ipl-data-analysis

July 14, 2023

[1]:
```
#importing libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

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

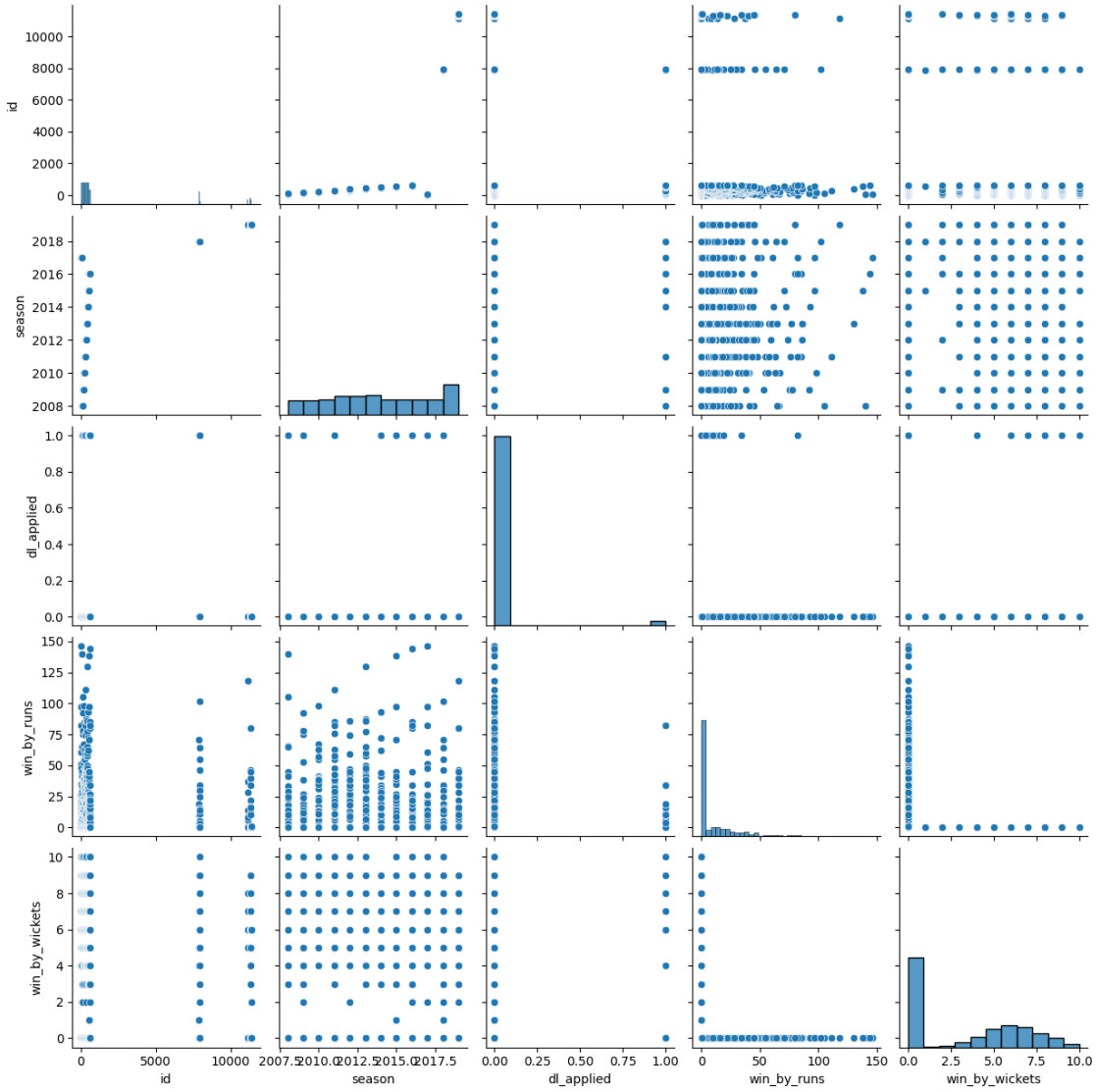
[1]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	35	0	Yuvraj Singh	Rajiv Gandhi International Stadium, Uppal
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	0	7	SPD Smith	Maharashtra Cricket Association Stadium
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	10	CA Lynn	Saurashtra Cricket Association Stadium
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	0	6	GJ Maxwell	Holkar Cricket Stadium
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	0	Royal Challengers Bangalore	15	0	KM Jadhav	M Chinnaswamy Stadium

	umpire1	umpire2	umpire3
0	AY Dandekar	NJ Llong	NaN
1	A Nand Kishore	S Ravi	NaN
2	Nitin Menon	CK Nandan	NaN
3	AK Chaudhary	C Shamshuddin	NaN
4	NaN	NaN	NaN

[101]:
```
# pairplot
sns.pairplot(df)
```

[101]: <seaborn.axisgrid.PairGrid at 0x780a5dec2e90>



[2]:
```
df.shape
```

[2]:
```
(756, 18)
```

[3]:
```
df['player_of_match'].value_counts()
# understanding the categorical column values occurences
```

[3]:
```
CH Gayle           21
AB de Villiers     20
RG Sharma          17
MS Dhoni           17
DA Warner          17
                   ..
PD Collingwood      1
NV Ojha             1
AC Voges            1
J Theron            1
S Hetmyer           1
Name: player_of_match, Length: 226, dtype: int64
```

[7]:
```
df['winner'].value_counts()
```

[7]:
```
Mumbai Indians                 109
Chennai Super Kings            100
Kolkata Knight Riders           92
Royal Challengers Bangalore     84
Kings XI Punjab                 82
Rajasthan Royals                75
Delhi Daredevils                67
Sunrisers Hyderabad             58
Deccan Chargers                 29
Gujarat Lions                   13
Pune Warriors                   12
Rising Pune Supergiant          10
Delhi Capitals                  10
Kochi Tuskers Kerala             6
Rising Pune Supergiants          5
Name: winner, dtype: int64
```

[8]:
```
df['city'].value_counts()
```

[8]:
```
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
Bloemfontein        2
Name: city, dtype: int64
```

[9]:
```
# top 7 players with most no of MOTM players
df['player_of_match'].value_counts()[0:7]
```

[9]:
```
CH Gayle          21
AB de Villiers    20
RG Sharma         17
MS Dhoni          17
DA Warner         17
YK Pathan         16
SR Watson         15
Name: player_of_match, dtype: int64
```

[10]:
```
list(df['player_of_match'].value_counts()[0:3].keys())
```

[10]:
```
['CH Gayle', 'AB de Villiers', 'RG Sharma']
```

[11]:
```
list(df['city'].value_counts()[0:3].keys())
```

[11]:
```
['Mumbai', 'Kolkata', 'Delhi']
```

[12]:
```
list(df['toss_winner'].value_counts()[0:3].keys())
```

[12]:
```
['Mumbai Indians', 'Kolkata Knight Riders', 'Chennai Super Kings']
```

[16]:
```
list(df['win_by_runs'].value_counts()[0:3].keys())
```

[16]:
```
[0, 14, 4]
```

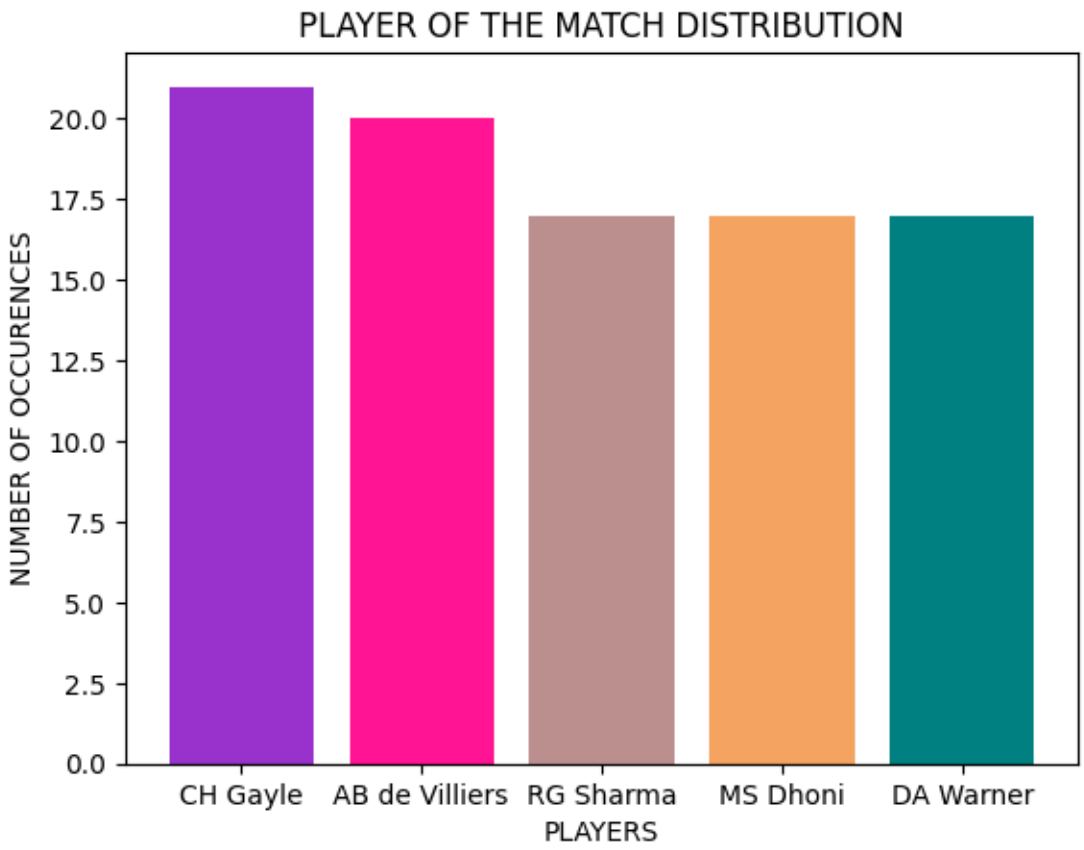
[17]:
```
list(df['win_by_wickets'].value_counts()[0:3].keys())
```

[17]:
```
[0, 6, 7]
```

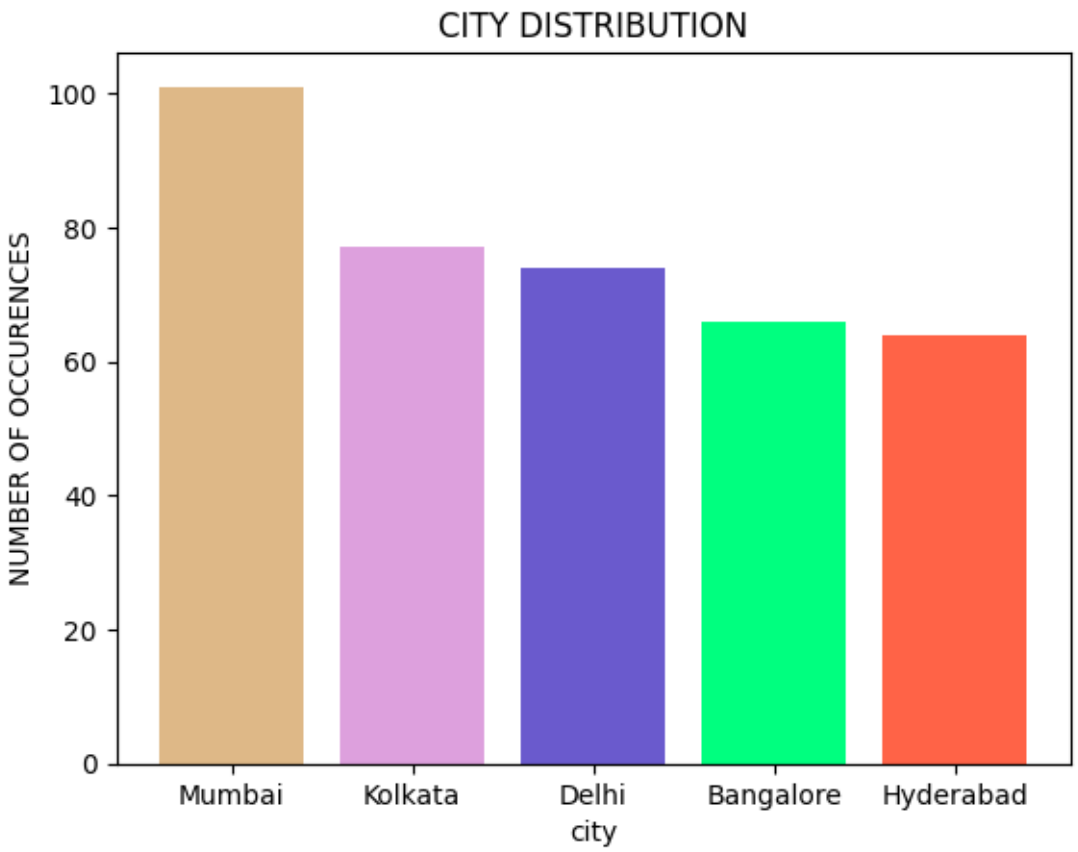
[75]:
```
#lineplot
sns.lineplot(x="season", y="winner" ,color='cyan',data=df)
```

[75]:
```
<Axes: xlabel='season', ylabel='winner'>
```

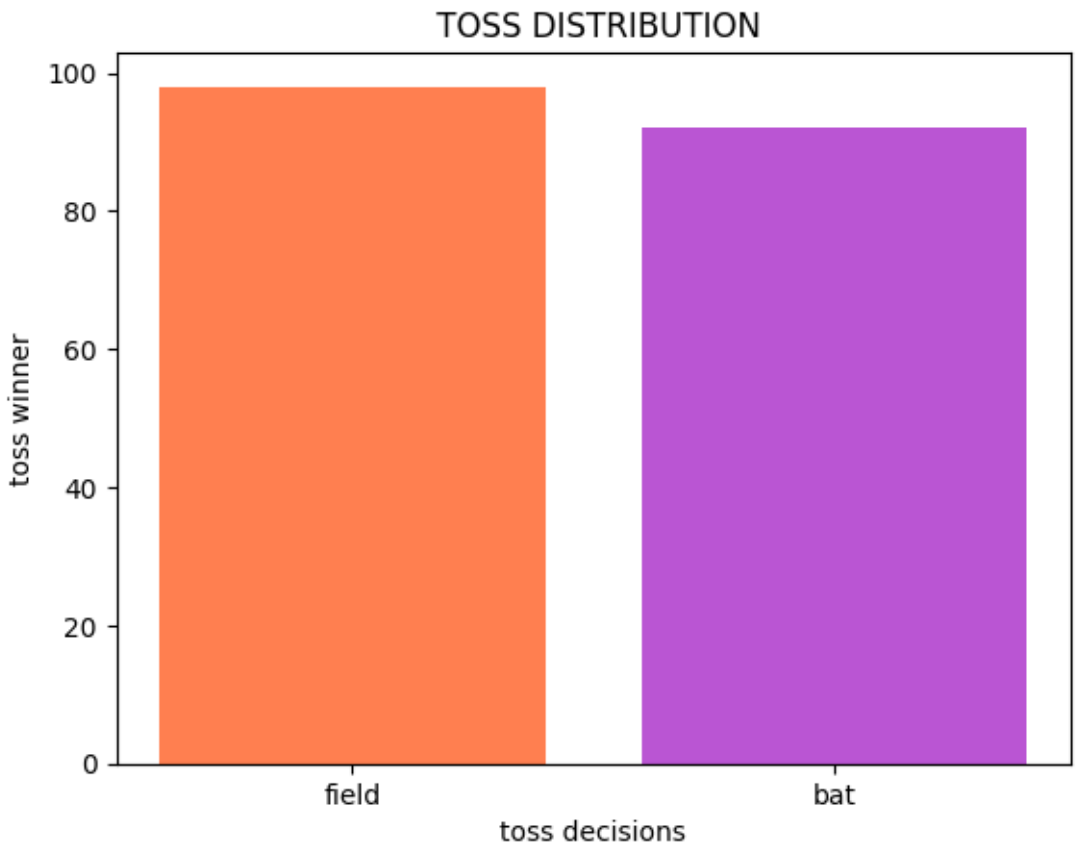
[92]:
```
#barplots
plt.bar(list(df['player_of_match'].value_counts()[0:5].
  ↪keys()),list(df['player_of_match'].value_counts()[0:
  ↪5]),color=['darkorchid','deeppink','rosybrown','sandybrown','teal'])
plt.xlabel("PLAYERS")
plt.ylabel("NUMBER OF OCCURENCES")
plt.title("PLAYER OF THE MATCH DISTRIBUTION")
plt.show()
```



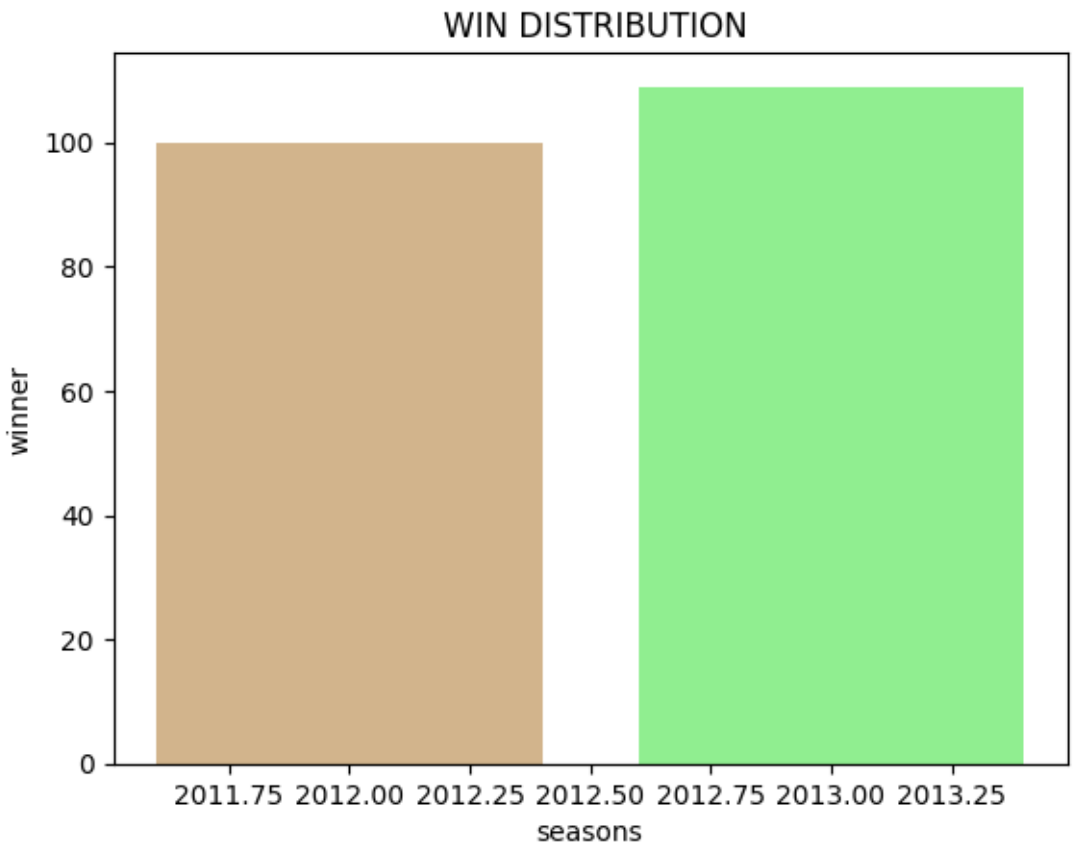
[94]:
```
plt.bar(list(df['city'].value_counts()[0:5].keys()),list(df['city'].
  ↪value_counts()[0:
  ↪5]),color=['burlywood','plum','slateblue','springgreen','tomato'])
plt.xlabel("city")
plt.ylabel("NUMBER OF OCCURENCES")
plt.title("CITY DISTRIBUTION")
plt.show()
```



[95]:
```
plt.bar(list(df['toss_decision'].value_counts()[0:2].
  ↪keys()),list(df['toss_winner'].value_counts()[0:
  ↪2]),color=['coral','mediumorchid'])
plt.xlabel("toss decisions")
plt.ylabel("toss winner")
plt.title(" TOSS DISTRIBUTION")
plt.show()
```



[96]:
```
plt.bar(list(df['season'].value_counts()[0:2].keys()),list(df['winner'].
  ↪value_counts()[0:2]),color=['lightgreen','tan'])
plt.xlabel("seasons")
plt.ylabel("winner")
plt.title(" WIN DISTRIBUTION")
plt.show()
```



[43]:
```
df['result'].value_counts()
```

[43]:
```
normal       743
tie            9
no result      4
Name: result, dtype: int64
```

In the above result no result denotes the cancellation of match due to unavoidable circumstances.

[44]:
```
df['toss_winner'].value_counts()
```

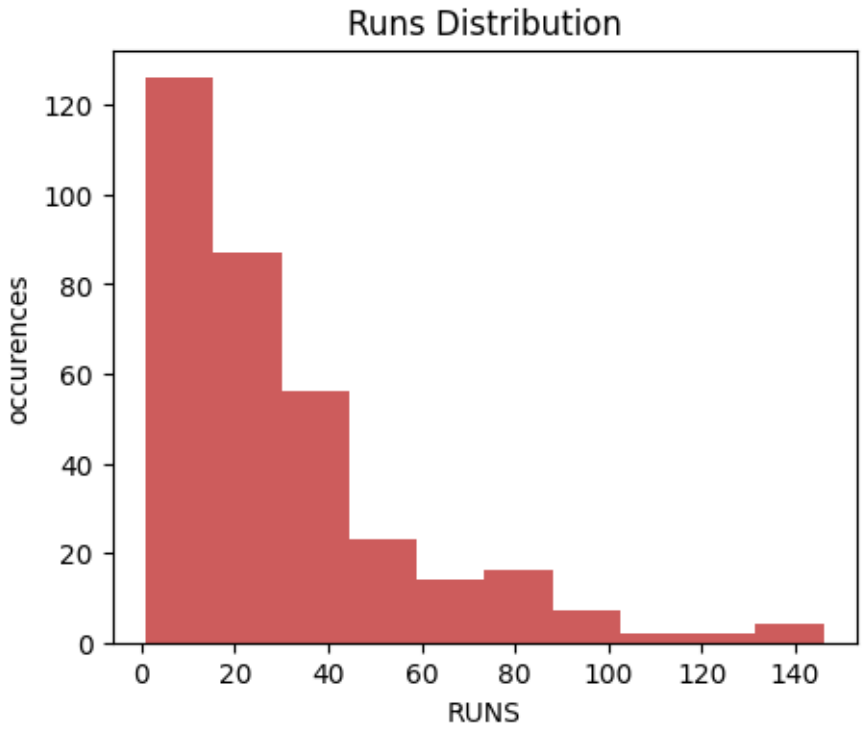
[44]:
```
Mumbai Indians                 98
Kolkata Knight Riders          92
Chennai Super Kings            89
Royal Challengers Bangalore    81
Kings XI Punjab                81
Delhi Daredevils               80
Rajasthan Royals               80
Sunrisers Hyderabad            46
Deccan Chargers                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
```

[45]:
```
# to describe the match winner by runs by assuming the runs to be a non-zero␣
↪value
batting_first = df[df['win_by_runs']!=0]
batting_first.head()
```

[45]:
```
    id  season       city        date                         team1  \
0    1    2017  Hyderabad  2017-04-05          Sunrisers Hyderabad
4    5    2017  Bangalore  2017-04-08  Royal Challengers Bangalore
8    9    2017       Pune  2017-04-11              Delhi Daredevils
13  14    2017    Kolkata  2017-04-15        Kolkata Knight Riders
14  15    2017      Delhi  2017-04-15              Delhi Daredevils

                         team2                 toss_winner toss_decision  \
0   Royal Challengers Bangalore  Royal Challengers Bangalore         field
4             Delhi Daredevils  Royal Challengers Bangalore           bat
8        Rising Pune Supergiant       Rising Pune Supergiant         field
13          Sunrisers Hyderabad          Sunrisers Hyderabad         field
14              Kings XI Punjab             Delhi Daredevils           bat

    result  dl_applied                       winner  win_by_runs  \
0   normal           0          Sunrisers Hyderabad           35
4   normal           0  Royal Challengers Bangalore           15
8   normal           0             Delhi Daredevils           97
13  normal           0        Kolkata Knight Riders           17
14  normal           0             Delhi Daredevils           51

    win_by_wickets player_of_match                                      venue  \
0                0    Yuvraj Singh  Rajiv Gandhi International Stadium, Uppal
4                0       KM Jadhav                      M Chinnaswamy Stadium
8                0       SV Samson    Maharashtra Cricket Association Stadium
13               0      RV Uthappa                                Eden Gardens
14               0     CJ Anderson                            Feroz Shah Kotla

        umpire1      umpire2 umpire3
0   AY Dandekar     NJ Llong     NaN
4           NaN          NaN     NaN
8   AY Dandekar       S Ravi     NaN
13  AY Dandekar     NJ Llong     NaN
14     YC Barde  Nitin Menon     NaN
```

[98]:
```
# analysing with a histogram
plt.figure(figsize=(5,4))
plt.hist(batting_first['win_by_runs'],color='indianred')
plt.title("Runs Distribution")
plt.xlabel('RUNS')
plt.ylabel("occurences")
plt.show()
```

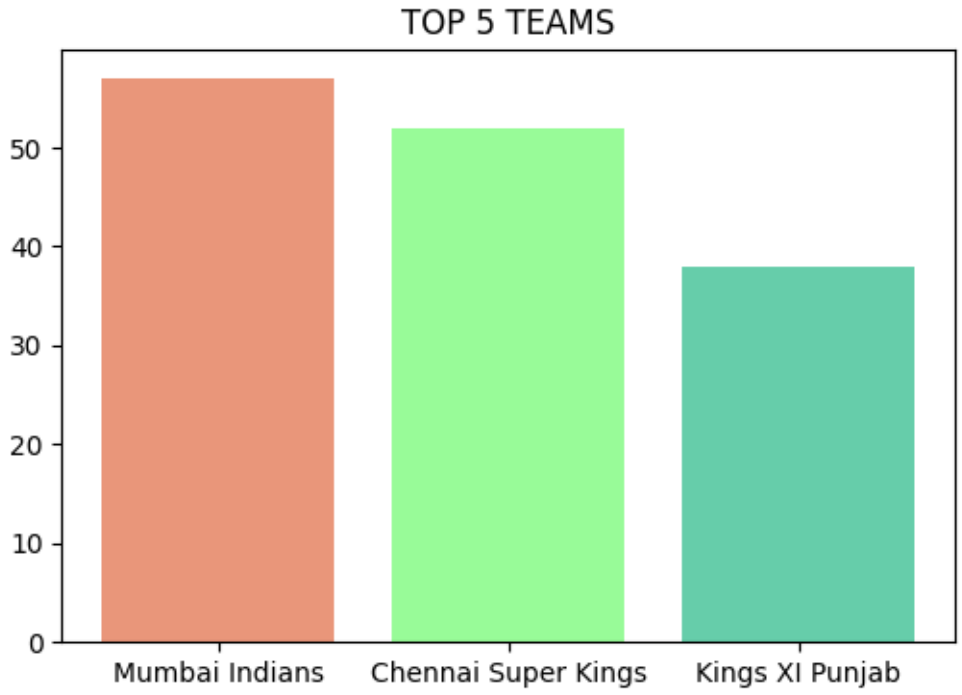


[48]:
```
batting_first['winner'].value_counts()
```

[48]:
```
Mumbai Indians                 57
Chennai Super Kings            52
Kings XI Punjab                38
Kolkata Knight Riders          36
Royal Challengers Bangalore    35
Sunrisers Hyderabad            30
Rajasthan Royals               27
Delhi Daredevils               25
Deccan Chargers                18
Pune Warriors                   6
Rising Pune Supergiant          5
Delhi Capitals                  3
Kochi Tuskers Kerala            2
```

```
Rising Pune Supergiants         2
Gujarat Lions                   1
Name: winner, dtype: int64
```

[50]:
```
# predicting the top 3 teams
plt.figure(figsize=(6,4))
plt.bar(list(batting_first['winner'].value_counts()[0:3].
 ↪keys()),list(batting_first['winner'].value_counts()[0:
 ↪3]),color=['darksalmon','palegreen','mediumaquamarine'])
plt.title("TOP 5 TEAMS")
plt.show()
```



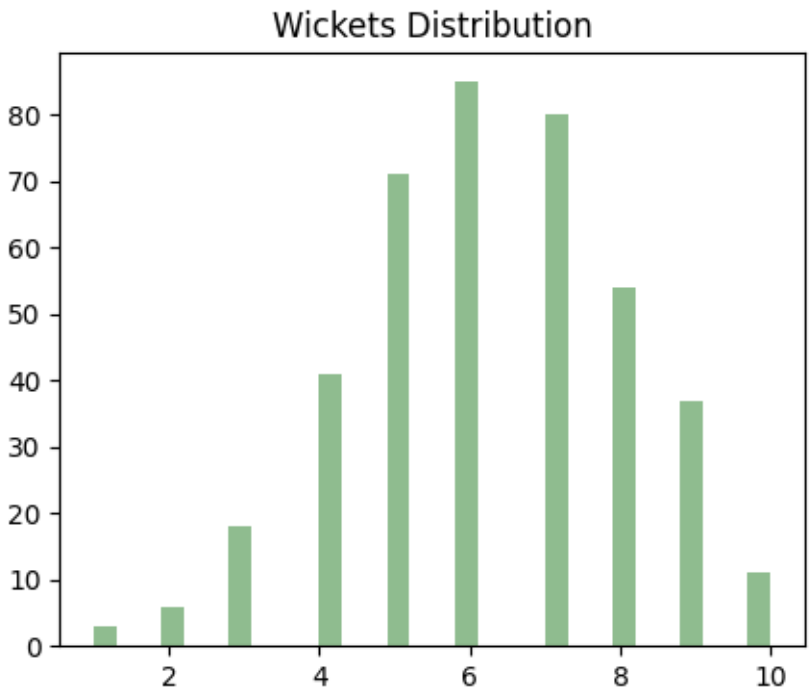
[51]:
```
batting_seconds = df[df['win_by_wickets']!=0]
batting_seconds.head()
```

[51]:

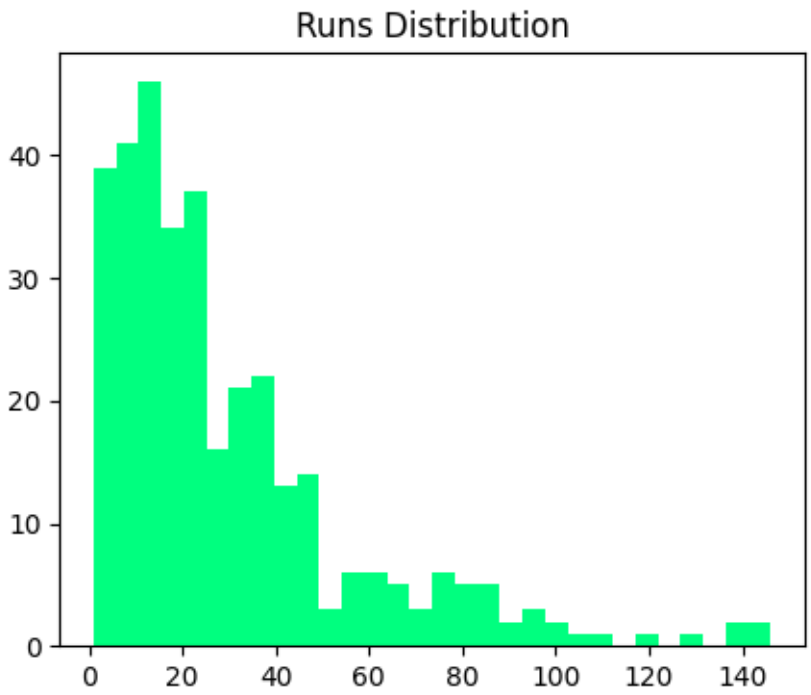
	id	season	city	date	team1	team2	toss_winner	toss_decision	result
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal
2	3	2017	Rajkot	2017-04-07	Gujarat Lions				
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant				
5	6	2017	Hyderabad	2017-04-09	Gujarat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

```
2   Kolkata Knight Riders   Kolkata Knight Riders         field  normal
3         Kings XI Punjab         Kings XI Punjab         field  normal
5     Sunrisers Hyderabad     Sunrisers Hyderabad         field  normal
6           Mumbai Indians          Mumbai Indians         field  normal

   dl_applied                 winner  win_by_runs  win_by_wickets  \
1           0  Rising Pune Supergiant            0               7
2           0   Kolkata Knight Riders            0              10
3           0         Kings XI Punjab            0               6
5           0     Sunrisers Hyderabad            0               9
6           0          Mumbai Indians            0               4

  player_of_match                                      venue         umpire1  \
1       SPD Smith    Maharashtra Cricket Association Stadium  A Nand Kishore
2         CA Lynn     Saurashtra Cricket Association Stadium     Nitin Menon
3      GJ Maxwell                      Holkar Cricket Stadium    AK Chaudhary
5     Rashid Khan  Rajiv Gandhi International Stadium, Uppal      A Deshmukh
6          N Rana                           Wankhede Stadium     Nitin Menon

         umpire2 umpire3
1         S Ravi     NaN
2      CK Nandan     NaN
3  C Shamshuddin     NaN
5        NJ Llong     NaN
6      CK Nandan     NaN
```

[52]:
```
plt.figure(figsize=(5,4))
plt.hist(batting_seconds['win_by_wickets'],color='darkseagreen',bins=30)
plt.title("Wickets Distribution")
plt.show()
```

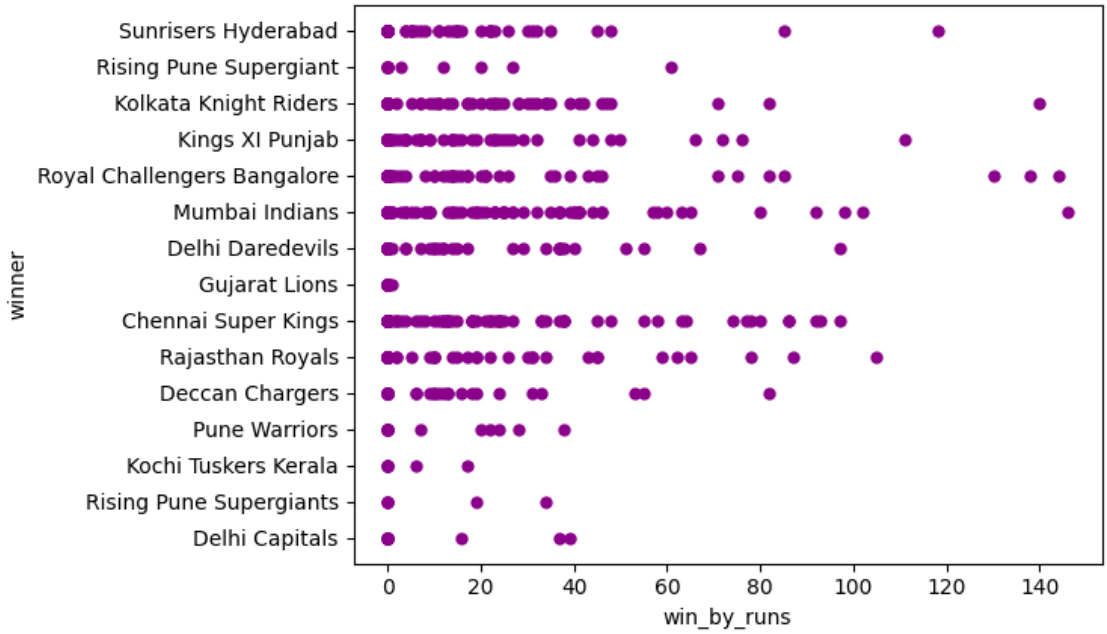


[60]:
```
plt.figure(figsize=(5,4))
plt.hist(batting_first['win_by_runs'],color='springgreen',bins=30)
plt.title("Runs Distribution")
plt.show()
```



[78]:
```
# scatter plot
sns.scatterplot(x="win_by_runs", y="winner", sizes=(1, 8),color='darkmagenta',␣
↪linewidth=0,data=df)
```

[78]: <Axes: xlabel='win_by_runs', ylabel='winner'>



[89]:
```
#rel plot
sns.relplot(x="win_by_wickets", y="winner",color = 'turquoise',data=df)
```

[89]: <seaborn.axisgrid.FacetGrid at 0x780a5a5c8f10>