27-07-2023

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

In [8]:
```
a=pd.read_csv(r"C:\Users\user\Downloads\7_uber.csv")
a
```

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 x 9 columns

In [9]:
```
a=a.head(200)
a
```

Out[9]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitu
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.738
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.728
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.740
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.790
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.744
...	...	...	...	...	...	
195	49202586	2014-05-28 01:00:00.00000073	14.5	2014-05-28 01:00:00 UTC	-74.005477	40.738
196	51452192	2009-05-12 10:32:00.000000154	24.0	2009-05-12 10:32:00 UTC	-73.981558	40.783
197	45317989	2012-08-07 20:53:18.0000001	10.5	2012-08-07 20:53:18 UTC	-73.965930	40.805
198	41858701	2009-09-24 16:21:42.0000001	8.9	2009-09-24 16:21:42 UTC	-73.952080	40.790
199	13472186	2011-04-03 00:01:40.0000002	14.1	2011-04-03 00:01:40 UTC	-74.000190	40.718

200 rows × 9 columns

In [10]:
```
a.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 9 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Unnamed: 0         200 non-null    int64  
 1   key                200 non-null    object 
 2   fare_amount        200 non-null    float64
 3   pickup_datetime    200 non-null    object 
 4   pickup_longitude   200 non-null    float64
 5   pickup_latitude    200 non-null    float64
 6   dropoff_longitude  200 non-null    float64
 7   dropoff_latitude   200 non-null    float64
 8   passenger_count    200 non-null    int64  
dtypes: float64(5), int64(2), object(2)
memory usage: 14.2+ KB
```

In [11]:
```
a.columns
```

Out[11]:
```
Index(['Unnamed: 0', 'key', 'fare_amount', 'pickup_datetime',
       'pickup_longitude', 'pickup_latitude', 'dropoff_longitude',
       'dropoff_latitude', 'passenger_count'],
      dtype='object')
```

In [12]:
```
a.head()
```

Out[12]:

	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

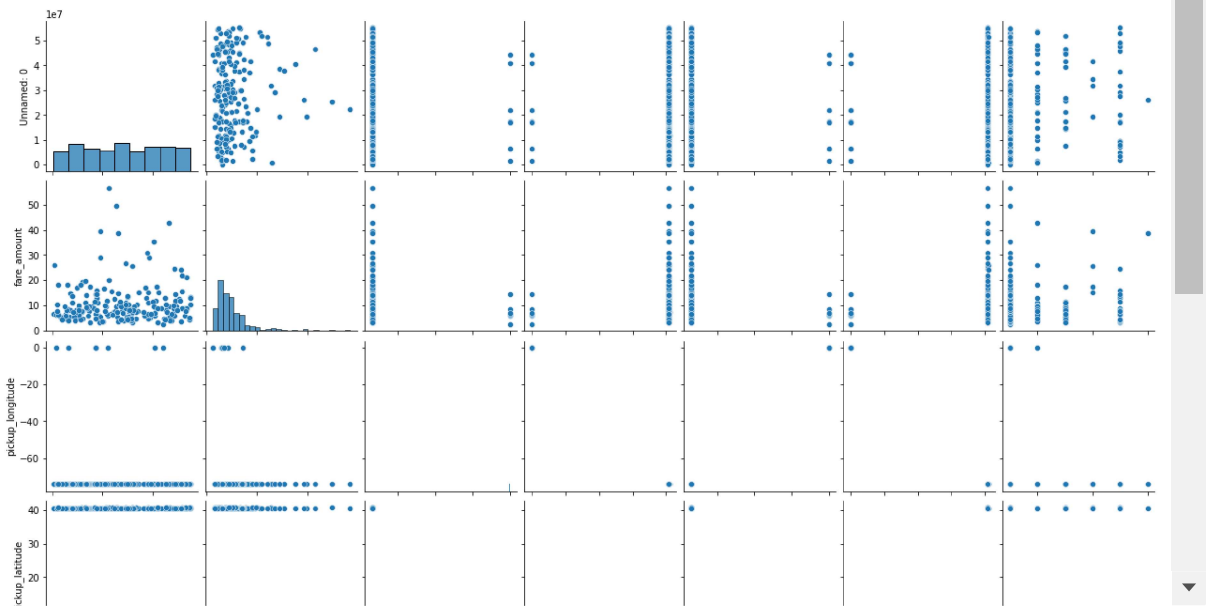
In [13]:
```
a.describe()
```

Out[13]:

	Unnamed: 0	fare_amount	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_l
count	2.000000e+02	200.000000	200.000000	200.000000	200.000000	200.
mean	2.779091e+07	10.620050	-71.388553	39.327046	-71.387016	39.
std	1.578378e+07	8.023976	13.629815	7.508297	13.629487	7.
min	2.268700e+05	2.500000	-74.015122	0.000000	-74.016152	0.
25%	1.418957e+07	6.000000	-73.992744	40.736897	-73.989370	40.
50%	2.799295e+07	8.100000	-73.982225	40.753583	-73.979274	40.
75%	4.126453e+07	12.125000	-73.968338	40.766672	-73.962785	40.
max	5.519870e+07	56.800000	0.001782	40.850558	0.000875	40.

In [37]:
```
sns.pairplot(a)
```

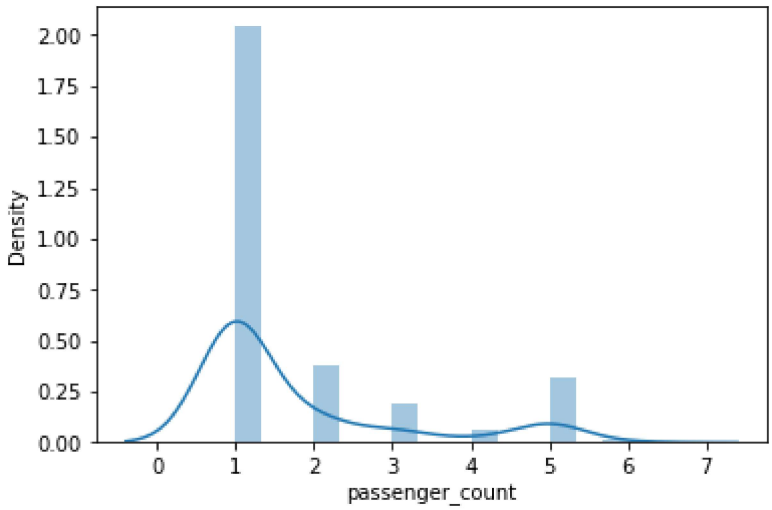
Out[37]:
```
<seaborn.axisgrid.PairGrid at 0x2a14c3d67c0>
```



In [38]:
```
sns.distplot(a['passenger_count'])
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: Fut
ureWarning: `distplot` is a deprecated function and will be removed in a futu
re version. Please adapt your code to use either `displot` (a figure-level fu
nction with similar flexibility) or `histplot` (an axes-level function for hi
stograms).
  warnings.warn(msg, FutureWarning)
```

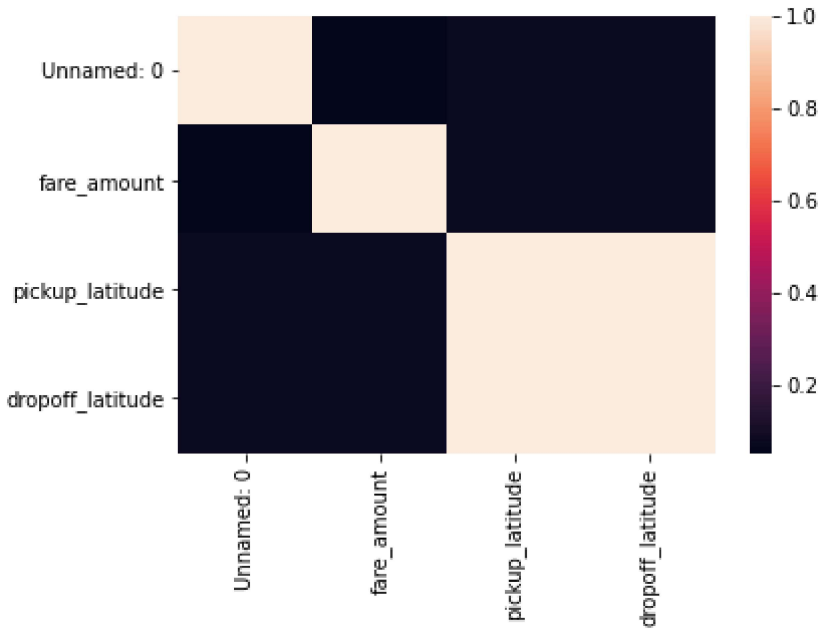
Out[38]:
```
<AxesSubplot:xlabel='passenger_count', ylabel='Density'>
```



In [39]:
```
x1=a[['Unnamed: 0',  'fare_amount', 'pickup_latitude','dropoff_latitude']]
```

In [36]:
```
sns.heatmap(x1.corr())
```

Out[36]: <AxesSubplot:>



In [40]:
```
x=a[['Unnamed: 0',  'fare_amount','pickup_latitude', 'dropoff_latitude']]
y=a['passenger_count']
```

In [41]:
```
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 [42]:
```
from sklearn.linear_model import LinearRegression
lr=LinearRegression()
lr.fit(x_train,y_train)
```

Out[42]: LinearRegression()

In [43]:
```
print(lr.intercept_)
```

```
1.0066736464296169
```

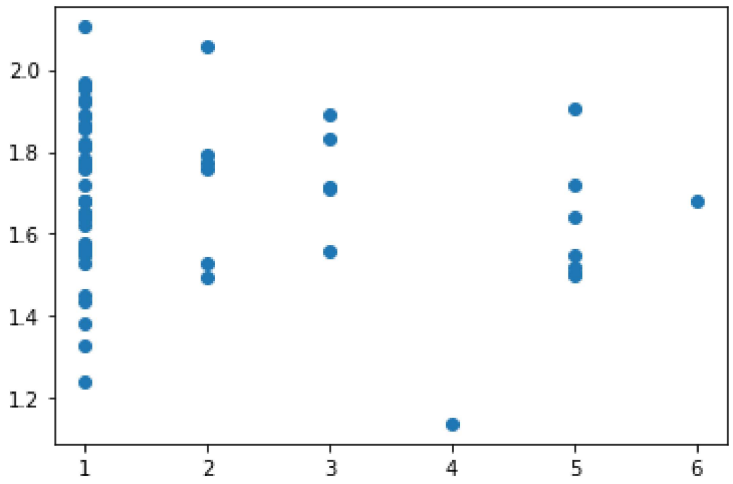
In [44]:
```
coeff=pd.DataFrame(lr.coef_,x.columns,columns=['Co-efficient'])
coeff
```

Out[44]:

	Co-efficient
Unnamed: 0	7.737940e-09
fare_amount	-4.287814e-03
pickup_latitude	4.864993e+00
dropoff_latitude	-4.851257e+00

In [45]:
```
prediction=lr.predict(x_test)
plt.scatter(y_test,prediction)
```

Out[45]:
```
<matplotlib.collections.PathCollection at 0x2a162707f40>
```



In [46]:
```
print(lr.score(x_test,y_test))
```

```
-0.07987275745940403
```