

```
In [13]: import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
import seaborn as sns

In [ ]: #eda on ipl data

In [72]: ipl=pd.read_csv('matches.csv')

In [75]: ipl.head(2)

Out[75]:
```

	id	Season	city	date	team1	team2	toss_winner	toss_decision	result	dt_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire2	umpire3
	0	1	IPL-2017	Hyderabad	05-04-2017	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	35	0	Yuvraj Singh	Rajiv Gandhi International Stadium, Uppal	AY Dandekar	NJ Llong
	1	2	IPL-2017	Pune	06-04-2017	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	0	7	SPD Smith	Maharashtra Association Stadium	A Nand Kishore	S Ravi

```


In [86]: ipl.shape
Out[86]: (756, 18)

In [76]: ipl['player_of_match'].value_counts()
#man of the match won by player

Out[76]:
```

CH Gayle	21
AB de Villiers	20
RG Sharma	17
MS Dhoni	17
DA Warner	17
..	..
PO Collingwood	1
NI Ojha	1
AC Voges	1
J Theron	1
S Hetmyer	1

```
Name: player_of_match, Length: 226, dtype: int64

In [77]: ipl['player_of_match'].value_counts()[0:5]
#top 5 man of the match won by player

Out[77]:
```

CH Gayle	21
AB de Villiers	20
RG Sharma	17
MS Dhoni	17
DA Warner	17

```
Name: player_of_match, dtype: int64

In [79]: ipl['result'].value_counts()

normal      743
tie          9
no result    4
Name: result, dtype: int64

In [80]: ipl['toss_winner'].value_counts()

Mumbai Indians      98
Kolkata Knight Riders      92
Chennai Super Kings      89
Royal Challengers Bangalore      81
Kings XI Punjab      81
Delhi Daredevils      80
Rajasthan Royals      80
Sunrisers Hyderabad      46
Deccan Chargers      43
Pune Warriors      20
Gujarat Lions      15
Delhi Capitals      10
Kochi Tuskers Kerala      8
Rising Pune Supergiants      7
Rising Pune Supergiant      6
Name: toss_winner, dtype: int64

In [82]: #q-how many time batting first team won the match
#if team is winning then 'win by run' column should not be 0
batting_first=ipl[ipl['win_by_runs']!=0]

In [84]: batting_first.head()

Out[84]:
```

	id	Season	city	date	team1	team2	toss_winner	toss_decision	result	dt_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire2	umpire3
	0	1	IPL-2017	Hyderabad	05-04-2017	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	35	0	Yuvraj Singh	Rajiv Gandhi International Stadium, Uppal	AY Dandekar	N Llong
	4	5	IPL-2017	Bangalore	08-04-2017	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	0	Royal Challengers Bangalore	15	0	KM Jadhav	M Chinnaswamy Stadium	NaN	NaN
	8	9	IPL-2017	Pune	11-04-2017	Delhi Daredevils	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Delhi Daredevils	97	0	SV Samson	Maharashtra Cricket Association Stadium	AY Dandekar	S Ravi
	13	14	IPL-2017	Kolkata	15-04-2017	Kolkata Knight Riders	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Kolkata Knight Riders	17	0	RV Uthappa	Eden Gardens	AY Dandekar	N Llong
	14	15	IPL-2017	Delhi	15-04-2017	Delhi Daredevils	Kings XI Punjab	Delhi Daredevils	bat	normal	0	Delhi Daredevils	51	0	CJ Anderson	Feroz Shah Kotla	YC Barde	Nitin Menon

```


In [85]: batting_first.shape
#337 time batting team won from total of 756 , so win batting first % is 44.5%
(337, 18)

Out[85]:

In [88]: 337/756*100

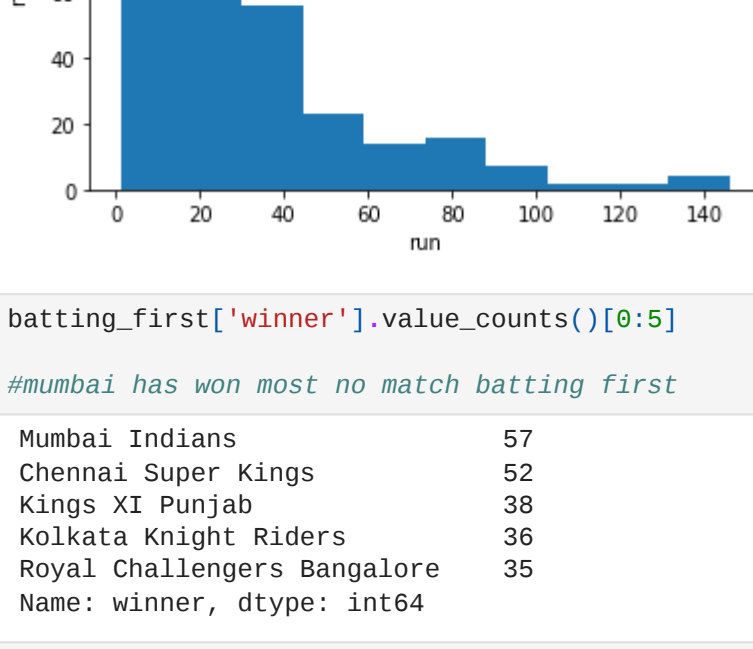
Out[88]: 44.576719576719576

In [ ]: #when numeric value is there we use histogram

In [93]: plt.hist(batting_first['win_by_runs'])
plt.title('distribution of run')
plt.xlabel('run')
plt.ylabel('match')
plt.show()

#there is only few team who who won matches by 140 run , most of the team is winning with margin 0-20

distribution of run
```



```


In [102]: batting_first['winner'].value_counts()[0:5]
#mumbai has won most no match batting first

Out[102]:
```

Mumbai Indians	57
Chennai Super Kings	52
Kings XI Punjab	38
Kolkata Knight Riders	36
Royal Challengers Bangalore	35

```
Name: winner, dtype: int64

In [108]: plt.figure(figsize=(20,10))
plt.pie(list(batting_first['winner'].value_counts()), labels=list(batting_first['winner'].value_counts().keys()), autopct='%0.1f%%')
plt.show()

Chennai Super Kings
15.4%
Mumbai Indians
36.9%
Kings XI Punjab
11.3%
Kolkata Knight Riders
10.7%
Royal Challengers Bangalore
10.4%
Sunrisers Hyderabad
8.9%
Rajasthan Royals
8.0%
Delhi Daredevils
7.4%
Deccan Chargers
5.3%
Pune Warriors
4.9%
Rising Pune Supergiant
1.9%
Delhi Capitals
1.5%
Kochi Tuskers Kerala
0.8%
Gujarat Lions
0.8%
Rising Pune Supergiants
0.8%

In [109]: #q-how many time bowling first team won the match
#if team is winning then win by wicket coloumn should not be 0
bowling_first=ipl[ipl['win_by_wickets']!=0]

In [110]: bowling_first.shape
(406, 18)

Out[110]:

In [111]: 406/756*100

Out[111]: 53.70370370370371

In [ ]: #####bowling first team win % is 53.7% more then batting first , so we can say toss doesn't matter much in ipl

In [ ]: #now look from team perspective

In [114]: bowling_first['winner'].value_counts()[0:5]

Out[114]:
```

Kolkata Knight Riders	56
Mumbai Indians	50
Royal Challengers Bangalore	48
Chennai Super Kings	48
Rajasthan Royals	46

```
Name: winner, dtype: int64

In [117]: ipl['team1'].value_counts()[0:5]
#kolkata has hosted 83 time

Out[117]:
```

Mumbai Indians	101
Kings XI Punjab	91
Chennai Super Kings	89
Royal Challengers Bangalore	85
Kolkata Knight Riders	83

```
Name: team1, dtype: int64

In [118]: ipl['team2'].value_counts()[0:5]
#kolkata has host 95 time

Out[118]:
```

Royal Challengers Bangalore	95
Kolkata Knight Riders	95
Delhi Daredevils	89
Mumbai Indians	86
Kings XI Punjab	85

```
Name: team2, dtype: int64

In [119]: #kolkata total match
83+95

Out[119]: 178

In [120]: #kolkata won 56 times while bowling first and 36 times while batting first

In [121]: ipl['city'].value_counts()
#mumbai hosted most matches

Out[121]:
```

Mumbai	101
Kolkata	77
Delhi	74
Bangalore	66
Hyderabad	64
Chennai	57
Jaipur	47
Chandigarh	46
Pune	38
Durban	15
Bengaluru	14
Visakhapatnam	13
Centurion	12
Ahmedabad	12
Rajkot	10
Mohali	10
Indore	9
Dharamsala	9
Johannesburg	8
Cuttack	7
Ranchi	7
Port Elizabeth	7
Cape Town	7
Abu Dhabi	7
Sharjah	6
Raipur	6
Kochi	5
Kanpur	4
Nagpur	3
Kimberley	3
East London	3
Bloomfontein	2

```
Name: city, dtype: int64

In [124]: ipl['Season'].value_counts()

Out[124]:
```

IPL-2013	76
IPL-2012	74
IPL-2011	73
IPL-2010	60
IPL-2014	60
IPL-2016	60
IPL-2018	60
IPL-2019	60
IPL-2017	59
IPL-2015	59
IPL-2008	58
IPL-2009	57

```
Name: Season, dtype: int64

In [ ]: #find out how many team won after winning toss

In [126]: ipl['toss_winner']==ipl['winner']

Out[126]:
```

0	False
1	True
2	True
3	True
4	True
...	...
751	False
752	False
753	True
754	True
755	True

```
Length: 756, dtype: bool

In [127]: sum(ipl['toss_winner']==ipl['winner'])

Out[127]: 393

In [128]: 393/756*100

Out[128]: 51.98412698412699

#51% so winning toss does not matter much

In [ ]:

In [ ]: ###eda on fifa

In [130]: fifa=pd.read_csv('CompleteDataset.csv')
C:\Users\India\AppData\Local\Temp\ipykernel_15352\4071428283.py:1: DtypeWarning: Columns (23,35) have mixed types. Specify dtype option on import or set low_m...
fifa=pd.read_csv('CompleteDataset.csv')

In [150]: fifa.head(5)

Out[150]:
```

Unnamed: 0	Name	Age	Photo	Nationality	Flag	Overall	Potential	Club	Club Logo	...	RB	RCB
0	Cristiano Ronaldo	32	https://cdn.sofia.org/48/18/players/20801.png	Portugal	https://cdn.sofia.org/flags/38.png	94	94	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	61.0	53.0
1	L. Messi	30	https://cdn.sofia.org/48/18/players/158023.png	Argentina	https://cdn.sofia.org/flags/52.png	93	93	FC Barcelona	https://cdn.sofia.org/24/18/teams/241.png	...	57.0	45.0
2	Neymar	25	https://cdn.sofia.org/48/18/players/190871.png	Brazil	https://cdn.sofia.org/flags/54.png	92	94	Paris Saint-Germain	https://cdn.sofia.org/24/18/teams/73.png	...	59.0	46.0
3	L. Suárez	30	https://cdn.sofia.org/48/18/players/176580.png	Uruguay	https://cdn.sofia.org/flags/60.png	92	92	FC Barcelona	https://cdn.sofia.org/24/18/teams/241.png	...	64.0	58.0
4	M. Neuer	31	https://cdn.sofia.org/48/18/players/167495.png	Germany	https://cdn.sofia.org/flags/21.png	92	92	FC Bayern Munich	https://cdn.sofia.org/24/18/teams/21.png	...	NaN	NaN

5 rows x 75 columns

```


In [145]: #top 5 country with max player
fifa['Nationality'].value_counts()[0:5]

Out[145]:
```

England	1630
Germany	1140
Spain	1019
France	978
Argentina	965

```
Name: Nationality, dtype: int64

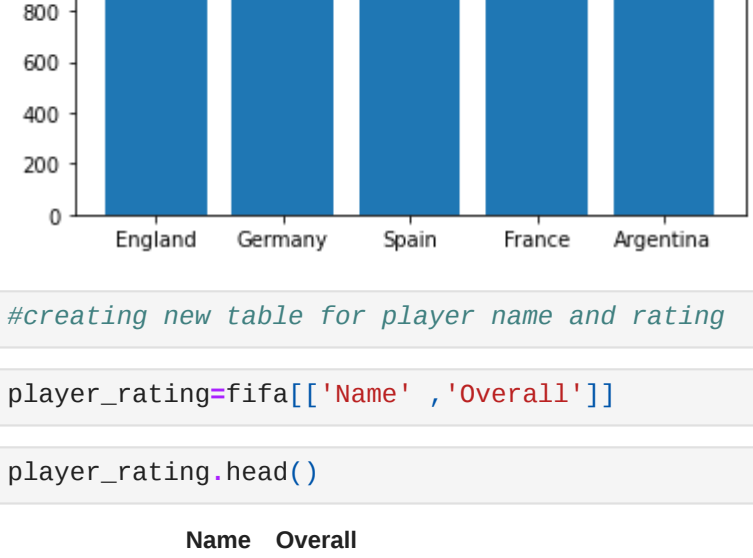
In [147]: fifa['Nationality'].value_counts()[0:5].keys()

Out[147]: Index(['England', 'Germany', 'Spain', 'France', 'Argentina'], dtype='object')

In [ ]: #in matplotlib when we create bar chart first parameter should be catogorical and 2nd should be numeric

In [149]: plt.bar(list(fifa['Nationality'].value_counts()[0:5].keys()),list(fifa['Nationality'].value_counts()[0:5] ) )

Out[149]: <BarContainer object of 5 artists>
```



```


In [ ]: #creating new table for player name and rating

In [151]: player_rating=fifa[['Name','Overall']]

In [152]: player_rating.head()

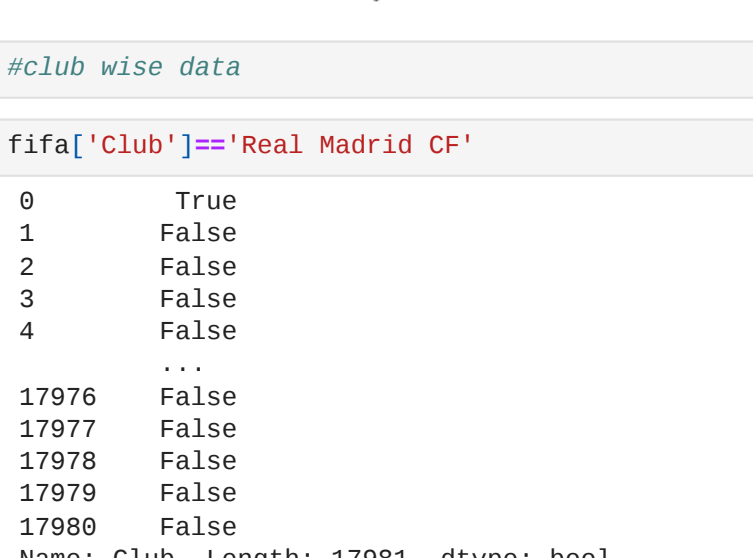
Out[152]:
```

	Name	Overall
0	Cristiano Ronaldo	94
1	L. Messi	93
2	Neymar	92
3	L. Suárez	92
4	M. Neuer	92

```


In [154]: plt.bar(list(player_rating['Name'][0:5]),list(player_rating['Overall'][0:5]) ,color=['b','y','g','r','r'])

Out[154]: <BarContainer object of 5 artists>
```



```


In [ ]: #club wise data

In [158]: fifa['Club']== 'Real Madrid CF'

Out[158]:
```

0	True
1	False
2	False
3	False
4	False
...	...
17976	False
17977	False
17978	False
17979	False
17980	False

```
Name: Club, Length: 17981, dtype: bool

In [159]: Real_madrid=fifa[fifa['Club']== 'Real Madrid CF']

In [160]: Real_madrid.head()

Out[160]:
```

Unnamed: 0	Name	Age	Photo	Nationality	Flag	Overall	Potential	Club	Club Logo	...	RB	RCB	RC
0	Cristiano Ronaldo	32	https://cdn.sofia.org/48/18/players/20801.png	Portugal	https://cdn.sofia.org/flags/38.png	94	94	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	61.0	53.0	8
8	T. Kroos	27	https://cdn.sofia.org/48/18/players/152521.png	Germany	https://cdn.sofia.org/flags/21.png	90	90	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	76.0	72.0	8
10	Sergio Ramos	31	https://cdn.sofia.org/48/18/players/155862.png	Spain	https://cdn.sofia.org/flags/45.png	90	90	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	84.0	87.0	7
14	L. Modrić	31	https://cdn.sofia.org/48/18/players/177003.png	Croatia	https://cdn.sofia.org/flags/10.png	89	89	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	78.0	72.0	8
15	G. Bale	27	https://cdn.sofia.org/48/18/players/173731.png	Wales	https://cdn.sofia.org/flags/50.png	89	89	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	72.0	67.0	8

5 rows x 75 columns

```


In [163]: Real_madrid.sort_values(by=['Age']).head()

Out[163]:
```

Unnamed: 0	Name	Age	Photo	Nationality	Flag	Overall	Potential	Club	Club Logo	...	RB	RCB	RC
4674	A. Hakimi	18	https://cdn.sofia.org/48/18/players/235212.png	Morocco	https://cdn.sofia.org/flags/129.png	71	81	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	70.0	66.0	8
11632	L. Zidane	19	https://cdn.sofia.org/48/18/players/240311.png	France	https://cdn.sofia.org/flags/18.png	64	73	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	NaN	NaN	8
7604	Óscar	19	https://cdn.sofia.org/48/18/players/239335.png	Spain	https://cdn.sofia.org/flags/45.png	68	81	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	51.0	41.0	8
1956	T. Hernández	19	https://cdn.sofia.org/48/18/players/232656.png	France	https://cdn.sofia.org/flags/18.png	75	85	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	74.0	72.0	7
3375	Borja Mayoral	20	https://cdn.sofia.org/48/18/players/228635.png	Spain	https://cdn.sofia.org/flags/45.png	72	84	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	43.0	31.0	8

5 rows x 75 columns

```


In [165]: Real_madrid.sort_values(by=['Age'], ascending=False).head()

Out[165]:
```

Unnamed: 0	Name	Age	Photo	Nationality	Flag	Overall	Potential	Club	Club Logo	...	RB	RCB	F
0	Cristiano Ronaldo	32	https://cdn.sofia.org/48/18/players/20801.png	Portugal	https://cdn.sofia.org/flags/38.png	94	94	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	61.0	53.0	8
10	Sergio Ramos	31	https://cdn.sofia.org/48/18/players/155862.png	Spain	https://cdn.sofia.org/flags/45.png	90	90	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	84.0	87.0	7
14	L. Modrić	31	https://cdn.sofia.org/48/18/players/177003.png	Croatia	https://cdn.sofia.org/flags/10.png	89	89	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	78.0	72.0	8
74	K. Navas	30	https://cdn.sofia.org/48/18/players/193041.png	Costa Rica	https://cdn.sofia.org/flags/72.png	85	85	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	NaN	NaN	8
467	Kiko Casilla	30	https://cdn.sofia.org/48/18/players/177644.png	Spain	https://cdn.sofia.org/flags/45.png	80	80	Real Madrid CF	https://cdn.sofia.org/24/18/teams/243.png	...	NaN	NaN	8

5 rows x 75 columns

```


In [168]: Real_madrid[['Name','Age']].sort_values(by=['Age'], ascending=False).head()

Out[168]:
```

	Name	Age
0	Cristiano Ronaldo	32
10	Sergio Ramos	31
14	L. Modrić	31
74	K. Navas	30
467	Kiko Casilla	30

```


In [170]: Real_madrid['Nationality'].value_counts()

Out[170]:
```

Spain	13
France	4
Croatia	2
Brazil	2
Portugal	1
Germany	1
Wales	1
Costa Rica	1
Morocco	1

```
Name: Nationality, dtype: int64
```