phython project

In [2]:
```
import pandas as pd
import seaborn as sns
```

In [9]:
```
tips=sns.load_dataset('tips')
```

Lets read the data using read csv

In [12]:
```
df=pd.read_csv(r"C:\Users\Deepika\Desktop\Deepi Nov\phython project\tips.csv")
df
```

Out[12]:

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4
...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

244 rows × 7 columns

First few rows to understand the dataset

In [5]:
```
tips.head()
```

Out[5]:

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

How many weekdays are present in the dataset

In [6]:
```
tips.day.unique()
```

Out[6]:
```
['Sun', 'Sat', 'Thur', 'Fri']
Categories (4, object): ['Thur', 'Fri', 'Sat', 'Sun']
```

Look at the last rows

In [7]:
```
tips.tail()
```

Out[7]:

	total_bill	tip	sex	smoker	day	time	size

239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

How many dinners and lunch are present in the dataset?

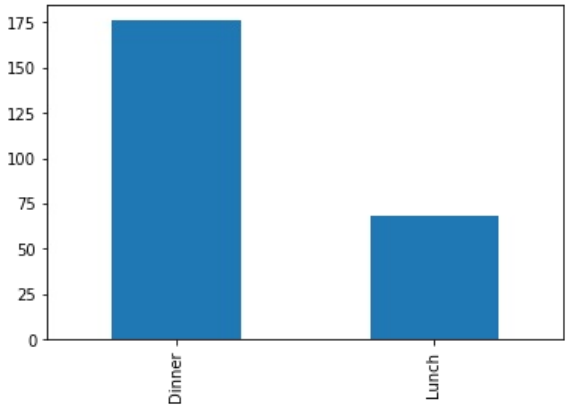
In [8]:
```
tips.time.value_counts()
```

Out[8]:
```
Dinner    176
Lunch      68
Name: time, dtype: int64
```

Let Plot the above values using matplotlib

In [15]:
```
tips.time.value_counts().plot(kind="bar")
```

Out[15]:
```
<AxesSubplot:>
```



what is the proportion of lunch and dinner in percentage?

In [16]:
```
tips.time.value_counts(normalize=True)
```

Out[16]:
```
Dinner    0.721311
Lunch     0.278689
Name: time, dtype: float64
```

when does people pay more tips (Ie:lunch or dinner)

In [18]:
```
tips.head()
```

Out[18]:

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

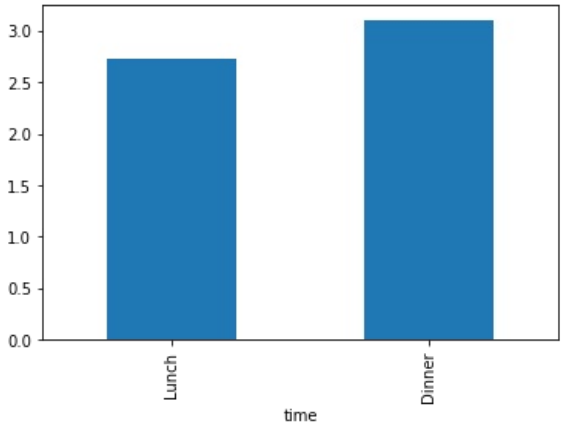
In [19]:
```
tips.groupby(['time'])['tip'].mean()
```

Out[19]:
```
time
Lunch     2.728088
Dinner    3.102670
Name: tip, dtype: float64
```

Let plot these values

In [20]:
```
tips.groupby(['time'])['tip'].mean().plot.bar()
```

Out[20]:
```
<AxesSubplot:xlabel='time'>
```



Calculate average tips value based on gender:

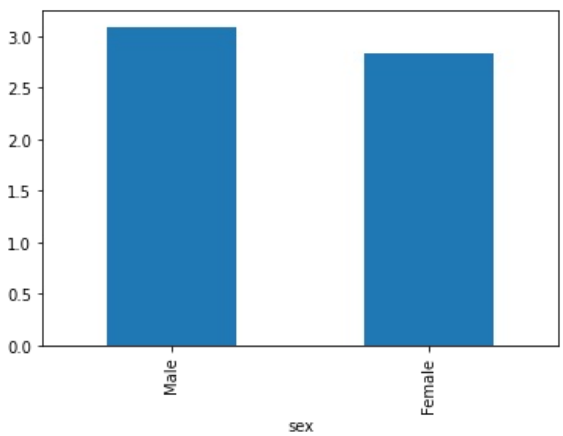
In [22]:
```
tips.groupby(['sex'])['tip'].mean()
```

Out[22]:
```
sex
Male      3.089618
Female    2.833448
Name: tip, dtype: float64
```

Lets plot the values

In [23]:
```
tips.groupby(['sex'])['tip'].mean().plot.bar()
```

Out[23]:
```
<AxesSubplot:xlabel='sex'>
```



Let's create a cross table for smoker and tips for average tips

Let's create a cross table for smoker and tips for average tips

In [25]:
```
tips.groupby(['smoker','sex'])['tip'].mean()
```

Out[25]:
```
smoker  sex
Yes     Male      3.051167
        Female    2.931515
No      Male      3.113402
        Female    2.773519
Name: tip, dtype: float64
```

pivot structure of cross table

In [26]:
```
tips.groupby(['smoker','sex'])['tip'].mean().unstack()
```

Out[26]:

sex	Male	Female
smoker		
Yes	3.051167	2.931515
No	3.113402	2.773519

Create a cross table for smoker and day for average tips

In [28]:
```
tips.groupby(['smoker','sex','day'])['tip'].mean()
```

Out[28]:
```
smoker  sex     day
Yes     Male    Thur    3.058000
                Fri     2.741250
                Sat     2.879259
                Sun     3.521333
        Female  Thur    2.990000
                Fri     2.682857
                Sat     2.868667
                Sun     3.500000
No      Male    Thur    2.941500
                Fri     2.500000
                Sat     3.256563
                Sun     3.115349
        Female  Thur    2.459600
                Fri     3.125000
                Sat     2.724615
                Sun     3.329286
Name: tip, dtype: float64
```

Pivot structure

In [29]:
```
tips.groupby(['smoker','sex','day'])['tip'].mean().unstack()
```

Out[29]:

day	Thur	Fri	Sat	Sun
smoker	**sex**			
Yes	**Male**	3.0580	2.741250	2.879259	3.521333
	Female	2.9900	2.682857	2.868667	3.500000
No	**Male**	2.9415	2.500000	3.256563	3.115349
	Female	2.4596	3.125000	2.724615	3.329286

Let's create a new column to find out the % of tip compare to total bill

In [30]:
```
tips['tips_percentage']=tips.tip/tips.total_bill
```

```
tips['tips_percentage']=tips.tip/tips.total_bill
```

let's look at the newly created column

In [31]:
```
tips.head()
```

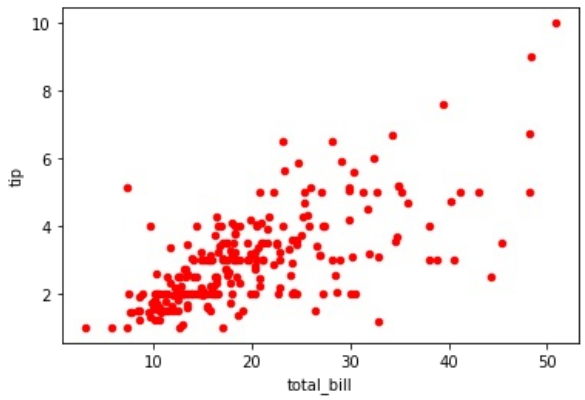
Out[31]:

	total_bill	tip	sex	smoker	day	time	size	tips_percentage
0	16.99	1.01	Female	No	Sun	Dinner	2	0.059447
1	10.34	1.66	Male	No	Sun	Dinner	3	0.160542
2	21.01	3.50	Male	No	Sun	Dinner	3	0.166587
3	23.68	3.31	Male	No	Sun	Dinner	2	0.139780
4	24.59	3.61	Female	No	Sun	Dinner	4	0.146808

let's visualize the relationship between total_bill and tips

In [37]:
```
tips.plot.scatter(x='total_bill',y='tip',color='red')
```

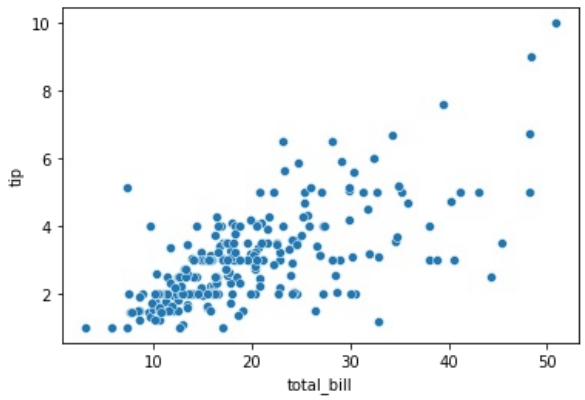
Out[37]:
```
<AxesSubplot:xlabel='total_bill', ylabel='tip'>
```



Visualize seaborn library

In [40]:
```
sns.scatterplot(x='total_bill',y='tip',data=tips)
```

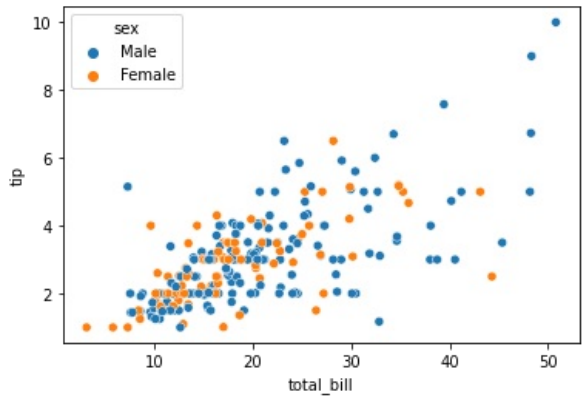
Out[40]:
```
<AxesSubplot:xlabel='total_bill', ylabel='tip'>
```



Seaborn visualization for gender tips

In [41]:
```
sns.scatterplot(x='total_bill',y='tip',data=tips,hue='sex')
```

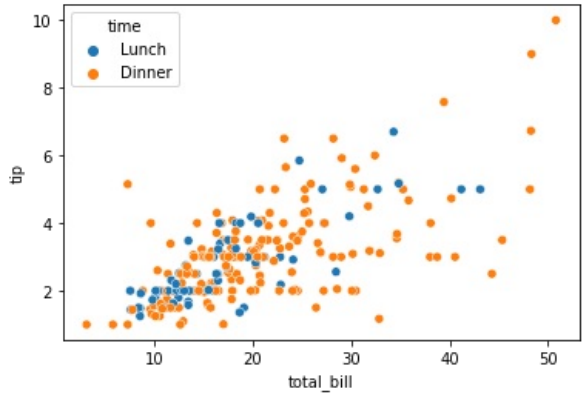
Out[41]: <AxesSubplot:xlabel='total_bill', ylabel='tip'>



Similarly let's see the visualization for time and tips

In [42]:
```
sns.scatterplot(x='total_bill',y='tip',data=tips,hue='time')
```

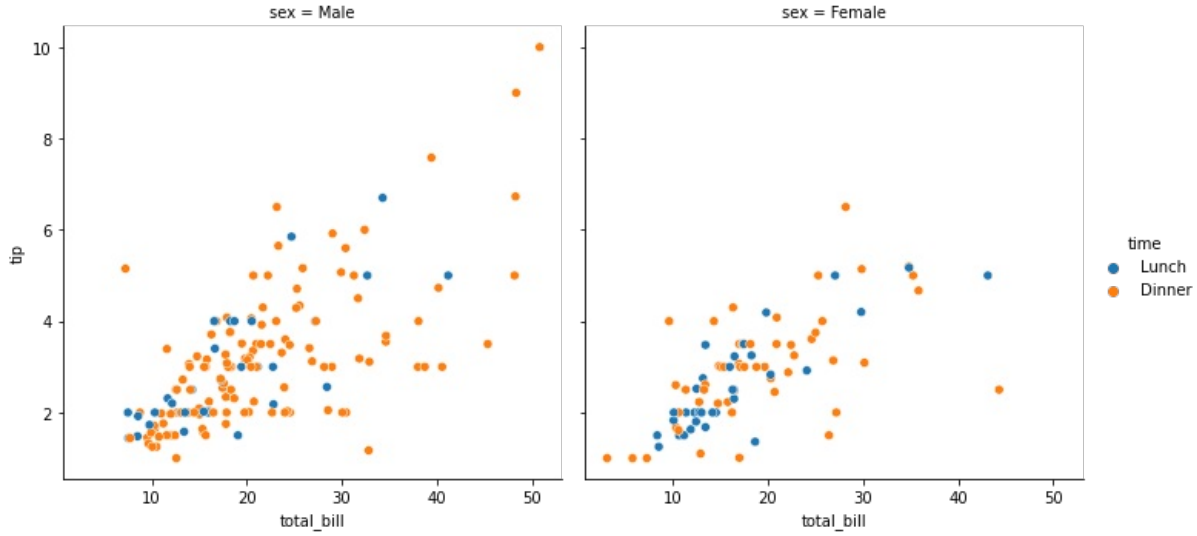
Out[42]: <AxesSubplot:xlabel='total_bill', ylabel='tip'>



Create the separate chart for gender based on the tips

In [44]:
```
sns.relplot(
    x='total_bill',y='tip',col='sex',data=tips,hue='time')
```

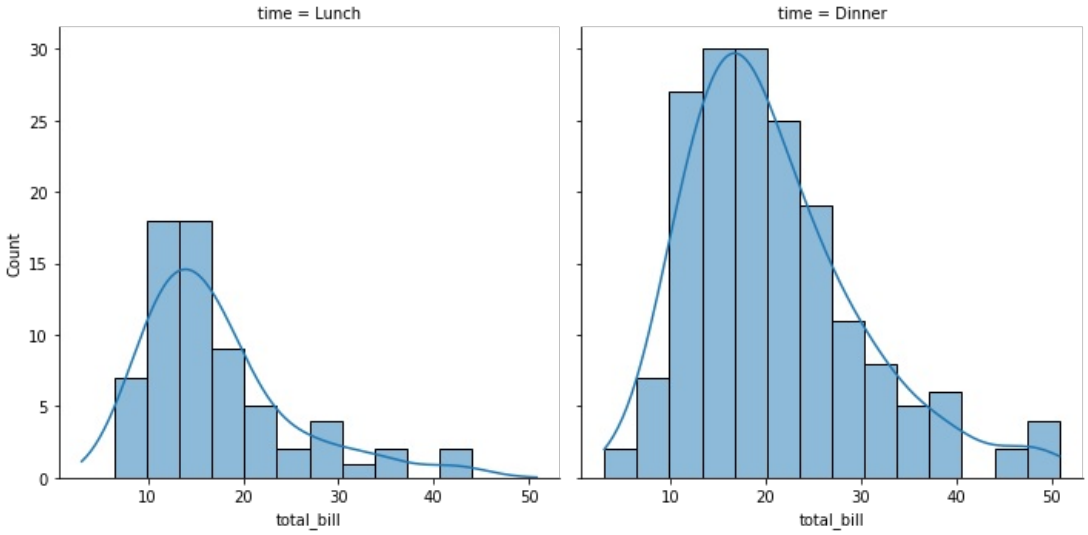
Out[44]: <seaborn.axisgrid.FacetGrid at 0x26818724700>



let's look at the distribution of total bill by time

In [45]:
```
sns.displot(
data=tips,x='total_bill',col='time',kde=True)
```

Out[45]:
```
<seaborn.axisgrid.FacetGrid at 0x2681871b130>
```

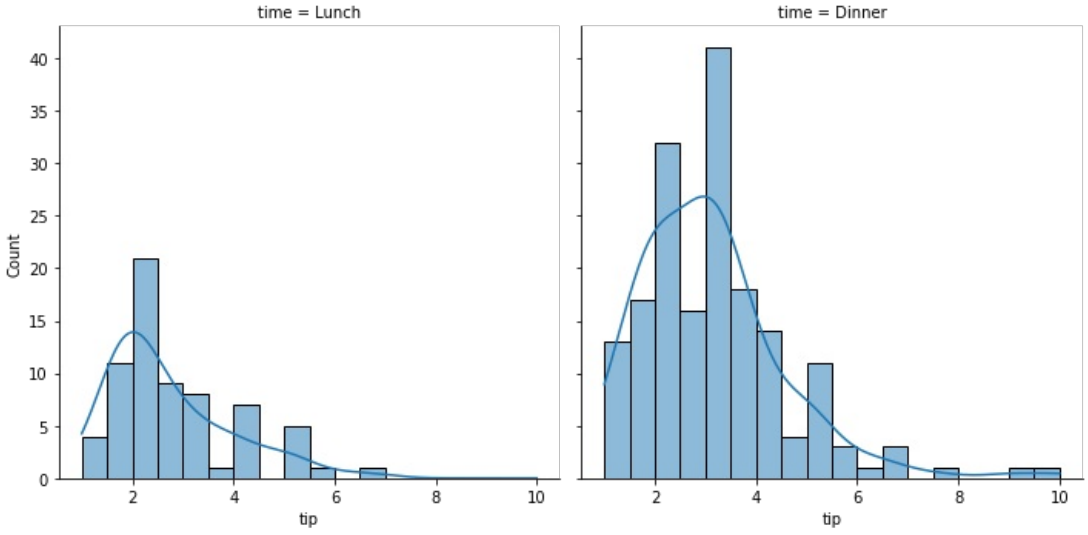


In [46]:
```
#clearly people pay more total bills during dinner
```

Distribution of tip by time

In [47]:
```
sns.displot(
data=tips,x='tip',col='time',kde=True)
```

Out[47]:
```
<seaborn.axisgrid.FacetGrid at 0x26816986850>
```



In [ ]: