In [1]:

```
1  import pandas as pd
2  import numpy as np
```

In [2]:

```
1  df = pd.read_csv("Match.csv")
2  df.head()
```

Out[2]:

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs	legbye_runs	noball_runs	penalty
0	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0
1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0
2	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0
3	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0
4	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0

5 rows × 21 columns

In [3]:

```
1  df.shape
```

Out[3]:

(69761, 21)

In [4]:

```
1  df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 69761 entries, 0 to 69760
Data columns (total 21 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   match_id         69761 non-null  int64
 1   inning           69761 non-null  int64
 2   batting_team     69761 non-null  object
 3   bowling_team     69761 non-null  object
 4   over             69761 non-null  int64
 5   ball             69761 non-null  int64
 6   batsman          69761 non-null  object
 7   non_striker      69761 non-null  object
 8   bowler           69761 non-null  object
 9   is_super_over    69761 non-null  int64
 10  wide_runs        69761 non-null  int64
 11  bye_runs         69761 non-null  int64
 12  legbye_runs      69761 non-null  int64
 13  noball_runs      69761 non-null  int64
 14  penalty_runs     69761 non-null  int64
 15  batsman_runs     69761 non-null  int64
 16  extra_runs       69761 non-null  int64
 17  total_runs       69761 non-null  int64
 18  player_dismissed 3516 non-null   object
 19  dismissal_kind   3516 non-null   object
 20  fielder          2529 non-null   object
dtypes: int64(13), object(8)
memory usage: 11.2+ MB
```

In [5]:

```
1  df[["player_dismissed"]].value_counts()
```

Out[5]:

```
player_dismissed
RG Sharma          59
SK Raina           59
AC Gilchrist       57
G Gambhir          56
Yuvraj Singh       55
                   ..
A Choudhary         1
BAW Mendis          1
H Das               1
GR Napier           1
Jaskaran Singh      1
Length: 354, dtype: int64
```

In [6]:

```
1  df[["fielder"]].value_counts()
```

Out[6]:

```
fielder
RV Uthappa        61
KD Karthik        57
MS Dhoni          55
AC Gilchrist      54
KC Sangakkara     46
                  ..
RA Jadeja (sub)    1
Misbah-ul-Haq      1
AD Mascarenhas     1
M Kaif (sub)       1
KS Williamson      1
Length: 352, dtype: int64
```

In [7]:

```
1  df[["dismissal_kind"]].value_counts()
```

Out[7]:

```
dismissal_kind
caught              2029
bowled               650
run out              379
lbw                  223
stumped              121
caught and bowled    106
hit wicket             4
retired hurt           4
dtype: int64
```
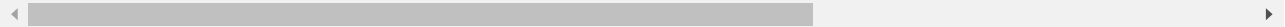
In [8]:

```
1  df.head()
```

Out[8]:

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs	legbye_runs	noball_runs	penalt
0	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
2	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
3	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
4	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	

5 rows × 21 columns

In [9]:

```
1  df.isnull().sum()
```

Out[9]:

```
match_id            0
inning              0
batting_team        0
bowling_team        0
over                0
ball                0
batsman             0
non_striker         0
bowler              0
is_super_over       0
wide_runs           0
bye_runs            0
legbye_runs         0
noball_runs         0
penalty_runs        0
batsman_runs        0
extra_runs          0
total_runs          0
player_dismissed    66245
dismissal_kind      66245
fielder             67232
dtype: int64
```

In [10]:

```
1  df["player_dismissed"].mode()
```

Out[10]:

```
0    RG Sharma
1     SK Raina
dtype: object
```

In [11]:

```
1  df["player_dismissed"].replace(np.nan,"RG Sharma",inplace=True)
```
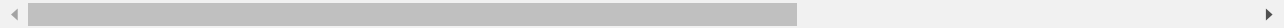
In [12]:

```
1  df
```

Out[12]:

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs	legbye_runs	noball_runs	p
0	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
2	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
3	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
4	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
69756	295	2	Deccan Chargers	Pune Warriors	16	4	JP Duminy	DT Christian	R Sharma	0	...	0	0	0	
69757	295	2	Deccan Chargers	Pune Warriors	16	5	JP Duminy	DT Christian	R Sharma	0	...	0	0	0	
69758	295	2	Deccan Chargers	Pune Warriors	16	6	JP Duminy	DT Christian	R Sharma	0	...	0	0	0	
69759	295	2	Deccan Chargers	Pune Warriors	16	7	JP Duminy	DT Christian	R Sharma	0	...	0	0	0	
69760	295	2	Deccan Chargers	Pune Warriors	17	1	DT Christian	JP Duminy	Yuvraj Singh	0	...	0	0	0	

69761 rows × 21 columns

In [13]:

```
1 df["dismissal_kind"].mode()
```

Out[13]:

```
0    caught
dtype: object
```

In [14]:

```
1 df["dismissal_kind"].replace(np.nan,"caught",inplace=True)
2 df
```

Out[14]:

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs	legbye_runs	noball_runs	p
0	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0
1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0
2	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0
3	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0
4	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
69756	295	2	Deccan Chargers	Pune Warriors	16	4	JP Duminy	DT Christian	R Sharma	0	...	0	0	0	0
69757	295	2	Deccan Chargers	Pune Warriors	16	5	JP Duminy	DT Christian	R Sharma	0	...	0	0	0	0
69758	295	2	Deccan Chargers	Pune Warriors	16	6	JP Duminy	DT Christian	R Sharma	0	...	0	0	0	0
69759	295	2	Deccan Chargers	Pune Warriors	16	7	JP Duminy	DT Christian	R Sharma	0	...	0	0	0	0
69760	295	2	Deccan Chargers	Pune Warriors	17	1	DT Christian	JP Duminy	Yuvraj Singh	0	...	0	0	0	0

69761 rows × 21 columns

In [15]:

```
1 df["fielder"].mode()
```

Out[15]:

```
0    RV Uthappa
dtype: object
```

In [16]:

```python
df["fielder"].replace(np.nan,"RV Uthappa",inplace=True)
df
```

Out[16]:

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs	legbye_runs	noball_runs	p
0	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
2	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
3	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
4	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
69756	295	2	Deccan Chargers	Pune Warriors	16	4	JP Duminy	DT Christian	R Sharma	0	...	0	0	0	
69757	295	2	Deccan Chargers	Pune Warriors	16	5	JP Duminy	DT Christian	R Sharma	0	...	0	0	0	
69758	295	2	Deccan Chargers	Pune Warriors	16	6	JP Duminy	DT Christian	R Sharma	0	...	0	0	0	
69759	295	2	Deccan Chargers	Pune Warriors	16	7	JP Duminy	DT Christian	R Sharma	0	...	0	0	0	
69760	295	2	Deccan Chargers	Pune Warriors	17	1	DT Christian	JP Duminy	Yuvraj Singh	0	...	0	0	0	

69761 rows × 21 columns

In [17]:

```python
df.isnull().sum()
```

Out[17]:

```
match_id           0
inning             0
batting_team       0
bowling_team       0
over               0
ball               0
batsman            0
non_striker        0
bowler             0
is_super_over      0
wide_runs          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
```

In [18]:

```
1  df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 69761 entries, 0 to 69760
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   match_id          69761 non-null  int64
 1   inning            69761 non-null  int64
 2   batting_team      69761 non-null  object
 3   bowling_team      69761 non-null  object
 4   over              69761 non-null  int64
 5   ball              69761 non-null  int64
 6   batsman           69761 non-null  object
 7   non_striker       69761 non-null  object
 8   bowler            69761 non-null  object
 9   is_super_over     69761 non-null  int64
 10  wide_runs         69761 non-null  int64
 11  bye_runs          69761 non-null  int64
 12  legbye_runs       69761 non-null  int64
 13  noball_runs       69761 non-null  int64
 14  penalty_runs      69761 non-null  int64
 15  batsman_runs      69761 non-null  int64
 16  extra_runs        69761 non-null  int64
 17  total_runs        69761 non-null  int64
 18  player_dismissed  69761 non-null  object
 19  dismissal_kind    69761 non-null  object
 20  fielder           69761 non-null  object
dtypes: int64(13), object(8)
memory usage: 11.2+ MB
```

In [19]:

```
1  df.head(3)
```

Out[19]:

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs	legbye_runs	noball_runs	penalty
0	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
2	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	

3 rows × 21 columns

Find Team wise runs scored.

In [20]:

```
1  df.groupby("batting_team")["total_runs"].sum()
```

Out[20]:

```
batting_team
Chennai Super Kings           9253
Deccan Chargers               8931
Delhi Daredevils             10654
Gujarat Lions                 2412
Kings XI Punjab              10763
Kochi Tuskers Kerala          1760
Kolkata Knight Riders         9924
Mumbai Indians               11202
Pune Warriors                 1657
Rajasthan Royals              8039
Rising Pune Supergiant        2470
Royal Challengers Bangalore  10221
Sunrisers Hyderabad           2349
Name: total_runs, dtype: int64
```

Find Total Runs of Virat Kohli

In [21]:

```
1  df[df["batsman"]=="V Kohli"]["batsman_runs"].sum()
```

Out[21]:

1428

Find Total Runs of Your Fav Player.

In [22]:
```python
df[df["batsman"]=="RG Sharma"]["batsman_runs"].sum()
```

Out[22]:

1794

Extra Runs Given By All Bowlers in a League.

In [23]:
```python
df.groupby("bowler")["extra_runs"].sum().sort_values(ascending=False)
```

Out[23]:

```
bowler
SL Malinga     158
RP Singh       135
I Sharma       118
P Kumar        111
IK Pathan      101
               ...
I Malhotra       0
B Chipli         0
Tejas Baroka     0
A Nel            0
TM Head          0
Name: extra_runs, Length: 285, dtype: int64
```

Extra Runs Given By TS Mills.

In [24]:
```python
df[df["bowler"]=="TS Mills"]["extra_runs"].sum()
```

Out[24]:

13

Catches Taken By Mandeep Singh.

In [25]:
```python
df[(df.dismissal_kind=="caught")&(df.fielder=="Mandeep Singh")].shape[0]
```

Out[25]:

7

How Many Times DA Warner is dismissed by Mohammed Shami.

In [26]:
```python
df[(df.player_dismissed=="DA Warner")&(df.bowler=="Mohammed Shami")]
```

Out[26]:

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs	legbye_runs	noball_runs
9029	39	1	Sunrisers Hyderabad	Delhi Daredevils	6	2	DA Warner	S Dhawan	Mohammed Shami	0	...	0	0	0

1 rows × 21 columns

Total Number of Wickets Taken by YS Chahal.

In [27]:
```python
df[(df.player_dismissed.notnull())&(df.bowler=="YS Chahal")].shape[0]
```

Out[27]:

271

Find Number of Stump Out done by MS Dhoni.

In [28]:

```python
df[(df.dismissal_kind=="stumped")&(df.fielder=="MS Dhoni")].shape[0]
```

Out[28]:

17

How Many Times S Dhawan Got Out by lbw.

In [29]:

```python
df[(df["batsman"]=="S Dhawan")&(df["dismissal_kind"]=="lbw")].shape[0]
```

Out[29]:

2

Total number sixes hit by MS Dhoni.

In [30]:

```python
df[(df["batsman"]=="MS Dhoni")&(df["batsman_runs"]==6)].shape[0]
```

Out[30]:

62

Total numbers of balls faced by MI Players.

In [31]:

```python
df[df["batting_team"]=="Mumbai Indians"].shape[0]
```

Out[31]:

8662

Total numbers of balls faced by KKR Players.

In [32]:

```python
df[df["batting_team"]=="Kolkata Knight Riders"].shape[0]
```

Out[32]:

7887

Find the total number of runs scored by Sachin Tendulkar.

In [33]:

```python
df[df.batsman=="SR Tendulkar"]["total_runs"].sum()
```

Out[33]:

1730

Find the total number of balls faced by Schin Tendulkar.

In [34]:

```python
df[df["batsman"]=="SR Tendulkar"].shape[0]
```

Out[34]:

1388

Find Sachin Tendulkar's dismissal types in IPL.

In [35]:

```python
df[df.player_dismissed=="SR Tendulkar"].dismissal_kind.value_counts()
```

Out[35]:

```
caught               23
bowled                7
lbw                   5
run out               2
caught and bowled     1
stumped               1
Name: dismissal_kind, dtype: int64
```

Find the bowlers who dismissed SR Tendulkar.

In [36]:

```python
df[df["player_dismissed"]=="SR Tendulkar"]["bowler"]
```

Out[36]:

```
22501     Joginder Sharma
23139       Shoaib Akhtar
23597       Shahid Afridi
24663            IK Pathan
25219            VY Mahesh
26239           SK Trivedi
29727              PP Ojha
30608            LR Shukla
31373             VS Malik
32028            AB Agarkar
33240            D du Preez
34298             RP Singh
34887           PJ Sangwan
35588       R Vinay Kumar
37502             SK Warne
37846             SK Raina
39836             R Bhatia
41253       AD Mascarenhas
42994             A Mishra
44154       R Vinay Kumar
45810             L Balaji
46897             RP Singh
47440            PP Chawla
48791             R Sharma
49947           T Thushara
50753            PP Chawla
52211           PJ Sangwan
53353             JH Kallis
54523             DW Steyn
55419            SB Jakati
60095             R Sharma
60650          DE Bollinger
61427             A Mishra
63078            AL Menaria
64590             RJ Harris
65497         Yuvraj Singh
66654            IK Pathan
67929              P Kumar
69098          Anand Rajan
Name: bowler, dtype: object
```

How many times was Suresh Raina dismissed.

In [37]:

```python
df[df.player_dismissed=="SK Raina"].shape[0]
```

Out[37]:

```
59
```

Display player details who hit the maximum no of sixes in the current IPL tournament.

In [38]:

```
1  df[df["batsman_runs"]==6]["batsman"].value_counts()
```

Out[38]:

```
SK Raina         83
YK Pathan        80
Yuvraj Singh     75
CH Gayle         72
RG Sharma        72
                 ..
RT Ponting        1
SD Chitnis        1
Sohail Tanvir     1
SP Goswami        1
WD Parnell        1
Name: batsman, Length: 234, dtype: int64
```

What is the Overall No Balls Runs for each team in the dataset?

In [40]:

```
1  df.groupby("batting_team")["noball_runs"].sum()
```

Out[40]:

```
batting_team
Chennai Super Kings          32
Deccan Chargers              45
Delhi Daredevils             42
Gujarat Lions                10
Kings XI Punjab              36
Kochi Tuskers Kerala         11
Kolkata Knight Riders        39
Mumbai Indians               42
Pune Warriors                 9
Rajasthan Royals             45
Rising Pune Supergiant        1
Royal Challengers Bangalore  34
Sunrisers Hyderabad           6
Name: noball_runs, dtype: int64
```

Which player scored the most runs in the dataset?

In [41]:

```
1  df.groupby("batsman")["total_runs"].sum().sort_values(ascending=False)
```

Out[41]:

```
batsman
SK Raina       2255
G Gambhir      2034
RG Sharma      1867
SR Tendulkar   1730
MS Dhoni       1650
                ...
C Nanda           0
S Kaul            0
Imran Tahir       0
SN Thakur         0
L Ablish          0
Name: total_runs, Length: 383, dtype: int64
```
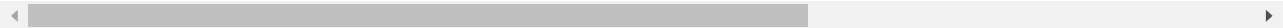
Create Your Own Question

In [42]:

```
1  df.head()
```

Out[42]:

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs	legbye_runs	noball_runs	penalty
0	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
2	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
3	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	
4	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	

5 rows × 21 columns

What is the Overall legbye runs for each team in the dataset?

In [43]:

```
1  df.groupby("batting_team")["legbye_runs"].sum()
```

Out[43]:

```
batting_team
Chennai Super Kings            156
Deccan Chargers                155
Delhi Daredevils               163
Gujarat Lions                   37
Kings XI Punjab                200
Kochi Tuskers Kerala            34
Kolkata Knight Riders          169
Mumbai Indians                 253
Pune Warriors                   25
Rajasthan Royals               157
Rising Pune Supergiant          33
Royal Challengers Bangalore    177
Sunrisers Hyderabad             32
Name: legbye_runs, dtype: int64
```

Find Number of Stump Out done by Yuvraj Singh.

In [44]:

```
1  df[(df.dismissal_kind=="stumped")&(df.fielder=="Yuvraj Singh")].shape[0]
```

Out[44]:

```
0
```

which bowler took the maximum number of wicket in catches

In [45]:

```
1  df[df["dismissal_kind"]=="caught"]["bowler"].value_counts().head(1)
```

Out[45]:

```
RP Singh    1235
Name: bowler, dtype: int64
```

which batsman hit the maximum number of 4

In [47]:

```
1  df[df["batsman_runs"]==4]["batsman"].value_counts().head(1)
```

Out[47]:

```
G Gambhir    230
Name: batsman, dtype: int64
```

how many wickets does SL Malinga took

In [48]:

```
1  df[(df.player_dismissed.notnull())&(df.bowler=="SL Malinga")].shape[0]
```

Out[48]:

1222

which batsman hit the maximum number of 6

In [50]:

```
1  df[df["batsman_runs"]==6]["batsman"].value_counts().head(1)
```

Out[50]:

```
SK Raina    83
Name: batsman, dtype: int64
```

what is the maximum number of penalty runs in database ?

In [51]:

```
1  df["penalty_runs"].max()
```

Out[51]:

5

display the name of batting teams in the dataset.

In [53]:

```
1  df["batting_team"].unique()
```

Out[53]:

```
array(['Sunrisers Hyderabad', 'Royal Challengers Bangalore',
       'Mumbai Indians', 'Rising Pune Supergiant', 'Gujarat Lions',
       'Kolkata Knight Riders', 'Kings XI Punjab', 'Delhi Daredevils',
       'Chennai Super Kings', 'Rajasthan Royals', 'Deccan Chargers',
       'Kochi Tuskers Kerala', 'Pune Warriors'], dtype=object)
```

In [ ]:

```
1
```