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

In [2]:
```
#read ipl.csv file and print first 5 records
df=pd.read_csv("ipl data.csv")
df.head(5)
```

Out[2]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	35	0
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	0	7
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	10
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	0	6
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	0	Royal Challengers Bangalore	15	0

In [3]:
```
#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]:
```
# find null values
df.isnull().sum()
```

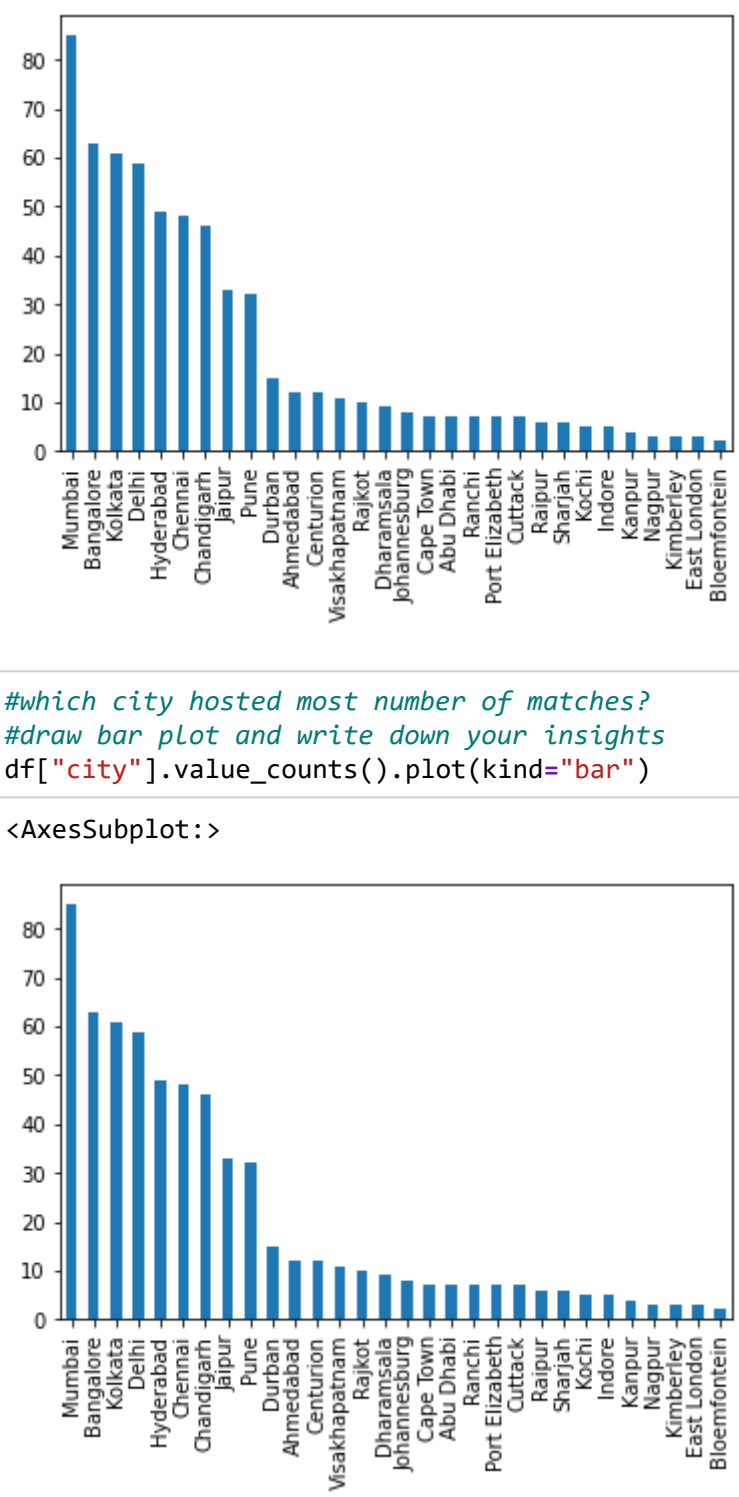
Out[4]:
```
id                 0
season             0
city               7
date               0
team1              0
team2              0
toss_winner        0
toss_decision      0
result             0
dl_applied         0
winner             3
win_by_runs        0
win_by_wickets     0
player_of_match    3
venue              0
umpire1            1
umpire2            1
umpire3          636
dtype: int64
```

In [5]:
```
#drop umpire3 column as it contains more than 75% of null values
df.drop(["umpire3"], axis=1, inplace=True)
```

In [6]:
```
#drop null values of city , winner ,player_of_match,umpire1,umpire2 column
df.dropna(subset=["city","winner","player_of_match","umpire1","umpire2"],inplace=True)
```

In [7]:
```
#The most matches played are in Mumbai and the least matches played are in Bloemfontein
df["city"].value_counts().plot(kind="bar")
```

Out[7]: <AxesSubplot:>



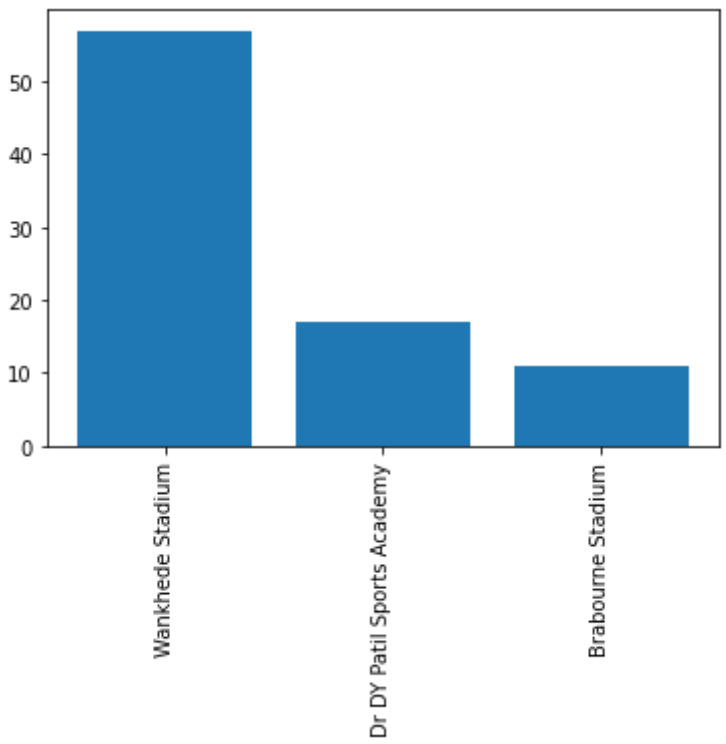
In [8]:
```
#which city hosted most number of matches?
#draw bar plot and write down your insights
df["city"].value_counts().plot(kind="bar")
```

Out[8]: <AxesSubplot:>

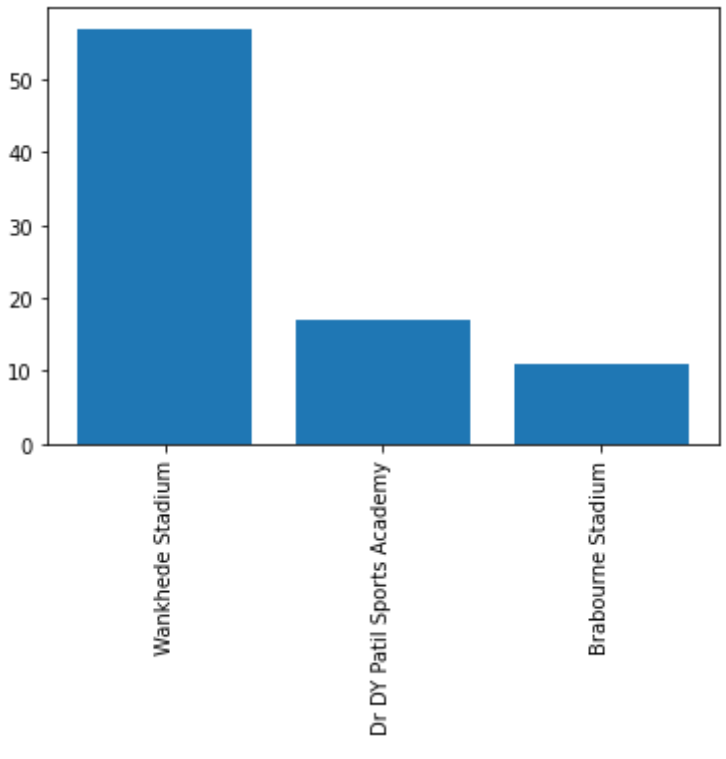
In [9]:
```
#find all venue of mumbai city
df[df["city"]=="Mumbai"]["venue"].unique()
```

Out[9]:
```
array(['Wankhede Stadium', 'Dr DY Patil Sports Academy',
       'Brabourne Stadium'], dtype=object)
```

In [10]:
```
#The most matches played in Mumbai were in Wankhede Stadium which were 57 matches in total.
a=df[df['city']=='Mumbai']['venue'].unique()
b=df[df['city']=='Mumbai']['venue'].value_counts()
plt.xticks(rotation=90)
plt.bar(a,b)
plt.show()
```



In [11]:
```
#now compare in which venue of mumbai most number of matches played (draw bar plot and  write down insights)
a=df[df['city']=='Mumbai']['venue'].unique()
b=df[df['city']=='Mumbai']['venue'].value_counts()
plt.xticks(rotation=90)
plt.bar(a,b)
plt.show()
```

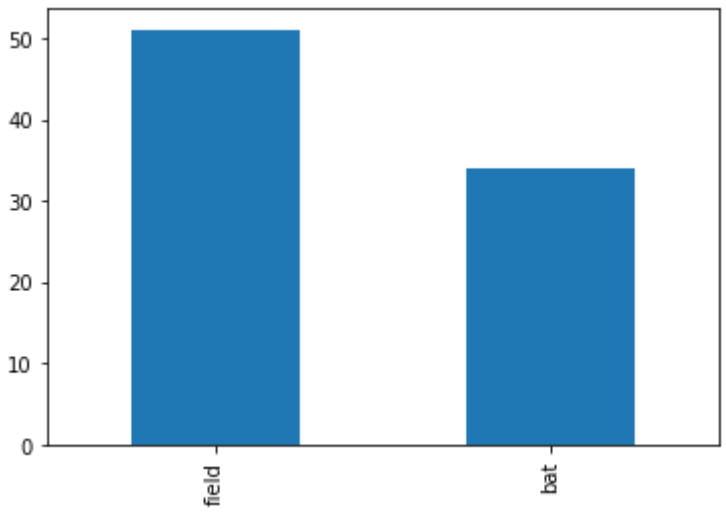


In [12]:
```
#what is the preferred choice after winning a toss in mumbai
df[df['city']=='Mumbai']['toss_decision'].value_counts()
```

Out[12]:
```
field    51
bat      34
Name: toss_decision, dtype: int64
```

In [13]:
```
#graphical representation of above question
df[df['city']=='Mumbai']['toss_decision'].value_counts().plot(kind='bar')
```

Out[13]: <AxesSubplot:>

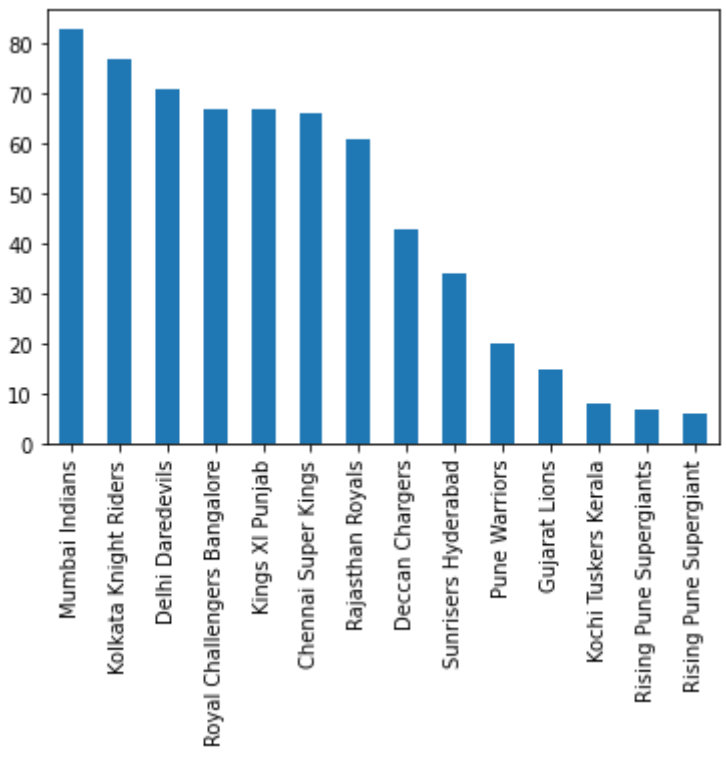


In [14]:
```
#which team won most number of toss
df['toss_winner'].value_counts()
```

Out[14]:
```
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 [15]:
```
#show graphical representation of above question
df['toss_winner'].value_counts().plot(kind='bar')
```

Out[15]: <AxesSubplot:>



In [16]:
```
#find what mumbai indians preferred after winning a toss?
df[df['toss_winner']=='Mumbai Indians']['toss_decision'].value_counts()
```

Out[16]:
```
field    44
bat      39
Name: toss_decision, dtype: int64
```

In [17]:
```
#head to head winning count of Mumbai Indians vs Chennai Super Kings
df[(df['team1']=='Chennai Super Kings')&(df['team2']=='Mumbai Indians')|((df['team2']=='Chennai Super Kings')&(df['team1
```

Out[17]:
```
Mumbai Indians         12
Chennai Super Kings     9
Name: winner, dtype: int64
```

In [18]:
```
#Which team won most of the matches in mumbai?
df[df['city']=='Mumbai']['winner'].value_counts()
```

Out[18]:
```
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 [27]:
```
#how many times each team won the toss and won the match
df[(df['toss_winner']==df['winner'])]['winner'].value_counts()
```

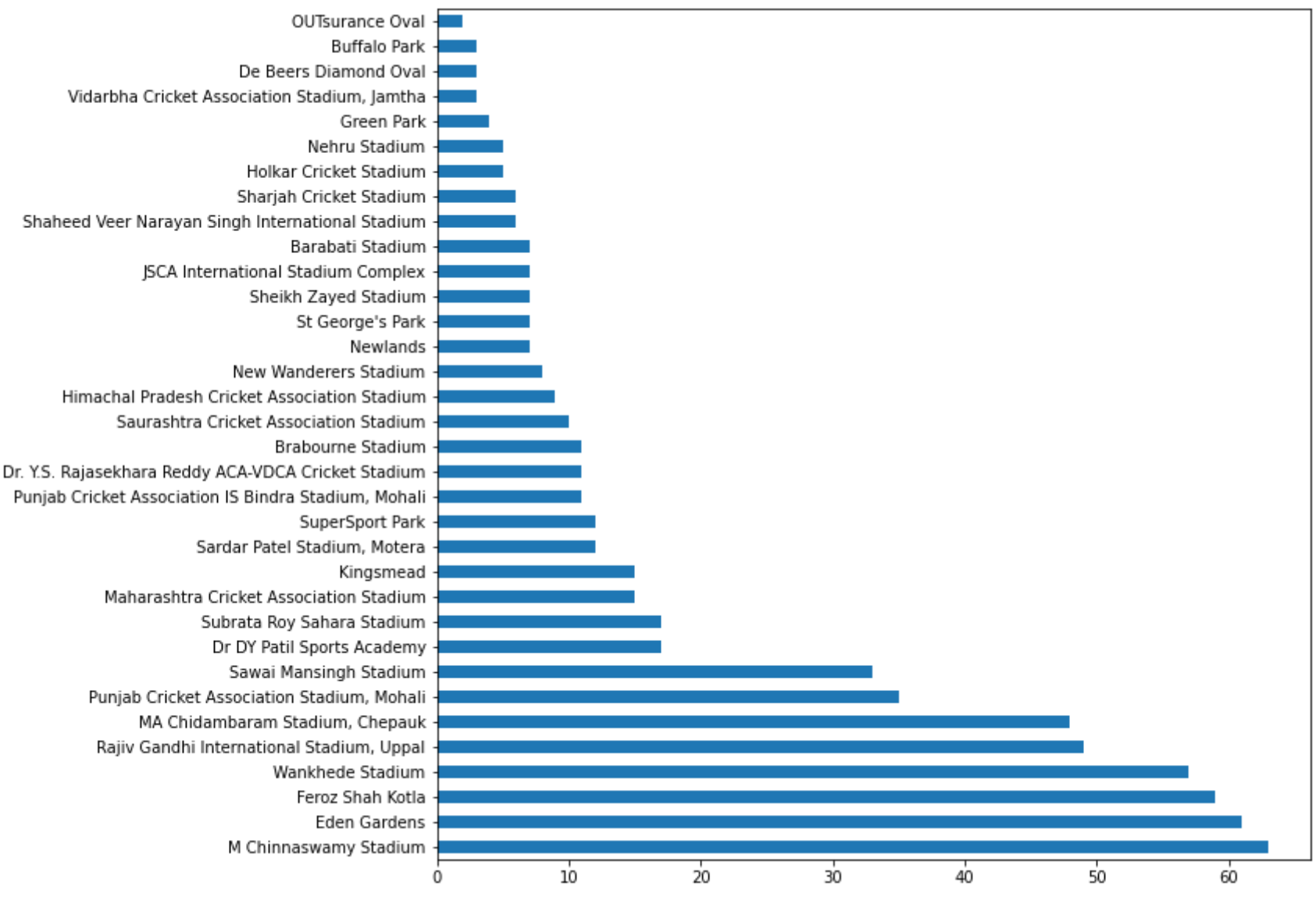
Out[27]:
```
Mumbai Indians                48
Kolkata Knight Riders         44
Chennai Super Kings           42
Rajasthan Royals              34
Delhi Daredevils              33
Royal Challengers Bangalore   33
Kings XI Punjab               27
Deccan Chargers               19
Sunrisers Hyderabad    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

In [29]:
```
#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_counts()
```

Out[29]:
```
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 [21]:
```
#which venue hosted most number of matches
df['venue'].value_counts().plot(kind="barh",figsize=(10,10))
```

Out[21]:
```
<AxesSubplot:>
```



In [22]:
```
#find how many matches chennai super kings played at M Chinnaswamy stadium?
a=df[(df['team1']=='Chennai Super Kings')|(df['team2']=='Chennai Super Kings')]['venue'].value_counts()
a['M Chinnaswamy Stadium']
```

Out[22]:
```
7
```

In [23]:
```
#who won most matches at M Chinnaswamy stadium?
df[df['venue']=='M Chinnaswamy Stadium']['winner'].value_counts()
```

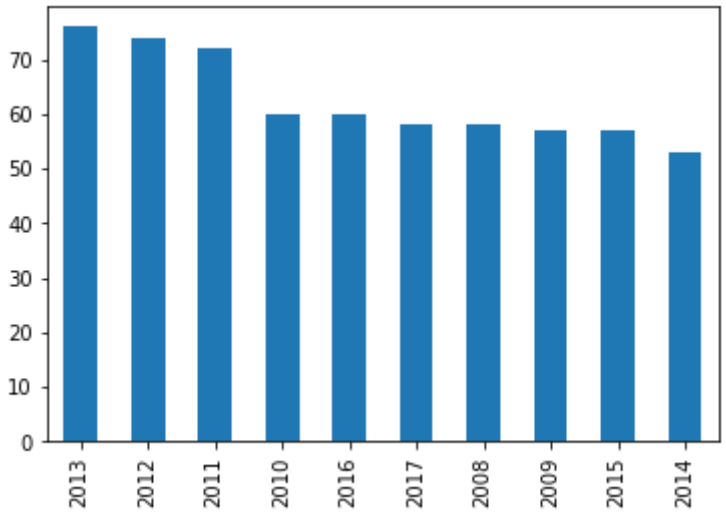
Out[23]:
```
Royal Challengers Bangalore    29
Mumbai Indians                  8
Kolkata Knight Riders           6
Kings XI Punjab                 5
Chennai Super Kings             4
Rajasthan Royals                3
Delhi Daredevils                3
Sunrisers Hyderabad             2
Rising Pune Supergiant          1
Gujarat Lions                   1
Deccan Chargers                 1
Name: winner, dtype: int64
```

year wise analysis

In [ ]:

In [24]:
```
#matches played in each year graphical Representation
df["season"].value_counts().plot(kind="bar")
```

Out[24]: <AxesSubplot:>



In [ ]:

In [25]:
```
#lets analyse ipl season held in 2013
#extract all the details of 2013 season
#here we will create yearwise groups (hint: use groupby() function)
a=df.groupby('season')
a.get_group(2013).head(5)
```

Out[25]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wicke
381	382	2013	Kolkata	2013-04-03	Delhi Daredevils	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	
382	383	2013	Bangalore	2013-04-04	Royal Challengers Bangalore	Mumbai Indians	Mumbai Indians	field	normal	0	Royal Challengers Bangalore	2	
383	384	2013	Hyderabad	2013-04-05	Sunrisers Hyderabad	Pune Warriors	Pune Warriors	field	normal	0	Sunrisers Hyderabad	22	
384	385	2013	Delhi	2013-04-06	Rajasthan Royals	Delhi Daredevils	Rajasthan Royals	bat	normal	0	Rajasthan Royals	5	
385	386	2013	Chennai	2013-04-06	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	normal	0	Mumbai Indians	9	

In [26]:
```
#which city hosted most number of matches in 2013
df[df['season']==2013]['city'].value_counts()
```

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

```