In [1]:

```
#import all libraries
import pandas as pd
import numpy as np
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.head()
```

Out[2]:

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

In [3]:

```
#check total number of columns,entries note down your findings

#Findings:-
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
```

In [4]:

```
#drop umpire3 column as it contains more than 75% of null values
df.drop(["umpire3"],axis=1,inplace=True)
```

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
dtype: int64
```

In [6]:

```
df
```

Out[6]:

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

636 rows × 17 columns

In [7]:

```
#drop null values of city , winner ,player_of_match,umpire1,umpire2 column
df.dropna(inplace=True)
```

In [8]:

```
df.isnull().sum()
```

Out[8]:

```
id                 0
season             0
city               0
date               0
team1              0
team2              0
toss_winner        0
toss_decision      0
result             0
dl_applied         0
winner             0
win_by_runs        0
win_by_wickets     0
player_of_match    0
venue              0
umpire1            0
umpire2            0
dtype: int64
```

In [9]:

```
df
```

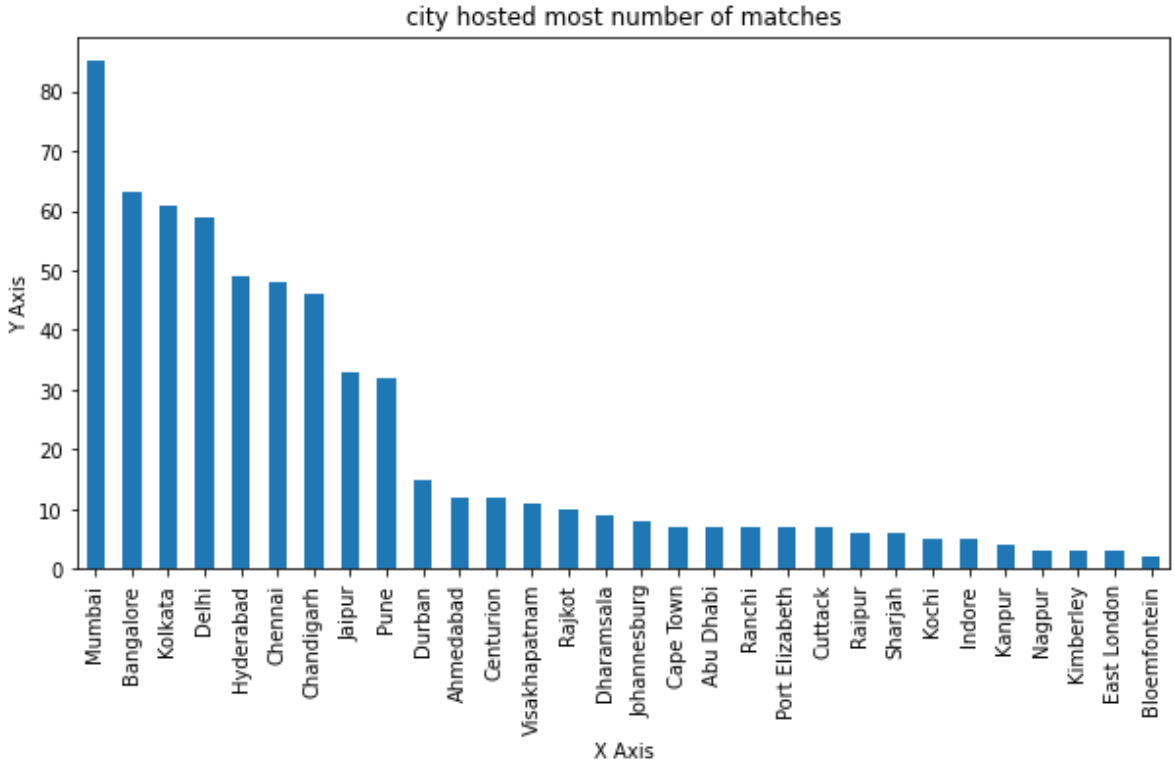
Out[9]:

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

625 rows × 17 columns

In [10]:

```
#which city hosted most number of matches?
#draw bar plot and write down your insights
plt.figure(figsize=(10,5))
df["city"].value_counts().plot(kind="bar")
plt.title("city hosted most number of matches")
plt.xlabel("X Axis")
plt.ylabel("Y Axis")
plt.show()
```



In [24]:

```
#find all venue of mumbai city
df[(df["city"]=="Mumbai")][["venue"]].shape
```

Out[24]:

```
(85, 1)
```

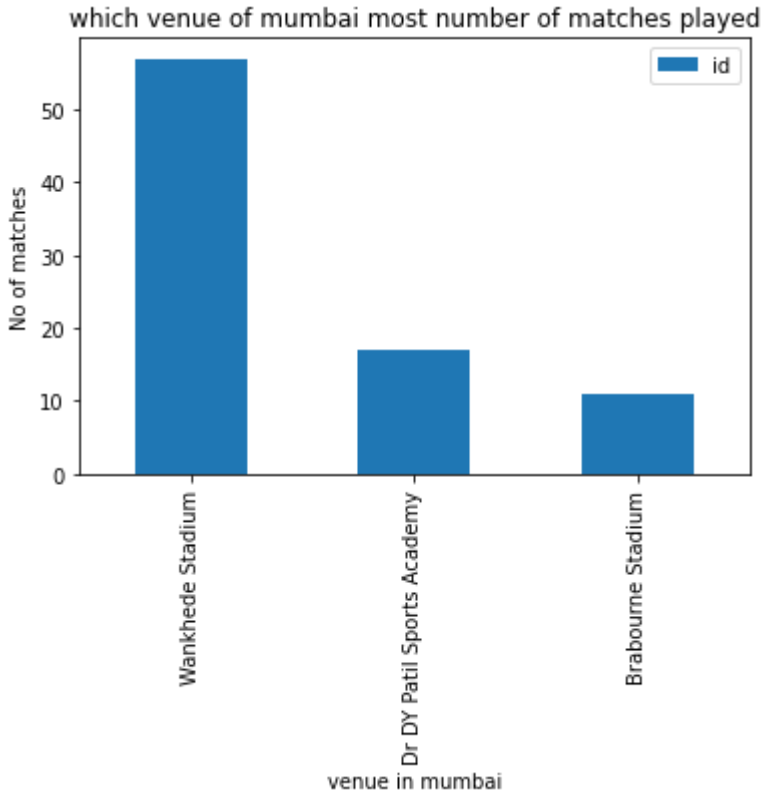
In [12]:

```
#now compare in which venue of mumbai most number of matches played (draw bar plot and  wri
plt.figure(figsize=(10,5))
a=df[df["city"]=="Mumbai"].venue
b=df.groupby(a)[["id"]].count().sort_values("id",ascending=False)
b.plot.bar()
plt.title("which venue of mumbai most number of matches played")
plt.xlabel("venue in mumbai")
plt.ylabel("No of matches")
```

Out[12]:

```
Text(0, 0.5, 'No of matches')
```

```
<Figure size 720x360 with 0 Axes>
```



In [13]:

```
df.head(2)
```

Out[13]:

	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	

In [14]:

```
#what is the preferred choice after winning a toss in mumbai
df['winner'].loc[(df['toss_decision']=='field')].value_counts()
```

Out[14]:

```
Mumbai Indians                 51
Kings XI Punjab                49
Royal Challengers Bangalore    46
Kolkata Knight Riders          43
Delhi Daredevils               34
Rajasthan Royals               33
Chennai Super Kings            28
Sunrisers Hyderabad            25
Deccan Chargers                15
Gujarat Lions                  11
Rising Pune Supergiant         10
Kochi Tuskers Kerala            6
Pune Warriors                   3
Rising Pune Supergiants         3
Name: winner, dtype: int64
```

In [15]:

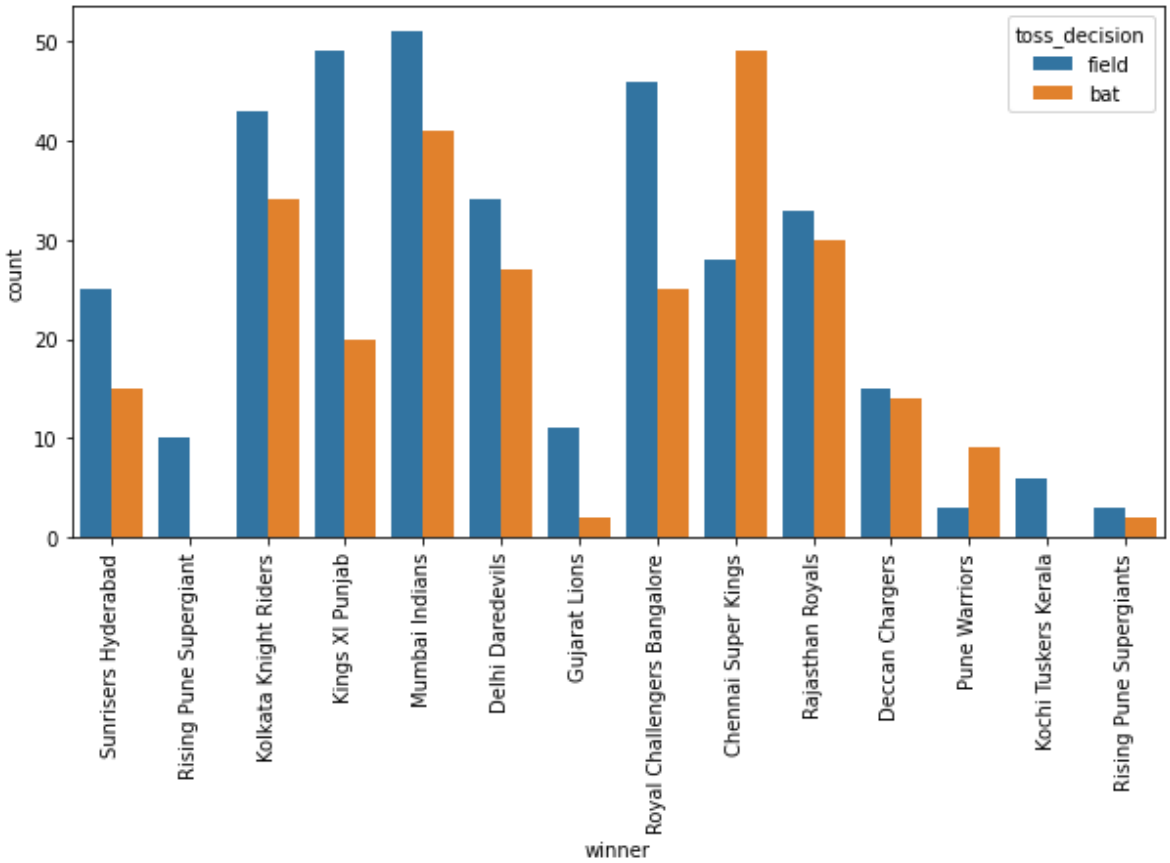
```
df['winner'].loc[(df['toss_decision']=='bat')].value_counts()
```

Out[15]:

```
Chennai Super Kings            49
Mumbai Indians                 41
Kolkata Knight Riders          34
Rajasthan Royals               30
Delhi Daredevils               27
Royal Challengers Bangalore    25
Kings XI Punjab                20
Sunrisers Hyderabad            15
Deccan Chargers                14
Pune Warriors                   9
Rising Pune Supergiants         2
Gujarat Lions                   2
Name: winner, dtype: int64
```

In [25]:

```
#graphical representation of above question
plt.figure(figsize=(10,5))
sns.countplot(x='winner',data=df,hue='toss_decision')
plt.xticks(rotation=90)
plt.show()
```



In [18]:

```
df.head(2)
```

Out[18]:

	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	

In [19]:

```
#which team won most number of toss :-Mumbai Indians
#df['toss_winner']=="Mumbai Indians"
df[(df["toss_winner"]=="Mumbai Indians")][["toss_decision"]].value_counts()
```

Out[19]:

```
toss_decision
field           44
bat             39
dtype: int64
```

In [20]:

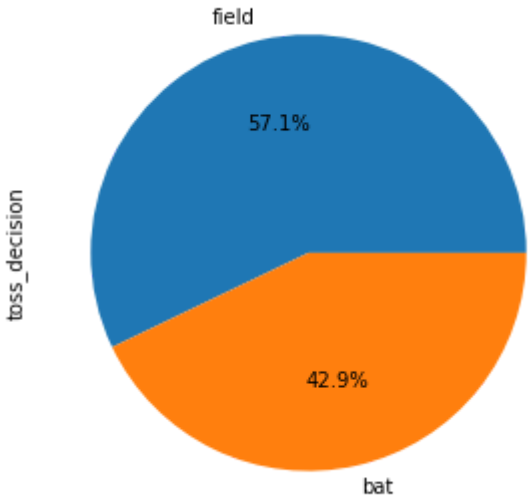
```
#show graphical representation of above question

plt.figure(figsize=(10,5))

df["toss_decision"].value_counts().plot.pie(autopct="%1.1f%%")
```

Out[20]:

```
<AxesSubplot:ylabel='toss_decision'>
```



In [26]:

```
#find what mumbai indians preferred after winning a toss?
df[(df["toss_winner"]=="Mumbai Indians")].shape
```

Out[26]:

```
(83, 17)
```

In [22]:

```
df.head(5)
```

Out[22]:

	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	

In [94]:

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

a=df[((df.team1=="Chennai Super Kings")|(df.team2=="Chennai Super Kings")&((df.team1=="Mumb
a[:2]
```

Out[94]:

```
Chennai Super Kings    48
Mumbai Indians         12
Name: winner, dtype: int64
```

In [37]:

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

Out[37]:

```
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 [38]:

```
#how many times each team won the toss and won the match in mumbai
df[df.toss_winner==df.winner].winner.value_counts()
```

Out[38]:

```
Mumbai Indians                 48
Kolkata Knight Riders          44
Chennai Super Kings            42
Rajasthan Royals               34
Delhi Daredevils               33
Royal Challengers Bangalore    33
Kings XI Punjab                27
Deccan Chargers                19
Sunrisers Hyderabad            16
Gujarat Lions                  10
Rising Pune Supergiant          5
Kochi Tuskers Kerala            4
Pune Warriors                   3
Rising Pune Supergiants         3
Name: winner, dtype: int64
```

In [59]:

```
#which venue hosted most number of matches
df["venue"].value_counts()
```

Out[59]:

```
M Chinnaswamy Stadium                                63
Eden Gardens                                         61
Feroz Shah Kotla                                     59
Wankhede Stadium                                     57
Rajiv Gandhi International Stadium, Uppal            49
MA Chidambaram Stadium, Chepauk                      48
Punjab Cricket Association Stadium, Mohali           35
Sawai Mansingh Stadium                               33
Dr DY Patil Sports Academy                           17
Subrata Roy Sahara Stadium                           17
Maharashtra Cricket Association Stadium              15
Kingsmead                                            15
Sardar Patel Stadium, Motera                         12
SuperSport Park                                      12
Punjab Cricket Association IS Bindra Stadium, Mohali 11
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium  11
Brabourne Stadium                                    11
Saurashtra Cricket Association Stadium               10
Himachal Pradesh Cricket Association Stadium          9
New Wanderers Stadium                                 8
Newlands                                              7
St George's Park                                      7
Sheikh Zayed Stadium                                  7
JSCA International Stadium Complex                    7
Barabati Stadium                                      7
Shaheed Veer Narayan Singh International Stadium      6
Sharjah Cricket Stadium                               6
Holkar Cricket Stadium                                5
Nehru Stadium                                         5
Green Park                                            4
Vidarbha Cricket Association Stadium, Jamtha          3
De Beers Diamond Oval                                 3
Buffalo Park                                          3
OUTsurance Oval                                       2
Name: venue, dtype: int64
```

In [46]:

```
#find how many matches chennai super kings played at M Chinnaswamy stadium?
#df[((df.team1=="chennai super kings") | (df.team2=="chennai super kings")) &(df.venue=="M
a=df[(df["venue"]=="M Chinnaswamy stadium") &((df["team1"]=="Chennai Super Kings")) |((df["
a
```

Out[46]:

```
(53, 17)
```

In [47]:

```
#who won most matches at M Chinnaswamy stadium?--Royal Challengers Bangalore
df[df["venue"]=="M Chinnaswamy Stadium"].winner.value_counts()
```

Out[47]:

```
Royal Challengers Bangalore    29
Mumbai Indians                  8
Kolkata Knight Riders           6
Kings XI Punjab                 5
Chennai Super Kings             4
Rajasthan Royals                3
Delhi Daredevils                3
Sunrisers Hyderabad             2
Rising Pune Supergiant          1
Gujarat Lions                   1
Deccan Chargers                 1
Name: winner, dtype: int64
```

In [48]:

```
#matches played in each year
df.season.value_counts()
```

Out[48]:

```
2013    76
2012    74
2011    72
2010    60
2016    60
2017    58
2008    58
2009    57
2015    57
2014    53
Name: season, dtype: int64
```

In [49]:

```
#which city hosted most number of matches in 2013
df[df["season"]==2013].city.value_counts()
```

Out[49]:

```
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 [60]:

```
#lets analyse ipl season held in 2013
#extract all the details of 2013 season

df[df["season"]==2013].shape
```

Out[60]:

```
(76, 17)
```

In [66]:

```
#here we will create yearwise groups (hint: use groupby() function)
#df.groupby("season").winner.value_counts()
df.groupby("season").winner.value_counts()
```

Out[66]:

```
season  winner
2008    Rajasthan Royals               13
        Kings XI Punjab                10
        Chennai Super Kings             9
        Delhi Daredevils                7
        Mumbai Indians                  7
                                       ..
2017    Sunrisers Hyderabad             8
        Kings XI Punjab                 7
        Delhi Daredevils                6
        Gujarat Lions                   4
        Royal Challengers Bangalore     2
Name: winner, Length: 84, dtype: int64
```

In [ ]: