Merge_Function (Merge two table)

In [72]:

```
import pandas as pd
```

In [73]:

```
delivery=pd.read_csv('deliveries.csv')
```

In [74]:

```
match=pd.read_csv('matches.csv')
```

In [75]:

```
new=delivery.merge(match,left_on='match_id',right_on='id')
```

In [76]:

```
print(delivery.shape)
print(match.shape)
```

```
(150460, 21)
(636, 18)
```

In [77]:

```
new.groupby('batsman')['batsman_runs'].sum()
```

Out[77]:

```
batsman
A Ashish Reddy     280
A Chandila           4
A Chopra            53
A Choudhary         25
A Flintoff          62
                  ... 
YV Takawale        192
Yashpal Singh       47
Younis Khan          3
Yuvraj Singh      2591
Z Khan             117
Name: batsman_runs, Length: 461, dtype: int64
```

How many runs make Batsman by particular season(Multi-Index Series(season,batsman))

In [78]:

```
new.groupby(['season','batsman'])['batsman_runs'].sum().reset_index()
```

Out[78]:

	season	batsman	batsman_runs
0	2008	A Chopra	42
1	2008	A Kumble	13
2	2008	A Mishra	37
3	2008	A Mukund	0
4	2008	A Nehra	3
...	...	...	...
1526	2017	Washington Sundar	9
1527	2017	YK Pathan	143
1528	2017	YS Chahal	13
1529	2017	Yuvraj Singh	252
1530	2017	Z Khan	4

1531 rows × 3 columns

In [79]:

```
new.groupby(['season','batsman'])['batsman_runs'].sum().sort_values(ascending=Fals
```

Out[79]:

	season	batsman	batsman_runs
0	2016	V Kohli	973
1	2016	DA Warner	848
2	2013	MEK Hussey	733
3	2012	CH Gayle	733
4	2013	CH Gayle	720
...	...	...	...
1526	2017	MM Patel	0
1527	2012	AC Blizzard	0
1528	2011	AB Dinda	0
1529	2017	AD Nath	0
1530	2008	L Balaji	0

1531 rows × 3 columns

Find only top scorer/top Batsman in every year

In [80]:

```
new.groupby(['season','batsman'])['batsman_runs'].sum().sort_values(ascending=Fals
```

Out[80]:

	season	batsman
10	2008	SE Marsh
14	2009	ML Hayden
9	2010	SR Tendulkar
11	2011	CH Gayle
3	2012	CH Gayle
2	2013	MEK Hussey
6	2014	RV Uthappa
17	2015	DA Warner
0	2016	V Kohli
7	2017	DA Warner

In [81]:

```
mask=delivery['batsman_runs']==6
six=delivery[mask]
#six.head()
six.shape
```

Out[81]:

```
(6523, 21)
```

In [82]:

```
pt=six.pivot_table(index='over',columns='batting_team',values='batsman_runs',aggfu
```

In [83]:

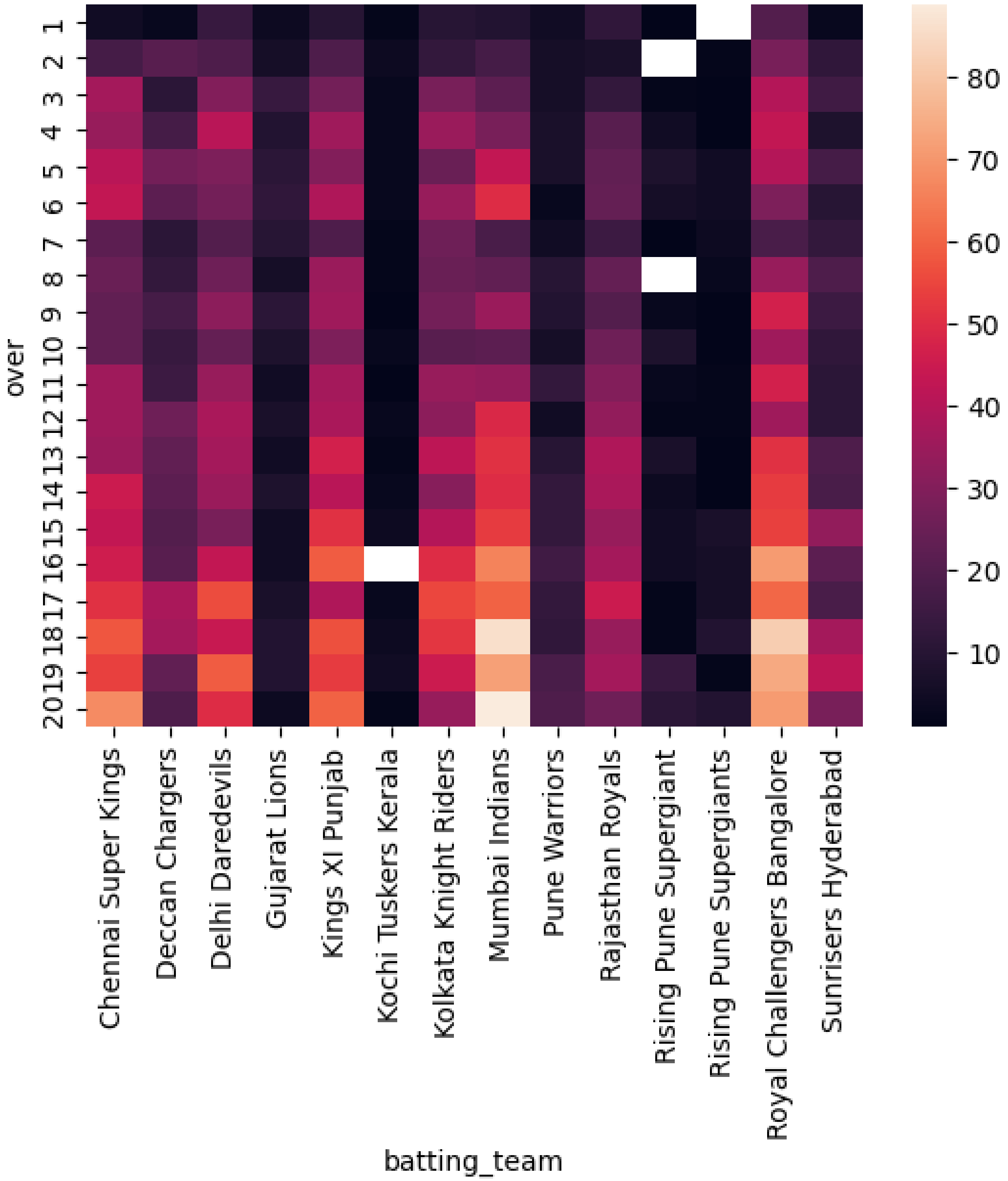
```
import seaborn as sns
```

In [84]:

```
sns.heatmap(pt)
```

Out[84]:

```
<Axes: xlabel='batting_team', ylabel='over'>
```


over
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Chennai Super Kings
Deccan Chargers
Delhi Daredevils
Gujarat Lions
Kings XI Punjab
Kochi Tuskers Kerala
Kolkata Knight Riders
Mumbai Indians
Pune Warriors
Rajasthan Royals
Rising Pune Supergiant
Rising Pune Supergiants
Royal Challengers Bangalore
Sunrisers Hyderabad
batting_team
80
70
60
50
40
30
20
10

In [85]:
```
match
```

Out[85]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal
...	...	...	...	...	...	...	...	...	...
631	632	2016	Raipur	2016-05-22	Delhi Daredevils	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal
632	633	2016	Bangalore	2016-05-24	Gujarat Lions	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal
633	634	2016	Delhi	2016-05-25	Sunrisers Hyderabad	Kolkata Knight Riders	Kolkata Knight Riders	field	normal
634	635	2016	Delhi	2016-05-27	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal
635	636	2016	Bangalore	2016-05-29	Sunrisers Hyderabad	Royal Challengers Bangalore	Sunrisers Hyderabad	bat	normal

636 rows × 18 columns

Co-relation

In [86]:

```
match.corr()
```

```
C:\Users\SHANTANU GARAIN\AppData\Local\Temp\ipykernel_19416\2923588360.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
  match.corr()
```

Out[86]:

	id	season	dl_applied	win_by_runs	win_by_wickets	umpire3
id	1.000000	0.471087	0.024281	-0.010263	-0.015510	NaN
season	0.471087	1.000000	0.004170	-0.016815	-0.000708	NaN
dl_applied	0.024281	0.004170	1.000000	-0.010893	-0.011640	NaN
win_by_runs	-0.010263	-0.016815	-0.010893	1.000000	-0.565181	NaN
win_by_wickets	-0.015510	-0.000708	-0.011640	-0.565181	1.000000	NaN
umpire3	NaN	NaN	NaN	NaN	NaN	NaN

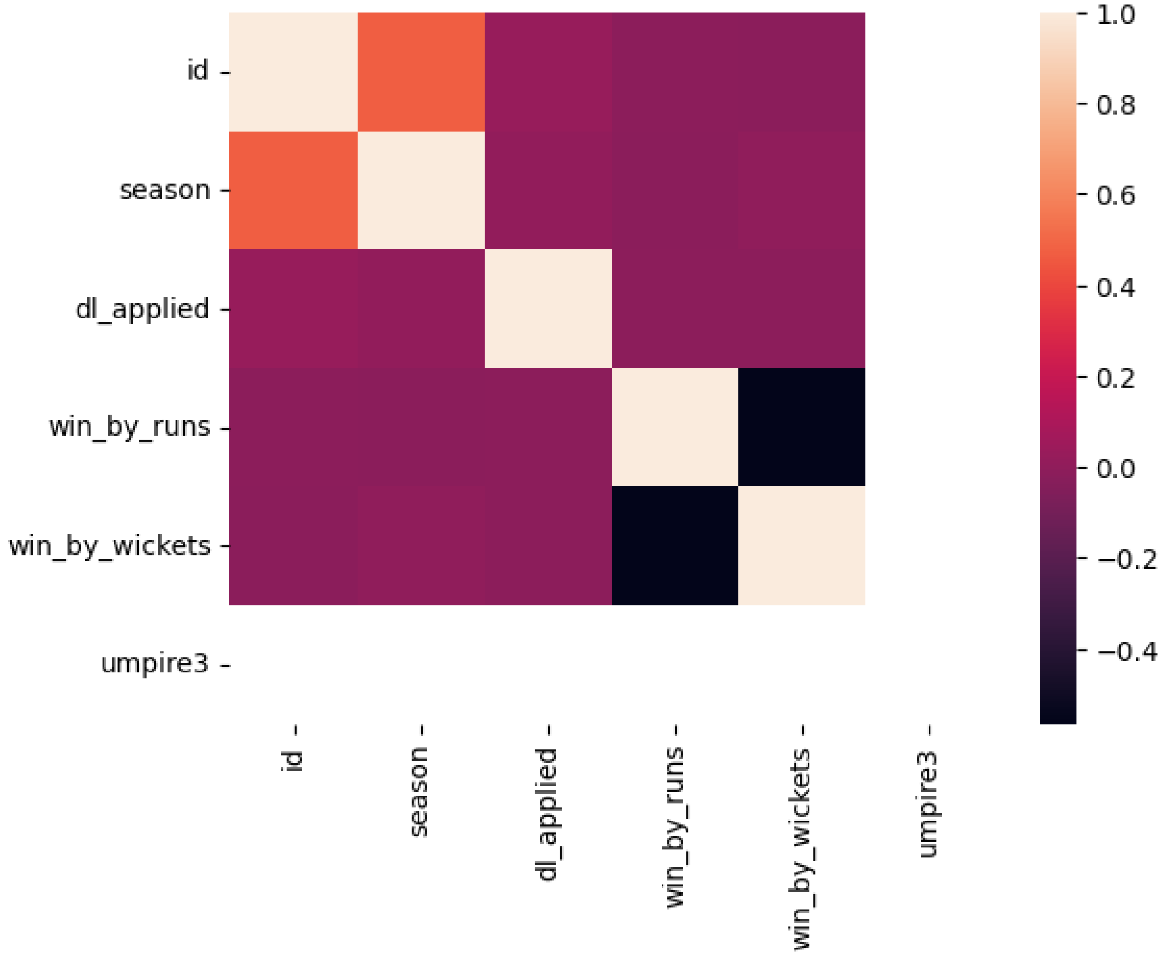
In [87]:

```
sns.heatmap(match.corr())
```

```
C:\Users\SHANTANU GARAIN\AppData\Local\Temp\ipykernel_19416\2946606924.py:1: Futur
eWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a
future version, it will default to False. Select only valid columns or specify the
value of numeric_only to silence this warning.
  sns.heatmap(match.corr())
```

Out[87]:

```
<Axes: >
```



Rename Cols

In [88]:

```
match.rename(columns={'city':'place','date':'dmom'})
```

Out[88]:

	id	season	place	dmom	team1	team2	toss_winner	toss_decision	result
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal
...	...	...	...	...	...	...	...	...	...
631	632	2016	Raipur	2016-05-22	Delhi Daredevils	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal
632	633	2016	Bangalore	2016-05-24	Gujarat Lions	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal
633	634	2016	Delhi	2016-05-25	Sunrisers Hyderabad	Kolkata Knight Riders	Kolkata Knight Riders	field	normal
634	635	2016	Delhi	2016-05-27	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal
635	636	2016	Bangalore	2016-05-29	Sunrisers Hyderabad	Royal Challengers Bangalore	Sunrisers Hyderabad	bat	normal

636 rows × 18 columns

In [89]:

```
match
```

Out[89]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal
...	...	...	...	...	...	...	...	...	...
631	632	2016	Raipur	2016-05-22	Delhi Daredevils	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal
632	633	2016	Bangalore	2016-05-24	Gujarat Lions	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal
633	634	2016	Delhi	2016-05-25	Sunrisers Hyderabad	Kolkata Knight Riders	Kolkata Knight Riders	field	normal
634	635	2016	Delhi	2016-05-27	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal
635	636	2016	Bangalore	2016-05-29	Sunrisers Hyderabad	Royal Challengers Bangalore	Sunrisers Hyderabad	bat	normal

636 rows × 18 columns

Set_index() and reset_index()

In [90]:

```
match.head()
```

Out[90]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_a
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	

In [91]:
```python
match.set_index('id', inplace=True)
```

In [92]:
```python
match['winner'].value_counts().reset_index()
```

Out[92]:

	index	winner
0	Mumbai Indians	92
1	Chennai Super Kings	79
2	Kolkata Knight Riders	77
3	Royal Challengers Bangalore	73
4	Kings XI Punjab	70
5	Rajasthan Royals	63
6	Delhi Daredevils	62
7	Sunrisers Hyderabad	42
8	Deccan Chargers	29
9	Gujarat Lions	13
10	Pune Warriors	12
11	Rising Pune Supergiant	10
12	Kochi Tuskers Kerala	6
13	Rising Pune Supergiants	5

match['winner'].value_counts().reset_index()

In [93]:
```python
match['winner'].value_counts()
```

Out[93]:

```
Mumbai Indians                 92
Chennai Super Kings            79
Kolkata Knight Riders          77
Royal Challengers Bangalore    73
Kings XI Punjab                70
Rajasthan Royals               63
Delhi Daredevils               62
Sunrisers Hyderabad            42
Deccan Chargers                29
Gujarat Lions                  13
Pune Warriors                  12
Rising Pune Supergiant         10
Kochi Tuskers Kerala            6
Rising Pune Supergiants         5
Name: winner, dtype: int64
```

In [94]:

```
delivery.shape
```

Out[94]:

```
(150460, 21)
```

In [95]:

```
match.shape
```

Out[95]:

```
(636, 17)
```

1. Dropping missing data using dropna()

In [96]:

```
match.dropna(axis=1, how='all').shape
```

Out[96]:

```
(636, 16)
```

In [97]:

```
match.dropna(subset=['umpire1']).shape
```

Out[97]:

```
(635, 17)
```

In [98]:

```
match.dropna(subset=['umpire1','umpire3']).shape
```

Out[98]:

```
(0, 17)
```

2. Filling missing values using fillna()

In [99]:

```
#atch['umpire1'].fillna("Not Specified")
```

In [104…

```
match['umpire1'].fillna(method='ffill')
```

Out[104]:

```
id
1           AY Dandekar
2        A Nand Kishore
3           Nitin Menon
4          AK Chaudhary
5          AK Chaudhary
              ...
632      A Nand Kishore
633        AK Chaudhary
634           M Erasmus
635           M Erasmus
636     HDPK Dharmasena
Name: umpire1, Length: 636, dtype: object
```

```
In [ ]:
```