2/19/2021 EDA sports IPL

Nagarkoti Mohit Singh

Data Science and Business Analytics Intern @ TSF

Exploratory Data Analysis : Sports (Indian Premier League)

```
Dataset: matches.csv (https://bit.ly/34SRn3b (https://bit.ly/34SRn3b))

In [1]: import pandas as pd import numpy as np import matplotlib.pyplot as plt import seaborn as sns %matplotlib inline
```

import warnings
warnings.filterwarnings('ignore')

Out[2]: city team1 team2 toss_winner toss_decision result dl_applied umpire2 umpire3 **0** 1 2017 Hyderabad 2017-04-05 Sunrisers Hyderabad Royal Challengers Bangalore Royal Challengers Bangalore Sunrisers Hyderabad Yuvraj Singh Rajiv Gandhi International Stadium, Uppal AY Dandekar **1** 2 2017 Pune 2017-04-06 Mumbai Indians Rising Pune Supergiant Rising Pune Supergiant Rising Pune Supergiant SPD Smith Maharashtra Cricket Association Stadium A Nand Kishore CA Lynn Saurashtra Cricket Association Stadium **2** 3 2017 Rajkot 2017-04-07 Kolkata Knight Riders Kolkata Knight Riders Gujarat Lions Kolkata Knight Riders field normal Indore 2017-04-08 GJ Maxwell Rising Pune Supergiant Kings XI Punjab Kings XI Punjab Kings XI Punjab Holkar Cricket Stadium AK Chaudhary C Shamshuddin **4** 5 2017 Bangalore 2017-04-08 Royal Challengers Bangalore KM Jadhav M Chinnaswamy Stadium Delhi Daredevils Royal Challengers Bangalore 0 Royal Challengers Bangalore

In [3]: Data = pd.read_csv("deliveries.csv")

Data.head()

| Match_in | Institute | Insti

5 rows × 21 columns

In [4]: season_data=data[['id','season','winner']]

complete_data=Data.merge(season_data,how='inner',left_on='match_id',right_on='id')

In [5]: data.columns.values

data.head()

'umpire3'], dtype=object)

In [6]: data = data.drop(columns=["umpire3"],axis=1)

Out[6]: city date team1 team2 toss_winner toss_decision result dl_applied winner win_by_runs win_by_wickets player_of_match umpire1 umpire2 id season venue **0** 1 2017 Hyderabad 2017-04-05 Sunrisers Hyderabad Royal Challengers Bangalore Royal Challengers Bangalore Sunrisers Hyderabad Yuvraj Singh Rajiv Gandhi International Stadium, Uppal NJ Llong Pune 2017-04-06 SPD Smith Maharashtra Cricket Association Stadium A Nand Kishore S Ravi **1** 2 2017 Mumbai Indians Rising Pune Supergiant Rising Pune Supergiant Rising Pune Supergiant field normal Rajkot 2017-04-07 Kolkata Knight Riders CK Nandan Gujarat Lions Kolkata Knight Riders field normal Kolkata Knight Riders Saurashtra Cricket Association Stadium Indore 2017-04-08 GJ Maxwell **3** 4 2017 Rising Pune Supergiant Kings XI Punjab Kings XI Punjab Holkar Cricket Stadium AK Chaudhary C Shamshuddin Kings XI Punjab field normal **4** 5 2017 Bangalore 2017-04-08 Royal Challengers Bangalore M Chinnaswamy Stadium Delhi Daredevils Royal Challengers Bangalore 0 Royal Challengers Bangalore KM Jadhav

In [7]: wins_per_season = data.groupby("season")["winner"].value_counts()

wins_per_season
Out[7]: season winner

2008 Rajasthan Royals 13 Kings XI Punjab 10 Chennai Super Kings Delhi Daredevils Mumbai Indians Kolkata Knight Riders Royal Challengers Bangalore Deccan Chargers Delhi Daredevils Deccan Chargers Royal Challengers Bangalore Chennai Super Kings Kings XI Punjab Rajasthan Royals Mumbai Indians Kolkata Knight Riders Mumbai Indians 11 Chennai Super Kings Deccan Chargers Royal Challengers Bangalore Delhi Daredevils Kolkata Knight Riders Rajasthan Royals Kings XI Punjab Chennai Super Kings 11 Mumbai Indians Royal Challengers Bangalore Kolkata Knight Riders Kings XI Punjab Deccan Chargers Royal Challengers Bangalore Kolkata Knight Riders Delhi Daredevils Mumbai Indians Rising Pune Supergiants Kings XI Punjab 12 Mumbai Indians Rising Pune Supergiant Kolkata Knight Riders Sunrisers Hyderabad Kings XI Punjab Delhi Daredevils Gujarat Lions Royal Challengers Bangalore Chennai Super Kings 10 Sunrisers Hyderabad Kolkata Knight Riders Rajasthan Royals Kings XI Punjab Mumbai Indians Royal Challengers Bangalore Delhi Daredevils Mumbai Indians Chennai Super Kings Delhi Capitals Kings XI Punjab Kolkata Knight Riders Sunrisers Hyderabad Rajasthan Royals

Royal Challengers Bangalore 5
Name: winner, Length: 100, dtype: int64

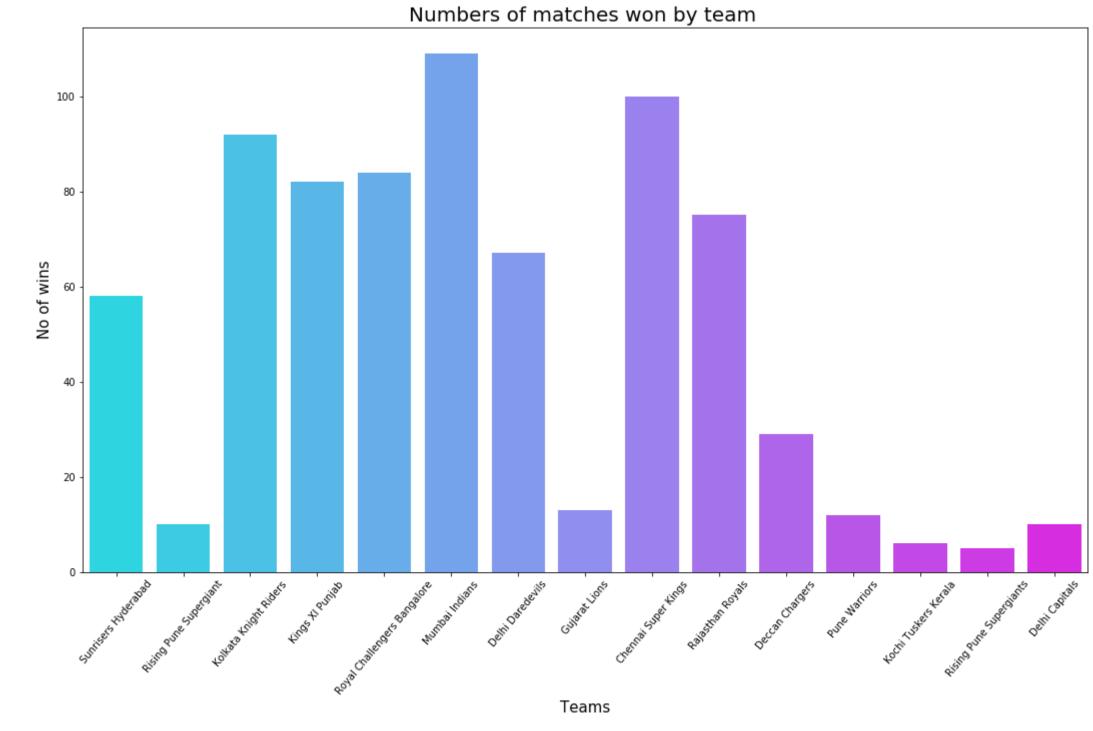
In [8]: plt.figure(figsize = (18,10))
sns.countplot('season',data=data,palette="winter")
plt.title("Number of Matches played in each IPL season",fontsize=20)

plt.xlabel("season",fontsize=15)
plt.ylabel('Matches',fontsize=15)

plt.show()

plt.show()

In [9]: plt.figure(figsize = (18,10))
 sns.countplot(x='winner',data=data, palette='cool')
 plt.title("Numbers of matches won by team ",fontsize=20)
 plt.xticks(rotation=50)
 plt.xlabel("Teams",fontsize=15)
 plt.ylabel("No of wins",fontsize=15)



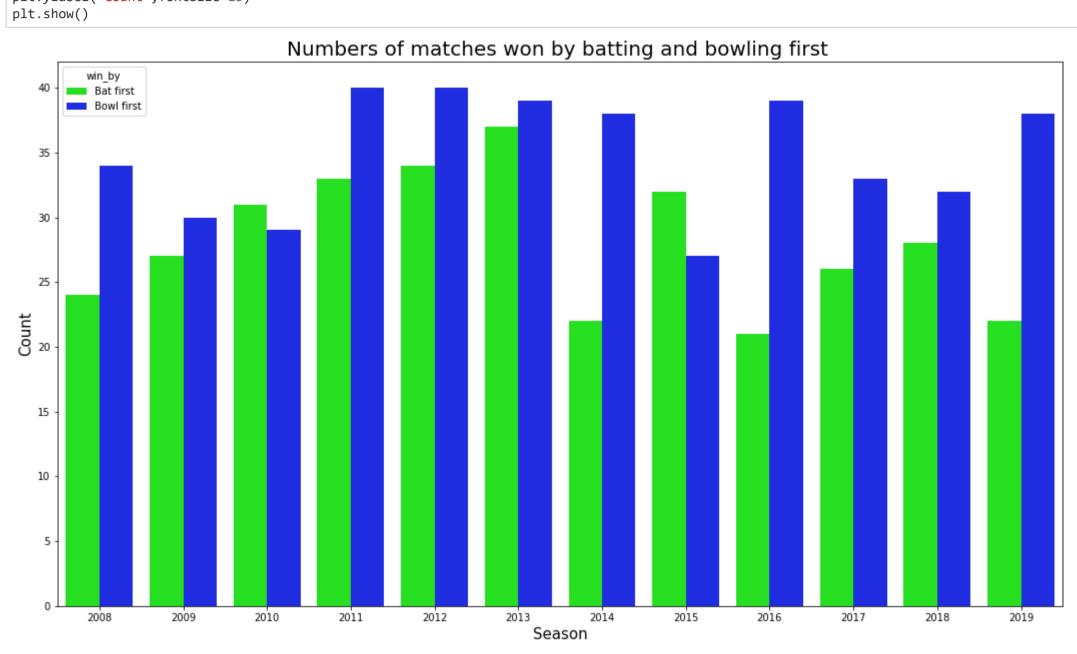
In [10]: data['win_by']=np.where(data['win_by_runs']>0,'Bat first','Bowl first')

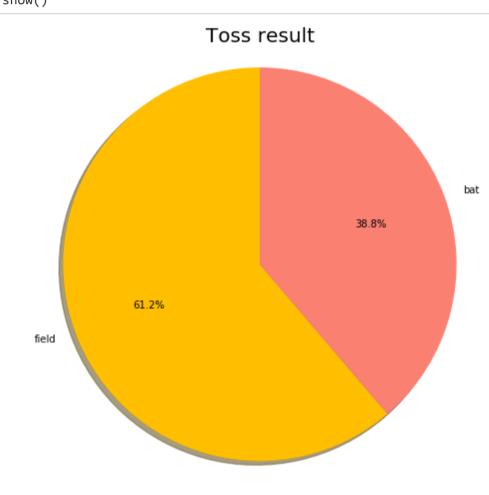
file:///C:/Users/hp/Desktop/EDA sports IPL.html

2/19/2021 EDA sports IPL

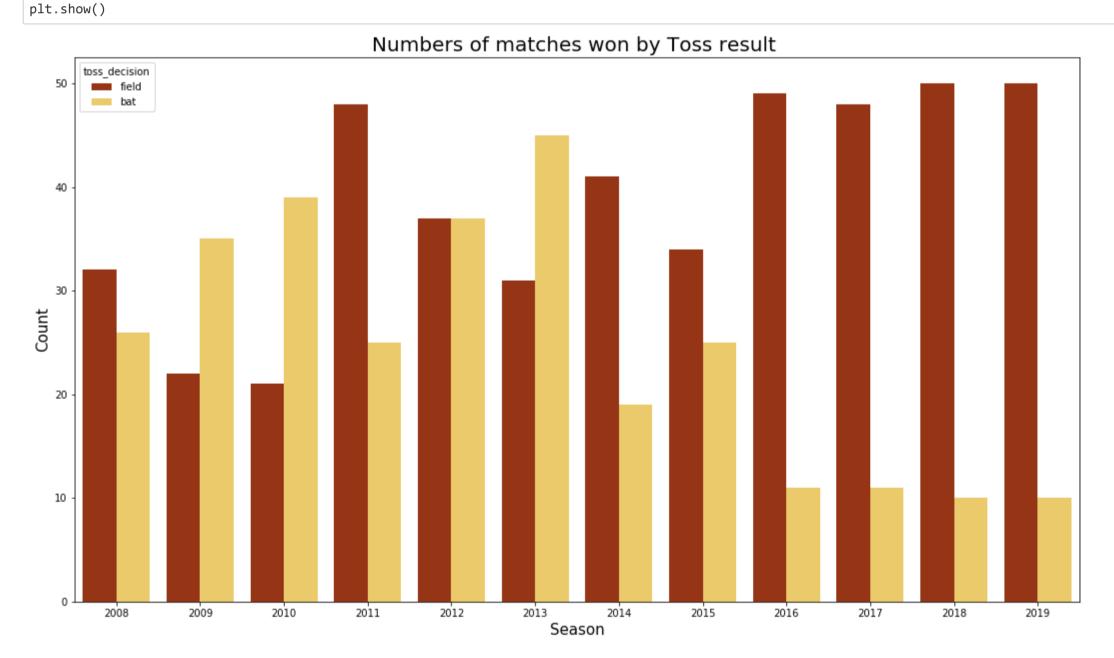
Bowl first Match Result 44.6% Bat first

In [12]: plt.figure(figsize = (18,10))
 sns.countplot('season',hue='win_by',data=data,palette='hsv')
 plt.title("Numbers of matches won by batting and bowling first ",fontsize=20)
 plt.xlabel("Season",fontsize=15)
 plt.ylabel("Count",fontsize=15)





In [14]: # we will plot graph on Numbers of matches won by Toss result
 plt.figure(figsize = (18,10))
 sns.countplot('season',hue='toss_decision',data=data,palette='afmhot')
 plt.title("Numbers of matches won by Toss result ",fontsize=20)
 plt.xlabel("Season",fontsize=15)
 plt.ylabel("Count",fontsize=15)



In [16]: # we will print winner season wise
final_matches=data.drop_duplicates(subset=['season'], keep='last')

final_matches[['season','winner']].reset_index(drop=True).sort_values('season')
Out[16]:

seasonwinner12008Rajasthan Royals22009Deccan Chargers32010Chennai Super Kings42011Chennai Super Kings52012Kolkata Knight Riders62013Mumbai Indians72014Kolkata Knight Riders82015Mumbai Indians92016Sunrisers Hyderabad02017Mumbai Indians102018Chennai Super Kings112019Mumbai Indians

plt.show()

Match Result

bat

66.7%

2/19/2021 EDA sports IPL In [18]: Toss=final_matches.toss_decision.value_counts() labels=np.array(Toss.index) sizes = Toss.values colors = ['#FFBF00', '#FA8072'] plt.figure(figsize = (10,8)) plt.pie(sizes, labels=labels, colors=colors, autopct='%1.1f%%', shadow=True,startangle=90) plt.title('Toss Result',fontsize=20) plt.axis('equal',fontsize=10) plt.show() Toss Result field 41.7% 58.3% In [19]: # we will print name of top player in IPL plt.figure(figsize = (18,10)) top_players = data.player_of_match.value_counts()[:10] fig, ax = plt.subplots() ax.set_ylim([0,20]) ax.set_ylabel("Count") ax.set_title("Top player of the match Winners") top_players.plot.bar() sns.barplot(x = top_players.index, y = top_players, orient='v', palette="hsv"); plt.show() <Figure size 1296x720 with 0 Axes> Top player of the match Winners 20.0 T 5 10.0 -In [20]: # We will print IPL Finals venues and winners along with the number of wins. final_matches.groupby(['city','winner']).size() Out[20]: city winner Kolkata Knight Riders 1 Bangalore Sunrisers Hyderabad Chennai Super Kings Kolkata Knight Riders Hyderabad Mumbai Indians Johannesburg Deccan Chargers Mumbai Indians Mumbai Chennai Super Kings Rajasthan Royals dtype: int64

In [21]: # we will print number of season won by teams
final_matches["winner"].value_counts()

```
Out[21]: Mumbai Indians 4
Chennai Super Kings 3
Kolkata Knight Riders 2
Deccan Chargers 1
Sunrisers Hyderabad 1
Rajasthan Royals 1
Name: winner, dtype: int64
```

In [22]: # we will print toss winner, toss decision, winner in final matches.
final_matches[['toss_winner','toss_decision','winner']].reset_index(drop=True)

```
Out[22]:
```

```
toss_winner toss_decision
                                                      winner
              Mumbai Indians
                                               Mumbai Indians
             Rajasthan Royals
                                     field Rajasthan Royals
2 Royal Challengers Bangalore
                                     field Deccan Chargers
         Chennai Super Kings
                                      bat Chennai Super Kings
          Chennai Super Kings
                                      bat Chennai Super Kings
          Chennai Super Kings
                                      bat Kolkata Knight Riders
              Mumbai Indians
                                      bat Mumbai Indians
         Kolkata Knight Riders
                                     field Kolkata Knight Riders
         Chennai Super Kings
                                     field Mumbai Indians
                                      bat Sunrisers Hyderabad
          Sunrisers Hyderabad
                                     field Chennai Super Kings
          Chennai Super Kings
              Mumbai Indians
                                      bat Mumbai Indians
```

In [23]: # we will print man of the match
final_matches[['winner','player_of_match']].reset_index(drop=True)

```
Out[23]:
                          winner player_of_match
                   Mumbai Indians
                                     KH Pandya
            1 Rajasthan Royals
                                      YK Pathan
            2 Deccan Chargers
                                      A Kumble
           3 Chennai Super Kings
                                       SK Raina
            4 Chennai Super Kings
           5 Kolkata Knight Riders
                                       MS Bisla
                  Mumbai Indians
                                      KA Pollard
                                     MK Pandey
            7 Kolkata Knight Riders
                                     RG Sharma
                   Mumbai Indians
                                     BCJ Cutting
           9 Sunrisers Hyderabad
```

10 Chennai Super Kings

Mumbai Indians

In [24]: len(final_matches[final_matches['toss_winner']==final_matches['winner']]['winner'])

SR Watson

JJ Bumrah

runs by fours fours

5776 1444

Out[24]: 9

In [25]: # we will print numbers of fours hit by team
four_data=complete_data[complete_data['batsman_runs']==4]

four_data=complete_data[complete_data[batsman_runs]==4]
four_data.groupby('batting_team')['batsman_runs'].agg([('runs by fours','sum'),('fours','count')])

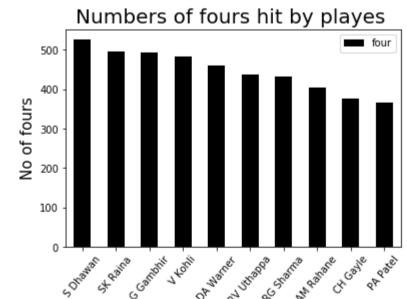
Out[25]:

```
batting_team
     Chennai Super Kings
                              8772 2193
                               3828 957
        Deccan Chargers
           Delhi Capitals
                               968 242
                               8632 2158
          Delhi Daredevils
           Gujarat Lions
                               1840 460
                               9832 2458
          Kings XI Punjab
     Kochi Tuskers Kerala
                               680 170
                               9736 2434
     Kolkata Knight Riders
          Mumbai Indians
                              10352 2588
                              2100 525
           Pune Warriors
        Rajasthan Royals
                              8140 2035
                               788 197
    Rising Pune Supergiant
                               684 171
   Rising Pune Supergiants
Royal Challengers Bangalore
                              9440 2360
```

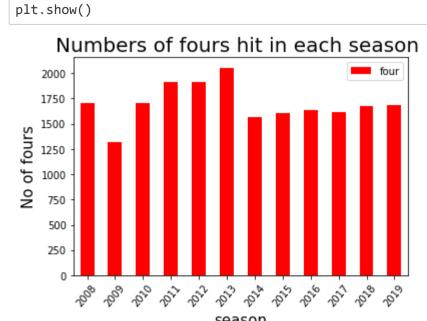
In [26]: # we will plot graph on four hit by players
 batsman_four=four_data.groupby('batsman')['batsman_runs'].agg([('four','count')]).reset_index().sort_values('four',ascending=0)
 ax=batsman_four.iloc[:10,:].plot('batsman','four',kind='bar',color='black')
 plt.title("Numbers of fours hit by playes ",fontsize=20)
 plt.xticks(rotation=50)
 plt.xlabel("Player name",fontsize=15)

plt.ylabel("No of fours",fontsize=15)
plt.show()

Sunrisers Hyderabad



In [27]: # we will plot graph on no of four hit in each season
 ax=four_data.groupby('season')['batsman_runs'].agg([('four','count')]).reset_index().plot('season','four',kind='bar',color = 'red')
 plt.title("Numbers of fours hit in each season ",fontsize=20)
 plt.xticks(rotation=50)
 plt.xlabel("season",fontsize=15)
 plt.ylabel("No of fours",fontsize=15)



file:///C:/Users/hp/Desktop/EDA sports IPL.html

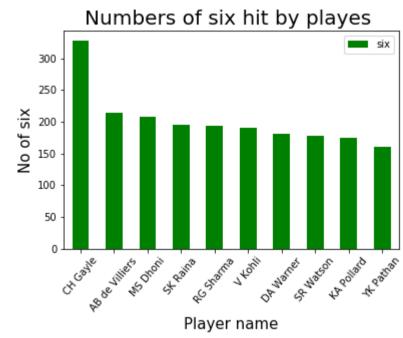
2/19/2021 EDA sports IPL

In [28]: # we will print no of sixes hit by team six_data=complete_data[complete_data['batsman_runs']==6] six_data.groupby('batting_team')['batsman_runs'].agg([('runs by six','sum'),('sixes','count')]) Out[28]: runs by six sixes batting_team Chennai Super Kings 5838 973 2400 400 **Deccan Chargers** 522 87 Delhi Capitals Delhi Daredevils 4806 801 **Gujarat Lions** 930 155

5856 976 Kings XI Punjab 318 53 Kochi Tuskers Kerala 5580 930 Kolkata Knight Riders Mumbai Indians 6576 1096 1176 196 **Pune Warriors** 4086 681 Rajasthan Royals 534 Rising Pune Supergiant Rising Pune Supergiants 408 68 6792 1132 Royal Challengers Bangalore Sunrisers Hyderabad 3198 533

In [29]: # we will plot graph of six hit by players
batsman_six=six_data.groupby('batsman')['batsman_runs'].agg([('six','count')]).reset_index().sort_values('six',ascending=0)
ax=batsman_six.iloc[:10,:].plot('batsman','six',kind='bar',color='green')
plt.title("Numbers of six hit by playes ",fontsize=20)

plt.xticks(rotation=50)
plt.xlabel("Player name",fontsize=15)
plt.ylabel("No of six",fontsize=15)
plt.show()



In [30]: # we will plot graph on no of six hit in each season
 ax=six_data.groupby('season')['batsman_runs'].agg([('six','count')]).reset_index().plot('season','six',kind='bar',color = 'blue')
 plt.title("Numbers of fours hit in each season ",fontsize=20)
 plt.xticks(rotation=50)

season

In [31]: # We will print the top 10 Leading run scorer in IPL
batsman_score=Data.groupby('batsman')['batsman_runs'].agg(['sum']).reset_index().sort_values('sum',ascending=False).reset_index(drop=True)
batsman_score=batsman_score.rename(columns={'sum':'batsman_runs'})
print("*** Top 10 Leading Run Scorer in IPL ***")

*** Top 10 Leading Run Scorer in IPL ***

plt.xlabel("season",fontsize=15)

plt.show()

plt.ylabel("No of fours",fontsize=15)

Out[31]: ______batsman batsman_runs

5434 V Kohli 5415 1 SK Raina 2 RG Sharma 4914 4741 3 DA Warner 4 S Dhawan 4632 5 CH Gayle 4560 6 MS Dhoni 4477 7 RV Uthappa 8 AB de Villiers 4428

batsman_score.iloc[:10,:]

In [33]: # we will print no of matches played by batsman
No_Matches_player= Data[["match_id","player_dismissed"]]

4223

No_Matches_player =No_Matches_player .groupby("player_dismissed")["match_id"].count().reset_index().sort_values(by="match_id",ascending=False).reset_index(drop=True)
No_Matches_player.columns=["batsman","No_of Matches"]
No_Matches_player .head(5)

Out[33]:

batsman No_of Matches

9 G Gambhir

 0
 SK Raina
 162

 1
 RG Sharma
 155

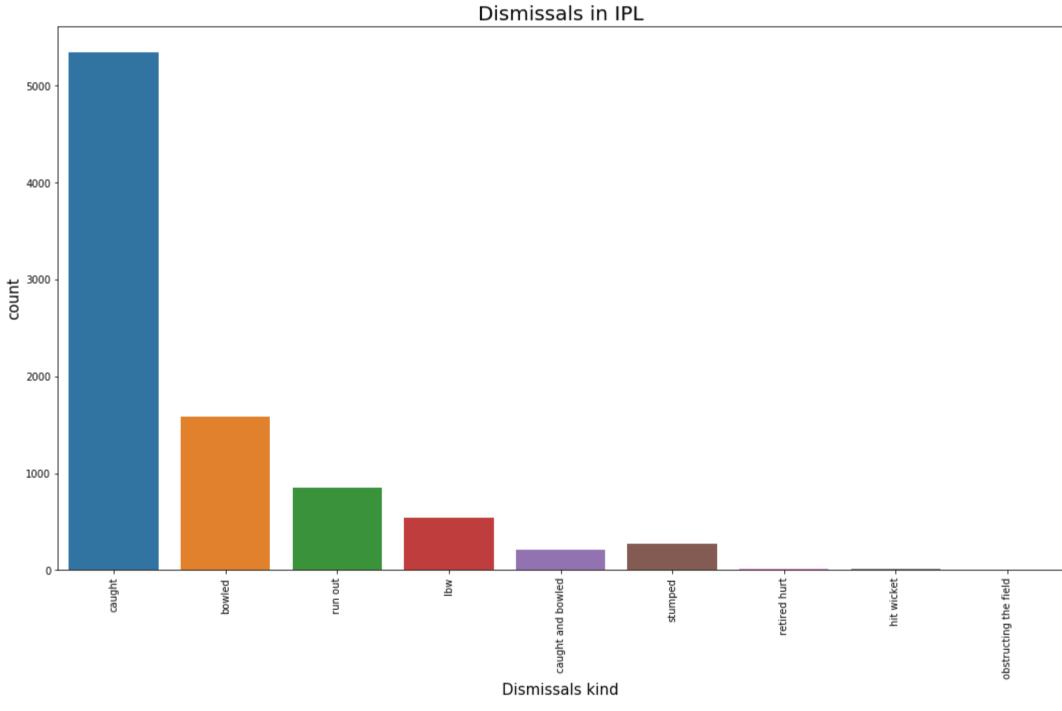
 2
 RV Uthappa
 153

 3
 V Kohli
 143

4 S Dhawan

In [34]: # Dismissals in IPL
plt.figure(figsize=(18,10))
ax=sns.countplot(Data.dismissal_kind)
plt.title("Dismissals in IPL",fontsize=20

ax=sns.countplot(Data.dismissal_kind)
plt.title("Dismissals in IPL",fontsize=20)
plt.xlabel("Dismissals kind",fontsize=15)
plt.ylabel("count",fontsize=15)
plt.xticks(rotation=90)
plt.show()



In [35]: wicket_data=Data.dropna(subset=['dismissal_kind'])
wicket_data=wicket_data[~wicket_data['dismissal_kind'].isin(['run out','retired hurt','obstructing the field'])]

In [36]: # we will print ipl most wicket taking bowlers
wicket_data.groupby('bowler')['dismissal_kind'].agg(['count']).reset_index().sort_values('count',ascending=False).reset_index(drop=True).iloc[:10,:]

 bowler count

 0
 SL Malinga
 170

 1
 A Mishra
 156

 2
 Harbhajan Singh
 150

 3
 PP Chawla
 149

 4
 DJ Bravo
 147

 5
 B Kumar
 133

 6
 R Ashwin
 125

 7
 SP Narine
 122

 8
 UT Yadav
 119

 9
 RA Jadeja
 108

Conclusion:

The highest number of match played in IPL season was 2013,2014,2015.

The highest number of match won by Mumbai Indians i.e 4 match out of 12 matches.

Teams which Bowl first has higher chances of winning then the team which bat first.

After winning toss more teams decide to do fielding first.

In finals teams which decide to do fielding first win the matches more then the team which bat first.

In finals most teams after winning toss decide to do fielding first.

Top player of match winning are CH gayle, AB de villers.

It is interesting that out of 12 IPL finals,9 times the team that won the toss was also the winner of IPL.

The highest number of four hit by player is Shikar Dhawan.

The highest number of six hit by player is CH gayle.

Top leading run scorer in IPL are Virat kholi, SK Raina, RG Sharma.

The highest number of matches played by player name are SK Raina.

The highest number of matches played by player name are SK Raina, RG Sharma.

Dismissals in IPL was most by Catch out

Dismissals in IPL was most by Catch out .

The IPL most wicket taken blower is SL Malinga.

Thank You!