```
In [1]:  #import all libraries
         import numpy as np
         import pandas as pd
         import matplotlib.pyplot as plt
         import seaborn as sns
```
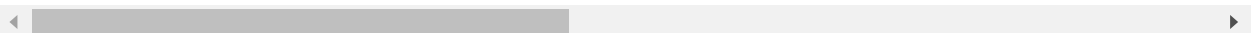
```
In [2]:  #read ipl.csv file and print first 5 records

         df=pd.read_csv("ipl data.csv")
         df
```

Out[2]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal
...	...	...	...	...	...	...	...	...	...
631	632	2016	Raipur	2016-05-22	Delhi Daredevils	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal
632	633	2016	Bangalore	2016-05-24	Gujarat Lions	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal
633	634	2016	Delhi	2016-05-25	Sunrisers Hyderabad	Kolkata Knight Riders	Kolkata Knight Riders	field	normal
634	635	2016	Delhi	2016-05-27	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal
635	636	2016	Bangalore	2016-05-29	Sunrisers Hyderabad	Royal Challengers Bangalore	Sunrisers Hyderabad	bat	normal

636 rows × 18 columns

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

Out[3]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	d
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	

In [4]:
```
#check total number of columns,entries note down your findings

#Findings:-
df.shape
```

Out[4]: `(636, 18)`

In [5]:
```
# find null values
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
```

```
In [6]:  #drop umpire3 column as it contains more than 75% of null values
         df.dropna(subset=['umpire3'])
```

Out[6]:

id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win

```
In [7]:  #drop null values of city , winner ,player_of_match,umpire1,umpire2 column
         df.dropna(subset=['city','winner','player_of_match','umpire1','umpire2'])
```

Out[7]:

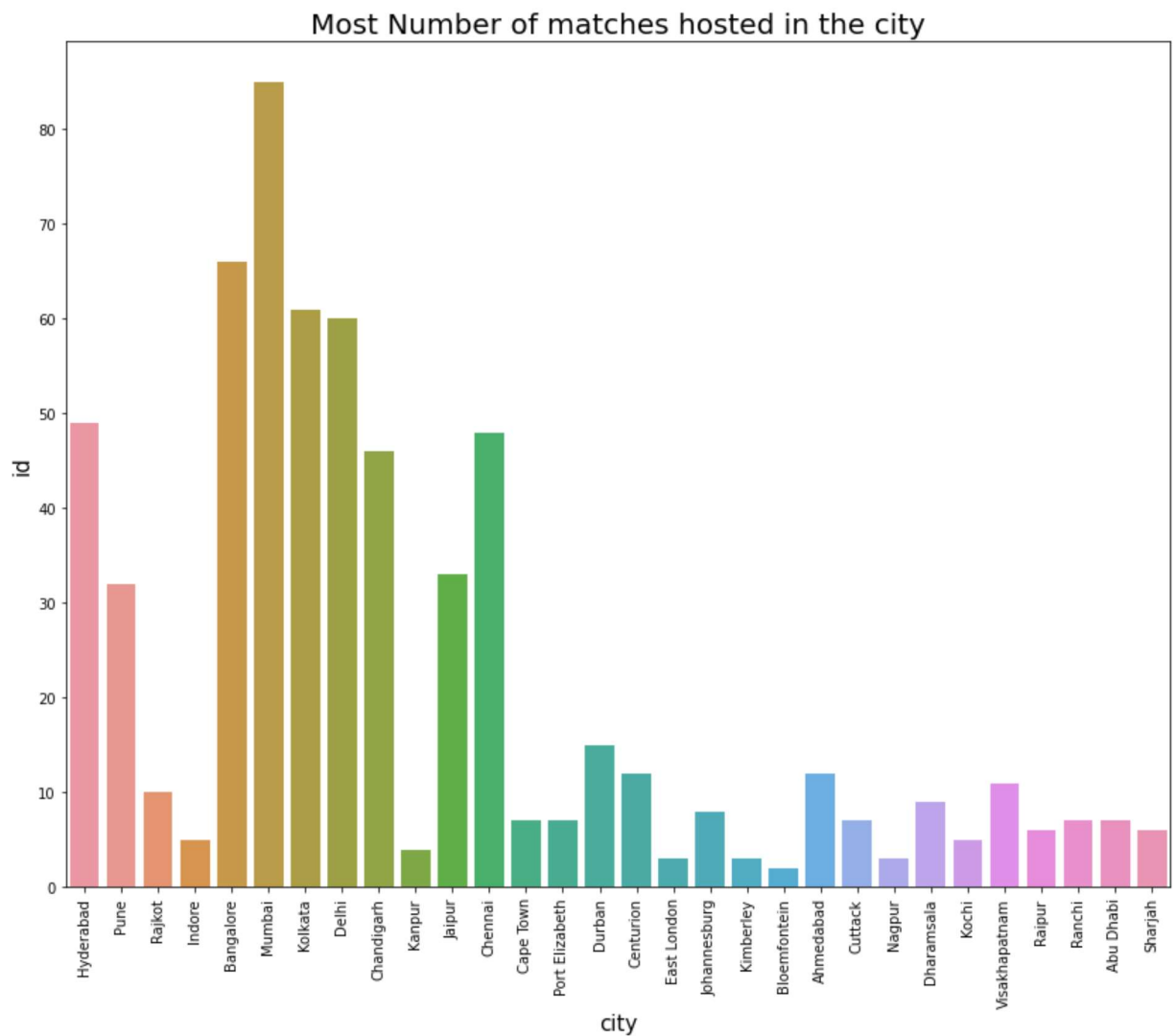|	id	season	city	date	team1	team2	toss_winner	toss_decision	result	d
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	
5	6	2017	Hyderabad	2017-04-09	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	
...	...	...	...	...	...	...	...	...	...	...
631	632	2016	Raipur	2016-05-22	Delhi Daredevils	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	
632	633	2016	Bangalore	2016-05-24	Gujarat Lions	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	
633	634	2016	Delhi	2016-05-25	Sunrisers Hyderabad	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	
634	635	2016	Delhi	2016-05-27	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	
635	636	2016	Bangalore	2016-05-29	Sunrisers Hyderabad	Royal Challengers Bangalore	Sunrisers Hyderabad	bat	normal	

625 rows × 18 columns

In [8]:
```
#which city hosted most number of matches?
a=df.groupby('city')[['id']].count().sort_values('id',ascending=False)[:1]
a
```

Out[8]:

city	id
Mumbai	85

In [9]:
```
#draw bar plot and write down your insights
plt.figure(figsize = (14,11))
sns.countplot(x='city',data=df)
plt.title('Most Number of matches hosted in the city',fontsize=20)
plt.xlabel('city',fontsize=15)
plt.ylabel('id',fontsize=15)
plt.xticks(rotation=90)
plt.show()
```
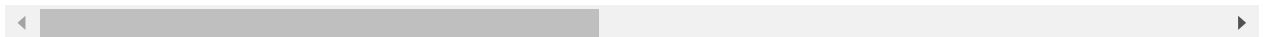
```
In [10]: #find all venue of mumbai city
         b=df[df.city=='Mumbai']
         b
```

Out[10]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_
6	7	2017	Mumbai	2017-04-09	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	normal	
9	10	2017	Mumbai	2017-04-12	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	normal	
15	16	2017	Mumbai	2017-04-16	Gujarat Lions	Mumbai Indians	Mumbai Indians	field	normal	
23	24	2017	Mumbai	2017-04-22	Mumbai Indians	Delhi Daredevils	Delhi Daredevils	field	normal	
27	28	2017	Mumbai	2017-04-24	Rising Pune Supergiant	Mumbai Indians	Mumbai Indians	field	normal	
...	...	...	...	...	...	...	...	...	...	
572	573	2015	Mumbai	2015-05-19	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	normal	
576	577	2016	Mumbai	2016-04-09	Mumbai Indians	Rising Pune Supergiants	Mumbai Indians	bat	normal	
584	585	2016	Mumbai	2016-04-16	Mumbai Indians	Gujarat Lions	Gujarat Lions	field	normal	
589	590	2016	Mumbai	2016-04-20	Royal Challengers Bangalore	Mumbai Indians	Mumbai Indians	field	normal	
599	600	2016	Mumbai	2016-04-28	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	normal	

85 rows × 18 columns

```
In [11]: b.venue.count()
```

Out[11]: 85

```
In [12]:   #now compare in which venue of mumbai most number of matches played (draw bar plo

           c=df[df.city== 'Mumbai']

           d= c.groupby('venue')[['id']].count().sort_values('id',ascending=False)[0:]
           d
```
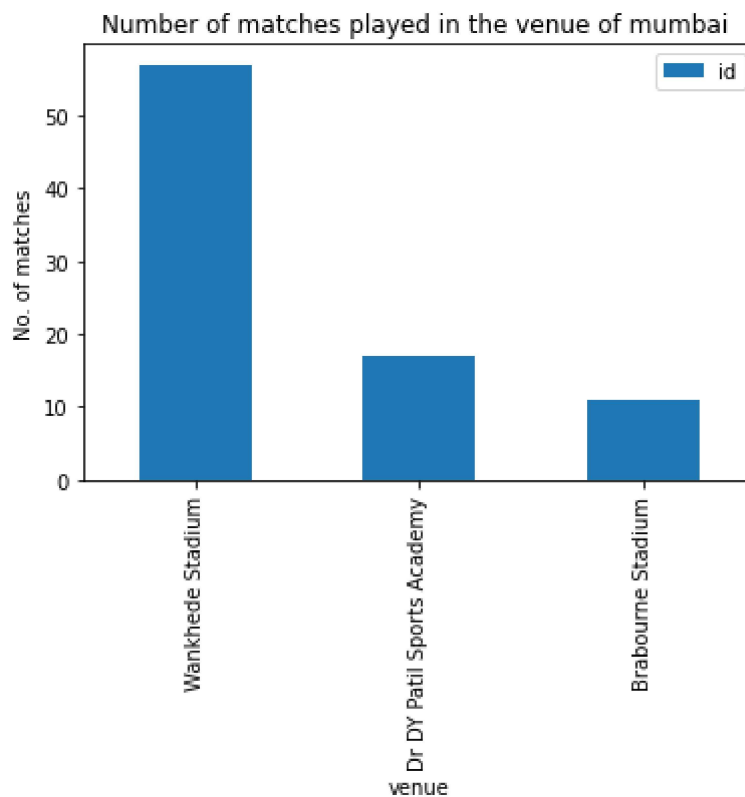
Out[12]:

venue	id
Wankhede Stadium	57
Dr DY Patil Sports Academy	17
Brabourne Stadium	11

```
In [13]:   d.plot.bar()
           plt.title('Number of matches played in the venue of mumbai ')
           plt.xlabel('venue')
           plt.ylabel('No. of matches')

           plt.show()
```
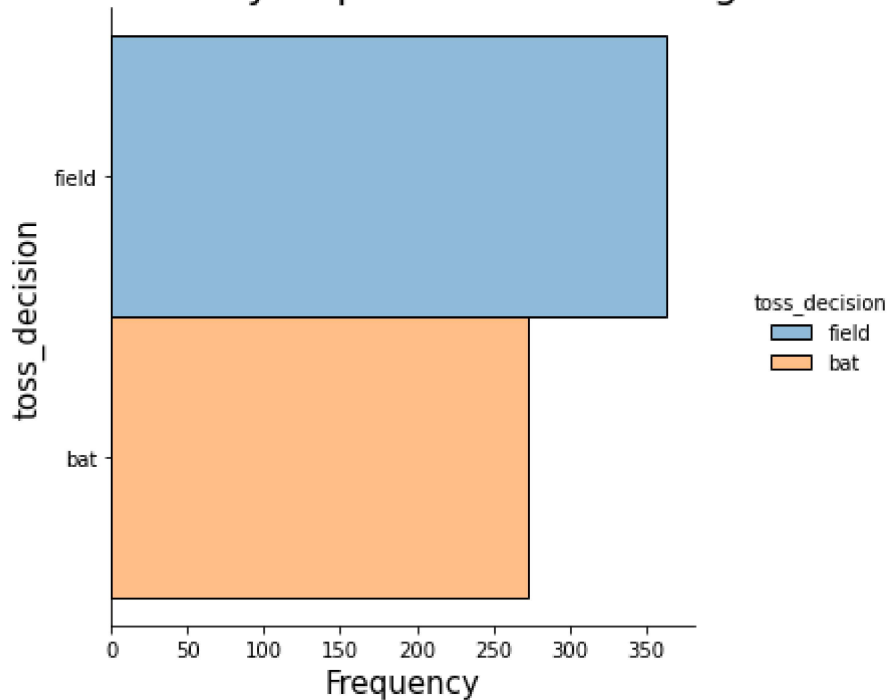


```
In [14]:   #what is the preferred choice after winning a toss in mumbai ?
           e=df[(df['city']=='Mumbai') & (df['toss_decision'] !='bat')]
           f= e['toss_decision'].value_counts()
           f
```

Out[14]:   field   51
           Name: toss_decision, dtype: int64

```
In [15]:  #graphical representation of above question
          plt.figure(figsize = (10,6))
          sns.displot(y='toss_decision',data=df,hue='toss_decision')
          plt.title('Decision taken by Captains after winning tosses ' , fontsize =20)
          plt.xlabel('Frequency',fontsize=15)
          plt.ylabel('toss_decision',fontsize= 15)
          plt.show()
```

<Figure size 720x432 with 0 Axes>


```
In [16]:  df['toss_decision'].head()

Out[16]:  0     field
          1     field
          2     field
          3     field
          4       bat
          Name: toss_decision, dtype: object
```

```
In [17]:  #which team won most number of toss :-Mumbai Indians

          g=df['toss_winner'].value_counts()[:1]
          g

Out[17]:  Mumbai Indians     85
          Name: toss_winner, dtype: int64
```
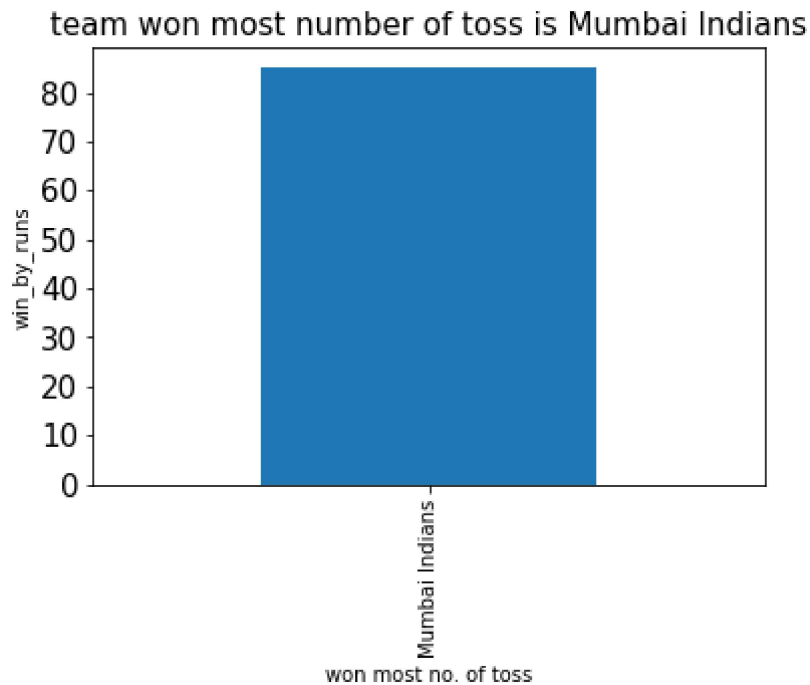
In [18]:
```
#show graphical representation of above questiong

g.plot.bar()
plt.title('team won most number of toss is Mumbai Indians', size =15)
plt.yticks(size = 15)
plt.xlabel('won most no. of toss')
plt.ylabel('win_by_runs')
```

Out[18]: Text(0, 0.5, 'win_by_runs')

In [19]:
```
#find what mumbai indians preferred after winning a toss?

e=df[(df['city']=='Mumbai') & (df['toss_decision'] !='bat')]
f= e['toss_decision'].value_counts()
f
```

Out[19]: field   51
         Name: toss_decision, dtype: int64

```python
#find what mumbai indians preferred after winning a toss?

plt.figure(figsize = (25,10))

sns.countplot('toss_winner', data = df, hue = 'toss_decision')

plt.title("Teams of Mumbai Indians preffered after winning toss", size =25)
plt.xticks(size = 15, rotation=90)
plt.yticks(size = 15)
plt.xlabel("Toss Winner",size=20)
plt.ylabel("Count", size =20)
```
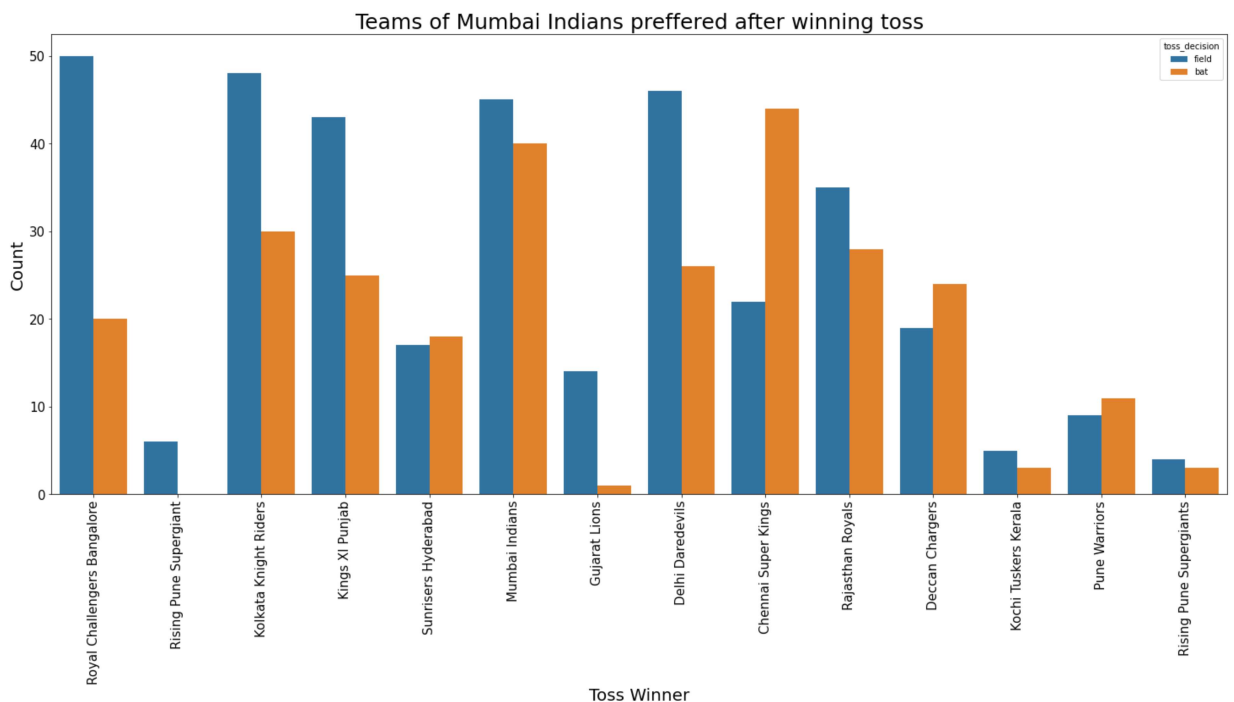
C:\apps\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the fo
llowing variable as a keyword arg: x. From version 0.12, the only valid positio
nal argument will be `data`, and passing other arguments without an explicit ke
yword will result in an error or misinterpretation.
  warnings.warn(

Out[20]: Text(0, 0.5, 'Count')

```
In [21]:  #head to head winning count of Mumbai Indians vs Chennai Super Kings

          h= df[((df.team1 == 'Mumbai Indians') & (df.team2 == 'Chennai Super Kings')) | ((
          h
```

Out[21]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applie
117	118	2009	Cape Town	2009-04-18	Mumbai Indians	Chennai Super Kings	Chennai Super Kings	field	normal	
161	162	2009	Port Elizabeth	2009-05-16	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	normal	
256	257	2011	Mumbai	2011-04-22	Mumbai Indians	Chennai Super Kings	Chennai Super Kings	field	normal	
385	386	2013	Chennai	2013-04-06	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	normal	
429	430	2013	Mumbai	2013-05-05	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	normal	
456	457	2013	Kolkata	2013-05-26	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	normal	
469	470	2014	NaN	2014-04-25	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	normal	
489	490	2014	Mumbai	2014-05-10	Mumbai Indians	Chennai Super Kings	Chennai Super Kings	field	normal	
514	515	2014	Mumbai	2014-05-28	Mumbai Indians	Chennai Super Kings	Chennai Super Kings	field	normal	
529	530	2015	Mumbai	2015-04-17	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	normal	
572	573	2015	Mumbai	2015-05-19	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	normal	
575	576	2015	Kolkata	2015-05-24	Mumbai Indians	Chennai Super Kings	Chennai Super Kings	field	normal	

```
In [22]:   #Which team won most of the matches in mumbai?
           df[(df.city)=='Mumbai']['winner'].value_counts()
```

Out[22]:   Mumbai Indians   45
           Chennai Super Kings   8
           Kings XI Punjab   5
           Royal Challengers Bangalore   5
           Rajasthan Royals   5
           Deccan Chargers   3
           Kolkata Knight Riders   3
           Pune Warriors   3
           Rising Pune Supergiant   2
           Delhi Daredevils   2
           Kochi Tuskers Kerala   1
           Sunrisers Hyderabad   1
           Rising Pune Supergiants   1
           Gujarat Lions   1
           Name: winner, dtype: int64

```
In [23]:   #how many times each team won the toss and won the match in mumbai
           df[(df.city)=='Mumbai']['toss_winner'].value_counts()
```

Out[23]:   Mumbai Indians   39
           Delhi Daredevils   9
           Chennai Super Kings   7
           Deccan Chargers   6
           Rajasthan Royals   6
           Royal Challengers Bangalore   4
           Kings XI Punjab   4
           Kolkata Knight Riders   3
           Pune Warriors   3
           Kochi Tuskers Kerala   2
           Sunrisers Hyderabad   1
           Gujarat Lions   1
           Name: toss_winner, dtype: int64

```
In [24]:   #which venue hosted most number of matches
           df['venue'].max()
```

Out[24]:   'Wankhede Stadium'

```
In [25]: #find how many matches chennai super kings played at M Chinnaswamy stadium?

k= df[(df['venue']== 'M Chinnaswamy Stadium') & ((df['team1'] == 'Chennai Super K
k
```

Out[25]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	c
73	74	2008	Bangalore	2008-04-28	Chennai Super Kings	Royal Challengers Bangalore	Chennai Super Kings	bat	normal	
191	192	2010	Bangalore	2010-03-23	Royal Challengers Bangalore	Chennai Super Kings	Chennai Super Kings	field	normal	
301	302	2011	Bangalore	2011-05-22	Chennai Super Kings	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	
378	379	2012	Bangalore	2012-05-23	Chennai Super Kings	Mumbai Indians	Mumbai Indians	field	normal	
451	452	2013	Bangalore	2013-05-18	Royal Challengers Bangalore	Chennai Super Kings	Chennai Super Kings	field	normal	
509	510	2014	Bangalore	2014-05-24	Royal Challengers Bangalore	Chennai Super Kings	Chennai Super Kings	field	normal	
537	538	2015	Bangalore	2015-04-22	Chennai Super Kings	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	

```
In [26]: #who won most matches at M Chinnaswamy stadium?--Royal Challengers Bangalore
r= df[(df.venue == 'M Chinnaswamy Stadium')].sort_values('id' , ascending=False)
r
```

Out[26]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl
635	636	2016	Bangalore	2016-05-29	Sunrisers Hyderabad	Royal Challengers Bangalore	Sunrisers Hyderabad	bat	normal	

```python
In [27]:   #matches played in each year
           df['season'].value_counts()
```

Out[27]:   2013   76
           2012   74
           2011   73
           2010   60
           2014   60
           2016   60
           2017   59
           2015   59
           2008   58
           2009   57
           Name: season, dtype: int64

```python
In [47]:   #which city hosted most number of matches in 2013
           z= df[(df['season']==2013)][('city')].value_counts()
           z
```

Out[47]:   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

```python
In [83]:   #lets analyse ipl season held in 2013

           z= df[(df['season']==2013)][('city')].value_counts()
           z
```

Out[83]:   Kolkata   8
           Bangalore   8
           Hyderabad   8
           Delhi   8
           Chennai   8
           Pune   8
           Jaipur   8
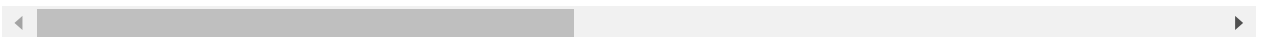   Mumbai   8
           Chandigarh   6
           Dharamsala   2
           Raipur   2
           Ranchi   2
           Name: city, dtype: int64
```

```
In [84]: #extract all the details of 2013 season

         df[(df['season']==2013)]
```

Out[84]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	d
381	382	2013	Kolkata	2013-04-03	Delhi Daredevils	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	
382	383	2013	Bangalore	2013-04-04	Royal Challengers Bangalore	Mumbai Indians	Mumbai Indians	field	normal	
383	384	2013	Hyderabad	2013-04-05	Sunrisers Hyderabad	Pune Warriors	Pune Warriors	field	normal	
384	385	2013	Delhi	2013-04-06	Rajasthan Royals	Delhi Daredevils	Rajasthan Royals	bat	normal	
385	386	2013	Chennai	2013-04-06	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	normal	
...	...	...	...	...	...	...	...	...	...	
452	453	2013	Hyderabad	2013-05-19	Kolkata Knight Riders	Sunrisers Hyderabad	Kolkata Knight Riders	bat	normal	
453	454	2013	Delhi	2013-05-21	Chennai Super Kings	Mumbai Indians	Chennai Super Kings	bat	normal	
454	455	2013	Delhi	2013-05-22	Sunrisers Hyderabad	Rajasthan Royals	Sunrisers Hyderabad	bat	normal	
455	456	2013	Kolkata	2013-05-24	Rajasthan Royals	Mumbai Indians	Rajasthan Royals	bat	normal	
456	457	2013	Kolkata	2013-05-26	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	normal	

76 rows × 18 columns

```
In [96]:  #here we will create yearwise groups (hint: use groupby() function)

          df.groupby('season').size()
```

Out[96]: season
          2008   58
          2009   57
          2010   60
          2011   73
          2012   74
          2013   76
          2014   60
          2015   59
          2016   60
          2017   59
          dtype: int64

In [ ]: