

# Tarvi Bhosle

## Data Science And Business Analytics Intern @ TSF GRP JULY2021.

Task 5: Exploratory Data Analysis- Sports (Indian Premier League)

Dataset: <https://bit.ly/34SRn3b>

```
In [ ]:

```

### Import Libraries

```
In [2]:
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

```

### Import Dataset

```
In [3]:
data1=pd.read_csv("C:\Users\ADMIN\Desktop\matches.csv")
data2=pd.read_csv("C:\Users\ADMIN\Desktop\deliveries.csv")

```

																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					</
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

```
In [4]:
data2=pd.read_csv("C:\Users\ADMIN\Desktop\deliveries.csv")
data2.head(5)

```

match_id	inning	battling_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs	legbye_runs	noball_runs	penalty_runs	batsman_runs	extra_runs	total_runs
0	1	1	1	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0	0	0	0
1	1	1	1	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0	0	0	0
2	1	1	1	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0	0	4	4
3	1	1	1	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0	0	0	0
4	1	1	1	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0	0	2	2

5 rows x 21 columns

```
In [5]:
data1.shape

```

```
Out[5]:
(756, 18)
```

```
In [6]:
data2.shape

```

```
Out[6]:
(179678, 21)
```

```
In [7]:
data1.info

```

```
Out[7]:
<bound method DataFrame.info of
0      1      2017      Hyderabad      4/5/2017      Sunrisers Hyderabad      field      normal      0      Sunrisers Hyderabad      35      0      0      0      0      0      0      0
1      2      2017      Pune      4/5/2017      Mumbai Indians      Rising Pune Supergiant      field      normal      0      Rising Pune Supergiant      0      7      0      0      0      0      0      0
2      3      2017      Rajkot      4/7/2017      Gujarat Lions      Kolkata Knight Riders      field      normal      0      Kolkata Knight Riders      0      10      0      0      0      0      0      0
3      4      2017      Indore      4/8/2017      Rising Pune Supergiant      Kings XI Punjab      field      normal      0      Kings XI Punjab      0      6      0      0      0      0      0      0
4      5      2017      Bangalore      4/8/2017      Royal Challengers Bangalore      Delhi Daredevils      bat      normal      0      Challengers Bangalore      15      0      0      0      0      0      2      2
...
751      11347      2019      Mumbai      5/5/2019      Kolkata Knight Riders      field
752      11312      2019      Chennai      7/5/2019      Chennai Super Kings      bat
753      11413      2019      Visakhapatnam      8/5/2019      Sunrisers Hyderabad      field
754      11414      2019      Visakhapatnam      10/5/2019      Delhi Capitals      field
755      11415      2019      Hyderabad      12/5/2019      Mumbai Indians      bat
...
team1      0      Royal Challengers Bangalore      Royal Challengers Bangalore      field
1      1      Rising Pune Supergiant      Rising Pune Supergiant      field
2      2      Kolkata Knight Riders      Kolkata Knight Riders      field
3      3      Kings XI Punjab      Kings XI Punjab      field
4      4      Delhi Daredevils      Royal Challengers Bangalore      bat
...
751      Mumbai Indians      Mumbai Indians      field
752      Mumbai Indians      Chennai Super Kings      bat
753      Delhi Capitals      Delhi Capitals      field
754      Chennai Super Kings      Chennai Super Kings      field
755      Chennai Super Kings      Mumbai Indians      bat
...
result      dl_applied      winner      win_by_runs      \
0      normal      0      Sunrisers Hyderabad      35
1      normal      0      Rising Pune Supergiant      0
2      normal      0      Kolkata Knight Riders      0
3      normal      0      Kings XI Punjab      0
4      normal      0      Royal Challengers Bangalore      15
...
751      normal      0      Mumbai Indians      0
752      normal      0      Mumbai Indians      0
753      normal      0      Delhi Capitals      0
754      normal      0      Chennai Super Kings      0
755      normal      0      Mumbai Indians      1
...
win_by_wickets      player_of_match      \
0      0      Yuvraj Singh
1      7      SFD Smith
2      10      CA Lynn
3      6      GJ Maxwell
4      0      KM Jadhav
...
751      0      HH Pandya
752      6      AS Yadav
753      2      RR Pant
754      6      F du Plessis
755      0      JJ Bumrah
...
venue      umpire1      \
0      Rajiv Gandhi International Stadium, Uppal      AY Dandekar
1      Maharashtra Cricket Association Stadium      A Nand Kishore
2      Saurashtra Cricket Association Stadium      Nitin Menon
3      Holkar Cricket Stadium      AK Chaudhary
4      M Chinnaswamy Stadium      NaN
...
751      Wankhede Stadium      Manda Kishore
752      M. A. Chidambaram Stadium      Nigel Llong
753      ACA-VDCA Stadium      RA Jadeja
754      ACA-VDCA Stadium      Sundaran Ravi
755      Rajiv Gandhi Intl. Cricket Stadium      Nitin Menon
...
umpire2      umpire3      \
0      NJ Llong      NaN
1      S Ravi      NaN
2      CK Nandan      NaN
3      C Shashidhin      NaN
4      NaN      NaN
...
751      0      Nandan      S Ravi
752      Nitin Menon      Ian Gould
753      NaN      NaN
754      Bruce Oxtford      Chettithody Shashidhin
755      Ian Gould      Nigel Llong
...
[756 rows x 18 columns]>

```

```
In [8]:
data2.info

```

```
Out[8]:
<bound method DataFrame.info of
0      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1
1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1
2      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1
3      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1
4      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1      1
...
179073      11415      2      Chennai Super Kings      Mumbai Indians
179074      11415      2      Chennai Super Kings      Mumbai Indians
179075      11415      2      Chennai Super Kings      Mumbai Indians
179076      11415      2      Chennai Super Kings      Mumbai Indians
179077      11415      2      Chennai Super Kings      Mumbai Indians
...
over      ball      batsman      non_striker      bowler      is_super_over      ...      \
0      1      1      DA Warner      S Dhawan      TS Mills      0      ...
1      1      2      DA Warner      S Dhawan      TS Mills      0      ...
2      1      3      DA Warner      S Dhawan      TS Mills      0      ...
3      1      4      DA Warner      S Dhawan      TS Mills      0      ...
4      1      5      DA Warner      S Dhawan      TS Mills      0      ...
...
179073      20      2      RA Jadeja      SR Watson      SL Malinga      0      ...
179074      20      2      SR Watson      RA Jadeja      SL Malinga      0      ...
179075      20      4      SR Watson      RA Jadeja      SL Malinga      0      ...
179076      20      5      SN Thakur      RA Jadeja      SL Malinga      0      ...
179077      20      6      SN Thakur      RA Jadeja      SL Malinga      0      ...
...
bye_runs      legbye_runs      noball_runs      penalty_runs      batsman_runs      \
0      0      0      0      0      0
1      0      0      0      0      0
2      0      0      0      0      4
3      0      0      0      0      0
4      0      0      0      0      0
...
179073      0      0      0      0      0
179074      0      0      0      0      2
179075      0      0      0      0      2
179076      0      0      0      0      0
179077      0      0      0      0      0
...
extra_runs      total_runs      player_dismissed      dismissal_kind      fielder      \
0      0      0      NaN      NaN      NaN
1      0      0      NaN      NaN      NaN
2      0      4      NaN      NaN      NaN
3      0      2      NaN      NaN      NaN
4      2      2      NaN      NaN      NaN
...
179073      0      0      NaN      NaN      NaN
179074      0      2      NaN      NaN      NaN
179075      0      1      SR Watson      run      out      KH Pandya
179076      0      2      NaN      NaN      NaN
179077      0      0      SN Thakur      lbw      NaN
...
[179678 rows x 21 columns]>

```

```
In [9]:
data1.describe()

```

```
Out[9]:
id      season      dl_applied      win_by_runs      win_by_wickets
count      756.000000      756.000000      756.000000      756.000000      756.000000
mean      179071.789517      2013.444444      0.006182      13.293069      3.360260
std      34664.478168      136689.95      0.156360      23.471144      3.367963
min      190.000000      2008.000000      0.000000      0.000000      0.000000
25%      189.750000      2011.000000      0.000000      0.000000      0.000000
50%      378.500000      2013.000000      0.000000      0.000000      0.000000
75%      567.250000      2016.000000      0.000000      19.000000      6.000000
max      11415.000000      2019.000000      1.000000      146.000000      10.000000

```

```
In [10]:
data2.describe()

```

```
Out[10]:
match_id      inning      over      ball      is_super_over      wide_runs      bye_runs      legbye_runs      noball_runs      penalty_runs      batsman_runs      extra_runs      tot
count      179078.000000      179078.000000      179078.000000      179078.000000      179078.000000      179078.000000      179078.000000      179078.000000      179078.000000      179078.000000      179078.000000      179078.000000
mean      1802.262957      1.482952      10.162488      3.615887      0.000452      0.036721      0.004936      0.021136      0.004183      0.000056      1.246864      0.067032      1
std      3472.322605      0.502074      5.677684      3.615887      0.021263      0.251161      0.116480      0.104908      0.070492      0.016709      1.668270      0.342553      1
min      190.000000      1.000000      1.000000      1.000000      0.000000      0.000000      0.000000      0.000000      0.000000      0.000000      0.000000      0.000000      0
25%      190.000000      1.000000      5.000000      2.000000      0.000000      0.000000      0.000000      0.000000      0.000000      0.000000      0.000000      0.000000      0
50%      379.000000      1.000000      10.000000      4.000000      0.000000      0.000000      0.000000      0.000000      0.000000      0.000000      1.000000      0.000000      1
75%      567.000000      2.000000      15.000000      5.000000      0.000000      0.000000      0.000000      0.000000      0.000000      0.000000      1.000000      0.000000      1
max      11415.000000      5.000000      20.000000      9.000000      1.000000      5.000000      4.000000      5.000000      5.000000      7.000000      7.000000      10.000000      10

```

### Finding Missing Values

```
In [11]:
data1.isnull().sum()

```

```
Out[11]:
id      0
season      0
city      0
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
umpire1      2
umpire2      2
umpire3      637
dtype: int64

```

```
In [12]:
data1=data1.drop(columns=["umpire3"],axis=1)

```

```
In [13]:
data2.isnull().sum()

```

```
Out[13]:
match_id      0
inning      0
battling_team      0
bowling_team      0
over      0
ball      0
batsman      0
non_striker      0
bowler      0
is_super_over      0
bye_runs      0
legbye_runs      0
noball_runs      0
penalty_runs      0
batsman_runs      0
extra_runs      0
total_runs      0
player_dismissed      0
dismissal_kind      0
fielder      0
dtype: int64

```

### Total Matches Played

```
In [14]:
data1["id"].count()

```

```
Out[14]:
756

```

### Total IPL Seasons

```
In [15]:
data1["season"].unique()

```

```
Out[15]:
array([2017, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2018, 2019], dtype=int64)
```

### IPL team won by scoring the maximum runs

```
In [16]:
data1.iloc[data1["win_by_runs"].idxmax()]

```

```
Out[16]:
id      season      city      date      team1      team2      toss_winner      toss_decision      result      dl_applied      winner      win_by_runs      win_by_wickets      player_of_match      venue      umpire1      umpire2      Name: 43, dtype: object

```

### IPL team won by taking the maximum wickets

```
In [17]:
data1.iloc[data1["win_by_wickets"].idxmax()]

```

```
Out[17]:
id      season      city      date      team1      team2      toss_winner      toss_decision      result      dl_applied      winner      win_by_runs      win_by_wickets      player_of_match      venue      umpire1      umpire2      Name: 2, dtype: object

```

### IPL team won by taking the minimum wickets

```
In [18]:
data1.iloc[data1["win_by_wickets"].idxmin()]

```

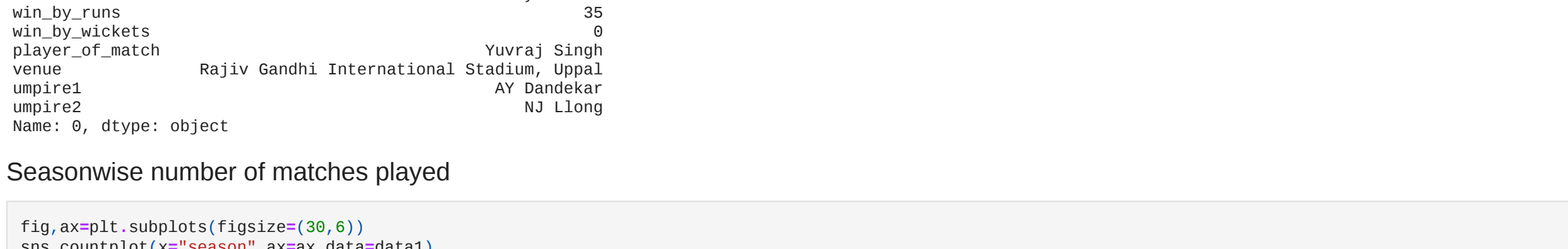
```
Out[18]:
id      season      city      date      team1      team2      toss_winner      toss_decision      result      dl_applied      winner      win_by_runs      win_by_wickets      player_of_match      venue      umpire1      umpire2      Name: 0, dtype: object

```

### Seasonwise number of matches played

```
In [19]:
fig=plt.figure(figsize=(25,6))
sns.countplot(x="season",ax=axis,data=data1)
plt.show()

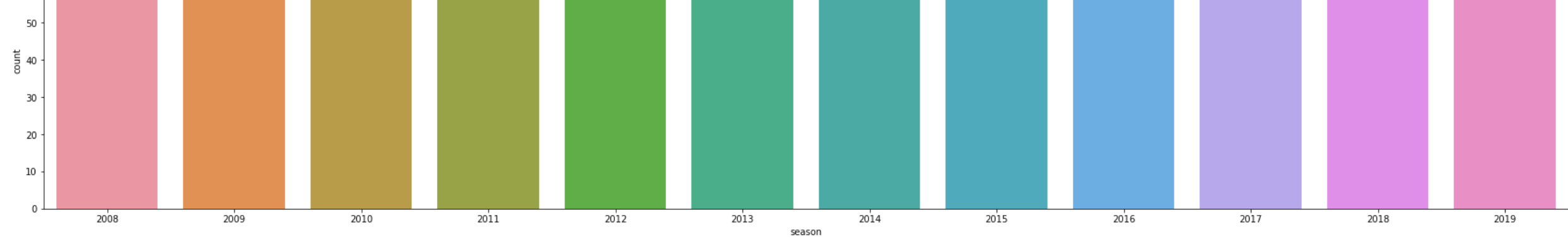
```



### Seasonwise number of teams played

```
In [20]:
data1.groupby("season")["team1"].nunique().plot(kind="bar",figsize=(12,5))

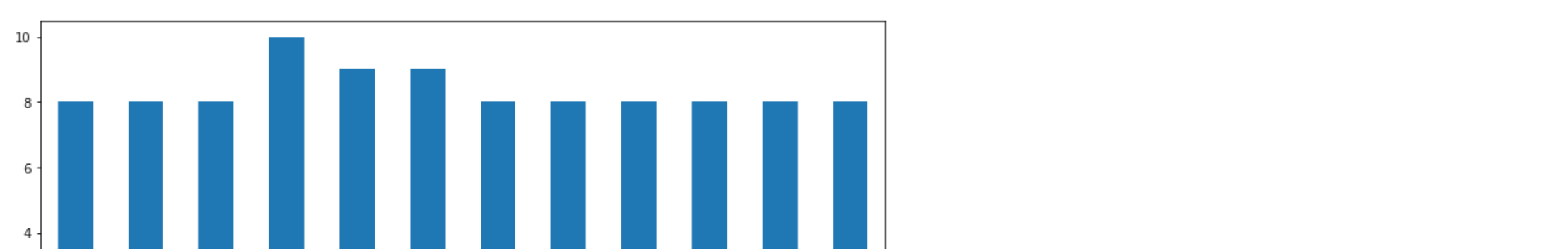
```



### Successful IPL team

```
In [21]:
sns.barplot(y=data1.winner.value_counts(),index=x=data1.winner.value_counts())

```



### Man of the match

```
In [22]:
data1["player_of_match"].value_counts()

```

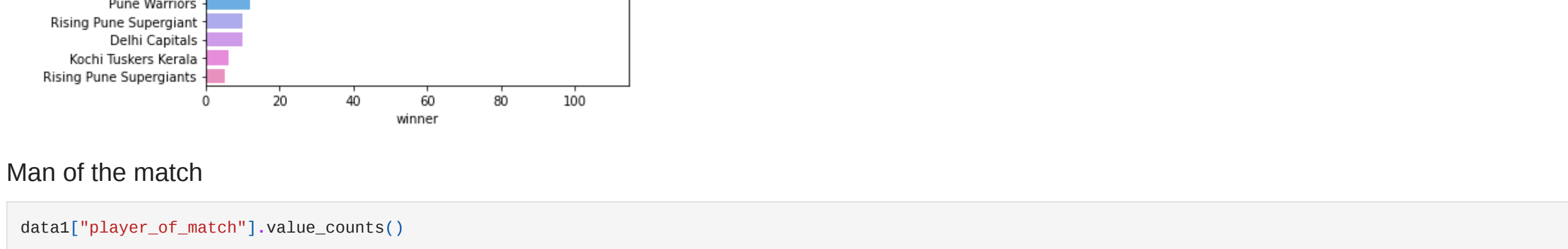
```
Out[22]:
CH Gayle      21
AB de Villiers      20
RG Sharma      17
MS Dhoni      17
DA Warner      17
PV Collingwood      1
NV Ojha      1
AC Voges      1
J Theron      1
S Hetmyer      1
Name: player_of_match, Length: 226, dtype: int64

```

### Venue for maximum matches

```
In [40]:
plt.figure(figsize=(25,15))
sns.barplot(y=data1.venue.value_counts(),index=x=data1.venue.value_counts())

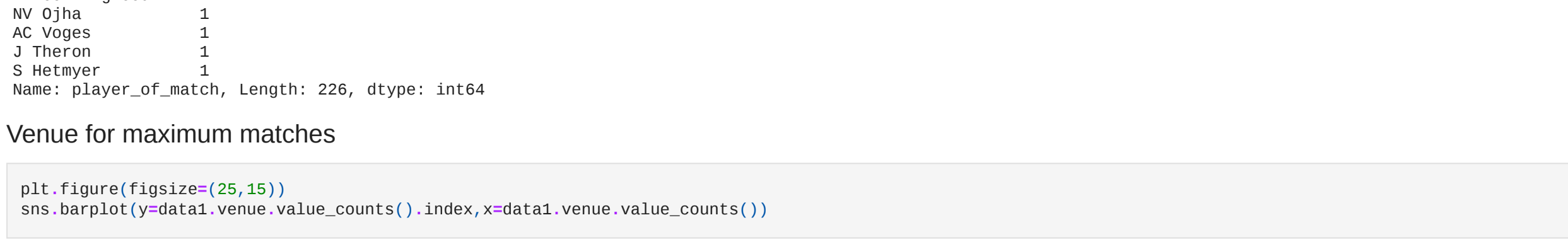
```



### Umpire for maximum matches

```
In [39]:
umpire=pd.concat([data1.umpire1,data1.umpire2])
plt.figure(figsize=(25,15))
sns.barplot(y=umpire.value_counts(),index=x=umpire.value_counts())

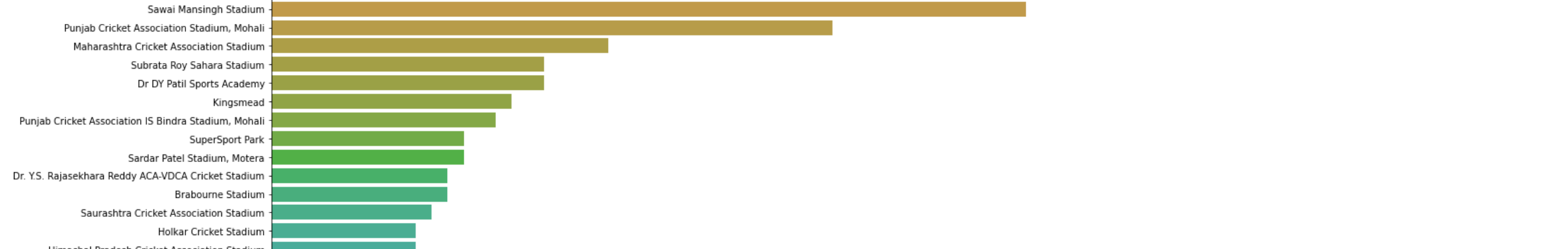
```



### Performance of the bowlers

```
In [25]:
data2.groupby("bowler")["player_dismissed"].count().sort_values(ascending=False).head(10).plot(kind="bar",figsize=(15,5))

```



### Top Run Getters

```
In [43]:
data2.groupby("batsman")["batsman_runs"].sum().sort_values(ascending=False).head(10).plot(kind="bar",figsize=(15,5))

```



### Decision after winning toss

```
In [38]:
data1.toss_decision.value_counts().len(data1)*100

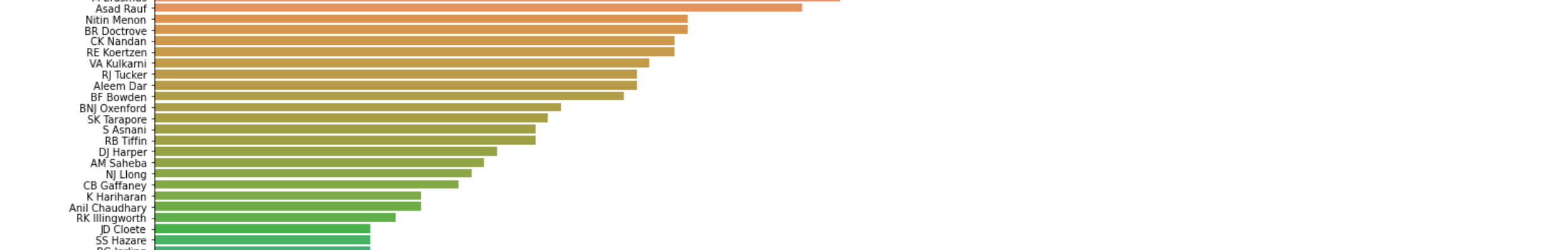
```

```
Out[38]:
field      61.243386
bat      38.756614
Name: toss_decision, dtype: float64

```

```
In [36]:
data1["toss_decision"].value_counts().plot(kind="pie",fontsize=14)

```



### Teams bat first or second after winning toss

```
In [42]:
plt.figure(figsize=(25,10))
sns.countplot(y="toss_winner",data=data1,hue="toss_decision")

```



### Conclusion

- 756 matches were played according to the dataset.
- From 2008 to 2019 IPL seasons are using to analyse.
- Mumbai Indians won by scoring the maximum runs ie.146.
- Kolkata Knight Riders won by taking maximum wickets ie.10.
- Sunrisers Hyderabad won by making minimum wickets ie.0.
- In season 2013 consisted the highest number of matches ever played followed by 2012 and 2011.
- Mumbai Indians is the most successful IPL team followed by Chennai Super Kings and Kolkata Knight Riders.
- CH 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 61% times teams who have won tosses have decided to chase down.

```
In [ ]:

```