Delhi Daredevils	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	6	SP Narine	Eden Gardens	S Ravi	SJA Taufel
382	383	2013	Bangalore	2013-04-04	Royal Challengers Bangalore	Mumbai Indians	Mumbai Indians	field	normal	0	Royal Challengers Bangalore	2	0	CH Gayle	M Chinnaswamy Stadium	VA Kulkarni	C Shamshuddin
383	384	2013	Hyderabad	2013-04-05	Sunrisers Hyderabad	Pune Warriors	Pune Warriors	field	normal	0	Sunrisers Hyderabad	22	0	A Mishra	Rajiv Gandhi International Stadium, Uppal	S Ravi	SJA Taufel
384	385	2013	Delhi	2013-04-06	Rajasthan Royals	Delhi Daredevils	Rajasthan Royals	bat	normal	0	Rajasthan Royals	5	0	R Dravid	Feroz Shah Kotla	S Das	C Shamshuddin
385	386	2013	Chennai	2013-04-06	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	normal	0	Mumbai Indians	9	0	KA Pollard	MA Chidambaram Stadium, Chepauk	M Erasmus	VA Kulkarni

In [29]:
```python
plt.figure(figsize=(10,10))
plt.title('IPL 2013 Analysis(Winning of Team based on Cities)')
sns.countplot(data=df13,y='winner',hue='city',palette='Set1')
plt.xticks(rotation=90)
plt.legend(loc=(1.05,0.5))
plt.show()
```


IPL 2013 Analysis(Winning of Team based on Cities)
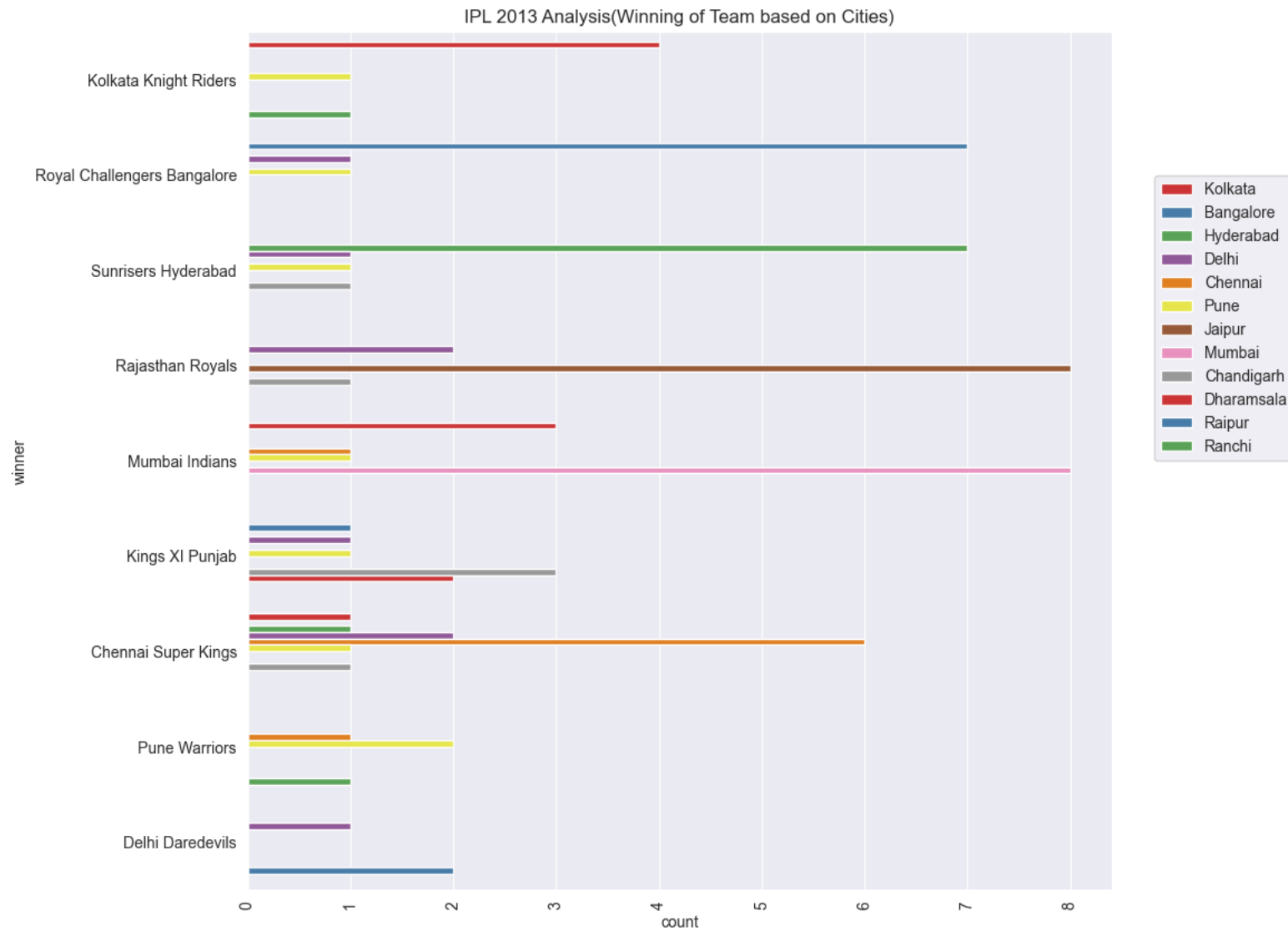
In [30]:
```python
#here we will create yearwise groups (hint: use groupby() function)
year=df.groupby('season')['winner'].value_counts()
year[2013]
```

Out[30]:
```
winner
Mumbai Indians                13
Chennai Super Kings           12
Rajasthan Royals              11
Sunrisers Hyderabad           10
Royal Challengers Bangalore    9
Kings XI Punjab                8
Kolkata Knight Riders          6
Pune Warriors                  4
Delhi Daredevils               3
Name: winner, dtype: int64
```

In [31]:
```python
plt.figure(figsize=(10,10))
plt.title('Team victory based on a  toss win')
sns.histplot(data=df13,y='winner',hue='toss_winner')
plt.show()
```


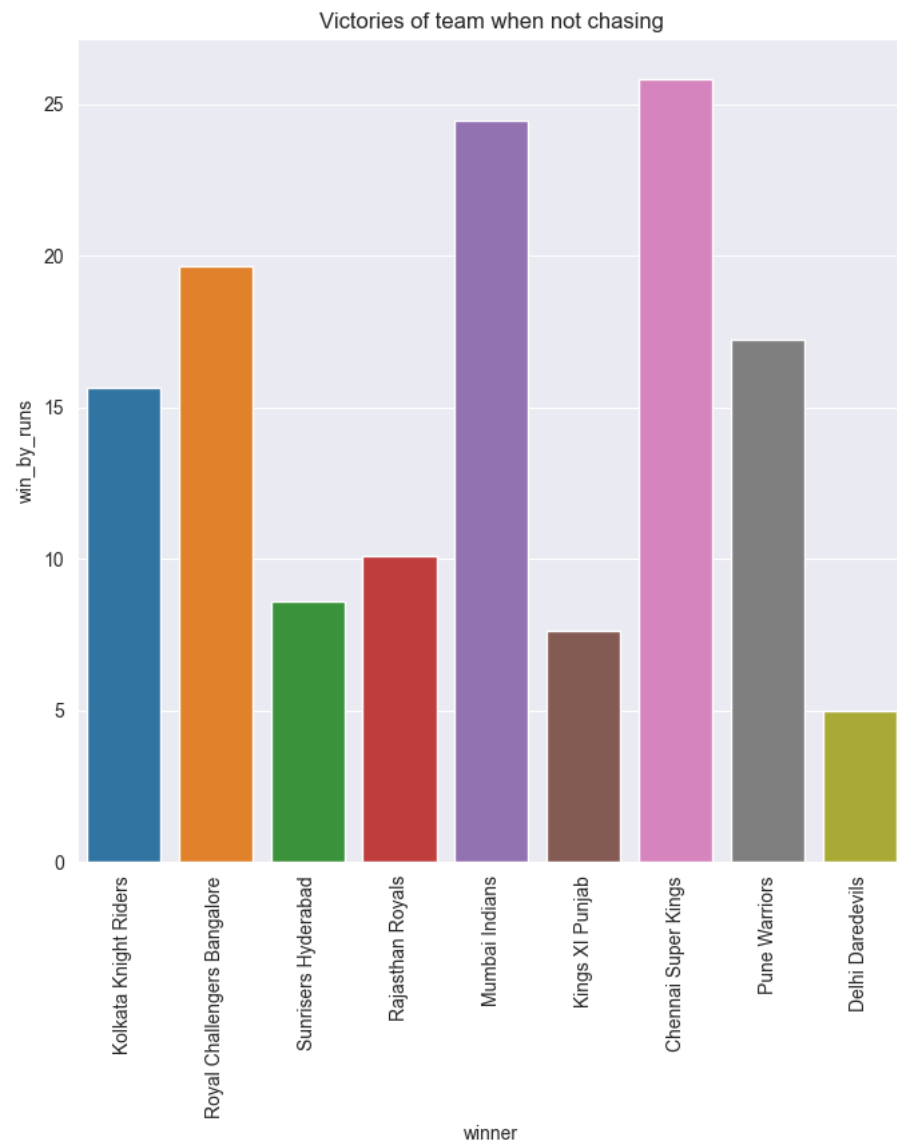Team victory based on a toss win
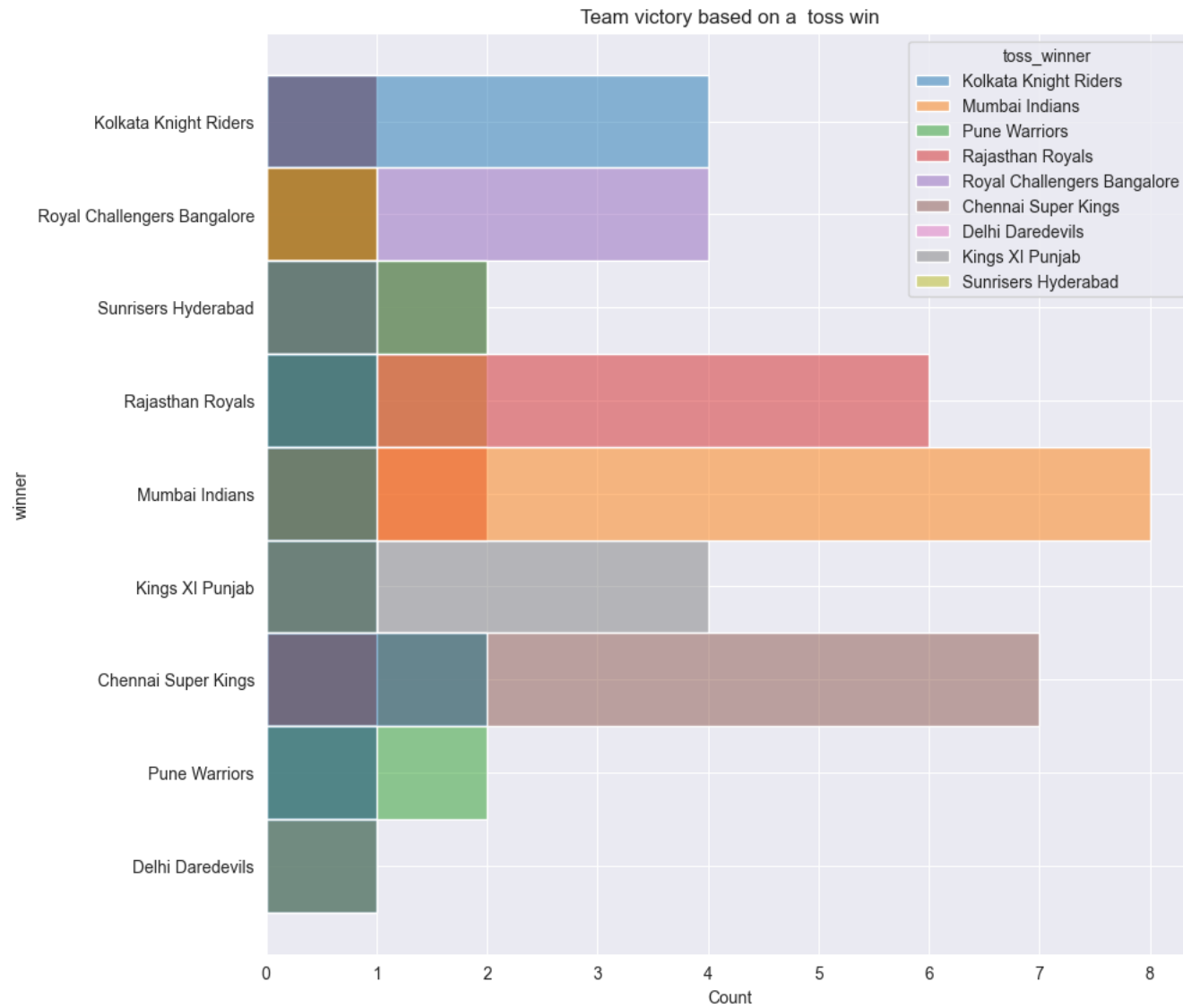
In [32]:
```python
plt.figure(figsize=(8,8))
plt.title("Victories of team While Chasing")
sns.barplot(data=df13,x='winner',y='win_by_wickets')
plt.xticks(rotation=90)
plt.show()
```


Victories of team While Chasing
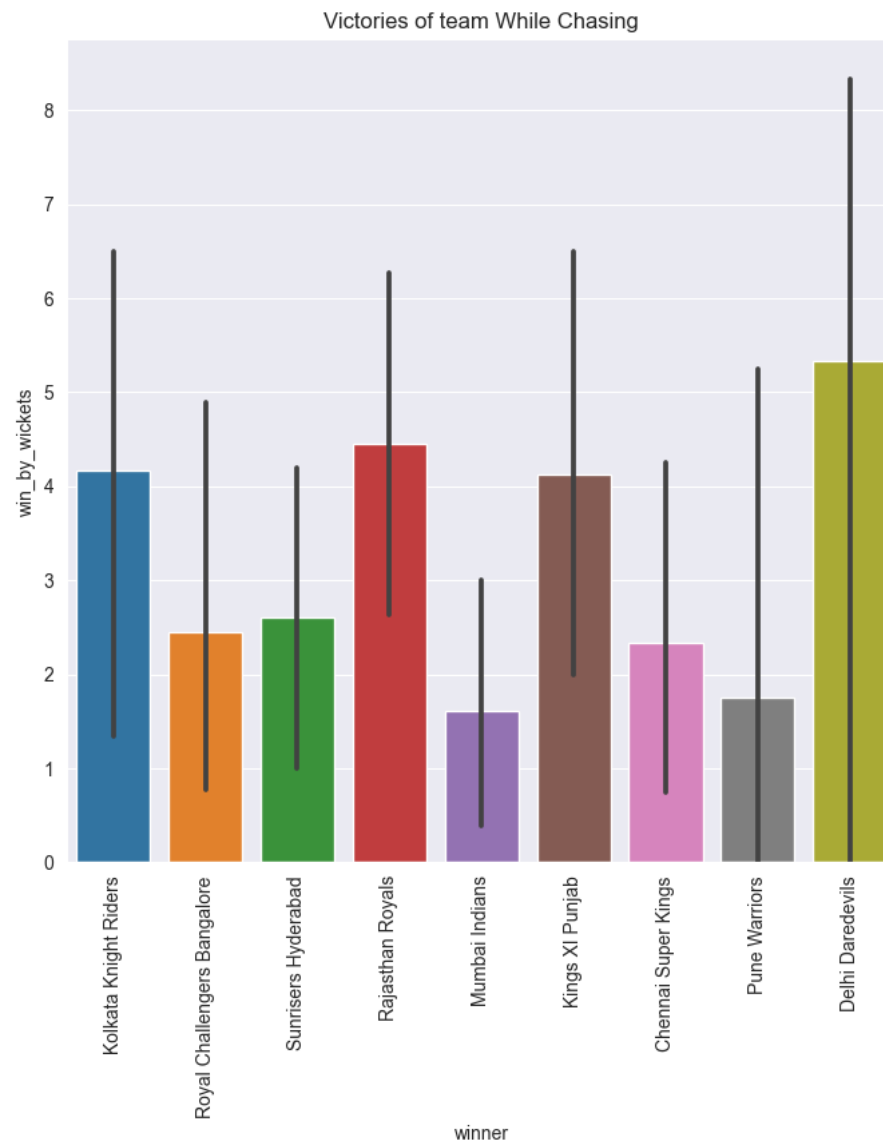
In [33]:
```python
plt.figure(figsize=(8,8))
plt.title('Victories of team when not chasing')
sns.barplot(data=df13,x='winner',y='win_by_runs',errorbar=None)
plt.xticks(rotation=90)
plt.show()
```


Victories of team when not chasing

In [ ]: