```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import seaborn as sns
import math

df = pd.read_csv(r"C:\Users\DELL\Desktop\IPL_Dataset.csv")

df
```

```
        match_id   season start_date                                venue  \
0         335982     2008   18-04-08                        M.Chinnaswamy
Stadium
1         335982     2008   18-04-08                        M.Chinnaswamy
Stadium
2         335982     2008   18-04-08                        M.Chinnaswamy
Stadium
3         335982     2008   18-04-08                        M.Chinnaswamy
Stadium
4         335982     2008   18-04-08                        M.Chinnaswamy
Stadium
...          ...      ...        ...
...
194349   1254060     2021   11-04-21   MA Chidambaram Stadium, Chepauk,
Chennai
194350   1254060     2021   11-04-21   MA Chidambaram Stadium, Chepauk,
Chennai
194351   1254060     2021   11-04-21   MA Chidambaram Stadium, Chepauk,
Chennai
194352   1254060     2021   11-04-21   MA Chidambaram Stadium, Chepauk,
Chennai
194353   1254060     2021   11-04-21   MA Chidambaram Stadium, Chepauk,
Chennai

        innings  ball           batting_team
bowling_team  \
0             1   0.1  Kolkata Knight Riders   Royal Challengers
Bangalore
1             1   0.2  Kolkata Knight Riders   Royal Challengers
Bangalore
2             1   0.3  Kolkata Knight Riders   Royal Challengers
Bangalore
3             1   0.4  Kolkata Knight Riders   Royal Challengers
Bangalore
4             1   0.5  Kolkata Knight Riders   Royal Challengers
Bangalore
...          ...   ...                            ...
...
```

```
194349          2   19.2     Sunrisers Hyderabad        Kolkata Knight
Riders
194350          2   19.3     Sunrisers Hyderabad        Kolkata Knight
Riders
194351          2   19.4     Sunrisers Hyderabad        Kolkata Knight
Riders
194352          2   19.5     Sunrisers Hyderabad        Kolkata Knight
Riders
194353          2   19.6     Sunrisers Hyderabad        Kolkata Knight
Riders

             striker   non_striker  ...  runs_off_bat   extras   wides
noballs  \
0          SC Ganguly   BB McCullum  ...             0        1       0
0
1         BB McCullum    SC Ganguly  ...             0        0       0
0
2         BB McCullum    SC Ganguly  ...             0        1       1
0
3         BB McCullum    SC Ganguly  ...             0        0       0
0
4         BB McCullum    SC Ganguly  ...             0        0       0
0
...               ...           ...  ...           ...      ...     ...
...
194349   Abdul Samad     MK Pandey   ...             0        0       0
0
194350   Abdul Samad     MK Pandey   ...             1        0       0
0
194351     MK Pandey   Abdul Samad   ...             1        0       0
0
194352   Abdul Samad     MK Pandey   ...             1        0       0
0
194353     MK Pandey   Abdul Samad   ...             6        0       0
0

        byes   legbyes  wicket_type player_dismissed run   over
0          0         1                                  1      0
1          0         0                                  0      0
2          0         0                                  1      0
3          0         0                                  0      0
4          0         0                                  0      0
...      ...       ...           ...                 ..    ...
194349     0         0                                  0     19
194350     0         0                                  1     19
194351     0         0                                  1     19
194352     0         0                                  1     19
194353     0         0                                  6     19

[194354 rows x 21 columns]
```

```
df. info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 194354 entries, 0 to 194353
Data columns (total 21 columns):
 #   Column            Non-Null Count    Dtype
---  ------            --------------    -----
 0   match_id          194354 non-null   int64
 1   season            194354 non-null   int64
 2   start_date        194354 non-null   object
 3   venue             194354 non-null   object
 4   innings           194354 non-null   int64
 5   ball              194354 non-null   float64
 6   batting_team      194354 non-null   object
 7   bowling_team      194354 non-null   object
 8   striker           194354 non-null   object
 9   non_striker       194354 non-null   object
 10  bowler            194354 non-null   object
 11  runs_off_bat      194354 non-null   int64
 12  extras            194354 non-null   int64
 13  wides             194354 non-null   int64
 14  noballs           194354 non-null   int64
 15  byes              194354 non-null   int64
 16  legbyes           194354 non-null   int64
 17  wicket_type       194354 non-null   object
 18  player_dismissed  194354 non-null   object
 19  run               194354 non-null   int64
 20  over              194354 non-null   int64
dtypes: float64(1), int64(11), object(9)
memory usage: 31.1+ MB

df['season'].unique()

array([2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2019,
       2018, 2020, 2021], dtype=int64)

df.iloc[0]              # isse hm ye analyise kiye ki pehla match kab huaa

match_id                          335982
season                              2008
start_date                      18-04-08
venue              M.Chinnaswamy Stadium
innings                                1
ball                                 0.1
batting_team         Kolkata Knight Riders
bowling_team     Royal Challengers Bangalore
striker                         SC Ganguly
non_striker                     BB McCullum
```

```
bowler                                        P Kumar
runs_off_bat                                        0
extras                                              1
wides                                               0
noballs                                             0
byes                                                0
legbyes                                             1
wicket_type
player_dismissed
run                                                 1
over                                                0
Name: 0, dtype: object
```
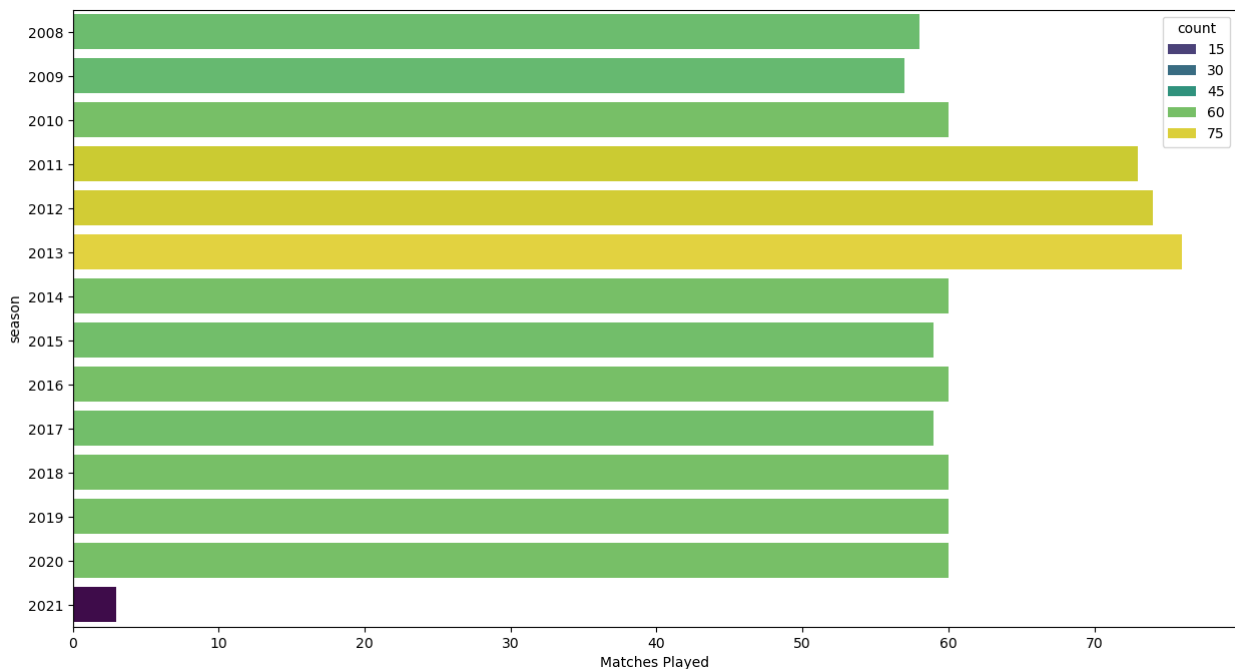
```python
plt.figure(figsize=(15,8))
data =
df.groupby(['match_id','season']).count().index.droplevel(level=0).val
ue_counts().sort_index()

sns.barplot(y=data.index,x=data,orient='h',hue=data,
palette="viridis")
plt.xlabel('Matches Played')
plt.ylabel('season')
plt.show()
```



```python
#How much match played in a venue

df.groupby(['venue','match_id']).count().droplevel(level=1).index.valu
e_counts()
```

```
venue
M.Chinnaswamy Stadium                                          80
Eden Gardens                                                   77
Arun Jaitley Stadium                                           74
Wankhede Stadium, Mumbai                                       74
Rajiv Gandhi International Stadium, Uppal                      64
MA Chidambaram Stadium, Chepauk, Chennai                      59
Punjab Cricket Association IS Bindra Stadium, Mohali          56
Sawai Mansingh Stadium                                        47
Dubai International Cricket Stadium                            33
Sheikh Zayed Stadium                                          29
Maharashtra Cricket Association Stadium                       21
Sharjah Cricket Stadium                                       18
Dr DY Patil Sports Academy                                    17
Subrata Roy Sahara Stadium                                    17
Kingsmead                                                     15
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium          13
SuperSport Park                                               12
Sardar Patel Stadium, Motera                                  12
Brabourne Stadium                                             11
Saurashtra Cricket Association Stadium                        10
Holkar Cricket Stadium                                         9
Himachal Pradesh Cricket Association Stadium                   9
New Wanderers Stadium                                          8
JSCA International Stadium Complex                             7
Barabati Stadium                                               7
St George's Park                                               7
Newlands                                                       7
Shaheed Veer Narayan Singh International Stadium              6
Nehru Stadium                                                  5
Green Park                                                     4
Buffalo Park                                                   3
Vidarbha Cricket Association Stadium, Jamtha                   3
De Beers Diamond Oval                                          3
OUTsurance Oval                                                2
Name: count, dtype: int64
```

```python
# which matches was played where?

df[['match_id', 'season', 'venue',
'start_date']].sort_values(by=['season', 'start_date'])
```

```
        match_id   season                                          venue
start_date
4022        336000   2008                       Sawai Mansingh Stadium
01-05-08
4023        336000   2008                       Sawai Mansingh Stadium
01-05-08
4024        336000   2008                       Sawai Mansingh Stadium
01-05-08
```

```
4025      336000    2008                    Sawai Mansingh Stadium
01-05-08
4026      336000    2008                    Sawai Mansingh Stadium
01-05-08
...         ...       ...                                        ...
...
194349   1254060    2021  MA Chidambaram Stadium, Chepauk, Chennai
11-04-21
194350   1254060    2021  MA Chidambaram Stadium, Chepauk, Chennai
11-04-21
194351   1254060    2021  MA Chidambaram Stadium, Chepauk, Chennai
11-04-21
194352   1254060    2021  MA Chidambaram Stadium, Chepauk, Chennai
11-04-21
194353   1254060    2021  MA Chidambaram Stadium, Chepauk, Chennai
11-04-21

[194354 rows x 4 columns]
```

where played IPL Match of 2010 and when?

```
df_2010 = df[df['season'] == 2010][['match_id', 'venue',
'start_date']].sort_values(by='start_date')
df_2010

        match_id                                     venue start_date
34343     419135                              Eden Gardens   01-04-10
34302     419135                              Eden Gardens   01-04-10
34303     419135                              Eden Gardens   01-04-10
34304     419135                              Eden Gardens   01-04-10
34305     419135                              Eden Gardens   01-04-10
...          ...                                       ...        ...
33822     419133  MA Chidambaram Stadium, Chepauk, Chennai   31-03-10
33821     419133  MA Chidambaram Stadium, Chepauk, Chennai   31-03-10
33820     419133  MA Chidambaram Stadium, Chepauk, Chennai   31-03-10
33832     419133  MA Chidambaram Stadium, Chepauk, Chennai   31-03-10
33719     419133  MA Chidambaram Stadium, Chepauk, Chennai   31-03-10

[14498 rows x 3 columns]
```

Most Run Scored by IPL Teams

```
df.groupby(['start_date','batting_team']).sum()
['run'].droplevel(level=0).sort_values(ascending=False)[:10]

batting_team
Royal Challengers Bangalore    263
Royal Challengers Bangalore    248
Chennai Super Kings            246
Kolkata Knight Riders          245
Chennai Super Kings            240
```

```
Royal Challengers Bangalore    235
Kings XI Punjab                232
Kolkata Knight Riders          232
Delhi Daredevils               231
Sunrisers Hyderabad            231
Name: run, dtype: int64
```

```python
# Most IPL Runs by a batsman

df.groupby('striker')
['runs_off_bat'].sum().sort_values(ascending=False).head(10)
```

```
striker
V Kohli            5921
SK Raina           5430
S Dhawan           5282
DA Warner          5257
RG Sharma          5251
AB de Villiers     4916
CH Gayle           4804
MS Dhoni           4632
RV Uthappa         4609
G Gambhir          4217
Name: runs_off_bat, dtype: int64
```

```python
# Most IPL Runs by a batsman in season by

data = df.groupby(['striker','season'])
['runs_off_bat'].sum().sort_values(ascending=False)[:10]
data
```

```
striker        season
V Kohli        2016     973
DA Warner      2016     848
KS Williamson  2018     735
MEK Hussey     2013     733
CH Gayle       2012     733
               2013     720
DA Warner      2019     692
AB de Villiers 2016     687
RR Pant        2018     684
KL Rahul       2020     676
Name: runs_off_bat, dtype: int64
```

```python
# Most IPL Runs by a batsman

plt.figure(figsize=(15,8))
data = df.groupby(['striker'])
['runs_off_bat'].sum().sort_values(ascending=False)[:10]
sns.barplot(y=data.index,x=data,orient='h',hue=data,palette="viridis")
plt.xlabel('Batsman')
```

```
plt.ylabel('Runs')
plt.show()
```
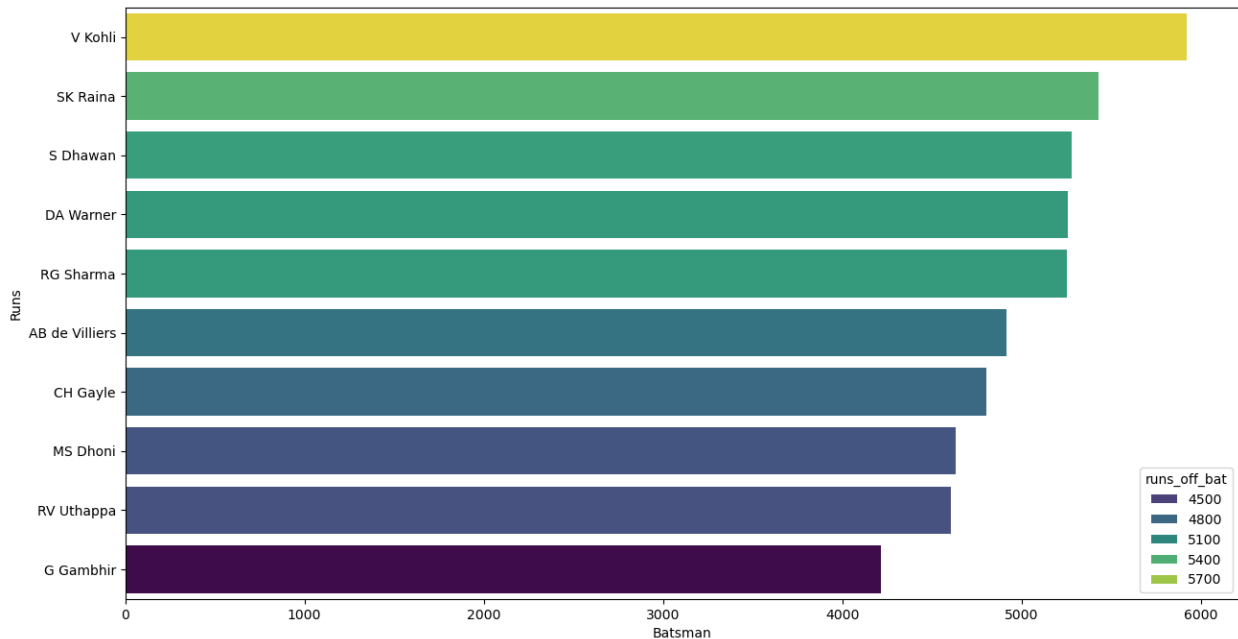


```
# avg Run by Team in powerplay

df[df['over']<6].groupby(['match_id','batting_team']).sum()
['run'].groupby('batting_team').mean().sort_values(ascending=False)
[2:]

batting_team
Delhi Capitals              48.176471
Sunrisers Hyderabad         47.856000
Rising Pune Supergiant      47.433333
Kings XI Punjab             47.126316
Kolkata Knight Riders       46.409326
Delhi Daredevils            45.714286
Deccan Chargers             45.560000
Mumbai Indians              45.529412
Chennai Super Kings         45.195531
Rajasthan Royals            44.912500
Royal Challengers Bangalore 44.826531
Pune Warriors               42.111111
Name: run, dtype: float64
```

```
#Most IPL Century by a Player

runs = df.groupby(['striker','match_id'])['runs_off_bat'].sum()
runs[runs>=100].droplevel(level=1).groupby('striker').count().sort_val
ues(ascending=False)[:10]
```

```
striker
CH Gayle          6
V Kohli           5
DA Warner         4
SR Watson         4
AB de Villiers    3
HM Amla           2
M Vijay           2
KL Rahul          2
SV Samson         2
S Dhawan          2
Name: runs_off_bat, dtype: int64
```

```python
#Most IPL half_Century by a Player

plt.figure(figsize=(15,8))
runs= df.groupby(['striker','start_date'])['runs_off_bat'].sum()
data =
runs[runs>=50].droplevel(level=1).groupby('striker').count().sort_valu
es(ascending=False)[:10]
sns.barplot(y=data.index,x=data,orient='h',hue=data,palette="viridis")
plt.xlabel('Half_Centuries')
plt.ylabel('Batsman')
plt.show()
```
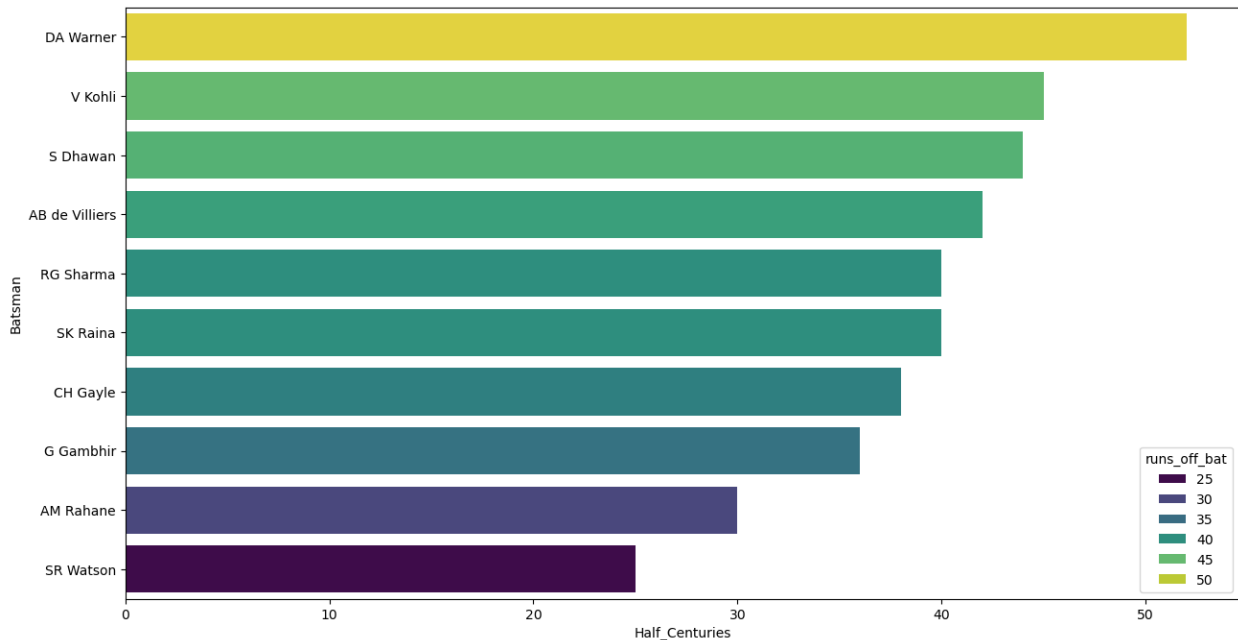


```python
# how much run by a player

df[df['striker'] == 'V Kohli']['runs_off_bat'].sum()
runs = df.groupby(['striker'])['runs_off_bat'].sum()
runs['V Kohli']
```

```
5921

df[df['striker'] == 'MK Pandey']['runs_off_bat'].sum()
runs = df.groupby(['striker'])['runs_off_bat'].sum()
runs['MK Pandey']

3338

# kis date ko kitan run banya

df[df['striker'] == 'V Kohli']['runs_off_bat'].sum()
runs = df.groupby(['striker','start_date'])['runs_off_bat'].sum()
runs.loc['V Kohli']

start_date
01-05-09    16
01-05-17    20
01-05-18    32
02-04-10    42
02-04-19    23
            ..
30-04-16    14
30-04-19    25
31-03-10    34
31-03-19     3
31-10-20     7
Name: runs_off_bat, Length: 185, dtype: int64

# kis year me kitna run banya

plt.figure(figsize=(15,8))
runs = df.groupby(['striker','season'])['runs_off_bat'].sum()
kohli_runs = runs.loc['V Kohli']

sns.barplot(x=kohli_runs.values, y=kohli_runs.index,
orient='h',palette="viridis")
plt.xlabel('Runs')
plt.ylabel('Season')
plt.title('V Kohli Total Runs Per Season')
plt.show()

C:\Users\DELL\AppData\Local\Temp\ipykernel_7560\126988445.py:5:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `y` variable to `hue` and set
`legend=False` for the same effect.

  sns.barplot(x=kohli_runs.values, y=kohli_runs.index,
orient='h',palette="viridis")
```

V Kohli Total Runs Per Season
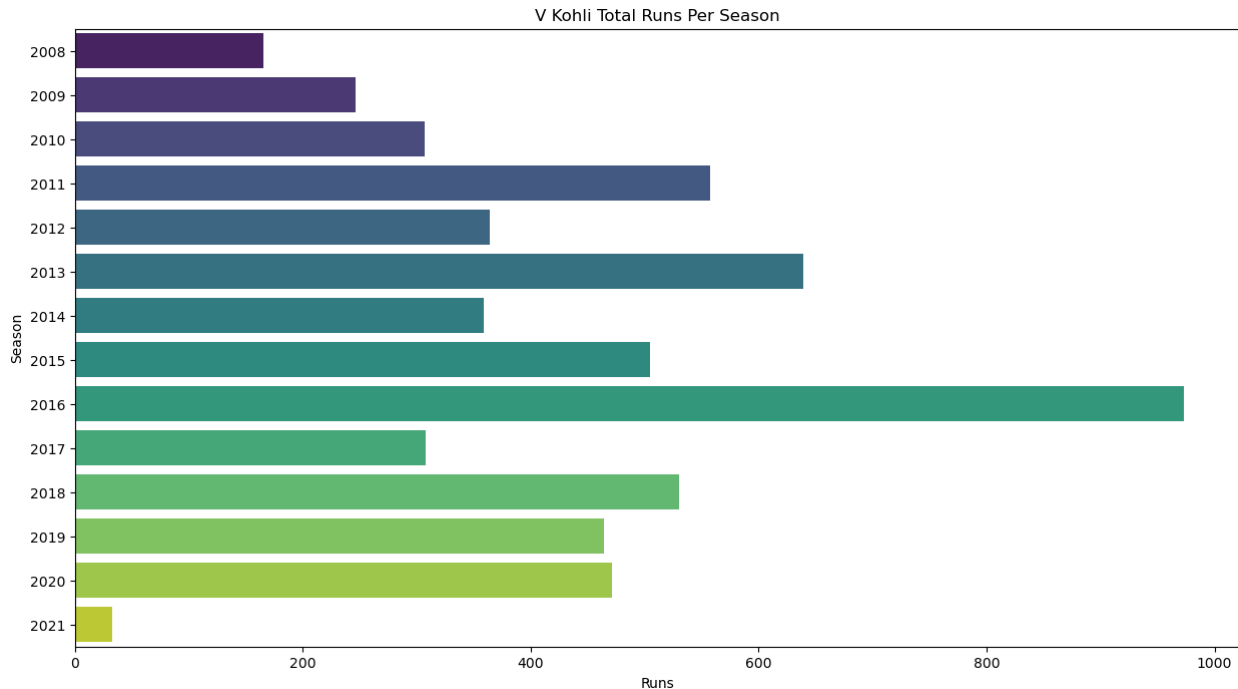
```
plt.figure(figsize=(15,8))
runs = df.groupby(['striker','season'])['runs_off_bat'].sum()
Pandey_runs = runs.loc['MK Pandey']

sns.barplot(x=Pandey_runs.values, y=Pandey_runs.index,
orient='h',palette="viridis")
plt.xlabel('Runs')
plt.ylabel('Season')
plt.title('MK Pandey Total Runs Per Season')
plt.show()

C:\Users\DELL\AppData\Local\Temp\ipykernel_7560\3709000249.py:5:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `y` variable to `hue` and set
`legend=False` for the same effect.

  sns.barplot(x=Pandey_runs.values, y=Pandey_runs.index,
orient='h',palette="viridis")
```

MK Pandey Total Runs Per Season
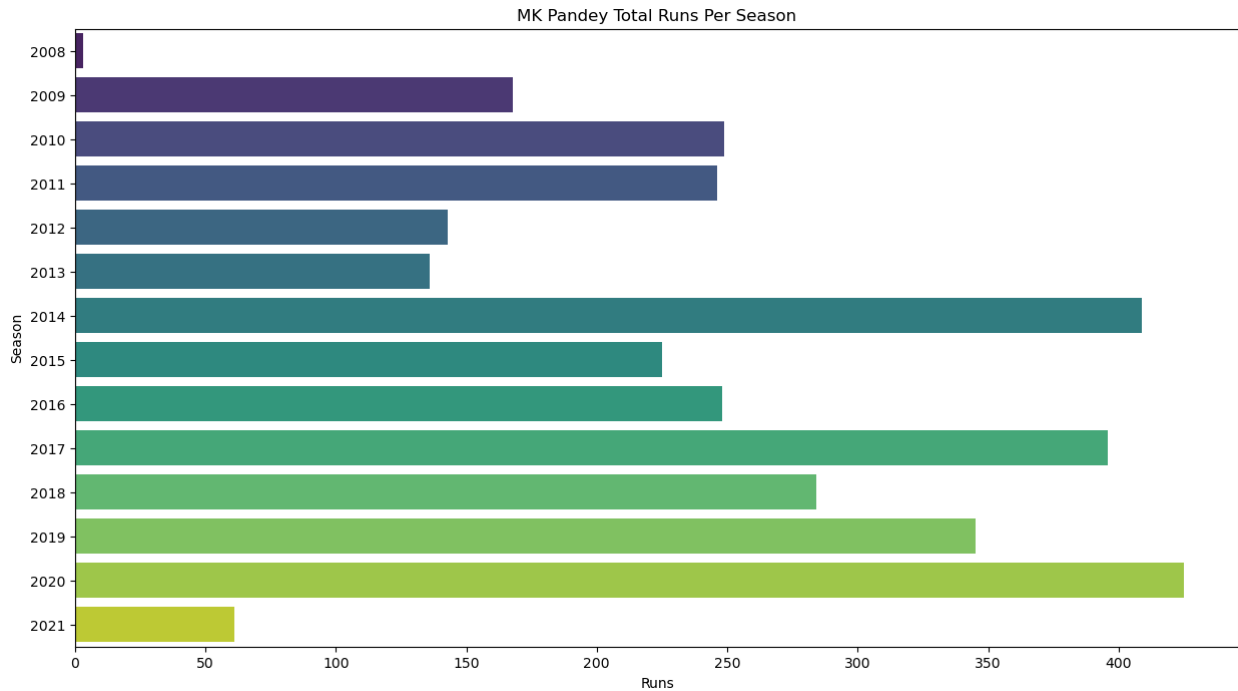
```
# No of sixes in ipl season

plt.figure(figsize=(15,8))
data = df[df['runs_off_bat'] ==6 ].groupby('season').count()
['match_id'].sort_values(ascending=False)

sns.barplot(y=data.index,x=data,orient='h',hue=
data,palette='viridis')
plt.xlabel('sixes')
plt.ylabel('season')
plt.show()
```
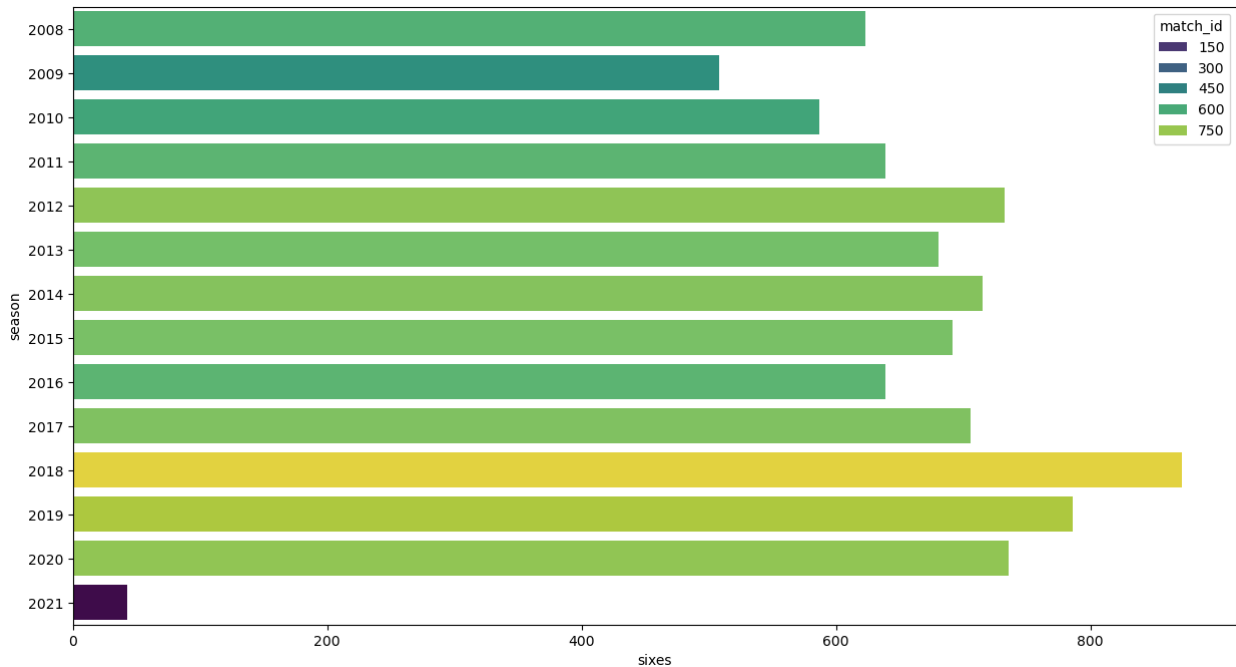
```
#top 15 players jisne sbse jiyada chhaka lagya

df[df['runs_off_bat'] == 6]
df[df['runs_off_bat'] ==
6].groupby('striker').size().sort_values(ascending=False).head(15)

striker
CH Gayle           351
AB de Villiers     239
MS Dhoni           216
RG Sharma          215
V Kohli            202
SK Raina           199
KA Pollard         199
DA Warner          195
SR Watson          190
RV Uthappa         163
YK Pathan          161
Yuvraj Singh       149
AT Rayudu          134
BB McCullum        130
AD Russell         129
dtype: int64

# sbse jiyada chakka lagane wla player in a year

# Top 15 player-season sixes
top_sixes = df[df['runs_off_bat'] == 6] \
                                    .groupby(['striker',
'season']) \
```

```
                                       .size().sort_values(ascending=
False) \
                                       .head(15).reset_index(name='si
xes')

# Plot
plt.figure(figsize=(12, 6))
sns.barplot(data=top_sixes, x='sixes', y=top_sixes['striker'] + " (" +
top_sixes['season'].astype(str) + ")", palette="viridis")
plt.xlabel('Number of Sixes')
plt.ylabel('Player (Season)')
plt.title('Top 15 Player-Season Six Hitters')
plt.tight_layout()
plt.show()
```
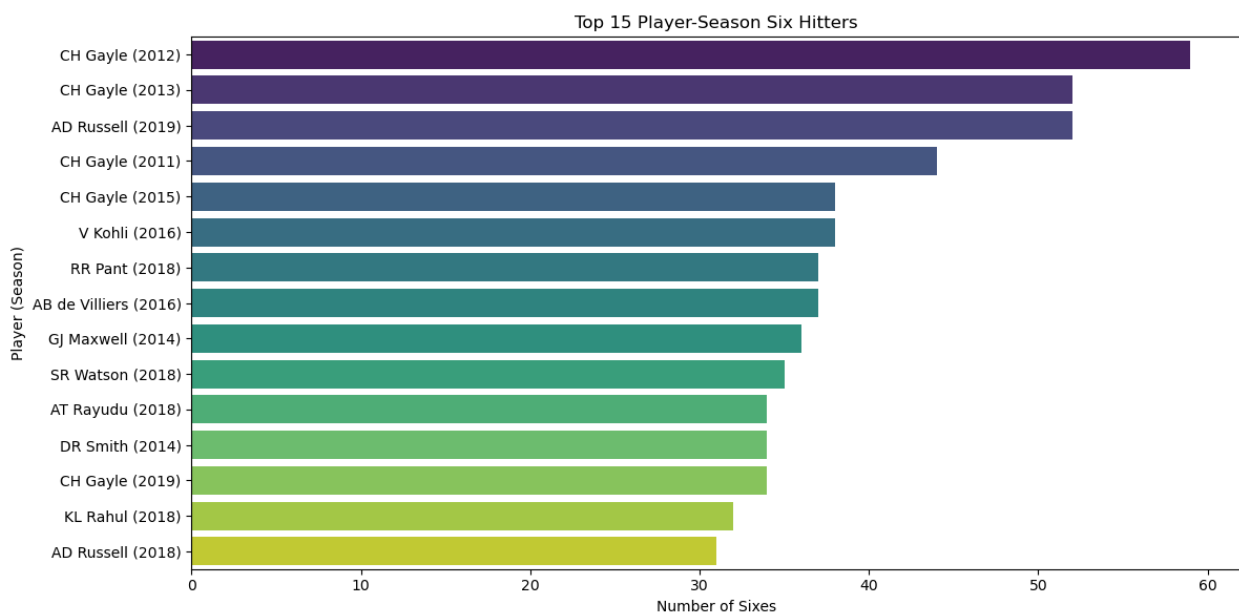
```
C:\Users\DELL\AppData\Local\Temp\ipykernel_7560\1859662846.py:9:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `y` variable to `hue` and set
`legend=False` for the same effect.

  sns.barplot(data=top_sixes, x='sixes', y=top_sixes['striker'] + " ("
+ top_sixes['season'].astype(str) + ")", palette="viridis")
```


Top 15 Player-Season Six Hitters

```
#ek sal me sbse jiyada chakka lagane wala

sixes2 = df[(df['runs_off_bat'] == 6) & (df['season'] == 2020)]
sixes2 = sixes2.groupby('striker').size().reset_index(name='sixes')
sixes2 = sixes2.sort_values(by='sixes', ascending=False).head(10)
print(sixes2)
```

```
plt.figure(figsize=(12, 6))
sns.barplot(data = sixes2, x= 'sixes', y= 'striker',palette='viridis')
plt.xlabel('Number of Sixes')
plt.ylabel('Striker')
plt.title("Top 10 Six Hitters in IPL 2020")
plt.tight_layout()
plt.show()
```

```
         striker  sixes
21    Ishan Kishan     30
68      SV Samson     26
42      N Pooran      25
19      HH Pandya     25
17     EJG Morgan     24
12      CH Gayle      24
0    AB de Villiers   23
33      KL Rahul      23
47      Q de Kock     22
30      KA Pollard    22
```

C:\Users\DELL\AppData\Local\Temp\ipykernel_7560\2099981772.py:7:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `y` variable to `hue` and set
`legend=False` for the same effect.

  sns.barplot(data = sixes2, x= 'sixes', y=
'striker',palette='viridis')

Top 10 Six Hitters in IPL 2020
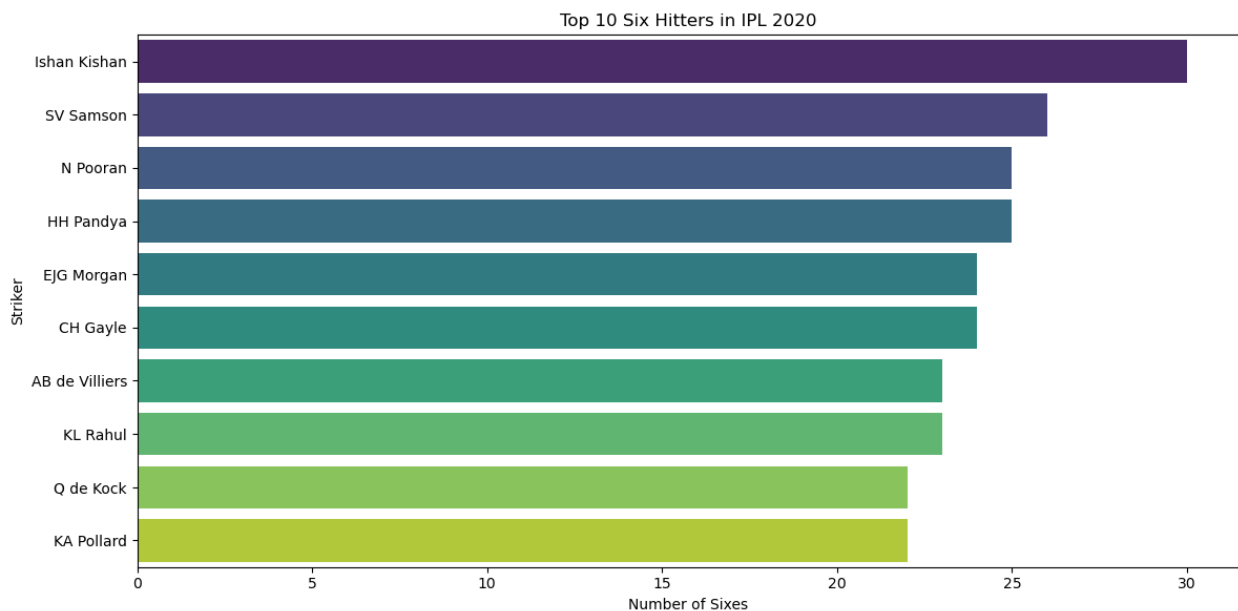
```
df[df['runs_off_bat'] == 6].groupby(['start_date','striker']).count()
['season'].sort_values(ascending=False).droplevel(level=0)[:10]

striker
CH Gayle          17
BB McCullum       13
CH Gayle          13
AB de Villiers    12
CH Gayle          12
CH Gayle          11
ST Jayasuriya     11
M Vijay           11
AD Russell        11
AC Gilchrist      10
Name: season, dtype: int64

# Hiest individual IPL score

df.groupby(['striker','start_date'])
['runs_off_bat'].sum().sort_values(ascending=False)[:10]

striker         start_date
CH Gayle        23-04-13      175
BB McCullum     18-04-08      158
AB de Villiers  10-05-15      133
KL Rahul        24-09-20      132
AB de Villiers  14-05-16      129
RR Pant         10-05-18      128
CH Gayle        17-05-12      128
M Vijay         03-04-10      127
DA Warner       30-04-17      126
V Sehwag        30-05-14      122
Name: runs_off_bat, dtype: int64

#most run conceded by a bowler in an inning

df.groupby(['bowler','start_date'])
['run'].sum().droplevel(level=1).sort_values(ascending=False)[:10]

bowler
Basil Thampi        70
Mujeeb Ur Rahman    66
I Sharma            66
Sandeep Sharma      66
UT Yadav            65
PJ Cummins          65
S Kaul              64
AS Rajpoot          64
TA Boult            63
AB Dinda            63
Name: run, dtype: int64
```
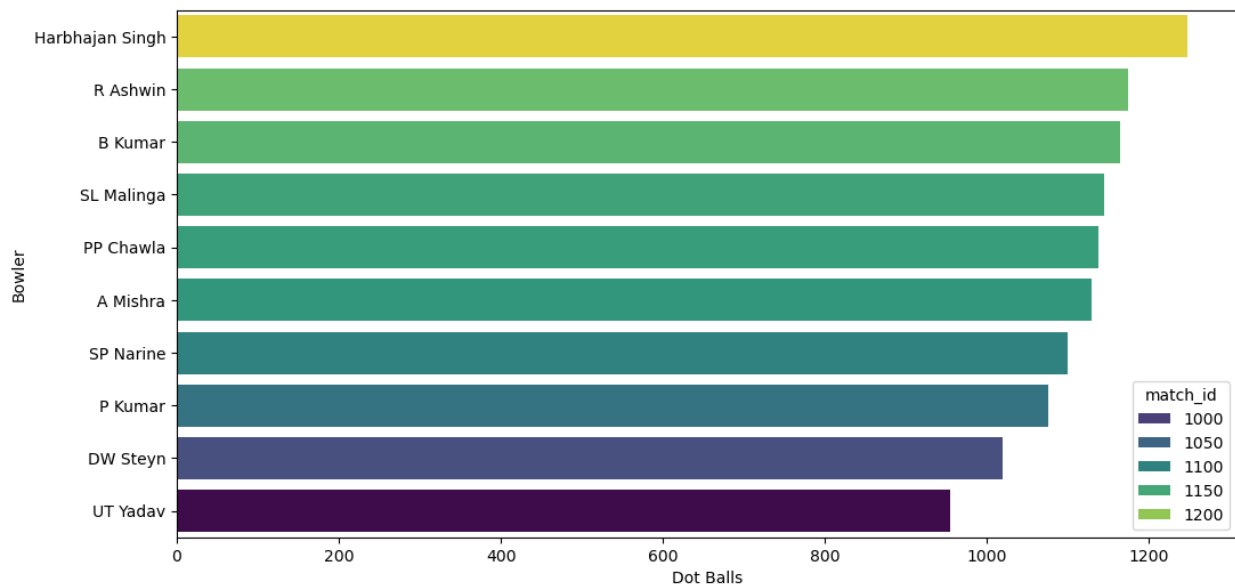
```python
# Most Dot ball by a Bowler

plt.figure(figsize=(12,6))
data = df[df['run'] == 0].groupby('bowler').count()
['match_id'].sort_values(ascending=False)[:10]

sns.barplot(y=data.index,x=data,orient='h',hue=data,palette='viridis')
plt.xlabel('Dot Balls')
plt.ylabel('Bowler')
plt.show()
```



```
data

bowler
Harbhajan Singh    1247
R Ashwin           1174
B Kumar            1165
SL Malinga         1144
PP Chawla          1137
A Mishra           1129
SP Narine          1100
P Kumar            1075
DW Steyn           1019
UT Yadav            955
Name: match_id, dtype: int64
```

```python
# most chacth out in ipl

df[df['wicket_type'] ==
'caught'].groupby('player_dismissed').size().sort_values(ascending=False)[:10]
```

```
player_dismissed
RV Uthappa        115
RG Sharma         112
SK Raina          110
V Kohli           105
KD Karthik         91
S Dhawan           86
Yuvraj Singh       84
G Gambhir          80
DA Warner          79
KA Pollard         78
dtype: int64

# suresh raina catch out

df_raina = df[df['player_dismissed'] == 'SK Raina']
df_raina['wicket_type'].value_counts().plot.pie(autopct='%1.1f%%')
plt.title('Dismissal_kinds',fontweight='bold',fontsize=15)
plt.show()
```


Dismissal_kinds
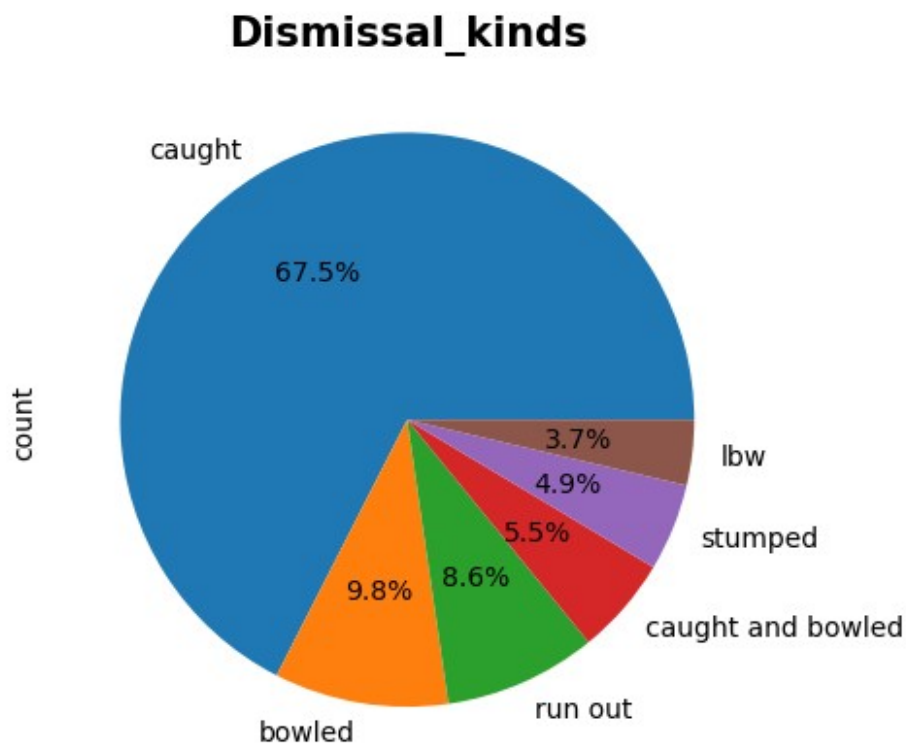
```
# Most IPL Wicket by A Bowler                    # by gaurav sir

lst = 'caught,bowled,lbw,stumped,caught and bowled,hit wicket'
```

```python
df[df['wicket_type'].apply(lambda x: True if x in lst and x !=' ' else
False)]['bowler'].value_counts()[:10]
```

```
bowler
SL Malinga          170
A Mishra            160
PP Chawla           156
DJ Bravo            154
Harbhajan Singh     150
R Ashwin            139
B Kumar             137
SP Narine           127
YS Chahal           121
UT Yadav            119
Name: count, dtype: int64
```

```python
#by me

lst = ['caught', 'bowled', 'lbw', 'stumped', 'caught and bowled', 'hit
wicket']

df[df['wicket_type'].isin(lst)]['bowler'].value_counts()[:10]
```

```
bowler
SL Malinga          170
A Mishra            160
PP Chawla           156
DJ Bravo            154
Harbhajan Singh     150
R Ashwin            139
B Kumar             137
SP Narine           127
YS Chahal           121
UT Yadav            119
Name: count, dtype: int64
```

```python
#most Miden over by a bowler

data = df.groupby(['start_date','bowler','over'])['run'].sum()
data = data[data.values == 0].droplevel(level=[0,2])
data.index.value_counts()[:10]
```

```
bowler
P Kumar            12
IK Pathan           9
DW Steyn            8
SL Malinga          8
DS Kulkarni         7
B Kumar             7
Sandeep Sharma      6
DJ Bravo            6
```

```
R Ashwin          5
DL Chahar         5
Name: count, dtype: int64
```

```python
# Most wicket by an IPL Team

plt.figure(figsize=(15,8))
lst = ['caught', 'bowled', 'lbw', 'stumped', 'caught and bowled', 'hit
wicket']
data = df[df['wicket_type'].isin(lst)]['bowling_team'].value_counts()
df.groupby(['batting_team'])
['extras'].agg('sum').sort_values(ascending=False)

sns.barplot(y=data.index,x=data,orient='h',
hue=data,palette='viridis')
plt.xlabel('Wicket')
plt.ylabel('Teams')
plt.show()
```
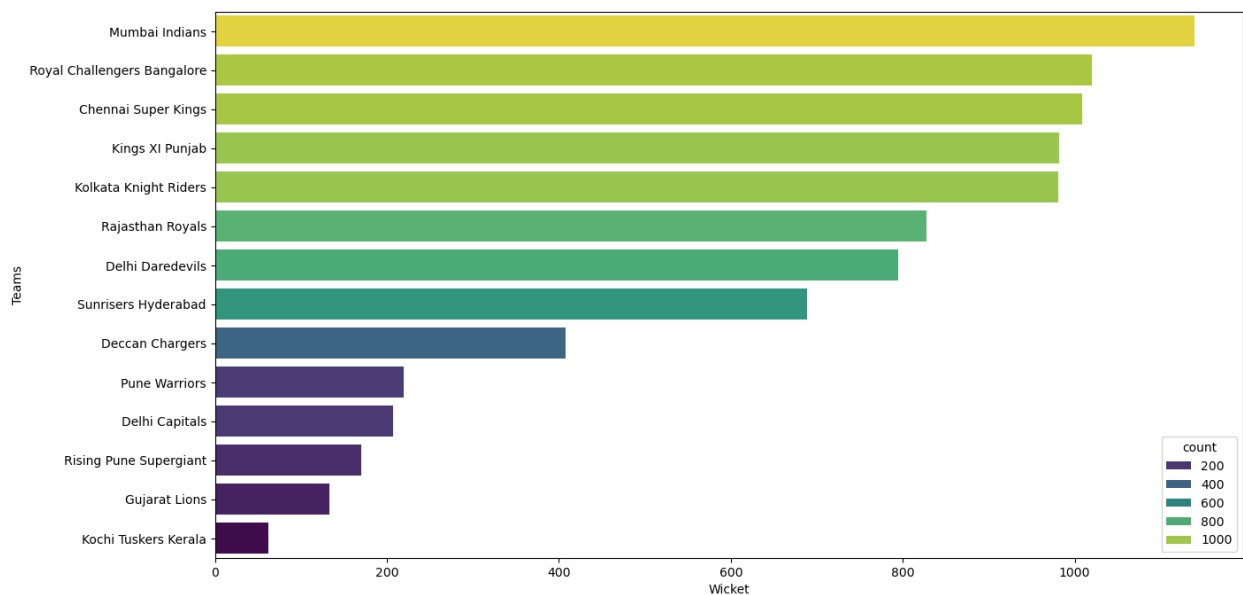


```python
# Most of No Balls by an iPL Team

df.groupby(['batting_team'])
['noballs'].agg('sum').sort_values(ascending=False)
```

```
batting_team
Royal Challengers Bangalore     106
Chennai Super Kings              96
Mumbai Indians                   96
Rajasthan Royals                 94
Kolkata Knight Riders            90
Delhi Daredevils                 73
```

```
Kings XI Punjab              71
Sunrisers Hyderabad          53
Deccan Chargers              49
Pune Warriors                24
Delhi Capitals               21
Gujarat Lions                17
Kochi Tuskers Kerala         11
Rising Pune Supergiant        8
Name: noballs, dtype: int64
```

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import math

df2 = pd.read_csv(r"C:\Users\DELL\Desktop\IPL_Matches 2008-2020.csv")
df2
```

```
          id        city        date player_of_match  \
0     335982   Bangalore    18-04-08      BB McCullum
1     335983  Chandigarh    19-04-08      MEK Hussey
2     335984       Delhi    19-04-08      MF Maharoof
3     335985      Mumbai    20-04-08      MV Boucher
4     335986     Kolkata    20-04-08       DJ Hussey
..       ...         ...         ...             ...
811  1216547       Dubai    28-09-20  AB de Villiers
812  1237177       Dubai    05-11-20       JJ Bumrah
813  1237178   Abu Dhabi    06-11-20   KS Williamson
814  1237180   Abu Dhabi    08-11-20       MP Stoinis
815  1237181       Dubai    10-11-20        TA Boult

                                      venue  neutral_venue  \
0                   M Chinnaswamy Stadium                0
1    Punjab Cricket Association Stadium, Mohali          0
2                         Feroz Shah Kotla                0
3                        Wankhede Stadium                0
4                           Eden Gardens                0
..                                      ...              ...
811      Dubai International Cricket Stadium             0
812      Dubai International Cricket Stadium             0
813               Sheikh Zayed Stadium                0
814               Sheikh Zayed Stadium                0
815      Dubai International Cricket Stadium             0

                     team1                        team2  \
0    Royal Challengers Bangalore        Kolkata Knight Riders
1              Kings XI Punjab          Chennai Super Kings
2              Delhi Daredevils            Rajasthan Royals
3               Mumbai Indians  Royal Challengers Bangalore
```

```
4              Kolkata Knight Riders                   Deccan Chargers
..                               ...                               ...
811  Royal Challengers Bangalore                    Mumbai Indians
812                Mumbai Indians                    Delhi Capitals
813  Royal Challengers Bangalore                Sunrisers Hyderabad
814                Delhi Capitals                Sunrisers Hyderabad
815                Delhi Capitals                    Mumbai Indians

                 toss_winner toss_decision                      winner  \
0    Royal Challengers Bangalore         field         Kolkata Knight
Riders
1              Chennai Super Kings           bat          Chennai Super
Kings
2                Rajasthan Royals           bat                  Delhi
Daredevils
3                  Mumbai Indians           bat  Royal Challengers
Bangalore
4                  Deccan Chargers           bat          Kolkata Knight
Riders
..                           ...           ...
...
811                Mumbai Indians         field  Royal Challengers
Bangalore
812                Delhi Capitals         field                Mumbai
Indians
813            Sunrisers Hyderabad         field              Sunrisers
Hyderabad
814                Delhi Capitals           bat                  Delhi
Capitals
815                Delhi Capitals           bat                  Mumbai
Indians

       result  result_margin eliminator method      umpire1
umpire2
0       runs           140.0          N    NaN   Asad Rauf        RE
Koertzen
1       runs            33.0          N    NaN   MR Benson         SL
Shastri
2     wickets            9.0          N    NaN   Aleem Dar  GA
Pratapkumar
3     wickets            5.0          N    NaN    SJ Davis          DJ
Harper
4     wickets            5.0          N    NaN   BF Bowden         K
Hariharan
..        ...            ...          ...   ...          ...
...
811      tie            NaN          Y    NaN  Nitin Menon          PR
Reiffel
```

```
812      runs               57.0           N    NaN  CB Gaffaney      Nitin
Menon
813   wickets               6.0            N    NaN   PR Reiffel            S
Ravi
814      runs               17.0           N    NaN   PR Reiffel            S
Ravi
815   wickets               5.0            N    NaN  CB Gaffaney      Nitin
Menon

[816 rows x 17 columns]
```

```python
# Who is the won the toss in IPL

TW = df2['toss_winner'].value_counts()
TW
```

```
toss_winner
Mumbai Indians               106
Kolkata Knight Riders         98
Chennai Super Kings           97
Royal Challengers Bangalore   87
Rajasthan Royals              87
Kings XI Punjab               85
Delhi Daredevils              80
Sunrisers Hyderabad           57
Deccan Chargers               43
Pune Warriors                 20
Delhi Capitals                20
Gujarat Lions                 15
Kochi Tuskers Kerala           8
Rising Pune Supergiants        7
Rising Pune Supergiant         6
Name: count, dtype: int64
```

```python
plt.figure(figsize=(15,8))
sns.barplot(x=TW.index,y=TW,hue=TW,palette='icefire')
plt.xticks(rotation=45)
plt.xlabel("Teams")
plt.ylabel("Number of Toss Wins")
plt.title("Toss Wins by IPL Teams (2008-2020)")
plt.show()
```

Toss Wins by IPL Teams (2008-2020)
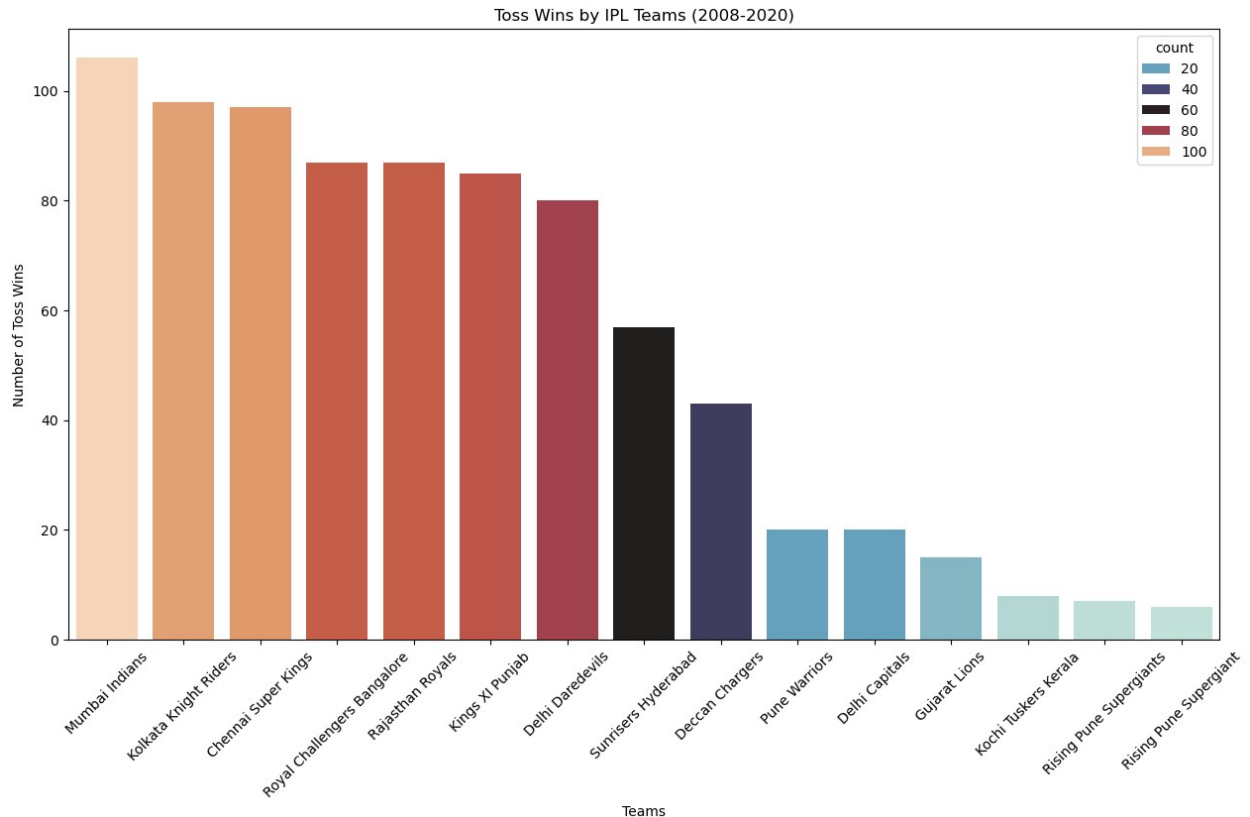
```
# After winning toss Which team choes Bating or Filding

# Extract season from date
df2['season'] = pd.to_datetime(df2['date']).dt.year
TD = df2.groupby(['season', 'toss_winner',
'toss_decision']).size().unstack()[:]
TD

C:\Users\DELL\AppData\Local\Temp\ipykernel_7560\2964906284.py:2:
UserWarning: Could not infer format, so each element will be parsed
individually, falling back to `dateutil`. To ensure parsing is
consistent and as-expected, please specify a format.
  df2['season'] = pd.to_datetime(df2['date']).dt.year
```

toss_decision	bat	field
season toss_winner		
2008 Chennai Super Kings	4.0	1.0
Deccan Chargers	4.0	5.0
Delhi Daredevils	2.0	4.0
Kings XI Punjab	4.0	4.0
Kolkata Knight Riders	6.0	NaN
...	...	...
2020 Kolkata Knight Riders	4.0	2.0
Mumbai Indians	4.0	4.0
Rajasthan Royals	2.0	5.0

```
         Royal Challengers Bangalore  4.0     2.0
         Sunrisers Hyderabad          4.0     7.0

[108 rows x 2 columns]

plt.figure(figsize=(15,8))
sns.countplot(x='season',hue='toss_decision',data=df2,palette='magma',
saturation=2)
plt.xticks(rotation=90,fontsize=10)
plt.yticks(fontsize=15)
plt.xlabel('\n Session',fontsize=15)
plt.ylabel('Count',fontsize=15)
plt.title('Toss decesion across
session',fontsize=15,fontweight='bold')
plt.show()
```

Toss decesion across session