

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

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

Out[2]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	40
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	40
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	40
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	40
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	40
...
199995	42598914	2012-10-28 10:49:00.00000053	3.0	2012-10-28 10:49:00 UTC	-73.987042	40
199996	16382965	2014-03-14 01:09:00.0000008	7.5	2014-03-14 01:09:00 UTC	-73.984722	40
199997	27804658	2009-06-29 00:42:00.00000078	30.9	2009-06-29 00:42:00 UTC	-73.986017	40
199998	20259894	2015-05-20 14:56:25.0000004	14.5	2015-05-20 14:56:25 UTC	-73.997124	40
199999	11951496	2010-05-15 04:08:00.00000076	14.1	2010-05-15 04:08:00 UTC	-73.984395	40

200000 rows × 9 columns



```
In [3]: df=data.head(100)
df
```

Out[3]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.7383
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.7282
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.7407
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.7906
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.7440
...
95	25431833	2015-04-11 08:47:47.0000001	9.5	2015-04-11 08:47:47 UTC	-73.978432	40.7523
96	44792012	2011-10-03 20:29:00.000000179	4.5	2011-10-03 20:29:00 UTC	-73.990055	40.7564
97	18571020	2010-04-26 03:12:44.0000001	3.3	2010-04-26 03:12:44 UTC	-73.982326	40.7313
98	37942404	2011-11-18 09:51:00.000000166	30.9	2011-11-18 09:51:00 UTC	-73.995888	40.7590
99	29024472	2009-08-30 14:03:55.0000002	26.9	2009-08-30 14:03:55 UTC	-73.990137	40.7560

100 rows × 9 columns




```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Unnamed: 0            100 non-null   int64
1   key                   100 non-null   object
2   fare_amount           100 non-null   float64
3   pickup_datetime       100 non-null   object
4   pickup_longitude      100 non-null   float64
5   pickup_latitude       100 non-null   float64
6   dropoff_longitude     100 non-null   float64
7   dropoff_latitude      100 non-null   float64
8   passenger_count       100 non-null   int64
dtypes: float64(5), int64(2), object(2)
memory usage: 7.2+ KB
```

```
In [5]: df.describe()
```

```
Out[5]:
```

	Unnamed: 0	fare_amount	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_
count	1.000000e+02	100.000000	100.000000	100.000000	100.000000	100
mean	2.810554e+07	11.065700	-71.019759	39.123621	-71.015479	39
std	1.635033e+07	9.029756	14.569902	8.026358	14.569028	8
min	2.268700e+05	2.500000	-74.013173	0.000000	-74.016152	0
25%	1.422691e+07	5.475000	-73.992601	40.733982	-73.989142	40
50%	2.710896e+07	8.100000	-73.982002	40.752764	-73.979396	40
75%	4.480811e+07	12.600000	-73.968615	40.765572	-73.960980	40
max	5.508597e+07	56.800000	0.000000	40.850558	0.000000	40

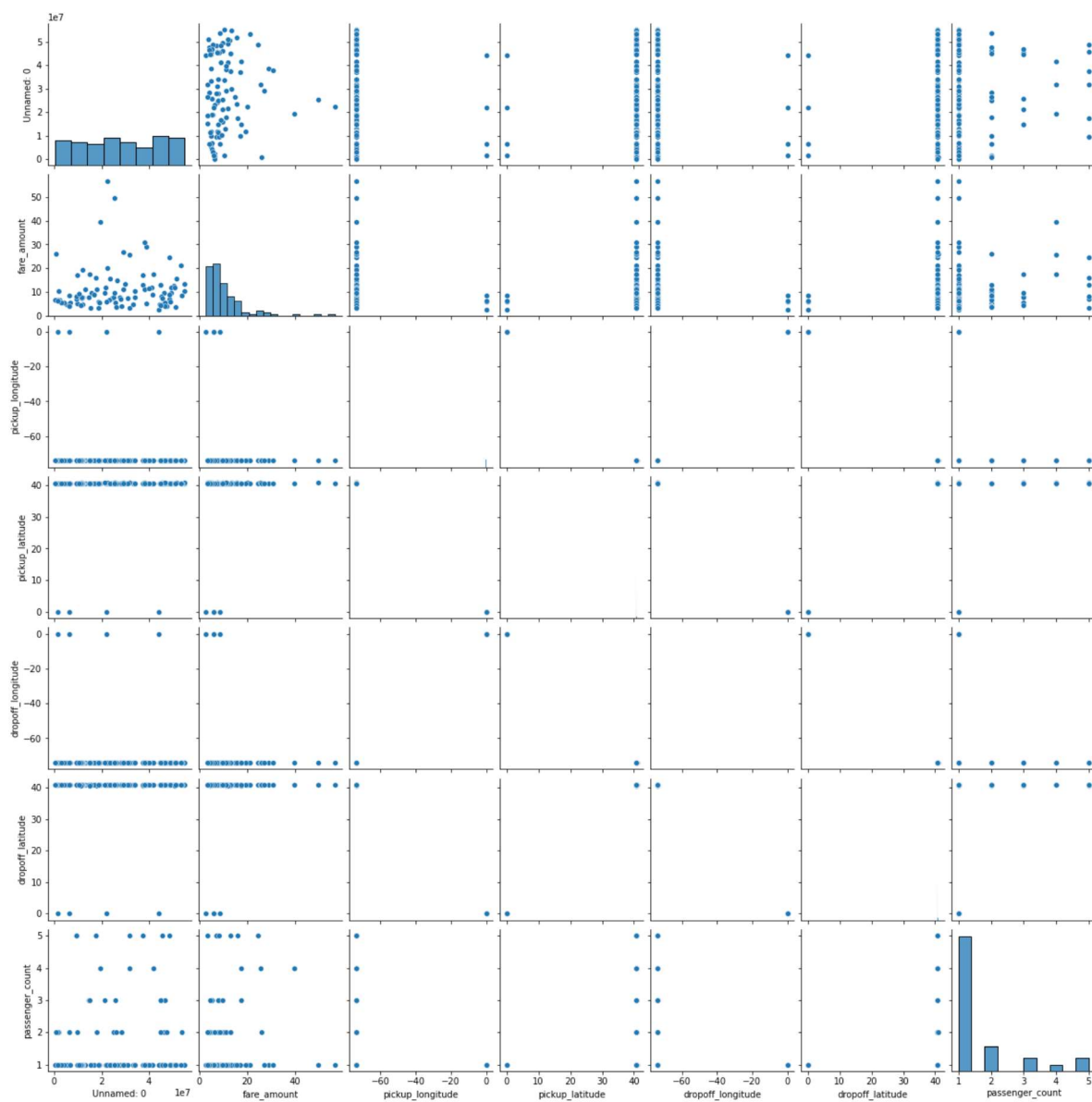


```
In [6]: df.columns
```

```
Out[6]: Index(['Unnamed: 0', 'key', 'fare_amount', 'pickup_datetime',  
              'pickup_longitude', 'pickup_latitude', 'dropoff_longitude',  
              'dropoff_latitude', 'passenger_count'],  
              dtype='object')
```

```
In [31]: sns.pairplot(df)
```

```
Out[31]: <seaborn.axisgrid.PairGrid at 0x1d75b5a34c0>
```



```
In [10]: da=data[['Unnamed: 0', 'fare_amount','pickup_longitude', 'pickup_latitude', 'dropoff_longitude', 'dropoff_latitude', 'passenger_count']]
da
```

Out[10]:

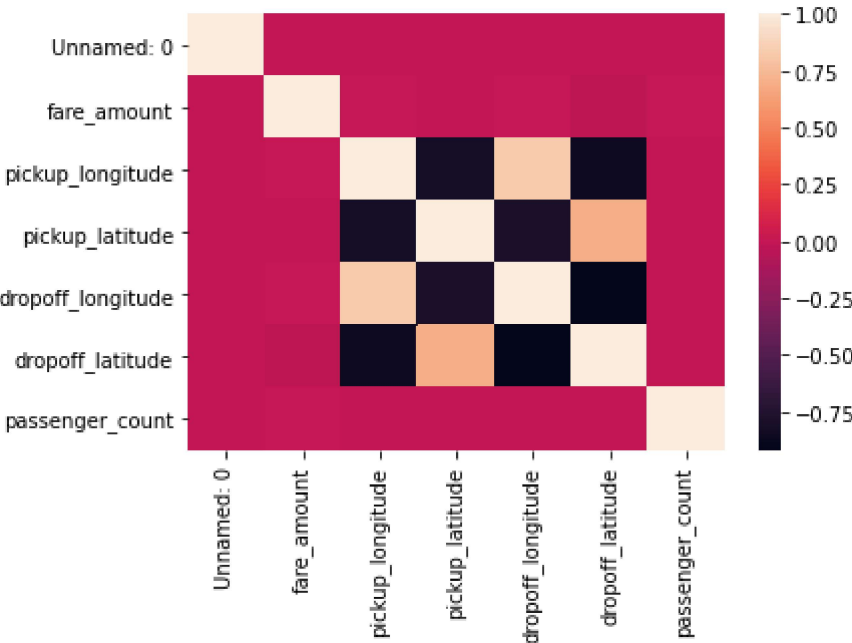
	Unnamed: 0	fare_amount	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude
0	24238194	7.5	-73.999817	40.738354	-73.999512	40.738354
1	27835199	7.7	-73.994355	40.728225	-73.994710	40.728225
2	44984355	12.9	-74.005043	40.740770	-73.962565	40.740770
3	25894730	5.3	-73.976124	40.790844	-73.965316	40.790844
4	17610152	16.0	-73.925023	40.744085	-73.973082	40.744085
...
199995	42598914	3.0	-73.987042	40.739367	-73.986525	40.739367
199996	16382965	7.5	-73.984722	40.736837	-74.006672	40.736837
199997	27804658	30.9	-73.986017	40.756487	-73.858957	40.756487
199998	20259894	14.5	-73.997124	40.725452	-73.983215	40.725452
199999	11951496	14.1	-73.984395	40.720077	-73.985508	40.720077

200000 rows × 7 columns



```
In [11]: sns.heatmap(da.corr())
```

Out[11]: <AxesSubplot:>



```
In [18]: x=da[['Unnamed: 0', 'fare_amount']]
y=da['passenger_count']
```

```
In [19]: from sklearn.model_selection import train_test_split

x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
```

```
In [20]: from sklearn.linear_model import LinearRegression

lr=LinearRegression()
lr.fit(x_train,y_train)
```

Out[20]: LinearRegression()

```
In [21]: print(lr.intercept_)

1.6635119197126038
```

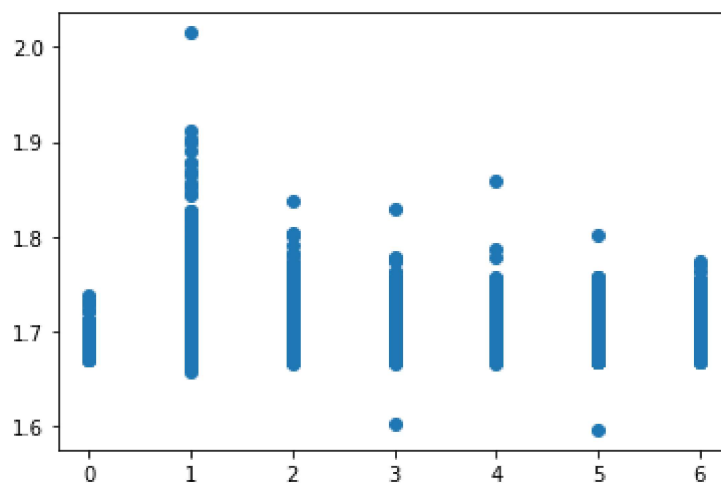
```
In [22]: coeff = pd.DataFrame(lr.coef_,x.columns,columns=['Co-efficient'])
coeff
```

Out[22]:

	Co-efficient
Unnamed: 0	2.959968e-10
fare_amount	1.359064e-03

```
In [23]: prediction=lr.predict(x_test)
plt.scatter(y_test,prediction)
```

Out[23]: <matplotlib.collections.PathCollection at 0x1d755c4e160>



```
In [24]: print(lr.score(x_test,y_test))

7.458853965869316e-05
```

In []:

In []: