Sristi Ganesh Bhalerao

Data Science And Business Analytics Intern @ TSF (ARP AUGUST 2021)

Task 3: Exploratory Data Analysis: Sports (Indian Premier League)

```
%matplotlib inline
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
plt.rcParams['figure.dpi'] = 150
```

```
%matplotlib inline
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
plt.rcParams['figure.dpi'] = 150
```

```
%cd "/content/drive/Colab Notebooks/Sparks Foundation/IPL"
```

```
/content/drive/Colab Notebooks/Sparks Foundation/IPL
```

```
!ls -lrtsh
```

```
total 4.0K
4.0K -rw-r-- 1 root root 3.9K Jun 13 16:50 IPL.ipynb
```

```
!unzip -n -q -d . *.zip
```

```
!ls -lrtsh
```

```
total 4.0K
4.0K -rw-r-- 1 root root 3.9K Jun 13 16:50 IPL.ipynb
```

```
!head -n 1 matches.csv
```

```
id,city,date,team1,team2,toss_winner,toss_decision,result,dl_applied,winner,win_by_runs,win_by_wickets,player_of_match,venue,umpire1,umpire3
```

```
!head -n 1 deliveries.csv
```

```
match_id,inning,batting_team,bowling_team,over,ball,batsman,non_striker,bowler,is_super_over,wide_runs,bye_runs,legbye_runs,noball_runs,penalty_runs,batsman_runs,extra_ru
```

```
df = pd.read_csv('matches.csv')
df.shape
```

```
(756, 18)
```

```
df.head(3)
```

	id	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire3
0	1	Sharjah	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Royal Challengers Bangalore	35	0	Yuzvendra Chahal	Sharjah Cricket Stadium	A Nand Kishore	S Ravi
1	2	Dubai	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	7	0	SPD Smith	Dubai International Cricket Stadium	Nitin Menon	CK Nandan
2	3	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	10	0	CA Lynn	Saurashtra Cricket Association Stadium	Nitin Menon	CK Nandan

```
df.tail(3)
```

	id	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire3
753	756	Bengaluru	2019-05-05	Royal Challengers Bangalore	Mumbai Indians	Royal Challengers Bangalore	bat	normal	0	Royal Challengers Bangalore	14	0	Q de Kock	M Chinnaswamy Stadium	Nitin Menon	CK Nandan
754	757	Jaipur	2019-05-06	Kings XI Punjab	Rajasthan Royals	Rajasthan Royals	field	normal	0	Rajasthan Royals	0	7	JJ Bumrah	Sawai Mansingh Stadium	Chettithody Shamshuddin	Anil Dandekar
755	758	Jaipur	2019-05-08	Rajasthan Royals	Mumbai Indians	Mumbai Indians	bat	normal	0	Mumbai Indians	6	0	AS Yadav	Sawai Mansingh Stadium	Chettithody Shamshuddin	Anil Dandekar

```
df.info()
```

```
<class 'DataFrame'>
RangeIndex: 756, 756 entries, 0 to 755
Data columns (total 18):
 id              756 non-null int64
 city            749 non-null object
 date            756 non-null object
 team1           756 non-null object
 team2           756 non-null object
 toss_winner     756 non-null object
 toss_decision   756 non-null object
 result          756 non-null object
 dl_applied      756 non-null int64
 winner          752 non-null object
 win_by_runs     756 non-null int64
 win_by_wickets  756 non-null int64
 player_of_match 752 non-null object
 venue           756 non-null object
 umpire1         754 non-null object
 umpire3         119 non-null object
dtypes: int64(6), object(12)
memory usage: 106.4+ KB
```

```
df.describe(include='all')
```

	id	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire3
count	756.000000	749	756	756	756	756	756	756	756.000000	752	756.000000	756.000000	752	756	754	119
unique	NaN	32	412	15	15	15	2	3	NaN	15	NaN	NaN	226	58	52	19
top	NaN	Mumbai	2019-04-24	Mumbai Indians	Mumbai Indians	Mumbai Indians	field	normal	NaN	Mumbai Indians	NaN	NaN	AB de Villiers	Eden Gardens	HDPK Dharmasena	Nitin Menon
freq	NaN	101	3	85	83	81	499	743	NaN	85	NaN	NaN	20	77	121	15
mean	1426.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.025794	NaN	13.283069	3.350265	NaN	NaN	NaN	NaN
std	1.090909	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.158486	NaN	23.471144	3.387963	NaN	NaN	NaN	NaN
min	1426.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.000000	NaN	0.000000	0.000000	NaN	NaN	NaN	NaN
25%	1427.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.000000	NaN	0.000000	0.000000	NaN	NaN	NaN	NaN
50%	1428.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.000000	NaN	7.000000	4.000000	NaN	NaN	NaN	NaN
75%	1429.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.000000	NaN	19.000000	6.000000	NaN	NaN	NaN	NaN
max	1430.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.000000	NaN	146.000000	10.000000	NaN	NaN	NaN	NaN

```
df.describe()
```

	id	dl_applied	win_by_runs	win_by_wickets
count	756.000000	756.000000	756.000000	756.000000
mean	1426.000000	0.025794	13.283069	3.350265
std	218.397925	0.158486	23.471144	3.387963
min	1426.000000	0.000000	0.000000	0.000000
25%	1427.000000	0.000000	0.000000	0.000000
50%	1428.000000	0.000000	7.000000	4.000000
75%	1429.000000	0.000000	19.000000	6.000000
max	1430.000000	1.000000	146.000000	10.000000

```
df.describe(include='O')
```

	city	date	team1	team2	toss_winner	toss_decision	result	winner	player_of_match	venue	umpire1	umpire3
count	749	756	756	756	756	756	756	752	752	756	754	119
unique	32	412	15	15	15	2	3	15	226	58	52	19
top	Mumbai	2019-04-24	Mumbai Indians	Mumbai Indians	Mumbai Indians	field	normal	Mumbai Indians	AB de Villiers	Eden Gardens	HDPK Dharmasena	Nitin Menon
freq	101	3	85	83	81	499	743	85	20	77	121	15

```
df.nunique()
```

```
id                 756
city                32
date               412
team1               15
team2               15
toss_winner         15
toss_decision        2
result               3
dl_applied           2
winner              15
win_by_runs         98
win_by_wickets      11
player_of_match    226
venue               58
umpire1             52
umpire3             19
dtype: int64
```

```
df.isnull().sum()
```

```
id                 0
city               7
date               0
team1              0
team2              0
toss_winner        0
toss_decision      0
result             0
dl_applied         0
winner             4
win_by_runs        0
win_by_wickets     0
player_of_match    4
venue              0
umpire1            2
umpire3          637
dtype: int64
```

```
df.drop(columns=['umpire3'],inplace=True)
```

```
df.shape
```

```
(756, 17)
```

```
df.columns
```

```
Index(['id', 'city', 'date', 'team1', 'team2', 'toss_winner',
       'toss_decision', 'result', 'dl_applied', 'winner', 'win_by_runs',
       'win_by_wickets', 'player_of_match', 'venue', 'umpire1'],
      dtype='object')
```

Finding Missing Values

```
df.isnull().sum()
```

```
id                 0
city               7
date               0
team1              0
team2              0
toss_winner        0
toss_decision      0
result             0
dl_applied         0
winner             4
win_by_runs        0
win_by_wickets     0
player_of_match    4
venue              0
umpire1            2
dtype: int64
```

```
df[df.winner.isnull()][['team1','team2','winner']]
```

	team1	team2	winner
300	Mumbai Indians	Kolkata Knight Riders	NaN
335	Delhi Daredevils	Kings XI Punjab	NaN
423	Kolkata Knight Riders	Rising Pune Supergiant	NaN
474	Royal Challengers Bangalore	Delhi Daredevils	NaN

```
df.dropna(inplace=True)
```

```
df.shape
```

```
(752, 17)
```

Total IPL Seasons

```
df.shape[0]*2/756
```

```
2.0
```

Total IPL Seasons

```
df.id.unique()
```

```
array([ 1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16, 17,
       18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34,
       35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51,
       52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68,
       69, 70, 71, 72, 73, 74, 75, 76]
```

IPL Team Wins by Scoring The Maximum Runs

```
df[df.win_by_runs==df.win_by_runs.max()]
```

	id	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire3
43	44	Mumbai	2017-05-20	Mumbai Indians	Delhi Daredevils	Delhi Daredevils	bat	normal	0	Mumbai Indians	146	0	LMP Simmons	Wankhede Stadium	Nitin Menon	CK Nandan

```
df[df.win_by_runs==df.win_by_runs.min()]
```

	id	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire3
563	564	Mumbai	2018-05-06	Mumbai Indians	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Royal Challengers Bangalore	1	0	Mohammed Siraj	Wankhede Stadium	Nitin Menon	CK Nandan
607	608	Delhi	2018-05-19	Delhi Daredevils	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	1	0	SP Narine	Feroz Shah Kotla	Nitin Menon	CK Nandan

IPL team win by taking the maximum wickets

```
df[df.win_by_wickets==df.win_by_wickets.max()]
```

	id	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire3
291	292	Kolkata	2017-05-21	Kolkata Knight Riders	Gujarat Lions	Gujarat Lions	field	normal	0	Gujarat Lions	0	10	CA Lynn	Eden Gardens	Nitin Menon	CK Nandan

IPL team win by taking the maximum wickets

```
df[df.win_by_wickets==df.win_by_wickets.min()]
```

	id	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire3
46	47	Bangalore	2017-05-23	Royal Challengers Bangalore	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	1	AD Russell	M Chinnaswamy Stadium	Nitin Menon	CK Nandan
475	476	Delhi	2019-05-12	Delhi Capitals	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	1	SP Narine	Feroz Shah Kotla	Ian Gould	Nitin Menon

```
df[df.win_by_wickets==df.win_by_wickets.max()]
```

	id	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire3
291	292	Kolkata	2017-05-21	Kolkata Knight Riders	Gujarat Lions	Gujarat Lions	field	normal	0	Gujarat Lions	0	10	CA Lynn	Eden Gardens	Nitin Menon	CK Nandan

IPL team win by taking the maximum wickets

```
df[df.win_by_wickets==df.win_by_wickets.min()]
```

	id	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire3
46	47	Bangalore	2017-05-23	Royal Challengers Bangalore	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	1	AD Russell	M Chinnaswamy Stadium	Nitin Menon	CK Nandan
475	476	Delhi	2019-05-12	Delhi Capitals	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	1	SP Narine	Feroz Shah Kotla	Ian Gould	Nitin Menon

```
df[df.win_by_wickets==df.win_by_wickets.max()]
```

	id	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire3
291	292	Kolkata	2017-05-21	Kolkata Knight Riders	Gujarat Lions	Gujarat Lions	field	normal	0	Gujarat Lions	0	10	CA Lynn	Eden Gardens	Nitin Menon	CK Nandan

Man of the match

```
df[df.player_of_match==df.value_counts()[0]]
```

```
id                                                   23
city                                                 Dubai
date                                            2020-09-22
team1                                   Kings XI Punjab
team2                                   Kings XI Punjab
toss_winner                             Kings XI Punjab
toss_decision                                       bat
result                                           normal
dl_antic                                              0
winner                                  Kings XI Punjab
win_by_runs                                          97
win_by_wickets                                        0
player_of_match                            Chris Gayle
venue                                   Dubai Sports City
umpire1                                     Ian Gould
umpire3                                     Nitin Menon
Name: 0, dtype: object
```

```
df.shape[0]/df.match_winner.value_counts()
```

```
NameError: name 'df.match_winner' is not defined
```

Venue for maximum matches

```
df[df.match_winner.isin(df.match_winner.value_counts()[:3])].venue.value_counts()
```

```
NameError: name 'df.match_winner' is not defined
```

Umpire for maximum matches

```
df[df.umpire1.isin(df.umpire1.value_counts()[:3])].umpire1.value_counts()
```

```
NameError: name 'df.umpire1' is not defined
```

Performance of the bowlers

```
df[df.player_of_match==df.player_of_match.value_counts()[:3]].sort_values(ascending=False).head(10).plot(kind='bar',figsize=(15,5))
```

```
NameError: name 'df.player_of_match' is not defined
```

Top Run Getters

```
df[df.player_of_match==df.player_of_match.value_counts()[:3].sort_values(ascending=False).head(10)].plot(kind='bar',figsize=(15,5))
```

```
NameError: name 'df.player_of_match' is not defined
```

Decision after winning toss

```
df[df.toss_decision.value_counts()[:2].sort_values(ascending=False).head(10)].plot(kind='bar',figsize=(15,5))
```

```
NameError: name 'df.toss_decision' is not defined
```

Teams lost first or second after winning toss

```
df[df.toss_decision.value_counts()[:2].sort_values(ascending=False).head(10)].toss_decision.value_counts()
```

```
NameError: name 'df.toss_decision' is not defined
```

Conclusion

- 756 matches were played according to the dataset.
- From 2008 to 2019 IPL seasons are using to analyze.
- Mumbai Indians wins by scoring the maximum runs at 146.
- Kolkata Knight Riders won by taking maximum wickets is 10.
- Sunrisers Hyderabad won by making minimum wickets is 1.
- In season 2013 contained the highest number of matches over played followed by 2012 and 2011.
- Mumbai Indians is the most successful IPL team followed by Chennai Super Kings and Kolkata Knight Riders.
- Chi Gayle is the man of the matches maximum time followed by AB de Villiers.
- The maximum number of matches were played in mumbai.
- Eden Gardens has hosted the maximum number of IPL matches followed by Wankhede Stadium and M Chinnaswamy Stadium.
- S Ravi has appeared in most number of matches as a umpire.
- Virat Kohli is the highest run getter in IPL.
- Upto 81% times teams who have won tosses have decided to chose down.