```
In [1]: # import all libraries
        import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
        import warnings as w
        w.filterwarnings("ignore")
```

```
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	resu
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
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	norma

In [3]:
```python
#check total number of columns,entries note down your 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]:
```python
# in this ipl dataset we have 18 columns and 636 records
# in this dataset only umpire3 column have float datatype and id, season, dl_a|
# integer datatype and rest of columns have object datatype
```

In [5]: ```python
# 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]: ```python
# there are 7 null values in city column, 3 null values in winner, player_of_m
# column, 636 null values in umpire3 column and rest of columns don't have any
```

In [7]: ```python
#drop umpire3 column as it contains more than 75% of null values
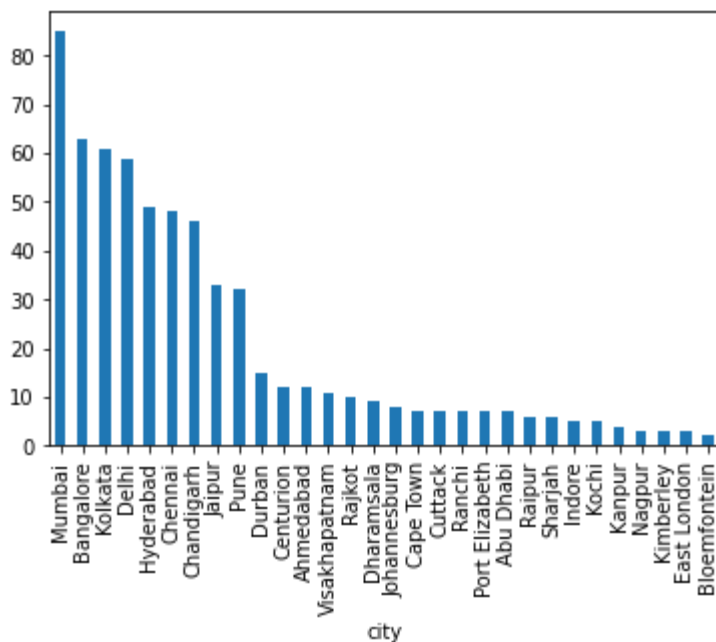df.drop("umpire3",inplace=True,axis=1)
```

In [9]: `df`

Out[9]:

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

636 rows × 17 columns

In [10]: 
```
#drop null values of city , winner ,player_of_match,umpire1,umpire2 columns
df.dropna(inplace=True)
```

In [18]: 
```python
#which city hosted most number of matches?
#draw bar plot and write down your insights
df.groupby(["city"])["id"].count().sort_values(ascending=False).plot(kind="bar
```

Out[18]: <AxesSubplot:xlabel='city'>

In [19]:
```python
# from the above graph mumbai city host most no. of matches and bloemfontein c
```

In [26]:
```python
#find all venue of mumbai city
df[df["city"]=="Mumbai"]["venue"].unique()
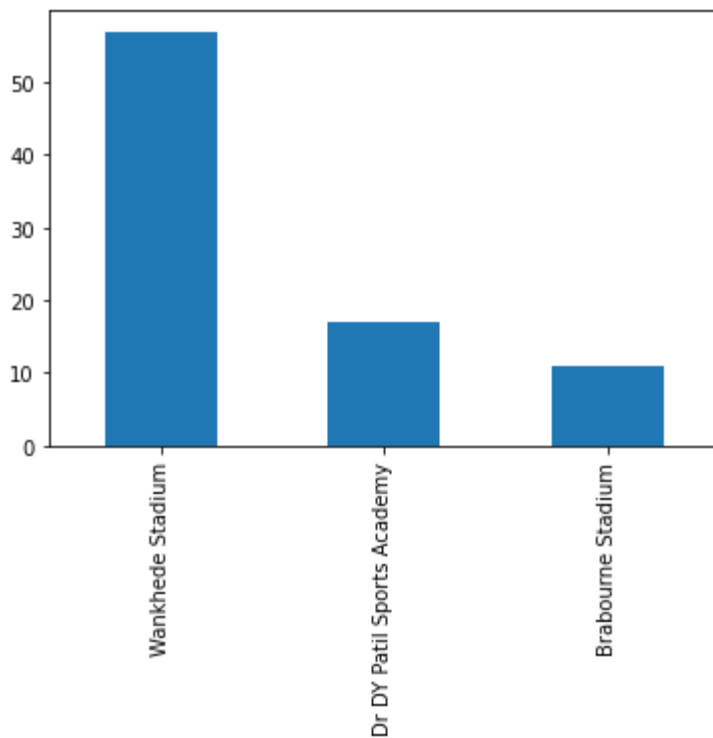```

Out[26]: array(['Wankhede Stadium', 'Dr DY Patil Sports Academy',
             'Brabourne Stadium'], dtype=object)

In [30]: 
```python
#now compare in which venue of mumbai most number of matches played (draw bar
a=df.groupby("city").get_group("Mumbai")["venue"].value_counts()
a
```

Out[30]:
```
Wankhede Stadium              57
Dr DY Patil Sports Academy    17
Brabourne Stadium             11
Name: venue, dtype: int64
```

```
In [31]:  a.plot(kind="bar")
```

Out[31]: <AxesSubplot:>


```
In [32]:  # in the mumbai city there are three venue   which is Wankhede Stadium, Dr DY P
          # and in this stadium 57 matches are played in Wankhede Stadium, 17 matches ar
          # 11 matches are plyed in Brabourne Stadium
```

```
In [36]:  #what is the preferred choice after winning a toss in mumbai
          b=df.groupby("city").get_group("Mumbai")["toss_decision"].value_counts()
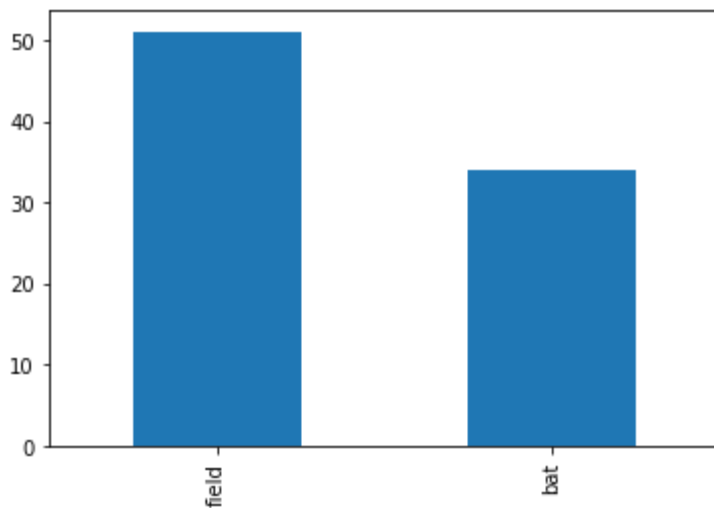          b
```

```
Out[36]:  field    51
          bat      34
          Name: toss_decision, dtype: int64
```

```
In [37]:  # from the given dataset it shows that mumbia prefered 51 times field and 34 t
```

```
In [38]:  #graphical representation of above question
```

In [40]:
```python
b.plot(kind="bar")
```

Out[40]: <AxesSubplot:>

In [44]:
```python
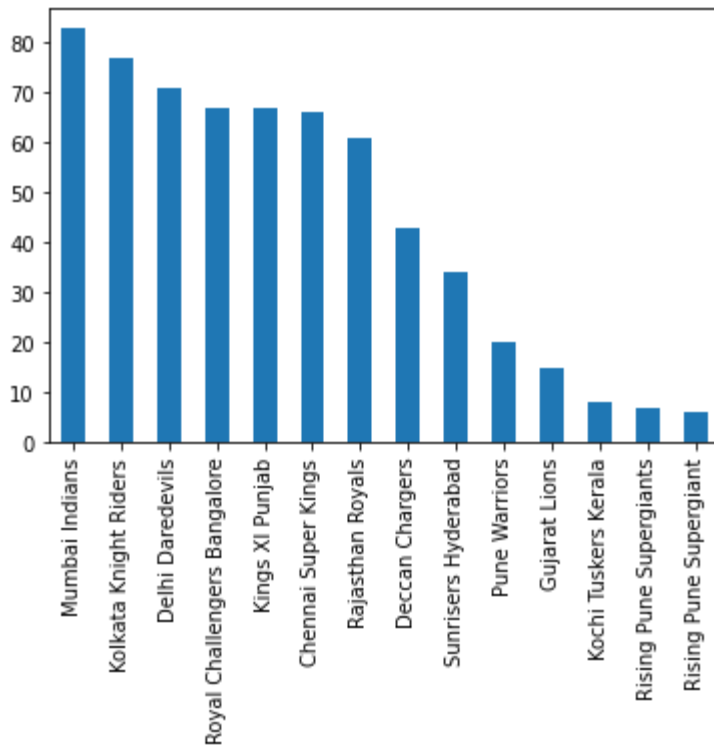#which team won most number of toss
c=df["toss_winner"].value_counts()
c
```

Out[44]:
```
Mumbai Indians                83
Kolkata Knight Riders         77
Delhi Daredevils              71
Royal Challengers Bangalore   67
Kings XI Punjab               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
Name: toss_winner, dtype: int64
```

In [45]:
```python
# Mumbai Indians won most no. of toss which is 83 times while Rising Pune Supe
```

In [48]:
```python
#show graphical representation of above question
c.plot(kind="bar")
```

Out[48]: <AxesSubplot:>

In [69]:
```python
#find what mumbai indians preferred after winning a toss?
df[df["toss_winner"]=="Mumbai Indians"]["toss_decision"].value_counts()
```

Out[69]:
```
field    44
bat      39
Name: toss_decision, dtype: int64
```

In [71]:
```python
# after winning toss Mumbai Indians prefer field for 44 times and bat for 39 t
```

In [75]:
```python
#Which team won most of the matches in mumbai?
```

In [74]:
```python
df.groupby("city").get_group("Mumbai")["winner"].value_counts()
```

Out[74]:
```
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 [81]:
```python
#how many times each team won the toss and won the match in mumbai
df[(df['toss_winner']==df['winner'])&(df['city']=='Mumbai')]['winner'].value_c
```

Out[81]:
```
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
```

In [82]:
```python
#which venue hosted most number of matches
df["venue"].value_counts()
```

Out[82]:
```
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 [87]:
```python
#find how many matches chennai super kings played at M Chinnaswamy stadium?
a=df[((df["team1"]=="Chennai Super Kings") | (df["team2"]=="Chennai Super King
len(a)
```

Out[87]: 7

In [89]:
```python
#who won most matches at M Chinnaswamy stadium?
df[df["venue"]=="M Chinnaswamy Stadium"]["winner"].value_counts()
# in Royal Challengers Bangalore won most no. of matches which is 29 in M Chin
```

Out[89]:
```
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 [90]:
```python
#matches played in each year
df["season"].value_counts()
```

Out[90]:
```
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 [ ]:
```python
#in 2013 matches played is 76, in 2012 matches played is 74, in 2011 matches p
#in 2016 matches played is 60, in 2017 matches played is 58, in 2008 matches p
#in 2015 matches played is 57, in 2014 matches played is 53,
```

In [97]:
```python
# which city hosted most number of matches in 2013
```

In [94]:
```python
df[df["season"]==2013]["city"].value_counts()
```

Out[94]:
```
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 [ ]:
```
# kolkata,Bangalore,Hydrabad,Delhi,Chennai,Pune,Jaiput,Mumbai cities host most
# while Chandigarh won 6 matches and in the end Raiput, Ranchi won least no. o
```

In [96]:
```
#lets analyse ipl season held in 2013
#extract all the details of 2013 season
#here we will create yearwise groups (hint: use groupby() function)
df.groupby('season').get_group(2013)
```

Out[96]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	re
381	382	2013	Kolkata	2013-04-03	Delhi Daredevils	Kolkata Knight Riders	Kolkata Knight Riders	field	no
382	383	2013	Bangalore	2013-04-04	Royal Challengers Bangalore	Mumbai Indians	Mumbai Indians	field	no
383	384	2013	Hyderabad	2013-04-05	Sunrisers Hyderabad	Pune Warriors	Pune Warriors	field	no
384	385	2013	Delhi	2013-04-06	Rajasthan Royals	Delhi Daredevils	Rajasthan Royals	bat	no
385	386	2013	Chennai	2013-04-06	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	no
...	...	...	...	...	...	...	...	...	
452	453	2013	Hyderabad	2013-05-19	Kolkata Knight Riders	Sunrisers Hyderabad	Kolkata Knight Riders	bat	no
453	454	2013	Delhi	2013-05-21	Chennai Super Kings	Mumbai Indians	Chennai Super Kings	bat	no
454	455	2013	Delhi	2013-05-22	Sunrisers Hyderabad	Rajasthan Royals	Sunrisers Hyderabad	bat	no
455	456	2013	Kolkata	2013-05-24	Rajasthan Royals	Mumbai Indians	Rajasthan Royals	bat	no
456	457	2013	Kolkata	2013-05-26	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	no

76 rows × 17 columns

In [ ]: