

```
In [27]: import pandas as pd
import numpy as np
import seaborn as sns
from matplotlib import pyplot as plt
```

```
In [4]: ipl_data=pd.read_csv("C:\\Users\\Prana\\Desktop\\DATA SCIENCE DATA\\CVC File\\IPL data\\IPL_Matches_2008_2022.csv")
ipl_data.head()
```

	ID	City	Date	Season	MatchNumber	Team1	Team2	Venue	TossWinner	TossDecision	SuperOver	WinningTeam	WonBy	Margin	method	Player_of_Match	Team1Players	Team2Players	Ump	
0	1312200	Ahmedabad	2022-05-29	2022	Final	Rajasthan Royals	Gujarat Titans	Narendra Modi Stadium, Ahmedabad	Rajasthan Royals		bat	N	Gujarat Titans	Wickets	7.0	NaN	HH Pandya	[YBK Jaswal, JC Butler, SV Samson, D ...]	[WP Saha, Shubman Gill, MS Wade, HH Pan...	CB Gaffa
1	1312199	Ahmedabad	2022-05-27	2022	Qualifier 2	Royal Challengers Bangalore	Rajasthan Royals	Narendra Modi Stadium, Ahmedabad	Rajasthan Royals		field	N	Rajasthan Royals	Wickets	7.0	NaN	JC Butler	[V Kohli, F du Plessis, RM Patidar, GJ ...]	[YBK Jaswal, JC Butler, SV Samson, D ...]	CB Gaffa
2	1312198	Kolkata	2022-05-25	2022	Eliminator	Royal Challengers Bangalore	Lucknow Super Giants	Eden Gardens, Kolkata	Lucknow Super Giants		field	N	Royal Challengers Bangalore	Runs	14.0	NaN	RM Patidar	[V Kohli, F du Plessis, RM Patidar, GJ ...]	[Q de Kock, KL Rahul, M Vohra, DJ Hooda...	Madanaga
3	1312197	Kolkata	2022-05-24	2022	Qualifier 1	Rajasthan Royals	Gujarat Titans	Eden Gardens, Kolkata	Gujarat Titans		field	N	Gujarat Titans	Wickets	7.0	NaN	DA Miller	[YBK Jaswal, JC Butler, SV Samson, D ...]	[WP Saha, Shubman Gill, MS Wade, HH Pan...	l Oxen
4	1304116	Mumbai	2022-05-22	2022	70	Sunrisers Hyderabad	Punjab Kings	Wankhede Stadium, Mumbai	Sunrisers Hyderabad		bat	N	Punjab Kings	Wickets	5.0	NaN	Harpreet Brar	[PK Garg, Abhishek Sharma, RA Tripathi, ...]	[JM Barstow, S Dhawan, M Shatrugh Khatri, ...]	Chaudh
In [5]:	#to see the last five rows ipl_data.tail()																			
Out[5]:	ID	City	Date	Season	MatchNumber	Team1	Team2	Venue	TossWinner	TossDecision	SuperOver	WinningTeam	WonBy	Margin	method	Player_of_Match	Team1Players	Team2Players	Ump	
945	335986	Kolkata	2008-04-20	2007/08	4	Kolkata Knight Riders	Deccan Chargers	Eden Gardens	Deccan Chargers		bat	N	Kolkata Knight Riders	Wickets	5.0	NaN	DJ Hussey	[WP Saha, BB McCullum, RT Poring, SC G...	[AC Ghohrni, Y Venugopal Rao, VVS Laxma...	Bow
946	335985	Mumbai	2008-04-20	2007/08	5	Mumbai Indians	Royal Challengers Bangalore	Wankhede Stadium	Mumbai Indians		bat	N	Royal Challengers Bangalore	Wickets	5.0	NaN	MV Boucher	[L Rohit, ST Jayasuriya, McCullum, DJ Thornely, ...]	[CS Chandrapaul, R Dravid, LRPL Taylor, ...]	SJ C
947	335984	Delhi	2008-04-19	2007/08	3	Delhi Daredevils	Rajasthan Royals	Feroz Shah Kotla	Rajasthan Royals		bat	N	Delhi Daredevils	Wickets	9.0	NaN	MF Maharoof	[G Gambhir, V Sehwag, S Dhawan, MK Tiwa...	[T Kohli, YK Pathani, SR Watson, M Kall...	Al
948	335983	Chandigarh	2008-04-19	2007/08	2	Kings XI Punjab	Chennai Super Kings	Punjab Cricket Association Stadium, Mohali	Chennai Super Kings		bat	N	Chennai Super Kings	Runs	33.0	NaN	MEK Hussey	[K Goel, JR Hopes, KC Sangakkara, Yuvra...	[PA Patel, ML Hayden, MEK Hussey, MS Dh...	Ben
949	335982	Bangalore	2008-04-18	2007/08	1	Royal Challengers Bangalore	Kolkata Knight Riders	M Chinnaswamy Stadium	Royal Challengers Bangalore		field	N	Kolkata Knight Riders	Runs	140.0	NaN	BB McCullum	[R David, W Jaffer, V Kohli, JH Kallia...	[SC Ganguly, BB McCullum, RT Poring, D...	/
In [7]:	#display the number of variable and observation ipl_data.shape																			
Out[7]:	(950, 20)																			
In [8]:	#information about data ipl_data.info()																			
	<class 'pandas.core.frame.DataFrame'> RangeIndex: 950 entries, 0 to 949 Data columns (total 20 columns): # Column Non-Null Count Dtype -- -- -- 0 ID 950 non-null int64 1 City 899 non-null object 2 Date 950 non-null object 3 Season 950 non-null object 4 MatchNumber 950 non-null object 5 Team1 950 non-null object 6 Team2 950 non-null object 7 Venue 950 non-null object 8 TossWinner 950 non-null object 9 TossDecision 950 non-null object 10 SuperOver 945 non-null object																			

```
In [5]: #to see the last five rows
ipl_data.tail()
```

Out[9]:	ID	0
	City	51
	Date	0
	Season	0
	MatchNumber	0
	Team1	0
	Team2	0
	Venue	0
	TossWinner	0
	TossDecision	0
	SuperOver	4
	WinningTeam	4
	WonBy	0
	Margin	18
	method	931
	Player_of_Match	4
	Team1Players	0
	Team2Players	0
	Umpire1	0
	Umpire2	0
	dtype:	int64
In [10]:	# display the variable names and their datatypes	
	ipl_data.dtypes	
Out[10]:	ID	int64
	City	object

```
In [7]: #display the number of variable and observation
ipl_data.shape
```

```
Out[7]: (950, 20)
```

```
In [8]: #information about data
ipl_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 950 entries, 0 to 949
Data columns (total 20 columns):
 #   Column              Non-Null Count  Dtype
---  --
 0   ID                   950 non-null    int64
 1   City                 899 non-null    object
 2   Date                 950 non-null    object
 3   Season              950 non-null    object
 4   MatchNumber         950 non-null    object
 5   Team1                950 non-null    object
 6   Team2                950 non-null    object
 7   Venue                950 non-null    object
 8   TossWinner           950 non-null    object
 9   TossDecision         950 non-null    object
10   SuperOver            946 non-null    object
11   WinningTeam          946 non-null    object
12   WonBy                950 non-null    object
13   Margin              932 non-null    float64
14   method               19 non-null     object
15   Player_of_Match     946 non-null    object
16   Team1Players         950 non-null    object
17   Team2Players         950 non-null    object
18   Umpire1              950 non-null    object
19   Umpire2              950 non-null    object
dtypes: float64(1), int64(1), object(18)
memory usage: 148.6+ KB
```

```
In [9]: # finding the null value in data
ipl_data.isnull().sum()
```

```
Out[9]: ID                0
City                51
Date                0
Season              0
MatchNumber        0
Team1               0
Team2              0
Venue              0
TossWinner          0
TossDecision        0
SuperOver           4
WinningTeam         0
WonBy               0
Margin             18
method              931
Player_of_Match     4
Team1Players        0
Team2Players        0
Umpire1             0
Umpire2             0
dtype: int64
```

```
In [10]: # display the variable names and their datatypes
ipl_data.dtypes
```

```
Out[10]: ID                int64
City                object
Date                object
Season              object
MatchNumber         object
Team1                object
Team2                object
Venue                object
TossWinner           object
TossDecision         object
SuperOver            object
WinningTeam          object
WonBy                object
Margin              float64
method              object
Player_of_Match     object
Team1Players         object
Team2Players         object
Umpire1              object
Umpire2              object
dtype: object
```

```
In [12]: # descriptive statistics
ipl_data.describe()
```

```
Out[12]:
```

	ID	Margin
count	9.500000e+02	932.000000
mean	8.304852e+05	17.056867
std	3.375678e+05	21.633109
min	3.359820e+05	1.000000
25%	5.012612e+05	6.000000
50%	8.297380e+05	8.000000
75%	1.175372e+06	19.000000
max	1.312200e+06	146.000000

```
In [13]: # descriptive statics for all features including categorical
ipl_data.describe(include='all')
```

```
Out[13]:
```

	ID	City	Date	Season	MatchNumber	Team1	Team2	Venue	TossWinner	TossDecision	SuperOver	WinningTeam	WonBy	Margin	method	Player_of_Match	Team1Players	Team2Players	U
count	9.500000e+02	899	950	950	950	950	950	950	950	946	946	950	932.000000	19	946	950	950	950	1
unique	NaN	33	706	15	80	18	18	49	18	2	2	18	4	NaN	1	262	932	937	

top	NaN	Mumbai	2014-05-10	2013	Final	Challengers Bangalore	Mumbai Indians	Eden Gardens	Mumbai Indians	field	N	Mumbai Indians	Wickets	NaN	D/L	AB de Villiers	[RO Gaidwar, F du Plessis, MS Wade, HH Pan...	[RV Uthappa, G Gambhir, SK ...]	Ch
-----	-----	--------	------------	------	-------	-----------------------	----------------	--------------	----------------	-------	---	----------------	---------	-----	-----	----------------	---	---------------------------------	----

freq	NaN	159	2	76	15	126	119	77	123	599	932	131	509	NaN	19	25	2	2	
mean	8.304852e+05	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	17.056867	NaN	NaN	NaN	NaN	NaN
std	3.375678e+05	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	21.633109	NaN	NaN	NaN	NaN	NaN
min	3.359820e+05	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.000000	NaN	NaN	NaN	NaN	NaN
25%	5.012612e+05	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	6.000000	NaN	NaN	NaN	NaN	NaN
50%	8.297380e+05	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	8.000000	NaN	NaN	NaN	NaN	NaN
75%	1.175372e+06	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	19.000000	NaN	NaN	NaN	NaN	NaN
max	1.312200e+06	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	146.000000	NaN	NaN	NaN	NaN	NaN

```
In [15]: ipl_data['City']=ipl_data['City'].fillna(ipl_data['City'].mode()[0])
```

```
In [16]: ipl_data['Margin']=ipl_data['Margin'].fillna(ipl_data['Margin'].mean())
```

```
In [17]: # dropping method columns it contain most missing value
ipl_data.drop('method',axis=1,inplace=True)
```

```
In [18]: # removing the missing value in data
ipl_data.dropna(inplace=True)
```

```
In [19]: #checking the duplicate values in data
ipl_data.duplicated().sum()
```

```
Out[19]: 0
```

```
In [20]: # checking the shape of data
ipl_data.shape
```

```
Out[20]: (946, 19)
```

```
In [21]: # EDA
# creating the empty columns to store year
ipl_data['Season']=1
```

```
In [22]: ipl_data.head()
```

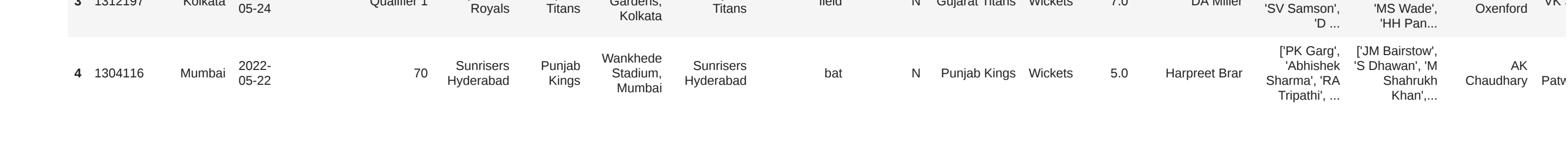
	ID	City	Date	Season	MatchNumber	Team1	Team2	Venue	TossWinner	TossDecision	SuperOver	WinningTeam	WonBy	Margin	Player_of_Match	Team1Players	Team2Players	Umpire1	U	
0	1312200	Ahmedabad	2022-05-29	2022	Final	Rajasthan Royals	Gujarat Titans	Narendra Modi Stadium, Ahmedabad	Rajasthan Royals		bat	N	Gujarat Titans	Wickets	7.0	HH Pandya	[YBK Jaswal, JC Butler, SV Samson, D ...]	[WP Saha, Shubman Gill, MS Wade, HH Pan...	CB Gaffaney	Nitin
1	1312199	Ahmedabad	2022-05-27	2022	Qualifier 2	Royal Challengers Bangalore	Rajasthan Royals	Narendra Modi Stadium, Ahmedabad	Rajasthan Royals		field	N	Rajasthan Royals	Wickets	7.0	JC Butler	[V Kohli, F du Plessis, RM Patidar, GJ ...]	[YBK Jaswal, JC Butler, SV Samson, D ...]	CB Gaffaney	Nitin
2	1312198	Kolkata	2022-05-25	2022	Eliminator	Royal Challengers Bangalore	Lucknow Super Giants	Eden Gardens, Kolkata	Lucknow Super Giants		field	N	Royal Challengers Bangalore	Runs	14.0	RM Patidar	[V Kohli, F du Plessis, RM Patidar, GJ ...]	[Q de Kock, KL Rahul, M Vohra, DJ Madanagopal Hooda...	J MA	MA
3	1312197	Kolkata	2022-05-24	2022	Qualifier 1	Rajasthan Royals	Gujarat Titans	Eden Gardens, Kolkata	Gujarat Titans		field	N	Gujarat Titans	Wickets	7.0	DA Miller	[YBK Jaswal, JC Butler, SV Samson, D ...]	[WP Saha, Shubman Gill, MS Wade, HH Pan...	BNJ Oxtenford	VK
4	1304116	Mumbai	2022-05-22	2022	70	Sunrisers Hyderabad	Punjab Kings	Wankhede Stadium, Mumbai	Sunrisers Hyderabad		bat	N	Punjab Kings	Wickets	5.0	Harpreet Brar	[PK Garg, Abhishek Sharma, RA Tripathi, ...]	[JM Barstow, S Dhawan, M Shatrugh Khatri, ...]	AK Chaudhary	Patv

```
In [23]: ipl_data['Season']=ipl_data['Date'].apply(lambda x:x.split("-")[0])
```

```
In [24]: ipl_data.head()
```

	ID	City	Date	Season	MatchNumber	Team1	Team2	Venue	TossWinner	TossDecision	SuperOver	WinningTeam	WonBy	Margin	Player_of_Match	Team1Players	Team2Players	Umpire1	U	
0	1312200	Ahmedabad	2022-05-29	2022	Final	Rajasthan Royals	Gujarat Titans	Narendra Modi Stadium, Ahmedabad	Rajasthan Royals		bat	N	Gujarat Titans	Wickets	7.0	HH Pandya	[YBK Jaswal, JC Butler, SV Samson, D ...]	[WP Saha, Shubman Gill, MS Wade, HH Pan...	CB Gaffaney	Nitin
1	1312199	Ahmedabad	2022-05-27	2022	Qualifier 2	Royal Challengers Bangalore	Rajasthan Royals	Narendra Modi Stadium, Ahmedabad	Rajasthan Royals		field	N	Rajasthan Royals	Wickets	7.0	JC Butler	[V Kohli, F du Plessis, RM Patidar, GJ ...]	[YBK Jaswal, JC Butler, SV Samson, D ...]	CB Gaffaney	Nitin
2	1312198	Kolkata	2022-05-25	2022	Eliminator	Royal Challengers Bangalore	Lucknow Super Giants	Eden Gardens, Kolkata	Lucknow Super Giants		field	N	Royal Challengers Bangalore	Runs	14.0	RM Patidar	[V Kohli, F du Plessis, RM Patidar, GJ ...]	[Q de Kock, KL Rahul, M Vohra, DJ Madanagopal Hooda...	J MA	MA
3	1312197	Kolkata	2022-05-24	2022	Qualifier 1	Rajasthan Royals	Gujarat Titans	Eden Gardens, Kolkata	Gujarat Titans		field	N	Gujarat Titans	Wickets	7.0	DA Miller	[YBK Jaswal, JC Butler, SV Samson, D ...]	[WP Saha, Shubman Gill, MS Wade, HH Pan...	BNJ Oxtenford	VK
4	1304116	Mumbai	2022-05-22	2022	70	Sunrisers Hyderabad	Punjab Kings	Wankhede Stadium, Mumbai	Sunrisers Hyderabad		bat	N	Punjab Kings	Wickets	5.0	Harpreet Brar	[PK Garg, Abhishek Sharma, RA Tripathi, ...]	[JM Barstow, S Dhawan, M Shatrugh Khatri, ...]	AK Chaudhary	Patv

```
In [29]: # Total number of matches till 2022
plt.figure(figsize=(8,6))
sns.countplot(y='Season',data=ipl_data,order=ipl_data['Season'].value_counts().index)
plt.title('Total Number of IPL Matches 2008 to 2022',fontsize=12)
plt.xlabel('Counts')
plt.ylabel('Season')
plt.grid(True)
plt.show()
```



```
In [33]: # Total Number Of Matches Win by Each Team till 2022
plt.figure(figsize=(8,6))
sns.countplot(y='WinningTeam',data=ipl_data,order=ipl_data['WinningTeam'].value_counts().index)
plt.title('Total Number of IPL Matches Win By Each Team',fontsize=12)
plt.xlabel('Counts')
plt.ylabel('Team')
plt.grid(True)
plt.show()
```



```
In [42]: # Top 10 Cities by Number of Matches
plt.figure(figsize=(20,20))
plt.pie(list(ipl_data['City'].value_counts()),labels=list(ipl_data['City'].value_counts().keys()))
```

```
lables=top_cities.index[0:10]
plt.legend(lables,loc='best')
plt.show()
```



```
In [61]: # Top 10 Man Of The Match
plt.figure(figsize=(20,7))
plt.bar(list(ipl_data['Player_of_Match'].value_counts()[0:10].keys()),list(ipl_data['Player_of_Match'].value_counts()[0:10]))
plt.title('Man Of The Match')
plt.xlabel('player name')
plt.grid(True)
plt.show()
```



```
In [66]: # Top 10 Toss Winner Team
plt.figure(figsize=(20,7))
plt.bar(list(ipl_data['TossWinner'].value_counts()[0:10].keys()),list(ipl_data['TossWinner'].value_counts()[0:10]),color='green')
plt.title('Toss Winner Team')
plt.xlabel('Team name')
plt.grid(True)
plt.show()
```



```
In [71]: # Top 10 1st Umpire to Features In Max Number Of Matches
plt.bar(list(ipl_data['Umpire1'].value_counts()[0:10].keys()),list(ipl_data['Umpire1'].value_counts()[0:10]),color='red')
plt.title('Top 10 Umpire 1')
plt.xlabel('Umpire name')
plt.grid(True)
plt.show()
```



```
In [69]: # Top 10 2nd Umpire to Features In Max Number Of Matches
plt.bar(list(ipl_data['Umpire2'].value_counts()[0:10].keys()),list(ipl_data['Umpire2'].value_counts()[0:10]),color='darkred')
plt.title('Top 10 Umpire 2')
plt.xlabel('Umpire name')
plt.grid(True)
plt.show()
```



```
In [ ]:
```