Statictics

```
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
```

File directory

```
In [2]: data=pd.read_csv(r"C:\Users\user\Desktop\Vicky\7_uber.csv")
```

All Mathematical function

```
In [3]: data.describe()
```

Out[3]:

	Unnamed: 0	fare_amount	pickup_longitude	pickup_latitude	dropoff_longitude	dropof
count	2.000000e+05	200000.000000	200000.000000	200000.000000	199999.000000	19999
mean	2.771250e+07	11.359955	-72.527638	39.935885	-72.525292	3
std	1.601382e+07	9.901776	11.437787	7.720539	13.117408	
min	1.000000e+00	-52.000000	-1340.648410	-74.015515	-3356.666300	-88
25%	1.382535e+07	6.000000	-73.992065	40.734796	-73.991407	4
50%	2.774550e+07	8.500000	-73.981823	40.752592	-73.980093	4
75%	4.155530e+07	12.500000	-73.967154	40.767158	-73.963658	4
max	5.542357e+07	499.000000	57.418457	1644.421482	1153.572603	87

To display the top portion of the dataset

In [4]: `data.head()`

Out[4]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitud
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.73835
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.72822
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.74077
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.79084
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.74408

◄ ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬ ►

To display the mean median mode of the dataset for only numerical value

In [6]:
```python
data1=data[["fare_amount","passenger_count"]]
print(data1.mean())
print(data1.mode())
print(data.median())
print()
```

```
fare_amount        11.359955
passenger_count     1.684535
dtype: float64
    fare_amount  passenger_count
0          6.5                1
Unnamed: 0          2.774550e+07
fare_amount         8.500000e+00
pickup_longitude   -7.398182e+01
pickup_latitude     4.075259e+01
dropoff_longitude  -7.398009e+01
dropoff_latitude    4.075304e+01
passenger_count     1.000000e+00
dtype: float64
```

To display the total of each columns

In [7]: `print(data1.sum())`

```
fare_amount        2271991.05
passenger_count     336907.00
dtype: float64
```

To displayn the minimum value

In [8]: `print(data.min())`

```
Unnamed: 0                                      1
key                       2009-01-01 01:15:22.0000006
fare_amount                                  -52.0
pickup_datetime            2009-01-01 01:15:22 UTC
pickup_longitude                        -1340.64841
pickup_latitude                          -74.015515
dropoff_longitude                         -3356.6663
dropoff_latitude                         -881.985513
passenger_count                                  0
dtype: object
```

To displayn theCummulative sum

In [9]: `print(data1.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
...               ...               ...
199995     2271924.05            336902
199996     2271931.55            336903
199997     2271962.45            336905
199998     2271976.95            336906
199999     2271991.05            336907

[200000 rows x 2 columns]
```

To count the total number of values in columns

In [10]: `print(data.count())`

```
Unnamed: 0           200000
key                  200000
fare_amount          200000
pickup_datetime      200000
pickup_longitude     200000
pickup_latitude      200000
dropoff_longitude    199999
dropoff_latitude     199999
passenger_count      200000
dtype: int64
```

In [11]: `print(data1.cov())`

```
                 fare_amount   passenger_count
fare_amount        98.045172          0.139296
passenger_count     0.139296          1.920986
```

```python
from scipy.stats import spearmanr
from scipy.stats import pearsonr
```

In [14]:

In [15]:
```python
data2=data[["pickup_longitude","pickup_latitude"]]
data2
```

Out[15]:

	pickup_longitude	pickup_latitude
0	-73.999817	40.738354
1	-73.994355	40.728225
2	-74.005043	40.740770
3	-73.976124	40.790844
4	-73.925023	40.744085
...	...	...
199995	-73.987042	40.739367
199996	-73.984722	40.736837
199997	-73.986017	40.756487
199998	-73.997124	40.725452
199999	-73.984395	40.720077

200000 rows × 2 columns

In [16]:
```python
print(spearmanr(data1,data2))
```

```
SpearmanrResult(correlation=array([[ 1.        ,  0.02329567,  0.0427319 , -
0.0735512 ],
       [ 0.02329567,  1.        , -0.01615785, -0.01575884],
       [ 0.0427319 , -0.01615785,  1.        ,  0.51336533],
       [-0.0735512 , -0.01575884,  0.51336533,  1.        ]]), pvalue=array
([[0.00000000e+000, 2.02039886e-025, 1.75772620e-081,
         6.38669717e-238],
       [2.02039886e-025, 0.00000000e+000, 4.95831337e-013,
        1.81556398e-012],
       [1.75772620e-081, 4.95831337e-013, 0.00000000e+000,
        0.00000000e+000],
       [6.38669717e-238, 1.81556398e-012, 0.00000000e+000,
        0.00000000e+000]]))
```

In [ ]:

In [ ]: