Task 4

Machine Learning and Deep learning Summer Internship

Registration id -SIRSS2290

Mohammad Sharique Salman

EDA on IPL dataset

In [6]:

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

Importing Data of Each ball

In [7]:

```python
df = pd.read_csv('deliveries.csv')
```

Having a look at our data

In [8]:

```python
df.head()
```

Out[8]:

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

5 rows × 21 columns

In [9]:

```python
filt = df['batsman'] == 'DA Warner'
df_warner = df[filt]
df_warner.shape
```
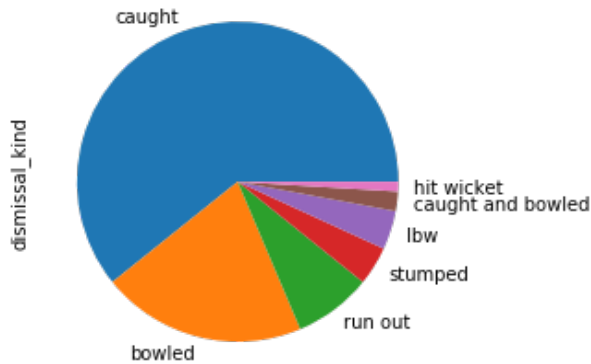
```
(2902, 21)
```

Visualizing how David Warner get's out using a Pie plot.

In [10]:

```
df_warner['dismissal_kind'].value_counts().plot.pie()
```

Out[10]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2ccb19b4a58>
```


In [11]:

```
len(df_warner[df_warner['batsman_runs']==4])
```

Out[11]:

```
401
```

In [12]:

```
def count(df,runs):
    return len(df_warner[df_warner['batsman_runs']==runs])*runs
```

In [13]:

```
print(count(df_warner,1))
```

```
997
```

Visualizing how Warner scores most of his runs

In [14]:

```
slices =[997,1604,960]
labels = [1,4,6]
explode =[0.,0.1,0]
plt.pie(slices, labels = labels, autopct='%1.1f%%', explode = explode)
```

Out[14]:

```
([<matplotlib.patches.Wedge at 0x2ccb266a048>,
  <matplotlib.patches.Wedge at 0x2ccb266a7f0>,
  <matplotlib.patches.Wedge at 0x2ccb266af28>],
 [Text(0.701226217262596, 0.8475150690244927, '1'),
  Text(-1.1993607480544137, -0.03916370802614444, '4'),
  Text(0.7285125276687221, -0.8241780736162116, '6')],
 [Text(0.3824870275977796, 0.4622809467406323, '28.0%'),
  Text(-0.6996271030317414, -0.02284549634858425, '45.0%'),
  Text(0.39737046963748474, -0.4495516765179336, '27.0%')])
```
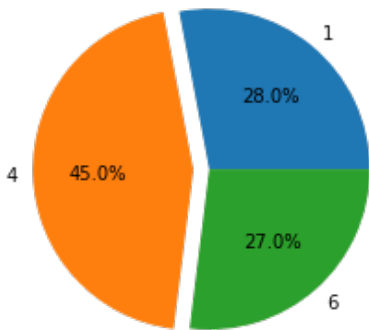
In [15]:

```python
df.head()
```

Out[15]:

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

5 rows × 21 columns

Assigning shorthands for each Team name

In [16]:

```python
df['bowling_team'].unique()
```

Out[16]:

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

In [17]:

```python
Teams = {
    'Royal Challengers Bangalore':'RCB', 'Sunrisers Hyderabad':'SRH',
    'Rising Pune Supergiant':'RPS', 'Mumbai Indians':'MI',
    'Kolkata Knight Riders':'KKR', 'Gujarat Lions':'GL', 'Kings XI Punjab':'KXIP',
    'Delhi Daredevils':'DD', 'Chennai Super Kings':'CSK', 'Rajasthan Royals':'RR',
    'Deccan Chargers':'DC', 'Kochi Tuskers Kerala':'KTK', 'Pune Warriors':'PW',
    'Rising Pune Supergiants':'RPS'
}
```

In [18]:

```
df['batting_team'] = df['batting_team'].map(Teams)
df['bowling_team'] = df['bowling_team'].map(Teams)
```

In [19]:

```
df.head()
```

Out[19]:

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs	legby
0	1	1	SRH	RCB	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	
1	1	1	SRH	RCB	1	2	DA Warner	S Dhawan	TS Mills	0	...	0	
2	1	1	SRH	RCB	1	3	DA Warner	S Dhawan	TS Mills	0	...	0	
3	1	1	SRH	RCB	1	4	DA Warner	S Dhawan	TS Mills	0	...	0	
4	1	1	SRH	RCB	1	5	DA Warner	S Dhawan	TS Mills	0	...	0	

5 rows × 21 columns

In [20]:

```
df.columns
```

Out[20]:

```
Index(['match_id', 'inning', 'batting_team', 'bowling_team', 'over', 'ball',
       'batsman', 'non_striker', 'bowler', 'is_super_over', 'wide_runs',
       'bye_runs', 'legbye_runs', 'noball_runs', 'penalty_runs',
       'batsman_runs', 'extra_runs', 'total_runs', 'player_dismissed',
       'dismissal_kind', 'fielder'],
      dtype='object')
```

In [21]:

```
runs = df.groupby(['match_id', 'inning', 'batting_team'])['total_runs'].sum().reset_inde
x()
runs.drop('match_id',axis = 1, inplace = True)
runs
```

Out[21]:

	inning	batting_team	total_runs
0	1	SRH	207
1	2	RCB	172
2	1	MI	184
3	2	RPS	187
4	1	GL	183
5	2	KKR	184
6	1	RPS	163
7	2	KXIP	164
8	1	RCB	157
9	2	DD	142
10	1	GL	135
11	2	SRH	140
12	1	KKR	178

13	inning	batting_team	total_runs
14	1	RCB	148
15	2	KXIP	150
16	1	DD	205
17	2	RPS	108
18	1	SRH	158
19	2	MI	159
20	1	KXIP	170
21	2	KKR	171
22	1	RCB	142
23	2	MI	145
24	1	RPS	171
25	2	GL	172
26	1	KKR	172
27	2	SRH	155
28	1	DD	188
29	2	KXIP	137
...	...	...	...
1254	1	KXIP	179
1255	2	SRH	180
1256	1	MI	206
1257	2	DD	126
1258	1	KKR	183
1259	2	RCB	186
1260	1	DD	121
1261	2	RPS	76
1262	1	RCB	211
1263	2	KXIP	120
1264	1	KKR	124
1265	2	GL	125
1266	1	SRH	158
1267	2	DD	161
1268	1	KXIP	172
1269	2	RPS	173
1270	1	MI	172
1271	2	GL	173
1272	1	KKR	171
1273	2	SRH	149
1274	1	DD	138
1275	2	RCB	139
1276	1	GL	158
1277	2	RCB	159
1278	1	SRH	162
1279	2	KKR	140
1280	1	GL	162

1281	inning	batting_team	total_runs
1282	1	SRH	208
1283	2	RCB	200

1284 rows × 3 columns

In [22]:

```
inning1 = runs[runs['inning'] == 1]
inning2 = runs[runs['inning'] == 2]
```

Taking a look at team totals in both the innings using boxplot, to get a better understanding of which team scores higher throughout the tournament.

In [26]:

```
sns.boxplot(x ='batting_team',y = 'total_runs', data = inning1)
```

Out[26]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2ccb2a8b4a8>
```
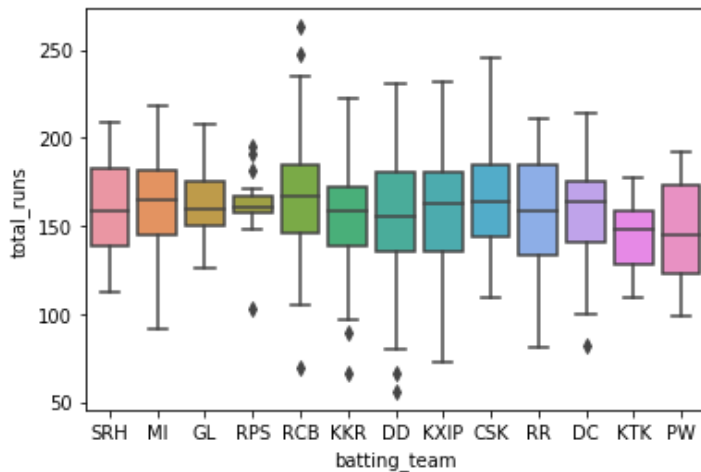

In [27]:

```
sns.boxplot(x ='batting_team',y = 'total_runs', data = inning2)
```

Out[27]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2ccb2c46400>
```
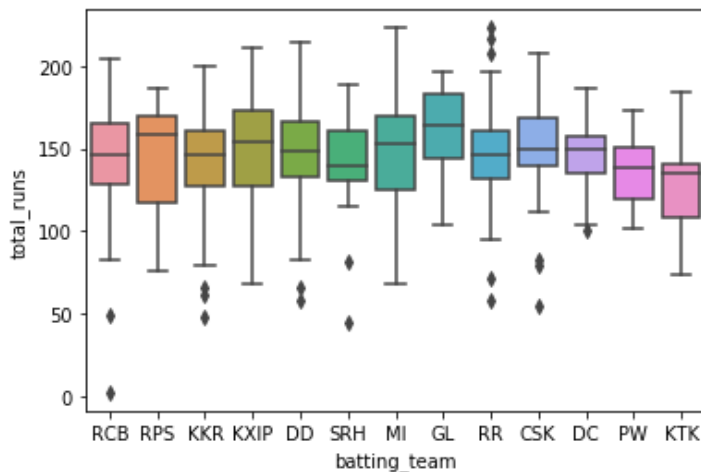

In [28]:

```
high_scores = df.groupby(['match_id','inning','batting_team','bowling_team'])['total_run
s'].sum().reset_index()
```

```
score_200 = high_scores[high_scores['total_runs'] >= 200]
```

```
score_200.head()
```

Out[30]:

	match_id	inning	batting_team	bowling_team	total_runs
0	1	1	SRH	RCB	207
16	9	1	DD	RPS	205
38	20	1	RCB	GL	213
62	32	1	SRH	KXIP	207
72	36	1	SRH	KKR	209

Taking a look at which team scores 200+ as team totals throughout the tournament with the frequencies of each team.

In [31]:

```
sns.countplot(score_200['batting_team'])
```

Out[31]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2ccb27cd7b8>
```
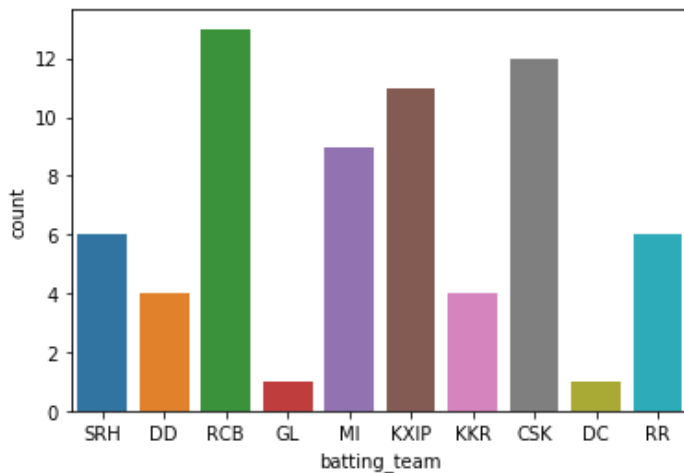

The Bowling teams that conceded 200+ runs in an innings

In [32]:

```
sns.countplot(score_200['bowling_team'])
```

Out[32]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2ccb27b9898>
```
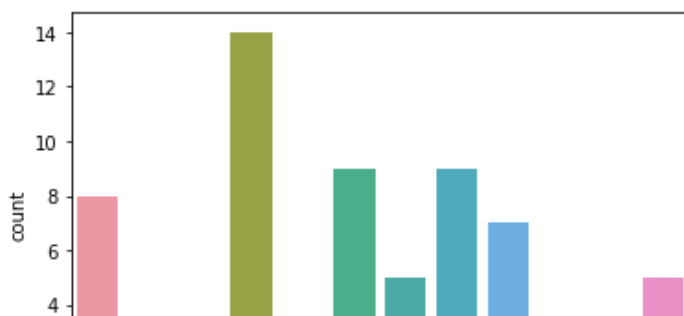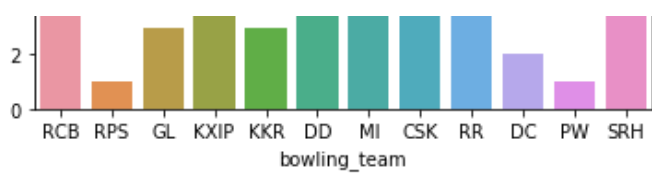
In [33]:

```
balls = df.groupby('batsman')['ball'].count().reset_index()
```

In [34]:

```
runs_total = df.groupby('batsman')['batsman_runs'].sum().reset_index()
runs_total.head(20)
```

Out[34]:

	batsman	batsman_runs
0	A Ashish Reddy	280
1	A Chandila	4
2	A Chopra	53
3	A Choudhary	25
4	A Flintoff	62
5	A Kumble	35
6	A Mishra	340
7	A Mithun	34
8	A Mukund	19
9	A Nehra	41
10	A Singh	2
11	A Symonds	974
12	A Uniyal	4
13	A Zampa	5
14	AA Bilakhia	69
15	AA Chavan	12
16	AA Jhunjhunwala	217
17	AA Noffke	9
18	AB Agarkar	179
19	AB Barath	42

In [35]:

```
fours =df[df['batsman_runs']==4]
fours.head(20)
```

Out[35]:

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs
2	1	1	SRH	RCB	1	3	DA Warner	S Dhawan	TS Mills	0	...	0
8	1	1	SRH	RCB	2	2	DA Warner	S Dhawan	A Choudhary	0	...	0
13	1	1	SRH	RCB	2	7	MC Henriques	S Dhawan	A Choudhary	0	...	0
30	1	1	SRH	RCB	5	5	MC Henriques	S Dhawan	S Aravind	0	...	0

32	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs
33	1	1	SRH	RCB	6	2	S Dhawan	MC Henriques	SR Watson	0	...	0
35	1	1	SRH	RCB	6	4	S Dhawan	MC Henriques	SR Watson	0	...	0
36	1	1	SRH	RCB	6	5	S Dhawan	MC Henriques	SR Watson	0	...	0
63	1	1	SRH	RCB	11	2	S Dhawan	MC Henriques	STR Binny	0	...	0
72	1	1	SRH	RCB	12	5	Yuvraj Singh	MC Henriques	S Aravind	0	...	0
74	1	1	SRH	RCB	13	1	Yuvraj Singh	MC Henriques	A Choudhary	0	...	0
76	1	1	SRH	RCB	13	3	Yuvraj Singh	MC Henriques	A Choudhary	0	...	0
81	1	1	SRH	RCB	14	1	Yuvraj Singh	MC Henriques	SR Watson	0	...	0
87	1	1	SRH	RCB	15	1	MC Henriques	Yuvraj Singh	S Aravind	0	...	0
101	1	1	SRH	RCB	17	3	Yuvraj Singh	DJ Hooda	TS Mills	0	...	0
109	1	1	SRH	RCB	18	5	Yuvraj Singh	DJ Hooda	A Choudhary	0	...	0
114	1	1	SRH	RCB	19	2	Yuvraj Singh	DJ Hooda	TS Mills	0	...	0
129	1	2	RCB	SRH	1	5	Mandeep Singh	CH Gayle	A Nehra	0	...	0
130	1	2	RCB	SRH	1	6	Mandeep Singh	CH Gayle	A Nehra	0	...	0
138	1	2	RCB	SRH	3	2	CH Gayle	Mandeep Singh	A Nehra	0	...	0

20 rows × 21 columns

In [36]:

```
runs_4 = fours.groupby('batsman')['batsman_runs'].count().reset_index()
runs_4.columns = ['batsman','4s']
runs_4
```

Out[36]:

	batsman	4s
0	A Ashish Reddy	16
1	A Chopra	7
2	A Choudhary	1
3	A Flintoff	5
4	A Kumble	3
5	A Mishra	27
6	A Mithun	4
7	A Mukund	1
8	A Nehra	3
9	A Symonds	74
10	A Zampa	1

	batsman	4s
11	AA Bilakhia	5
12	AA Chavan	2
13	AA Jhunjhunwala	19
14	AA Noffke	1
15	AB Agarkar	13
16	AB Barath	5
17	AB Dinda	2
18	AB McDonald	9
19	AB de Villiers	287
20	AC Blizzard	21
21	AC Gilchrist	239
22	AC Thomas	1
23	AC Voges	15
24	AD Mascarenhas	5
25	AD Mathews	44
26	AD Nath	1
27	AD Russell	48
28	AG Murtaza	1
29	AG Paunikar	9
...	...	...
338	TM Dilshan	140
339	TM Head	12
340	TR Birt	9
341	UA Birla	1
342	UBT Chand	32
343	UT Khawaja	14
344	UT Yadav	8
345	Umar Gul	1
346	V Kohli	384
347	V Sehwag	334
348	V Shankar	12
349	VH Zol	3
350	VR Aaron	2
351	VRV Singh	1
352	VS Malik	1
353	VVS Laxman	33
354	Vishnu Vinod	1
355	W Jaffer	14
356	WA Mota	2
357	WD Parnell	4
358	WP Saha	130
359	WPUJC Vaas	2
360	Y Gnaneswara Rao	3
361	Y Nagar	20
362	Y Venugopal Rao	77

	batsman	4s
363	YK Pathan	240
364	YV Takawale	26
365	Yashpal Singh	5
366	Yuvraj Singh	205
367	Z Khan	11

368 rows × 2 columns

In [37]:

```
six =df[df['batsman_runs']==6]
six
```

Out[37]:

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	by
10	1	1	SRH	RCB	2	4	DA Warner	S Dhawan	A Choudhary	0	...	
47	1	1	SRH	RCB	8	4	MC Henriques	S Dhawan	TM Head	0	...	
75	1	1	SRH	RCB	13	2	Yuvraj Singh	MC Henriques	A Choudhary	0	...	
89	1	1	SRH	RCB	15	3	Yuvraj Singh	MC Henriques	S Aravind	0	...	
91	1	1	SRH	RCB	15	5	MC Henriques	Yuvraj Singh	S Aravind	0	...	
105	1	1	SRH	RCB	18	1	DJ Hooda	Yuvraj Singh	A Choudhary	0	...	
115	1	1	SRH	RCB	19	3	Yuvraj Singh	DJ Hooda	TS Mills	0	...	
121	1	1	SRH	RCB	20	3	BCJ Cutting	DJ Hooda	SR Watson	0	...	
124	1	1	SRH	RCB	20	6	BCJ Cutting	DJ Hooda	SR Watson	0	...	
142	1	2	RCB	SRH	3	6	CH Gayle	Mandeep Singh	A Nehra	0	...	
144	1	2	RCB	SRH	4	2	CH Gayle	Mandeep Singh	B Kumar	0	...	
164	1	2	RCB	SRH	7	2	CH Gayle	TM Head	DJ Hooda	0	...	
180	1	2	RCB	SRH	9	6	KM Jadhav	TM Head	BCJ Cutting	0	...	
219	1	2	RCB	SRH	15	6	SR Watson	STR Binny	Rashid Khan	0	...	
221	1	2	RCB	SRH	16	2	STR Binny	SR Watson	BCJ Cutting	0	...	
241	1	2	RCB	SRH	19	4	TS Mills	YS Chahal	B Kumar	0	...	
246	1	2	RCB	SRH	20	3	A Choudhary	YS Chahal	BCJ Cutting	0	...	
270	2	1	MI	RPS	4	4	JC Buttler	PA Patel	BA Stokes	0	...	
271	2	1	MI	RPS	4	5	JC Buttler	PA Patel	BA Stokes	0	...	
283	2	1	MI	RPS	6	5	JC Buttler	RG Sharma	DL Chahar	0	...	
299	2	1	MI	RPS	9	3	N Rana	AT Rayudu	A Zampa	0	...	
323	2	1	MI	RPS	13	2	N Rana	KH Pandya	BA Stokes	0	...	
351	2	1	MI	RPS	17	5	KA Pollard	HH Pandya	Imran Tahir	0	...	
365	2	1	MI	RPS	19	6	TG Southee	HH Pandya	BA Stokes	0	...	

366	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	by
367	2	1	MI	RPS	20	2	HH Pandya	TG Southee	AB Dinda	0	...	
368	2	1	MI	RPS	20	3	HH Pandya	TG Southee	AB Dinda	0	...	
370	2	1	MI	RPS	20	5	HH Pandya	TG Southee	AB Dinda	0	...	
388	2	2	RPS	MI	3	3	AM Rahane	MA Agarwal	TG Southee	0	...	
406	2	2	RPS	MI	6	3	AM Rahane	SPD Smith	MJ McClenaghan	0	...	
...	...	...	...	...	...	...	...	...	...	...	...	
150136	635	2	SRH	GL	8	1	DA Warner	Yuvraj Singh	SK Raina	0	...	
150171	635	2	SRH	GL	13	6	DA Warner	NV Ojha	S Kaushik	0	...	
150180	635	2	SRH	GL	15	3	DA Warner	NV Ojha	DR Smith	0	...	
150183	635	2	SRH	GL	15	6	NV Ojha	DA Warner	DR Smith	0	...	
150192	635	2	SRH	GL	17	3	Bipul Sharma	DA Warner	P Kumar	0	...	
150200	635	2	SRH	GL	18	5	Bipul Sharma	DA Warner	DS Kulkarni	0	...	
150207	635	2	SRH	GL	19	6	Bipul Sharma	DA Warner	DJ Bravo	0	...	
150235	636	1	SRH	RCB	5	1	DA Warner	S Dhawan	SR Watson	0	...	
150237	636	1	SRH	RCB	5	3	S Dhawan	DA Warner	SR Watson	0	...	
150247	636	1	SRH	RCB	6	7	DA Warner	S Dhawan	CH Gayle	0	...	
150256	636	1	SRH	RCB	8	3	DA Warner	MC Henriques	Iqbal Abdulla	0	...	
150280	636	1	SRH	RCB	12	3	Yuvraj Singh	DA Warner	CJ Jordan	0	...	
150298	636	1	SRH	RCB	15	3	Yuvraj Singh	DJ Hooda	YS Chahal	0	...	
150319	636	1	SRH	RCB	18	5	BCJ Cutting	Bipul Sharma	SR Watson	0	...	
150331	636	1	SRH	RCB	20	2	BCJ Cutting	B Kumar	SR Watson	0	...	
150332	636	1	SRH	RCB	20	3	BCJ Cutting	B Kumar	SR Watson	0	...	
150335	636	1	SRH	RCB	20	6	BCJ Cutting	B Kumar	SR Watson	0	...	
150345	636	2	RCB	SRH	2	4	CH Gayle	V Kohli	BB Sran	0	...	
150358	636	2	RCB	SRH	4	2	CH Gayle	V Kohli	BB Sran	0	...	
150362	636	2	RCB	SRH	4	6	CH Gayle	V Kohli	BB Sran	0	...	
150365	636	2	RCB	SRH	5	3	CH Gayle	V Kohli	BCJ Cutting	0	...	
150375	636	2	RCB	SRH	7	1	CH Gayle	V Kohli	MC Henriques	0	...	
150384	636	2	RCB	SRH	8	4	CH Gayle	V Kohli	BCJ Cutting	0	...	
150391	636	2	RCB	SRH	9	4	CH Gayle	V Kohli	MC Henriques	0	...	
150392	636	2	RCB	SRH	9	5	CH Gayle	V Kohli	MC Henriques	0	...	
150395	636	2	RCB	SRH	10	2	V Kohli	CH Gayle	Mustafizur Rahman	0	...	

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	by
150413	636	2	RCB	SRH	13	2	V Kohli	AB de Villiers	BB Sran	0	...	
150428	636	2	RCB	SRH	15	5	SR Watson	KL Rahul	MC Henriques	0	...	
150440	636	2	RCB	SRH	17	5	STR Binny	Sachin Baby	Mustafizur Rahman	0	...	
150453	636	2	RCB	SRH	19	6	Sachin Baby	CJ Jordan	Mustafizur Rahman	0	...	

6523 rows × 21 columns

In [38]:

```
runs_6 = six.groupby('batsman')['batsman_runs'].count().reset_index()
runs_6.columns = ['batsman','6s']
runs_6
```

Out[38]:

	batsman	6s
0	A Ashish Reddy	15
1	A Choudhary	1
2	A Flintoff	2
3	A Mishra	5
4	A Mithun	1
5	A Nehra	1
6	A Symonds	41
7	AA Jhunjhunwala	5
8	AB Agarkar	5
9	AB Barath	1
10	AB McDonald	4
11	AB de Villiers	158
12	AC Blizzard	2
13	AC Gilchrist	92
14	AC Thomas	1
15	AC Voges	3
16	AD Mascarenhas	1
17	AD Mathews	29
18	AD Russell	37
19	AJ Finch	59
20	AJ Tye	3
21	AL Menaria	18
22	AM Nayar	20
23	AM Rahane	60
24	AN Ahmed	1
25	AP Dole	1
26	AP Majumdar	2
27	AP Tare	11
28	AR Patel	28
29	AS Raut	7

	batsman	6s
271	Sunny Singh	1
272	T Henderson	1
273	T Kohli	1
274	TG Southee	3
275	TL Suman	26
276	TM Dilshan	24
277	TM Head	8
278	TR Birt	2
279	TS Mills	1
280	UBT Chand	9
281	UT Khawaja	3
282	UT Yadav	3
283	Umar Gul	5
284	V Kohli	160
285	V Sehwag	106
286	V Shankar	1
287	VH Zol	1
288	VR Aaron	2
289	VVS Laxman	5
290	Vishnu Vinod	1
291	W Jaffer	3
292	WD Parnell	1
293	WP Saha	57
294	WPUJC Vaas	3
295	Y Nagar	9
296	Y Venugopal Rao	37
297	YK Pathan	149
298	YV Takawale	3
299	Yuvraj Singh	141
300	Z Khan	2

301 rows × 2 columns

Now looking at each player, total runs scored, balls played, and boundaries.

In [39]:

```
player = pd.concat([runs_total, balls.iloc[:,1], runs_4.iloc[:,1], runs_6.iloc[:,1]], axis=1)
player
```

Out[39]:

	batsman	batsman_runs	ball	4s	6s
0	A Ashish Reddy	280	196	16.0	15.0
1	A Chandila	4	7	7.0	1.0
2	A Chopra	53	75	1.0	2.0

	batsman	batsman_runs	ball	4s	6s
3	A Choudhary	25	20	5.0	5.0
4	A Flintoff	62	57	3.0	1.0
5	A Kumble	35	49	27.0	1.0
6	A Mishra	340	382	4.0	41.0
7	A Mithun	34	26	1.0	5.0
8	A Mukund	19	23	3.0	5.0
9	A Nehra	41	63	74.0	1.0
10	A Singh	2	10	1.0	4.0
11	A Symonds	974	781	5.0	158.0
12	A Uniyal	4	7	2.0	2.0
13	A Zampa	5	8	19.0	92.0
14	AA Bilakhia	69	88	1.0	1.0
15	AA Chavan	12	11	13.0	3.0
16	AA Jhunjhunwala	217	218	5.0	1.0
17	AA Noffke	9	12	2.0	29.0
18	AB Agarkar	179	160	9.0	37.0
19	AB Barath	42	43	287.0	59.0
20	AB Dinda	26	50	21.0	3.0
21	AB McDonald	123	103	239.0	18.0
22	AB de Villiers	3486	2402	1.0	20.0
23	AC Blizzard	120	91	15.0	60.0
24	AC Gilchrist	2069	1555	5.0	1.0
25	AC Thomas	20	19	44.0	1.0
26	AC Voges	181	143	1.0	2.0
27	AD Mascarenhas	79	79	48.0	11.0
28	AD Mathews	724	599	1.0	28.0
29	AD Nath	20	28	9.0	7.0
...	...	...	...	...	...
431	V Kohli	4423	3494	NaN	NaN
432	V Pratap Singh	0	1	NaN	NaN
433	V Sehwag	2728	1833	NaN	NaN
434	V Shankar	101	77	NaN	NaN
435	VH Zol	29	26	NaN	NaN
436	VR Aaron	46	56	NaN	NaN
437	VRV Singh	4	3	NaN	NaN
438	VS Malik	7	7	NaN	NaN
439	VS Yeligati	2	2	NaN	NaN
440	VVS Laxman	282	271	NaN	NaN
441	VY Mahesh	15	28	NaN	NaN
442	Vishnu Vinod	19	26	NaN	NaN
443	W Jaffer	130	129	NaN	NaN
444	WA Mota	56	76	NaN	NaN
445	WD Parnell	63	80	NaN	NaN
446	WP Saha	1557	1219	NaN	NaN
447	WPUJC Vaas	81	73	NaN	NaN

	batsman	batsman_runs	ball	4s	6s
448	Washington Sundar	9	12	NaN	NaN
449	X Thalaivan Sargunam	10	17	NaN	NaN
450	Y Gnaneswara Rao	19	18	NaN	NaN
451	Y Nagar	285	271	NaN	NaN
452	Y Venugopal Rao	985	865	NaN	NaN
453	YA Abdulla	0	1	NaN	NaN
454	YK Pathan	2922	2076	NaN	NaN
455	YS Chahal	15	29	NaN	NaN
456	YV Takawale	192	183	NaN	NaN
457	Yashpal Singh	47	67	NaN	NaN
458	Younis Khan	3	7	NaN	NaN
459	Yuvraj Singh	2591	2050	NaN	NaN
460	Z Khan	117	141	NaN	NaN

461 rows × 5 columns

Replacing NA values with 0

In [40]:

```
player.fillna(0,inplace = True)
```

In [41]:

```
player
```

Out[41]:

	batsman	batsman_runs	ball	4s	6s
0	A Ashish Reddy	280	196	16.0	15.0
1	A Chandila	4	7	7.0	1.0
2	A Chopra	53	75	1.0	2.0
3	A Choudhary	25	20	5.0	5.0
4	A Flintoff	62	57	3.0	1.0
5	A Kumble	35	49	27.0	1.0
6	A Mishra	340	382	4.0	41.0
7	A Mithun	34	26	1.0	5.0
8	A Mukund	19	23	3.0	5.0
9	A Nehra	41	63	74.0	1.0
10	A Singh	2	10	1.0	4.0
11	A Symonds	974	781	5.0	158.0
12	A Uniyal	4	7	2.0	2.0
13	A Zampa	5	8	19.0	92.0
14	AA Bilakhia	69	88	1.0	1.0
15	AA Chavan	12	11	13.0	3.0
16	AA Jhunjhunwala	217	218	5.0	1.0
17	AA Noffke	9	12	2.0	29.0
18	AB Agarkar	179	160	9.0	37.0
19	AB Barath	42	43	287.0	59.0

	batsman	batsman_runs	ball	4s	6s
20	AB Dinda	26	50	210	39
21	AB McDonald	123	103	239.0	18.0
22	AB de Villiers	3486	2402	1.0	20.0
23	AC Blizzard	120	91	15.0	60.0
24	AC Gilchrist	2069	1555	5.0	1.0
25	AC Thomas	20	19	44.0	1.0
26	AC Voges	181	143	1.0	2.0
27	AD Mascarenhas	79	79	48.0	11.0
28	AD Mathews	724	599	1.0	28.0
29	AD Nath	20	28	9.0	7.0
...	...	...	...	...	...
431	V Kohli	4423	3494	0.0	0.0
432	V Pratap Singh	0	1	0.0	0.0
433	V Sehwag	2728	1833	0.0	0.0
434	V Shankar	101	77	0.0	0.0
435	VH Zol	29	26	0.0	0.0
436	VR Aaron	46	56	0.0	0.0
437	VRV Singh	4	3	0.0	0.0
438	VS Malik	7	7	0.0	0.0
439	VS Yeligati	2	2	0.0	0.0
440	VVS Laxman	282	271	0.0	0.0
441	VY Mahesh	15	28	0.0	0.0
442	Vishnu Vinod	19	26	0.0	0.0
443	W Jaffer	130	129	0.0	0.0
444	WA Mota	56	76	0.0	0.0
445	WD Parnell	63	80	0.0	0.0
446	WP Saha	1557	1219	0.0	0.0
447	WPUJC Vaas	81	73	0.0	0.0
448	Washington Sundar	9	12	0.0	0.0
449	X Thalaivan Sargunam	10	17	0.0	0.0
450	Y Gnaneswara Rao	19	18	0.0	0.0
451	Y Nagar	285	271	0.0	0.0
452	Y Venugopal Rao	985	865	0.0	0.0
453	YA Abdulla	0	1	0.0	0.0
454	YK Pathan	2922	2076	0.0	0.0
455	YS Chahal	15	29	0.0	0.0
456	YV Takawale	192	183	0.0	0.0
457	Yashpal Singh	47	67	0.0	0.0
458	Younis Khan	3	7	0.0	0.0
459	Yuvraj Singh	2591	2050	0.0	0.0
460	Z Khan	117	141	0.0	0.0

461 rows × 5 columns

Adding Batsman Strike Rate to the data

```
player['strike_rate'] = (player['batsman_runs']/player['ball'])*100
```

```
player
```

Out[43]:

	batsman	batsman_runs	ball	4s	6s	strike_rate
0	A Ashish Reddy	280	196	16.0	15.0	142.857143
1	A Chandila	4	7	7.0	1.0	57.142857
2	A Chopra	53	75	1.0	2.0	70.666667
3	A Choudhary	25	20	5.0	5.0	125.000000
4	A Flintoff	62	57	3.0	1.0	108.771930
5	A Kumble	35	49	27.0	1.0	71.428571
6	A Mishra	340	382	4.0	41.0	89.005236
7	A Mithun	34	26	1.0	5.0	130.769231
8	A Mukund	19	23	3.0	5.0	82.608696
9	A Nehra	41	63	74.0	1.0	65.079365
10	A Singh	2	10	1.0	4.0	20.000000
11	A Symonds	974	781	5.0	158.0	124.711908
12	A Uniyal	4	7	2.0	2.0	57.142857
13	A Zampa	5	8	19.0	92.0	62.500000
14	AA Bilakhia	69	88	1.0	1.0	78.409091
15	AA Chavan	12	11	13.0	3.0	109.090909
16	AA Jhunjhunwala	217	218	5.0	1.0	99.541284
17	AA Noffke	9	12	2.0	29.0	75.000000
18	AB Agarkar	179	160	9.0	37.0	111.875000
19	AB Barath	42	43	287.0	59.0	97.674419
20	AB Dinda	26	50	21.0	3.0	52.000000
21	AB McDonald	123	103	239.0	18.0	119.417476
22	AB de Villiers	3486	2402	1.0	20.0	145.129059
23	AC Blizzard	120	91	15.0	60.0	131.868132
24	AC Gilchrist	2069	1555	5.0	1.0	133.054662
25	AC Thomas	20	19	44.0	1.0	105.263158
26	AC Voges	181	143	1.0	2.0	126.573427
27	AD Mascarenhas	79	79	48.0	11.0	100.000000
28	AD Mathews	724	599	1.0	28.0	120.868114
29	AD Nath	20	28	9.0	7.0	71.428571
...	...	...	...	...	...	...
431	V Kohli	4423	3494	0.0	0.0	126.588437
432	V Pratap Singh	0	1	0.0	0.0	0.000000
433	V Sehwag	2728	1833	0.0	0.0	148.827059
434	V Shankar	101	77	0.0	0.0	131.168831
435	VH Zol	29	26	0.0	0.0	111.538462
436	VR Aaron	46	56	0.0	0.0	82.142857

	batsman	batsman_runs	ball	4s	6s	strike_rate
437	VRV Singh	4	3	0.0	0.0	133.333333
438	VS Malik	7	7	0.0	0.0	100.000000
439	VS Yeligati	2	2	0.0	0.0	100.000000
440	VVS Laxman	282	271	0.0	0.0	104.059041
441	VY Mahesh	15	28	0.0	0.0	53.571429
442	Vishnu Vinod	19	26	0.0	0.0	73.076923
443	W Jaffer	130	129	0.0	0.0	100.775194
444	WA Mota	56	76	0.0	0.0	73.684211
445	WD Parnell	63	80	0.0	0.0	78.750000
446	WP Saha	1557	1219	0.0	0.0	127.727646
447	WPUJC Vaas	81	73	0.0	0.0	110.958904
448	Washington Sundar	9	12	0.0	0.0	75.000000
449	X Thalaivan Sargunam	10	17	0.0	0.0	58.823529
450	Y Gnaneswara Rao	19	18	0.0	0.0	105.555556
451	Y Nagar	285	271	0.0	0.0	105.166052
452	Y Venugopal Rao	985	865	0.0	0.0	113.872832
453	YA Abdulla	0	1	0.0	0.0	0.000000
454	YK Pathan	2922	2076	0.0	0.0	140.751445
455	YS Chahal	15	29	0.0	0.0	51.724138
456	YV Takawale	192	183	0.0	0.0	104.918033
457	Yashpal Singh	47	67	0.0	0.0	70.149254
458	Younis Khan	3	7	0.0	0.0	42.857143
459	Yuvraj Singh	2591	2050	0.0	0.0	126.390244
460	Z Khan	117	141	0.0	0.0	82.978723

461 rows × 6 columns

Thank you

In [ ]: