

EDA on IPL Dataset

Importing libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

Importing IPL dataset

```
data=pd.read_csv("/matches.csv")
data.head(5)
```

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



```
data.tail()
```

	id	Season	city	date	team1	team2	toss_winner	toss_decisi
751	11347	IPL-2019	Mumbai	05-05-2019	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	fi
752	11412	IPL-2019	Chennai	07-05-2019	Chennai Super Kings	Mumbai Indians	Chennai Super Kings	
753	11413	IPL-2019	Visakhapatnam	08-05-2019	Sunrisers Hyderabad	Delhi Capitals	Delhi Capitals	fi
754	11414	IPL-2019	Visakhapatnam	10-05-2019	Delhi Capitals	Chennai Super Kings	Chennai Super Kings	fi
755	11415	IPL-2019	Hyderabad	12-05-2019	Mumbai Indians	Chennai Super Kings	Mumbai Indians	

How big is the dataset? (Rows and columns)



```
data.shape
```

```
(756, 18)
```

```
data.info
```

```
<bound method DataFrame.info of          id    Season  ...      umpire2
0         1  IPL-2017  ...      NJ Llong      NaN
1         2  IPL-2017  ...      S Ravi      NaN
2         3  IPL-2017  ...      CK Nandan      NaN
3         4  IPL-2017  ...  C Shamshuddin      NaN
4         5  IPL-2017  ...           NaN      NaN
..      ...      ...      ...      ...
751    11347  IPL-2019  ...      O Nandan      S Ravi
752    11412  IPL-2019  ...      Nitin Menon      Ian Gould
753    11413  IPL-2019  ...           NaN      NaN
754    11414  IPL-2019  ...  Bruce Oxenford  Chettithody Shamshuddin
755    11415  IPL-2019  ...      Ian Gould      Nigel Llong

[756 rows x 18 columns]>
```



Data Pre-processing: Finding out NaN values

```
data.isna().any()
```

```
id          False
Season      False
city        True
date        False
team1       False
team2       False
```


```

toss_winner      False
toss_decision    False
result           False
dl_applied       False
winner           True
win_by_runs      False
win_by_wickets   False
player_of_match  True
venue            False
umpire1          True
umpire2          True
umpire3          True
dtype: bool

```

Statistical Description of dataset

```
data.describe()
```

	id	dl_applied	win_by_runs	win_by_wickets	
count	756.000000	756.000000	756.000000	756.000000	
mean	1792.178571	0.025132	13.283069	3.350529	
std	3464.478148	0.156630	23.471144	3.387963	
min	1.000000	0.000000	0.000000	0.000000	
25%	189.750000	0.000000	0.000000	0.000000	
50%	378.500000	0.000000	0.000000	4.000000	
75%	567.250000	0.000000	19.000000	6.000000	
max	11415.000000	1.000000	146.000000	10.000000	

How many matches (in total) were played according to the dataset?

```
data['id'].count()
```

```
756
```

How many IPL seasons are we using to analyse?

```
data['Season'].unique()
```

```

array(['IPL-2017', 'IPL-2008', 'IPL-2009', 'IPL-2010', 'IPL-2011',
      'IPL-2012', 'IPL-2013', 'IPL-2014', 'IPL-2015', 'IPL-2016',
      'IPL-2018', 'IPL-2019'], dtype=object)

```

Which IPL team won by scoring the maximum runs?

```
data.iloc[data['win_by_runs'].idxmax()]

id          44
Season      IPL-2017
city        Delhi
date        06-05-2017
team1       Mumbai Indians
team2       Delhi Daredevils
toss_winner Delhi Daredevils
toss_decision field
result      normal
dl_applied  0
winner      Mumbai Indians
win_by_runs 146
win_by_wickets 0
player_of_match LMP Simmons
venue          Feroz Shah Kotla
umpire1        Nitin Menon
umpire2        CK Nandan
umpire3        NaN
Name: 43, dtype: object
```

Which IPL team won by consuming maximum wickets?

```
data.iloc[data['win_by_wickets'].idxmax()]

id          3
Season      IPL-2017
city        Rajkot
date        07-04-2017
team1       Gujarat Lions
team2       Kolkata Knight Riders
toss_winner Kolkata Knight Riders
toss_decision field
result      normal
dl_applied  0
winner      Kolkata Knight Riders
win_by_runs 0
win_by_wickets 10
player_of_match CA Lynn
venue          Saurashtra Cricket Association Stadium
umpire1        Nitin Menon
umpire2        CK Nandan
umpire3        NaN
Name: 2, dtype: object
```

Which IPL team won by taking minimum wickets?

```
data.iloc[data['win_by_wickets'].idxmin()]

id          1
Season      IPL-2017
city        Hyderabad
date        05-04-2017
```

```

team1                Sunrisers Hyderabad
team2                Royal Challengers Bangalore
toss_winner          Royal Challengers Bangalore
toss_decision        field
result               normal
dl_applied           0
winner               Sunrisers Hyderabad
win_by_runs          35
win_by_wickets       0
player_of_match      Yuvraj Singh
venue                Rajiv Gandhi International Stadium, Uppal
umpire1              AY Dandekar
umpire2              NJ Llong
umpire3              NaN
Name: 0, dtype: object

```

Which season consisted of the highest number of matches ever played?

```

fig_dims = (20, 4)
fig, ax = plt.subplots(figsize=fig_dims)
sns.countplot(x='Season', ax=ax, data=data)
plt.show()

```



Which is the most successful IPL team with all the data at hand?

```

data1 = data.winner.value_counts()
sns.barplot(y = data1.index, x = data1)

```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9123b35ed0>
```

What is the probability of winning a match if the toss was won?

Kolkata Knight Riders



```
probability_of_win = data['toss_winner'] == data['winner']
```

```
probability_of_win.groupby(probability_of_win).size()
```

```
False    363
True     393
dtype: int64
```

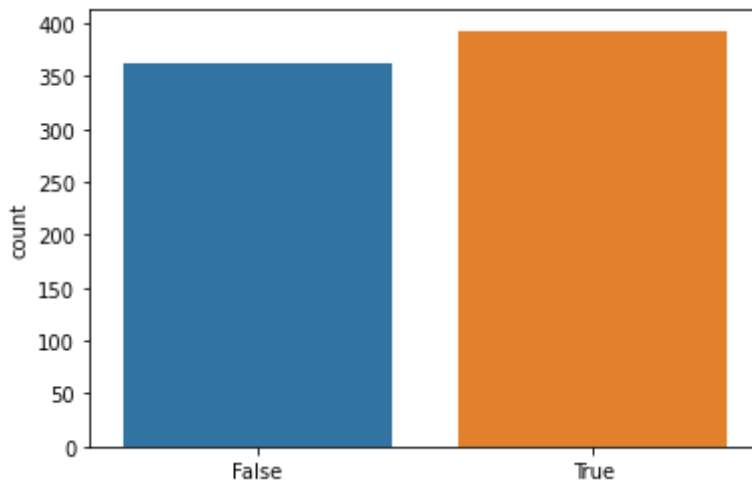
Kochi Tuskers Kerala



```
sns.countplot(probability_of_win)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pas
FutureWarning
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f91235b0bd0>
```



Setting a higher row width

```
pd.set_option('max_rows', 99999)
pd.set_option('max_colwidth', 400)
pd.describe_option('max_colwidth')
```

```
display.max_colwidth : int or None
```

The maximum width in characters of a column in the repr of a pandas data structure. When the column overflows, a "..." placeholder is embedded in the output. A 'None' value means unlimited. [default: 50] [currently: 400]

Highest wins by teams per season

```
data.groupby('Season')['winner'].value_counts()
```

```
IPL-2013  Mumbai Indians          13
          Chennai Super Kings     12
          Rajasthan Royals        11
          Sunrisers Hyderabad     10
          Royal Challengers Bangalore 9
```

```

Royal Challengers Bangalore      5
Kings XI Punjab                  8
Kolkata Knight Riders            6
Pune Warriors                    4
Delhi Daredevils                 3
IPL-2014 Kings XI Punjab          12
IPL-2014 Kolkata Knight Riders    11
IPL-2014 Chennai Super Kings     10
IPL-2014 Mumbai Indians          7
IPL-2014 Rajasthan Royals        7
IPL-2014 Sunrisers Hyderabad     6
IPL-2014 Royal Challengers Bangalore 5
IPL-2014 Delhi Daredevils        2
IPL-2015 Chennai Super Kings     10
IPL-2015 Mumbai Indians          10
IPL-2015 Royal Challengers Bangalore 8
IPL-2015 Kolkata Knight Riders    7
IPL-2015 Rajasthan Royals        7
IPL-2015 Sunrisers Hyderabad     7
IPL-2015 Delhi Daredevils        5
IPL-2015 Kings XI Punjab          3
IPL-2016 Sunrisers Hyderabad     11
IPL-2016 Gujarat Lions           9
IPL-2016 Royal Challengers Bangalore 9
IPL-2016 Kolkata Knight Riders    8
IPL-2016 Delhi Daredevils        7
IPL-2016 Mumbai Indians          7
IPL-2016 Rising Pune Supergiants  5
IPL-2016 Kings XI Punjab          4
IPL-2017 Mumbai Indians          12
IPL-2017 Rising Pune Supergiant   10
IPL-2017 Kolkata Knight Riders    9
IPL-2017 Sunrisers Hyderabad     8
IPL-2017 Kings XI Punjab          7
IPL-2017 Delhi Daredevils        6
IPL-2017 Gujarat Lions           4
IPL-2017 Royal Challengers Bangalore 3
IPL-2018 Chennai Super Kings     11
IPL-2018 Sunrisers Hyderabad     10
IPL-2018 Kolkata Knight Riders    9
IPL-2018 Rajasthan Royals        7
IPL-2018 Kings XI Punjab          6
IPL-2018 Mumbai Indians          6
IPL-2018 Royal Challengers Bangalore 6
IPL-2018 Delhi Daredevils        5
IPL-2019 Mumbai Indians          11
IPL-2019 Chennai Super Kings     10
IPL-2019 Delhi Capitals          10
IPL-2019 Kings XI Punjab          6
IPL-2019 Kolkata Knight Riders    6
IPL-2019 Sunrisers Hyderabad     6
IPL-2019 Rajasthan Royals        5
IPL-2019 Royal Challengers Bangalore 5
Name: winner, dtype: int64

```

```
data['toss_decision'].value_counts()
```

```

field    463
bat      293

```

Name: toss_decision, dtype: int64

Man of the match - Highest to lowest (in won matches)

```
data['player_of_match'].value_counts()
```

R Bhatia	1
EJG Morgan	1
Mustafizur Rahman	1
S Curran	1
PD Collingwood	1
SM Katich	1
WPUJC Vaas	1
RJ Harris	1
MN Samuels	1
Z Khan	1
JDP Oram	1
LH Ferguson	1
MS Bisla	1
S Anirudha	1
S Badrinath	1
S Sohal	1
RA Tripathi	1
KV Sharma	1
AD Mathews	1
Imran Tahir	1
J Botha	1
Washington Sundar	1
DL Vettori	1
HH Gibbs	1
TG Southee	1
B Stanlake	1
Kuldeep Yadav	1
YS Chahal	1
JEC Franklin	1
LJ Wright	1
Q de Kock	1
DP Nannes	1
M Ur Rahman	1
MR Marsh	1
SA Yadav	1
TM Dilshan	1
P Parameswaran	1
BCJ Cutting	1
Mohammed Shami	1
K Rabada	1
Mohammed Siraj	1
KK Cooper	1
DJG Sammy	1
A Joseph	1
A Singh	1
CR Brathwaite	1
S Gill	1
MF Maharroof	1
JD Ryder	1
A Chandila	1
AC Voges	1
P Shaw	1
Shoaib Akhtar	1


```
CRD Fernando          1
SB Wagh                1
K Paul                 1
GJ Bailey              1
Name: player_of_match, dtype: int64
```

In which city were the number of matches played?

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

```
Mumbai          101
Kolkata          77
Delhi            74
Bangalore        66
Hyderabad        64
Chennai          57
Jaipur           47
Chandigarh       46
Pune             38
Durban           15
Bengaluru        14
Visakhapatnam    13
Ahmedabad        12
Centurion        12
Mohali           10
Rajkot           10
Indore           9
Dharamsala       9
Johannesburg     8
Port Elizabeth   7
Abu Dhabi        7
Cuttack          7
Cape Town        7
Ranchi           7
Raipur           6
Sharjah          6
Kochi            5
Kanpur           4
Nagpur           3
Kimberley        3
East London      3
Bloemfontein     2
Name: city, dtype: int64
```

✓ 0s completed at 10:32 PM

● ✕