5_Probability - Correlation

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
dataset = pd.read_csv('tips.csv')
```

In [3]:
```
dataset.head(3)
```

Out[3]:

	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

In [4]:
```
dataset.isnull().sum()
```

Out[4]:
```
total_bill    0
tip           0
sex           0
smoker        0
day           0
time          0
size          0
dtype: int64
```

In [5]:
```
# To check datatypes in dataset
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 244 entries, 0 to 243
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   total_bill  244 non-null    float64
 1   tip         244 non-null    float64
 2   sex         244 non-null    object 
 3   smoker      244 non-null    object 
 4   day         244 non-null    object 
 5   time        244 non-null    object 
 6   size        244 non-null    int64  
dtypes: float64(2), int64(1), object(4)
memory usage: 13.5+ KB
```

In [6]:
```
dataset.select_dtypes("float64" ,"int64")
```

Out[6]:

	total_bill	tip
0	16.99	1.01
1	10.34	1.66
2	21.01	3.50
3	23.68	3.31
4	24.59	3.61
...	...	...
239	29.03	5.92
240	27.18	2.00
241	22.67	2.00
242	17.82	1.75
243	18.78	3.00

244 rows × 2 columns

In [10]:
```
data_cor = dataset.select_dtypes("float64" ,"int64").corr()
data_cor
```

Out[10]:

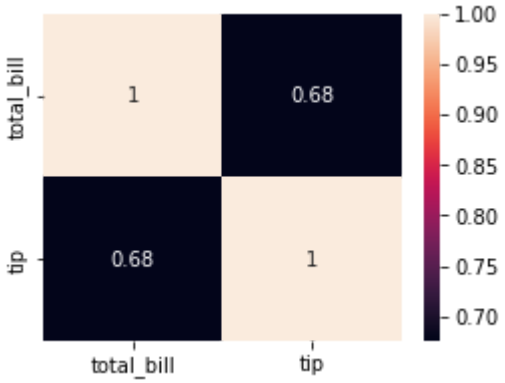
	total_bill	tip
total_bill	1.000000	0.675734
tip	0.675734	1.000000

In [11]:
```
data_cov = dataset.select_dtypes("float64" ,"int64").cov()
data_cov
```

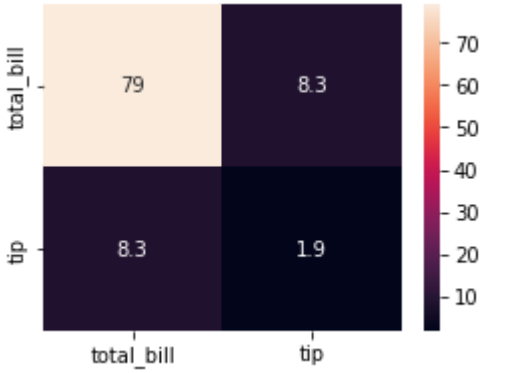
Out[11]:

	total_bill	tip
total_bill	79.252939	8.323502
tip	8.323502	1.914455

In [16]:
```
plt.figure(figsize=(4,3))
sns.heatmap(data_cor, annot=True)
plt.show()
```



In [17]:
```
plt.figure(figsize=(4,3))
sns.heatmap(data_cov, annot=True)
plt.show()
```



In [ ]: