In [36]:

IPL Data Analysis by SIDDHANT NEGI

In [37]:

```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
import seaborn as sns
```

In [38]:

```
ipl = pd.read_csv('matches.csv')
```

In [39]:

```
ipl.head()
```

Out[39]:

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

In [40]:

```
ipl.shape
```

Out[40]:

```
(636, 18)
```

In [42]:

```
# MAN OF THE MATCH
ipl['player_of_match'].value_counts()
```

Out[42]:

```
CH Gayle             18
YK Pathan            16
AB de Villiers       15
DA Warner            15
SK Raina             14
RG Sharma            14
MS Dhoni             13
G Gambhir            13
AM Rahane            12
MEK Hussey           12
V Sehwag             11
V Kohli              11
DR Smith             11
JH Kallis            10
SR Watson            10
SE Marsh              9
KA Pollard            9
A Mishra              9
SR Tendulkar          8
AT Rayudu             7
AC Gilchrist          7
RV Uthappa            7
RA Jadeja             7
BJ Hodge              6
Harbhajan Singh       6
A Nehra               6
AD Russell            6
UT Yadav              6
M Vijay               6
SP Narine             6
                     ..
HV Patel              1
WPUJC Vaas            1
KV Sharma             1
JC Buttler            1
RP Singh              1
Shoaib Akhtar         1
BCJ Cutting           1
SP Goswami            1
SB Jakati             1
J Theron              1
YS Chahal             1
TL Suman              1
R Bhatia              1
GJ Bailey             1
TM Dilshan            1
DP Nannes             1
EJG Morgan            1
M Kartik              1
AC Voges              1
JDP Oram              1
MA Agarwal            1
PP Ojha               1
A Singh               1
MN Samuels            1
LJ Wright             1
MS Bisla              1
VR Aaron              1
KMDN Kulasekara       1
S Sreesanth           1
Washington Sundar     1
Name: player_of_match, Length: 201, dtype: int64
```

In [43]:

```
ipl['player_of_match'].value_counts()[0:5]
```

Out[43]:

```
CH Gayle          18
YK Pathan         16
AB de Villiers    15
DA Warner         15
SK Raina          14
Name: player_of_match, dtype: int64
```

In [44]:

```
list(ipl['player_of_match'].value_counts()[0:5].keys())
```

Out[44]:

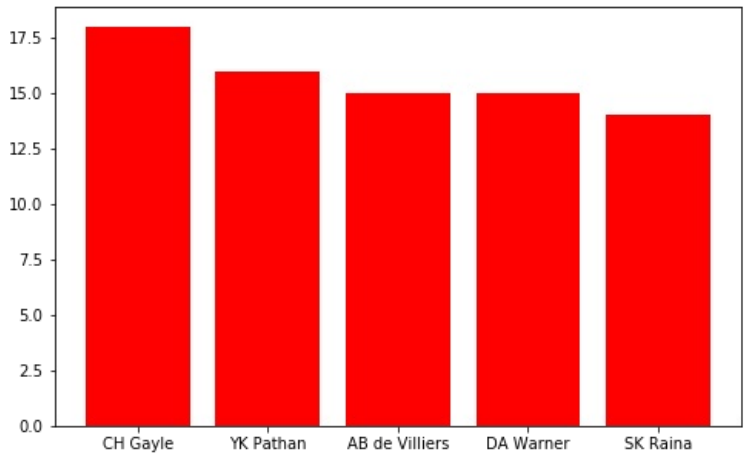
```
['CH Gayle', 'YK Pathan', 'AB de Villiers', 'DA Warner', 'SK Raina']
```

In [45]:

```
plt.figure(figsize=(8,5))
plt.bar(list(ipl['player_of_match'].value_counts()[0:5].keys()),list(ipl['player_of_match'].value_counts()[0:5]),
color="r")
```

Out[45]:

```
<BarContainer object of 5 artists>
```



In [46]:

```
ipl['result'].value_counts()
```

Out[46]:

```
normal       626
tie            7
no result      3
Name: result, dtype: int64
```

In [47]:

```
ipl['toss_winner'].value_counts()
```

Out[47]:

```
Mumbai Indians                 85
Kolkata Knight Riders          78
Delhi Daredevils               72
Royal Challengers Bangalore    70
Kings XI Punjab                68
Chennai Super Kings            66
Rajasthan Royals               63
Deccan Chargers                43
Sunrisers Hyderabad            35
Pune Warriors                  20
Gujarat Lions                  15
Kochi Tuskers Kerala            8
Rising Pune Supergiants         7
Rising Pune Supergiant          6
Name: toss_winner, dtype: int64
```

In [11]:

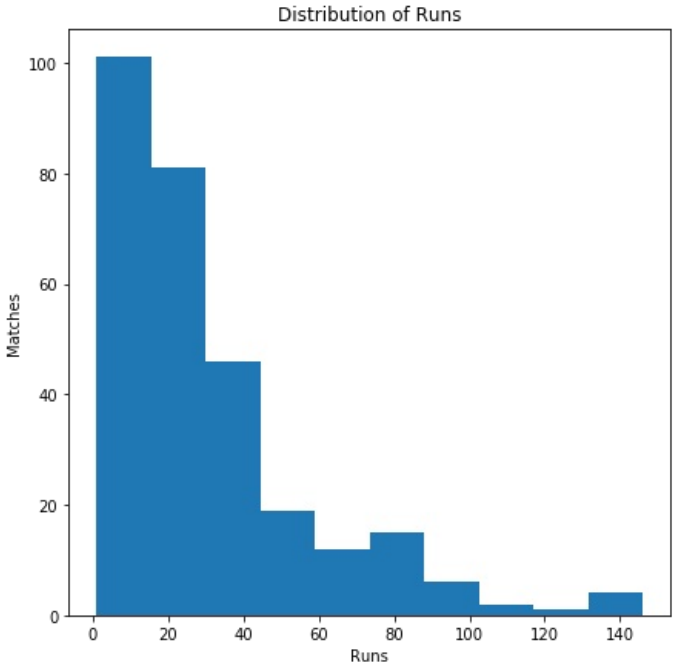
```
batting_first = ipl[ipl['win_by_runs']!=0]
batting_first.head()
```

Out[11]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	35	
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	0	Royal Challengers Bangalore	15	
8	9	2017	Pune	2017-04-11	Delhi Daredevils	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Delhi Daredevils	97	
13	14	2017	Kolkata	2017-04-15	Kolkata Knight Riders	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Kolkata Knight Riders	17	
14	15	2017	Delhi	2017-04-15	Delhi Daredevils	Kings XI Punjab	Delhi Daredevils	bat	normal	0	Delhi Daredevils	51	

In [12]:

```
plt.figure(figsize=(7,7))
plt.hist(batting_first['win_by_runs'])
plt.title('Distribution of Runs')
plt.xlabel('Runs')
plt.ylabel('Matches')
plt.show()
```



In [13]:

```
batting_first['winner'].value_counts()
```

Out[13]:

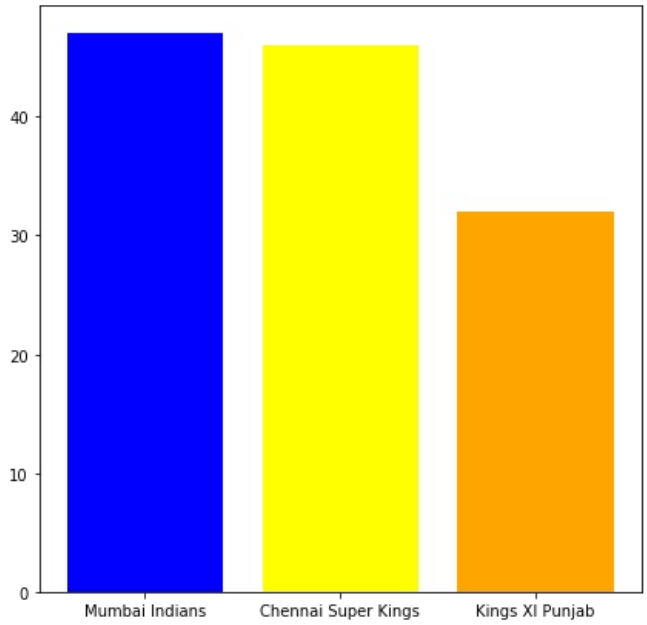
```
Mumbai Indians                 47
Chennai Super Kings            46
Kings XI Punjab                32
Kolkata Knight Riders          31
Royal Challengers Bangalore    30
Sunrisers Hyderabad            23
Rajasthan Royals               23
Delhi Daredevils               21
Deccan Chargers                18
Pune Warriors                   6
Rising Pune Supergiant          5
Rising Pune Supergiants         2
Kochi Tuskers Kerala            2
Gujarat Lions                   1
Name: winner, dtype: int64
```

In [14]:

```
plt.figure(figsize=(7,7))
plt.bar(list(batting_first['winner'].value_counts()[0:3].keys()),list(batting_first['winner'].value_counts()[0:3]
),color=["blue","yellow","orange"])
```

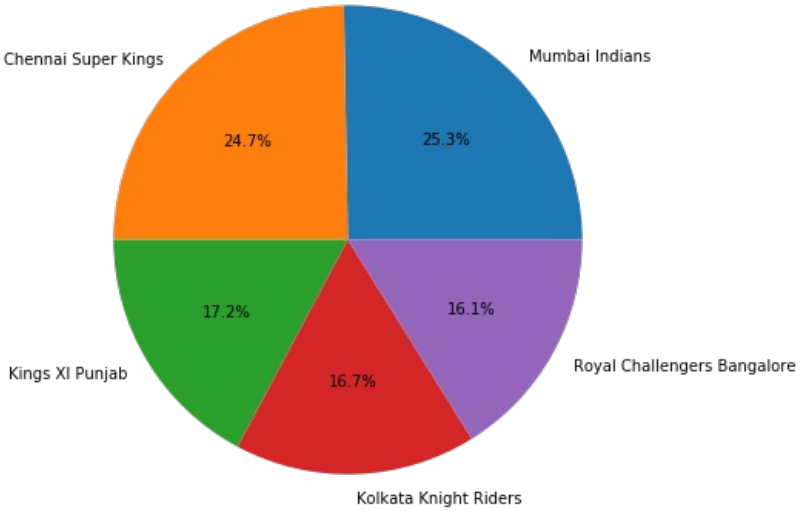
Out[14]:

```
<BarContainer object of 3 artists>
```



In [15]:

```
plt.figure(figsize=(7,7))
plt.pie(list(batting_first['winner'].value_counts()[0:5]), labels=list(batting_first['winner'].value_counts()[0:5
].keys()),autopct = "%0.1f%%")
plt.show()
```



In [16]:

```
ipl['season'].value_counts()
```

Out[16]:

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

In [17]:

```
ipl['city'].value_counts()[0:5]
```

Out[17]:

```
Mumbai       85
Bangalore    66
Kolkata      61
Delhi        60
Hyderabad    49
Name: city, dtype: int64
```

In [18]:

```
batting_second= ipl[ipl['win_by_wickets']!=0]
batting_second.head()
```

Out[18]:

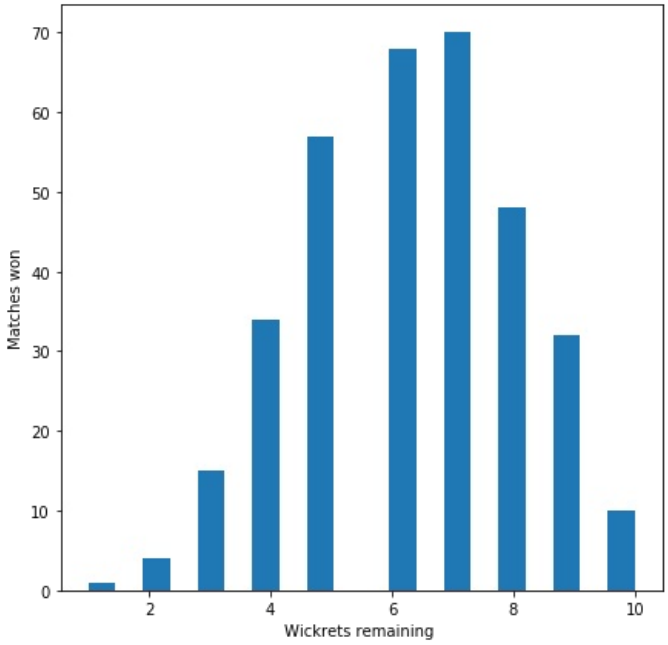
	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	0	
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	0	
5	6	2017	Hyderabad	2017-04-09	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Sunrisers Hyderabad	0	
6	7	2017	Mumbai	2017-04-09	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	normal	0	Mumbai Indians	0	

In [19]:

```
plt.figure(figsize=(7,7))
plt.hist(batting_second['win_by_wickets'],bins=20)
plt.xlabel('Wickrets remaining ')
plt.ylabel('Matches won')
```

Out[19]:

```
Text(0, 0.5, 'Matches won')
```

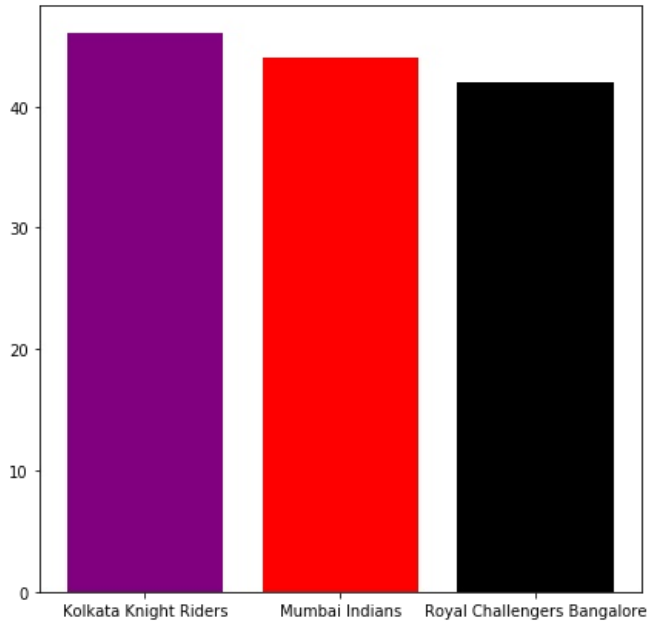


In [20]:

```
plt.figure(figsize=(7,7))
plt.bar(list(batting_second['winner'].value_counts()[0:3].keys()),list(batting_second['winner'].value_counts()[0:3]),color=["purple","red","black"])
```

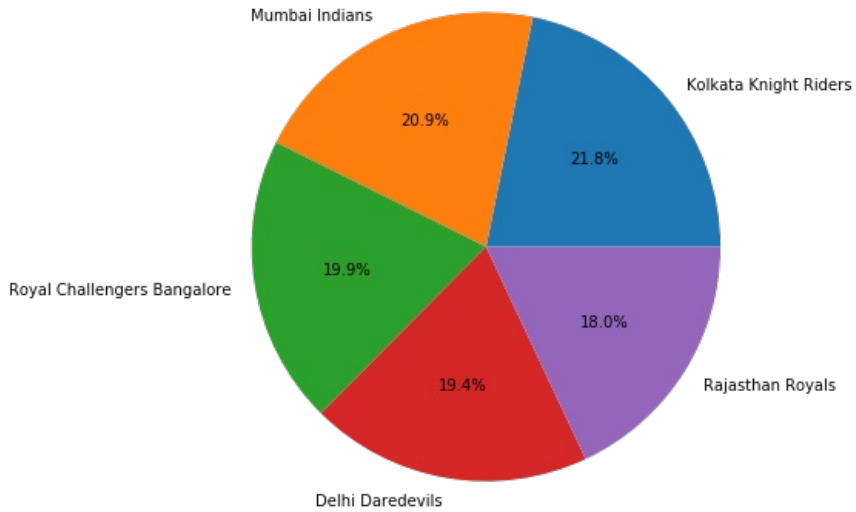
Out[20]:

```
<BarContainer object of 3 artists>
```



In [21]:

```
plt.figure(figsize=(7,7))
plt.pie(list(batting_second['winner'].value_counts()[0:5]), labels=list(batting_second['winner'].value_counts()[0:5].keys()),autopct = "%0.1f%%")
plt.show()
```



In [22]:

```
ipl['season'].value_counts()
```

Out[22]:

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

In [23]:

```
ipl['city'].value_counts()
```

Out[23]:

```
Mumbai            85
Bangalore         66
Kolkata           61
Delhi             60
Hyderabad         49
Chennai           48
Chandigarh        46
Jaipur            33
Pune              32
Durban            15
Centurion         12
Ahmedabad         12
Visakhapatnam     11
Rajkot            10
Dharamsala         9
Johannesburg       8
Ranchi             7
Port Elizabeth     7
Cape Town          7
Cuttack            7
Abu Dhabi          7
Sharjah            6
Raipur             6
Kochi              5
Indore             5
Kanpur             4
Kimberley          3
Nagpur             3
East London        3
Bloemfontein       2
Name: city, dtype: int64
```

In [24]:

```
np.sum(ipl['toss_winner']==ipl['winner'])
```

Out[24]:

```
325
```

In [25]:

```
325/636
```

Out[25]:

```
0.5110062893081762
```

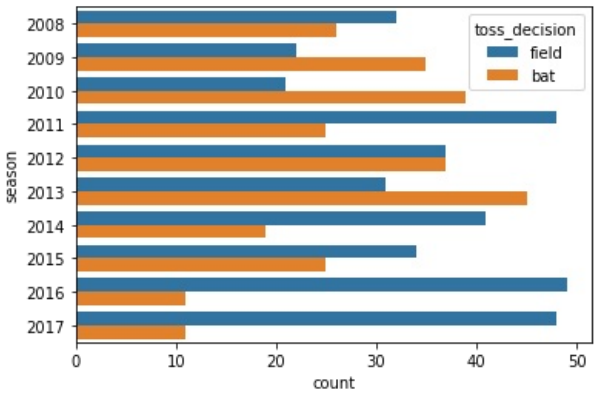
In [26]:

```
x=ipl.iloc[[ipl['win_by_runs'].idxmax()]]
print('It was in', x['season'].item(),',','when the biggest score difference was', x['win_by_runs'].item(), ',',
'between', x['team1'].item(), '&', x['team2'].item(),'.' ,
      ' ','The match was won by:', x['winner'].item())
```

```
It was in 2017 , when the biggest score difference was 146 , between Mumbai Indians & Delhi Daredevi
ls .   The match was won by: Mumbai Indians
```

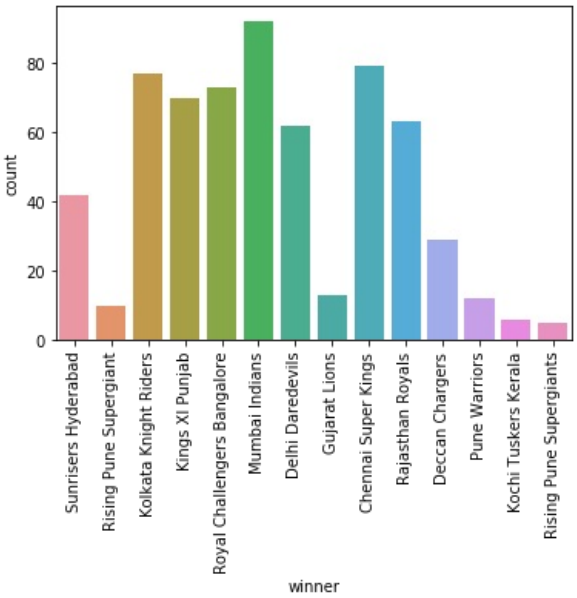
In [27]:

```
#--> The graph will show the decision of fielding or batting if a team wins the toss.
sns.countplot(y='season',hue='toss_decision',data=ipl)
plt.show()
```



In [28]:

```
#--> Team that won the most
sns.countplot(x='winner', data=ipl)
plt.xticks(rotation='vertical')
plt.show()
```



In [29]:

```
#read Delivery file
delivery=pd.read_csv('deliveries.csv')
 #print only the first 5 rows
delivery.head(5)
```

Out[29]:

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs	legbye_runs
0	1	1	Kolkata Knight Riders	Royal Challengers Bangalore	1	1	SC Ganguly	BB McCullum	P Kumar	0	...	0	1
1	1	1	Kolkata Knight Riders	Royal Challengers Bangalore	1	2	BB McCullum	SC Ganguly	P Kumar	0	...	0	0
2	1	1	Kolkata Knight Riders	Royal Challengers Bangalore	1	3	BB McCullum	SC Ganguly	P Kumar	0	...	0	0
3	1	1	Kolkata Knight Riders	Royal Challengers Bangalore	1	4	BB McCullum	SC Ganguly	P Kumar	0	...	0	0
4	1	1	Kolkata Knight Riders	Royal Challengers Bangalore	1	5	BB McCullum	SC Ganguly	P Kumar	0	...	0	0

5 rows × 21 columns

In [30]:

```
#--> finding the number of matches played in total
print ("Total number of matches played:", len(ipl))
#--> Location of Matches, name of teams that played and the name of unique umpires
print(' \n Location for all matches: \n',ipl['city'].unique(), ' \n \n Teams :',ipl['team1'].unique(), '\n \nTota
l umpires ',ipl['umpire1'].unique())
```

```
Total number of matches played: 636

 Location for all matches:
 ['Hyderabad' 'Pune' 'Rajkot' 'Indore' 'Bangalore' 'Mumbai' 'Kolkata'
 'Delhi' 'Chandigarh' 'Kanpur' 'Jaipur' 'Chennai' 'Cape Town'
 'Port Elizabeth' 'Durban' 'Centurion' 'East London' 'Johannesburg'
 'Kimberley' 'Bloemfontein' 'Ahmedabad' 'Cuttack' 'Nagpur' 'Dharamsala'
 'Kochi' 'Visakhapatnam' 'Raipur' 'Ranchi' 'Abu Dhabi' 'Sharjah' nan]

 Teams : ['Sunrisers Hyderabad' 'Mumbai Indians' 'Gujarat Lions'
 'Rising Pune Supergiant' 'Royal Challengers Bangalore'
 'Kolkata Knight Riders' 'Delhi Daredevils' 'Kings XI Punjab'
 'Chennai Super Kings' 'Rajasthan Royals' 'Deccan Chargers'
 'Kochi Tuskers Kerala' 'Pune Warriors' 'Rising Pune Supergiants']

Total umpires  ['AY Dandekar' 'A Nand Kishore' 'Nitin Menon' 'AK Chaudhary' nan
 'A Deshmukh' 'KN Ananthapadmanabhan' 'YC Barde' 'S Ravi' 'CB Gaffaney'
 'M Erasmus' 'NJ Llong' 'CK Nandan' 'Asad Rauf' 'MR Benson' 'Aleem Dar'
 'SJ Davis' 'BF Bowden' 'IL Howell' 'DJ Harper' 'RE Koertzen'
 'BR Doctrove' 'AV Jayaprakash' 'BG Jerling' 'HDPK Dharmasena' 'S Asnani'
 'GAV Baxter' 'SS Hazare' 'K Hariharan' 'SL Shastri' 'SK Tarapore'
 'SJA Taufel' 'S Das' 'AM Saheba' 'PR Reiffel' 'JD Cloete' 'VA Kulkarni'
 'BNJ Oxenford' 'C Shamshuddin' 'RK Illingworth' 'RM Deshpande'
 'K Srinath' 'SD Fry' 'PG Pathak' 'K Bharatan']
```

In [32]:

```
#--> Cleaning:remove the column with no data or consists of NaN
del ipl['umpire3']
ipl.head(4)
```

Out[32]:

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

In [33]:

```
a= ipl['player_of_match'].value_counts()
b= a.idxmax()

print(' Who has the highest man of the match awards?\n', b)

c= ipl['winner'].value_counts()
d=c.idxmax()
print(' Which team has won the most?\n', d)
```

```
 Who has the highest man of the match awards?
 CH Gayle
 Which team has won the most?
 Mumbai Indians
```

In [ ]:

```

```