Day 6

Uber Dataset

In [1]:

```
import numpy as np
import pandas as pd
```

In [2]:

```
d=pd.read_csv(r"c:\Users\user\Downloads\7_uber.csv")
d
```

Out[2]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	picku
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	
...	...	...	...	...	...	
199995	42598914	2012-10-28 10:49:00.00000053	3.0	2012-10-28 10:49:00 UTC	-73.987042	
199996	16382965	2014-03-14 01:09:00.0000008	7.5	2014-03-14 01:09:00 UTC	-73.984722	
199997	27804658	2009-06-29 00:42:00.00000078	30.9	2009-06-29 00:42:00 UTC	-73.986017	
199998	20259894	2015-05-20 14:56:25.0000004	14.5	2015-05-20 14:56:25 UTC	-73.997124	
199999	11951496	2010-05-15 04:08:00.00000076	14.1	2010-05-15 04:08:00 UTC	-73.984395	

200000 rows × 9 columns

Mean,median,mode,describe

In [3]:

```
data=pd.DataFrame(d[['fare_amount','passenger_count']][0:500])
data
```

Out[3]:

	fare_amount	passenger_count
0	7.5	1
1	7.7	1
2	12.9	1
3	5.3	3
4	16.0	5
...	...	...
495	25.7	1
496	8.0	1
497	10.5	2
498	5.5	1
499	10.0	1

500 rows × 2 columns

In [4]:

```
print(data.mean())
```

```
fare_amount        10.70872
passenger_count     1.66400
dtype: float64
```

In [5]:

```
print(data.median())
```

```
fare_amount        8.1
passenger_count    1.0
dtype: float64
```

In [6]:

```
data.fillna(value=1)
```

Out[6]:

	fare_amount	passenger_count
0	7.5	1
1	7.7	1
2	12.9	1
3	5.3	3
4	16.0	5
...	...	...
495	25.7	1
496	8.0	1
497	10.5	2
498	5.5	1
499	10.0	1

500 rows × 2 columns

In [7]:

```
print(data.mode())
```

```
   fare_amount  passenger_count
0          6.5                1
```

In [8]:

```
print(data.describe())
```

```
       fare_amount  passenger_count
count   500.000000       500.000000
mean     10.708720         1.664000
std       8.334145         1.267405
min       2.500000         0.000000
25%       6.000000         1.000000
50%       8.100000         1.000000
75%      12.500000         2.000000
max      57.330000         6.000000
```

Sum,cumsum,count,min,max

In [9]:

```
print(data.sum())
```

```
fare_amount        5354.36
passenger_count     832.00
dtype: float64
```

In [10]:

```
print(data.cumsum())
```

```
     fare_amount  passenger_count
0           7.50                1
1          15.20                2
2          28.10                3
3          33.40                6
4          49.40               11
..           ...              ...
495      5320.36              827
496      5328.36              828
497      5338.86              830
498      5344.36              831
499      5354.36              832

[500 rows x 2 columns]
```

In [11]:

```
print(data.count())
```

```
fare_amount        500
passenger_count    500
dtype: int64
```

In [12]:

```
print(data.min())
```

```
fare_amount        2.5
passenger_count    0.0
dtype: float64
```

In [13]:

```
print(data.max())
```

```
fare_amount        57.33
passenger_count     6.00
dtype: float64
```

covariance and correlation (spearman and pearsons)

In [14]:

```
data1=data['fare_amount'][0:10]
data1
```

Out[14]:

```
0     7.5
1     7.7
2    12.9
3     5.3
4    16.0
5     4.9
6    24.5
7     2.5
8     9.7
9    12.5
Name: fare_amount, dtype: float64
```

In [15]:

```
data2=data['passenger_count'][0:10]
data2
```

Out[15]:

```
0    1
1    1
2    1
3    3
4    5
5    1
6    5
7    1
8    1
9    1
Name: passenger_count, dtype: int64
```

In [16]:

```
from numpy import cov
print(cov(data1,data2))
```

```
[[41.74055556  7.67777778]
 [ 7.67777778  2.88888889]]
```

In [18]:

```
from scipy.stats import pearsonr
print(pearsonr(data1,data2))
```

```
(0.6991832347843764, 0.024444145792245162)
```

In [19]:

```
from scipy.stats import spearmanr
print(spearmanr(data1,data2))
```

```
SpearmanrResult(correlation=0.509395451638894, pvalue=0.1326052475011008)
```

In [ ]:

```
```