Name – Shradharghya Shankar Adhikary

College – Indian Institute of Technology, Madras (600036)

Branch – Applied Mechanics (M.S.)

Year – 3rd

About Dataset – A dataset consist of 244 rows and 8 column was taken. I tried to perform EDA on different columns (example –total_bill, tip, sex, smoker, day, time, size, bill_group) and studied their effect on each other.

In [1]:
```
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```
df=pd.read_excel('https://drive.google.com/uc?export=download&id=1I8NBVu-c_Jfm8hx5An5uCnWu-5pFq3zP')
```

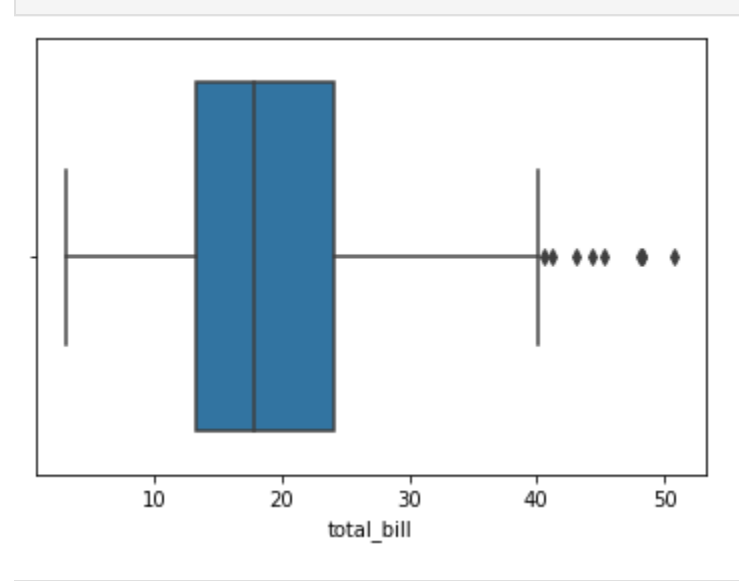
In [3]:
```
df
```

Out[3]:

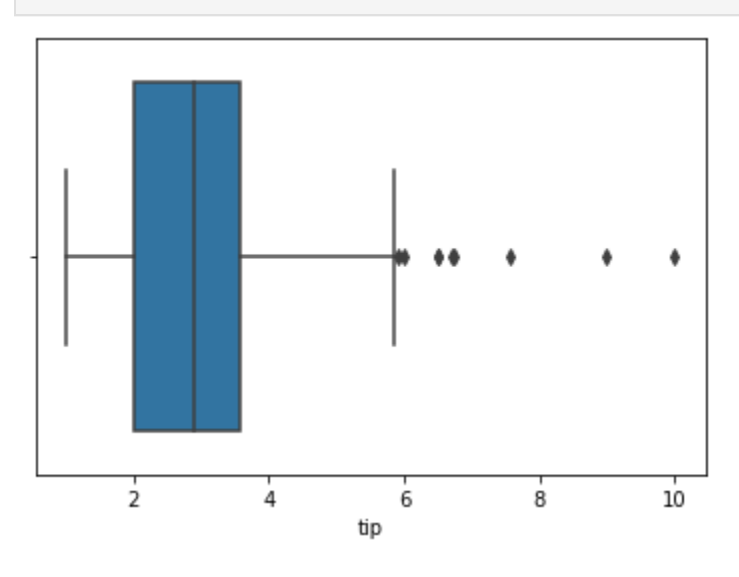
	total_bill	tip	sex	smoker	day	time	size	bill_group
0	16.99	1.01	Female	No	Sun	Dinner	2	(12.618, 17.392]
1	10.34	1.66	Male	No	Sun	Dinner	3	(7.844, 12.618]
2	21.01	3.50	Male	No	Sun	Dinner	3	(17.392, 22.166]
3	23.68	3.31	Male	No	Sun	Dinner	2	(22.166, 26.94]
4	24.59	3.61	Female	No	Sun	Dinner	4	(22.166, 26.94]
...	...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3	(26.94, 31.714]
240	27.18	2.00	Female	Yes	Sat	Dinner	2	(26.94, 31.714]
241	22.67	2.00	Male	Yes	Sat	Dinner	2	(22.166, 26.94]
242	17.82	1.75	Male	No	Sat	Dinner	2	(17.392, 22.166]
243	18.78	3.00	Female	No	Thur	Dinner	2	(17.392, 22.166]

244 rows × 8 columns

In [4]:
```
sns.boxplot(df.total_bill)
plt.show()
```



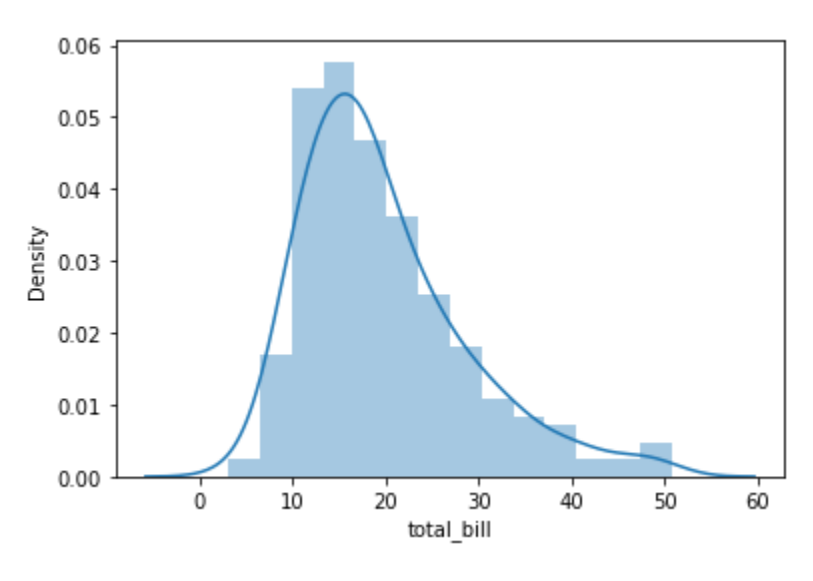
In [5]:
```
sns.boxplot(df.tip)
plt.show()
```



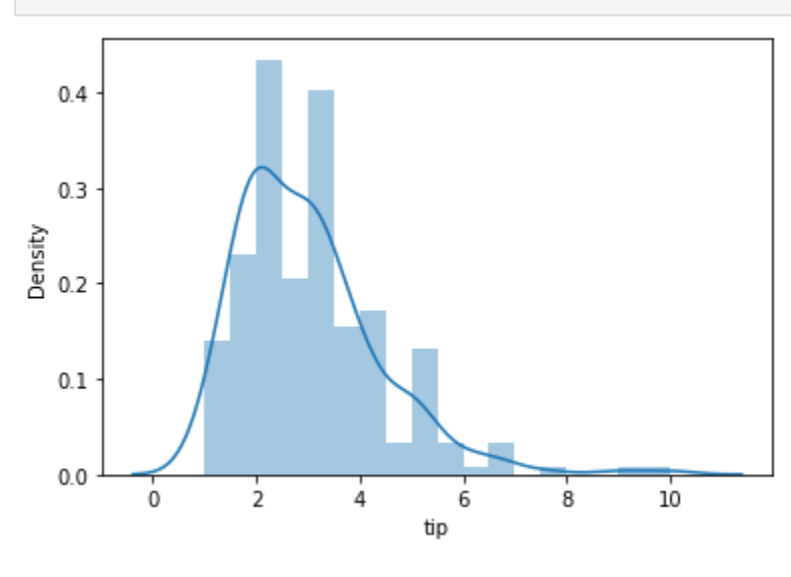
Conclusion1: From the above Box Plot we can see that there are some outlier's in the data

In [6]:
```
sns.distplot(df.total_bill)
```

Out[6]:
```
<AxesSubplot:xlabel='total_bill', ylabel='Density'>
```



In [7]:
```
sns.distplot(df.tip)
plt.show()
```



Conclusion2: From the above distribution plot we conclude that our data is following Normal Distribution

In [8]:
```
df
```

Out[8]:

	total_bill	tip	sex	smoker	day	time	size	bill_group
0	16.99	1.01	Female	No	Sun	Dinner	2	(12.618, 17.392]
1	10.34	1.66	Male	No	Sun	Dinner	3	(7.844, 12.618]
2	21.01	3.50	Male	No	Sun	Dinner	3	(17.392, 22.166]
3	23.68	3.31	Male	No	Sun	Dinner	2	(22.166, 26.94]
4	24.59	3.61	Female	No	Sun	Dinner	4	(22.166, 26.94]
...	...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3	(26.94, 31.714]
240	27.18	2.00	Female	Yes	Sat	Dinner	2	(26.94, 31.714]
241	22.67	2.00	Male	Yes	Sat	Dinner	2	(22.166, 26.94]
242	17.82	1.75	Male	No	Sat	Dinner	2	(17.392, 22.166]
243	18.78	3.00	Female	No	Thur	Dinner	2	(17.392, 22.166]

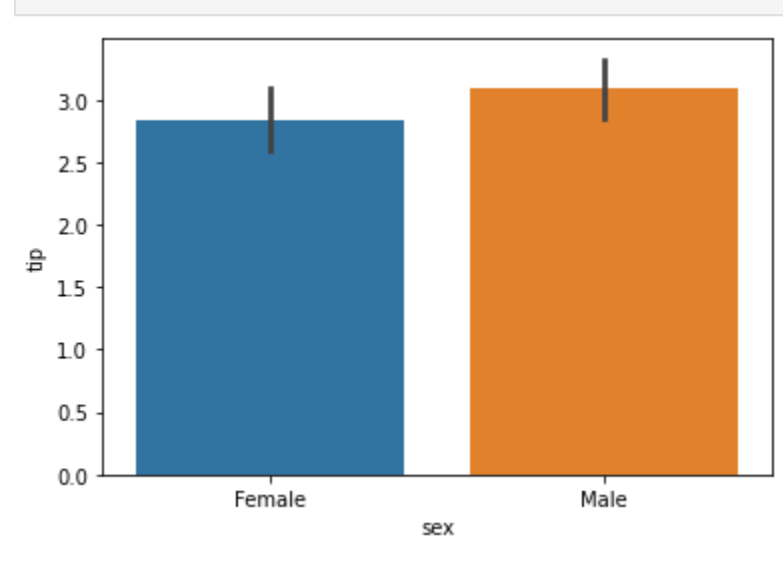
244 rows × 8 columns

In [9]:
```
df[['tip','sex']].groupby('sex').mean()
```

Out[9]:

sex	tip
Female	2.833448
Male	3.089618

In [10]:
```
sns.barplot(df['sex'],df.tip)
plt.show()
```



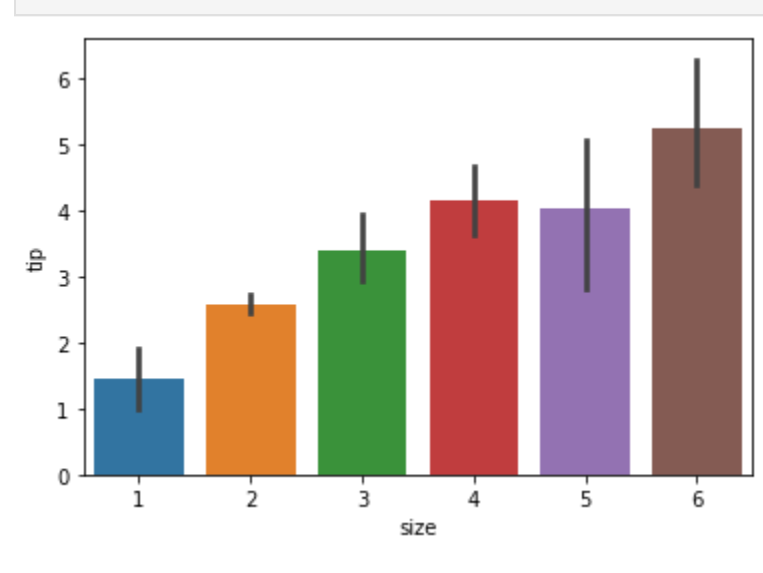
Conclusion3: From the above bar graph we can conclude male are giving more tip than female

In [11]:
```
df[['tip','size']].groupby('size').mean()
```

Out[11]:

size	tip
1	1.437500
2	2.582308
3	3.393158
4	4.135405
5	4.028000
6	5.225000

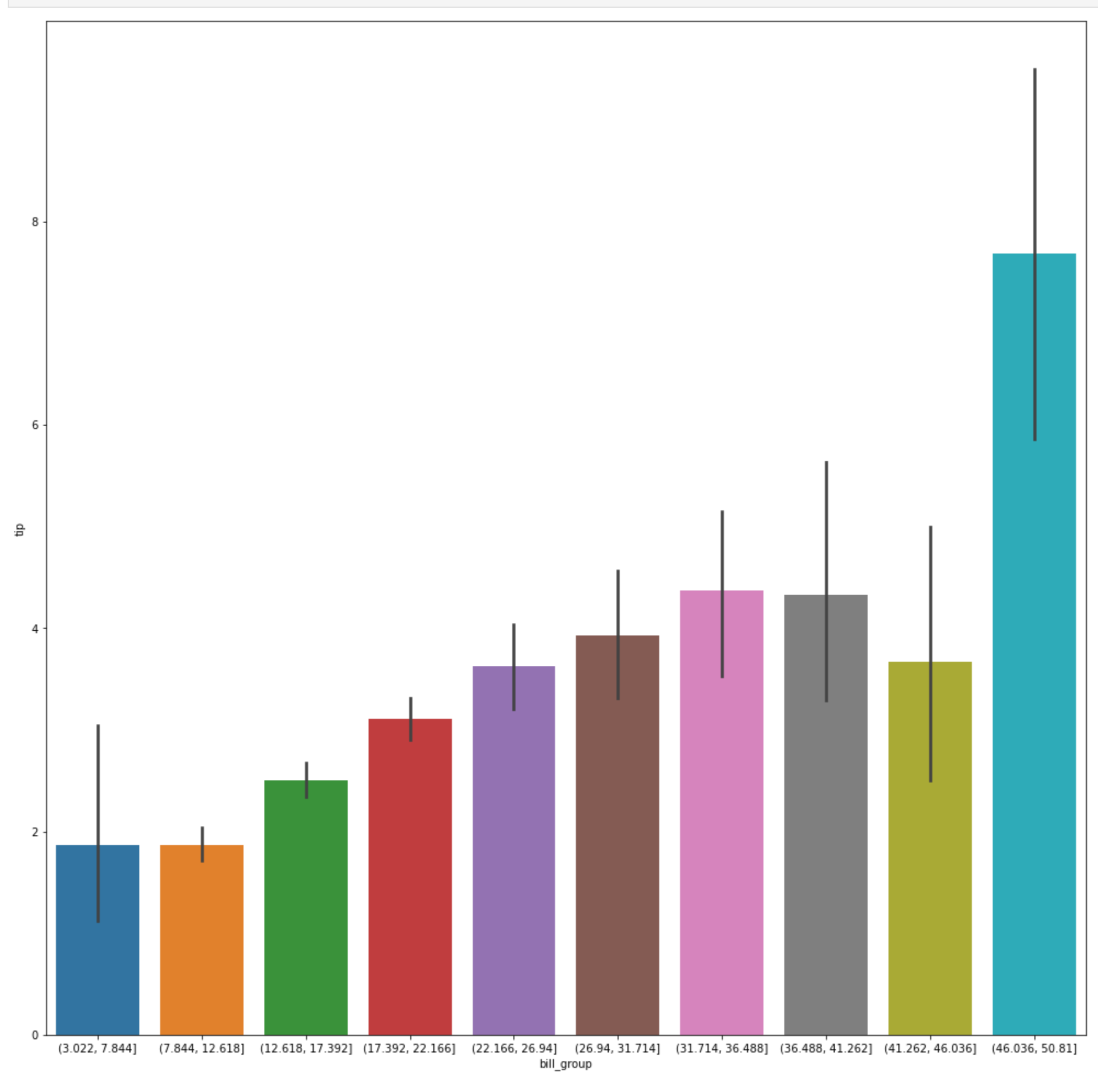
In [12]:
```
sns.barplot(df['size'],df.tip)
plt.show()
```



Conclusion4: From the above bar graph we can conclude that, more number of people in a group gives more tip

In [13]:
```
df['bill_group']=pd.cut(df['total_bill'],10)
```

In [14]:
```
plt.figure(figsize=(17,17))
sns.barplot(df['bill_group'],df.tip)
plt.show()
```



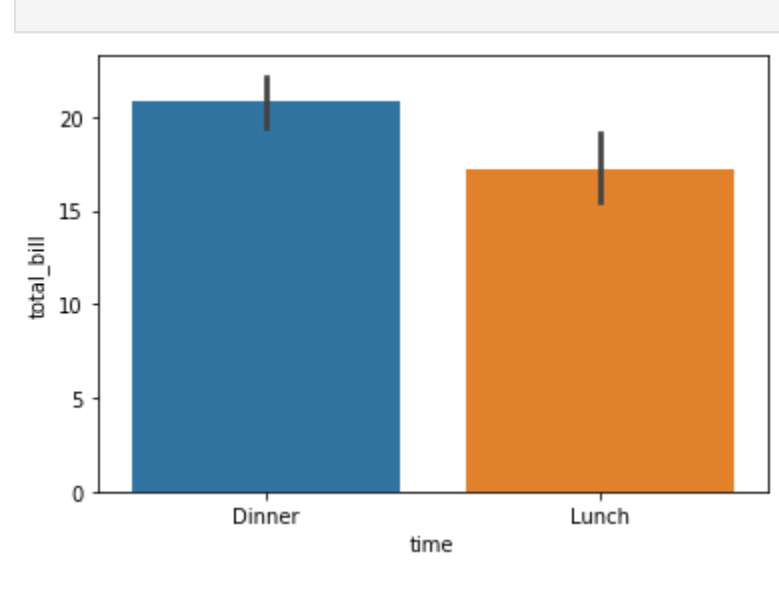
Conclusion5: From the above bar graph we conclude that, with increasing bill amount, tip is also increasing

In [15]:
```
df[['total_bill','time']].groupby('time').mean()
```

Out[15]:

time	total_bill
Dinner	20.797159
Lunch	17.168676

In [16]:
```
sns.barplot(df['time'],df.total_bill)
plt.show()
```



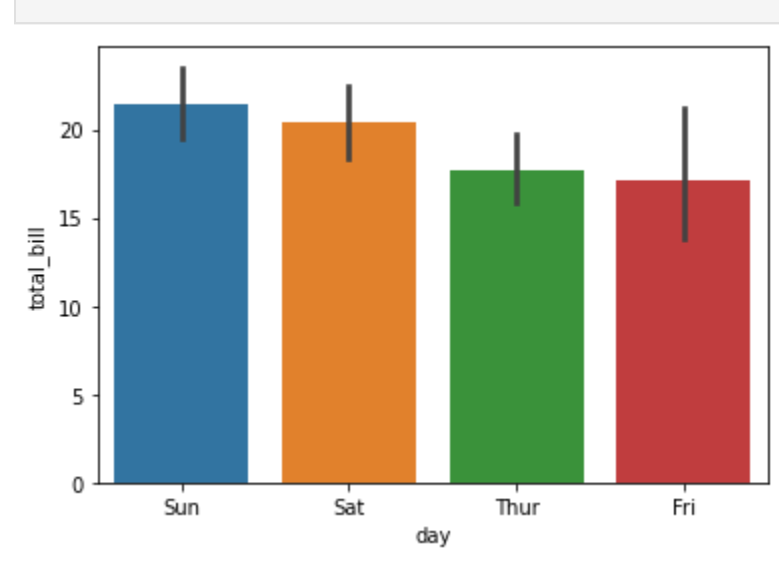
Conclusion6: From the above bar graph we cnclude that in Dinner time bill amount is more

In [17]:
```
df[['total_bill','day']].groupby('day').mean()
```

Out[17]:

day	total_bill
Fri	17.151579
Sat	20.441379
Sun	21.410000
Thur	17.682742

In [18]:
```
sns.barplot(df['day'],df.total_bill)
plt.show()
```



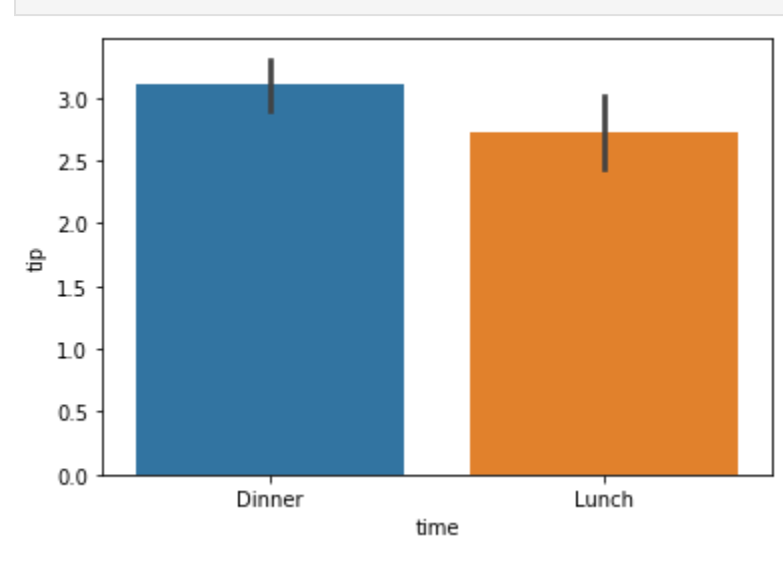
Conclusion7: From the above bar graph we can conclude that in weekend bill amount is more

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

Out[19]:

time	tip
Dinner	3.102670
Lunch	2.728088

In [20]:
```
sns.barplot(df['time'],df.tip)
plt.show()
```



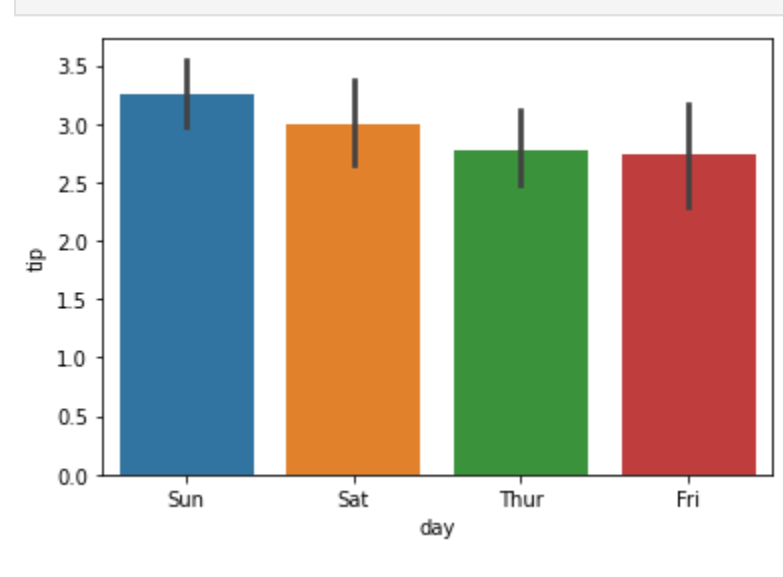
Conclusion8: From the above bar graph we can conclude that in Dinner time people give more tip

In [21]:
```
df[['tip','day']].groupby('day').mean()
```

Out[21]:

day	tip
Fri	2.734737
Sat	2.993103
Sun	3.255132
Thur	2.771452

In [22]:
```
sns.barplot(df['day'],df.tip)
plt.show()
```



Conclusion9: From the above bar graph we can conclude that in weekend people give more tip

In [23]:
```
df
```

Out[23]:

	total_bill	tip	sex	smoker	day	time	size	bill_group
0	16.99	1.01	Female	No	Sun	Dinner	2	(12.618, 17.392]
1	10.34	1.66	Male	No	Sun	Dinner	3	(7.844, 12.618]
2	21.01	3.50	Male	No	Sun	Dinner	3	(17.392, 22.166]
3	23.68	3.31	Male	No	Sun	Dinner	2	(22.166, 26.94]
4	24.59	3.61	Female	No	Sun	Dinner	4	(22.166, 26.94]
...	...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3	(26.94, 31.714]
240	27.18	2.00	Female	Yes	Sat	Dinner	2	(26.94, 31.714]
241	22.67	2.00	Male	Yes	Sat	Dinner	2	(22.166, 26.94]
242	17.82	1.75	Male	No	Sat	Dinner	2	(17.392, 22.166]
243	18.78	3.00	Female	No	Thur	Dinner	2	(17.392, 22.166]

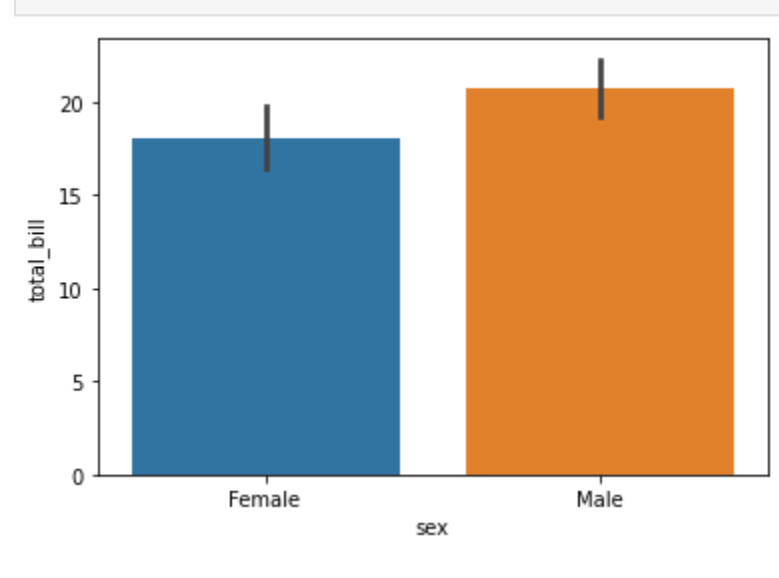
244 rows × 8 columns

In [24]:
```
df[['total_bill','sex']].groupby('sex').mean()
```

Out[24]:

sex	total_bill
Female	18.056897
Male	20.744076

In [25]:
```
sns.barplot(df['sex'],df.total_bill)
plt.show()
```



Conclusion 10: From the above bar graph we can conclude that Male are paying more bill than female

In [26]:
```
df
```

Out[26]:

	total_bill	tip	sex	smoker	day	time	size	bill_group
0	16.99	1.01	Female	No	Sun	Dinner	2	(12.618, 17.392]
1	10.34	1.66	Male	No	Sun	Dinner	3	(7.844, 12.618]
2	21.01	3.50	Male	No	Sun	Dinner	3	(17.392, 22.166]
3	23.68	3.31	Male	No	Sun	Dinner	2	(22.166, 26.94]
4	24.59	3.61	Female	No	Sun	Dinner	4	(22.166, 26.94]
...	...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3	(26.94, 31.714]
240	27.18	2.00	Female	Yes	Sat	Dinner	2	(26.94, 31.714]
241	22.67	2.00	Male	Yes	Sat	Dinner	2	(22.166, 26.94]
242	17.82	1.75	Male	No	Sat	Dinner	2	(17.392, 22.166]
243	18.78	3.00	Female	No	Thur	Dinner	2	(17.392, 22.166]

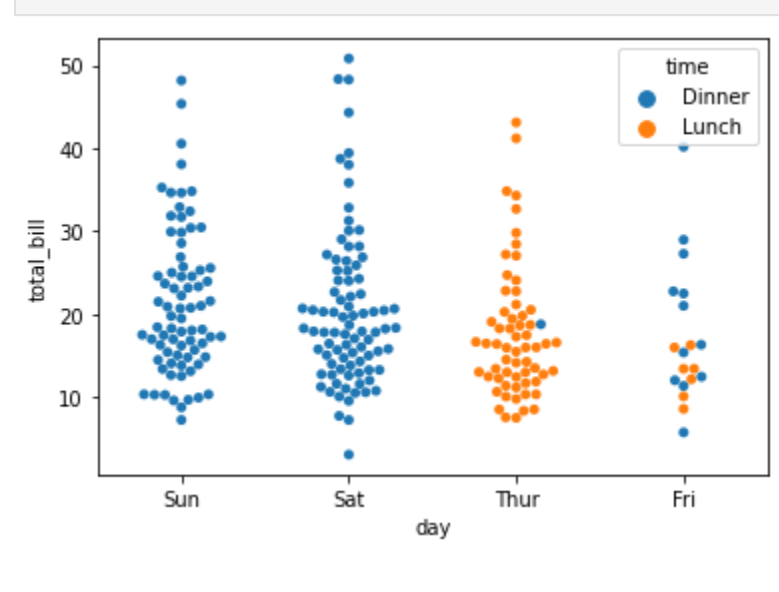
244 rows × 8 columns

In [27]:
```
df[['size','time']].groupby('time').mean()
```

Out[27]:

time	size
Dinner	2.630682
Lunch	2.411765

In [28]:
```
sns.swarmplot(x='day',y='total_bill',hue='time',data=df)
plt.show()
```



Conclusion 11: From the above plot we can conclude that people prefer to go for dinner on resturant in weekends

Conclusion 12: From the above plot we can conclude that on week days people come for lunch more than dinner

Conclusions – Following conclusions were made by EDA performance.

1. There are some outlier's in the data.
2. Our data is following Normal Distribution.
3. Male are giving more tip than female.
4. More number of people in a group gives more tip.
5. With increasing bill amount, tip is also increasing.
6. In Dinner time bill amount is more.
7. In weekend bill amount is more.
8. In Dinner time people give more tip.
9. In weekend people give more tip.
10. Male are paying more bill than female.
11. People prefer to go for dinner on restaurant in weekends.
12. On week days people come for lunch more than dinner.