kaviyadevi 20106064

In [2]:
```
#to import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [5]:
```
#to import dataset
data1=pd.read_csv(r"C:\Users\user\Downloads\7_uber - 7_uber.csv")
data1
```

Out[5]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	drop
0	24238194	2015-05-07 19:52:06	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.738354	
1	27835199	2009-07-17 20:04:56	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.728225	
2	44984355	2009-08-24 21:45:00	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.740770	
3	25894730	2009-06-26 08:22:21	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.790844	
4	17610152	2014-08-28 17:47:00	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.744085	
...	...	...	...	...	...	...	...
199995	42598914	2012-10-28 10:49:00	3.0	2012-10-28 10:49:00 UTC	-73.987042	40.739367	
199996	16382965	2014-03-14 01:09:00	7.5	2014-03-14 01:09:00 UTC	-73.984722	40.736837	
199997	27804658	2009-06-29 00:42:00	30.9	2009-06-29 00:42:00 UTC	-73.986017	40.756487	
199998	20259894	2015-05-20 14:56:25	14.5	2015-05-20 14:56:25 UTC	-73.997124	40.725452	
199999	11951496	2010-05-15 04:08:00	14.1	2010-05-15 04:08:00 UTC	-73.984395	40.720077	

200000 rows × 9 columns

In [6]:
```
#to display top 5 rows
data=data1.head(200)
data
```

Out[6]:

key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_la
2015-05-07 :52:06	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.738354	-73.999512	40.7
2009-07-17 :04:56	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.728225	-73.994710	40.7
2009-08-24 :45:00	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.740770	-73.962565	40.7
2009-06-26 :22:21	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.790844	-73.965316	40.8
2014-08-28 :47:00	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.744085	-73.973082	40.7
...	...	...	...	...	...	
2014-05-28 :00:00	14.5	2014-05-28 01:00:00 UTC	-74.005477	40.738575	-73.972722	40.7
2009-05-12 :32:00	24.0	2009-05-12 10:32:00 UTC	-73.981558	40.783752	-73.900931	40.8
2012-08-07 :53:18	10.5	2012-08-07 20:53:18 UTC	-73.965930	40.805358	-73.949923	40.7
2009-09-24 :21:42	8.9	2009-09-24 16:21:42 UTC	-73.952080	40.790119	-73.963637	40.7
2011-04-03 :01:40	14.1	2011-04-03 00:01:40 UTC	-74.000190	40.718336	-73.956801	40.7

ns

DATA CLEANING AND PREPROCESSING

In [7]:
```
data.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 [8]:
```
#to display summary of statistics
data.describe()
```

Out[8]:

	Unnamed: 0	fare_amount	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_lati
count	2.000000e+02	200.000000	200.000000	200.000000	200.000000	200.00
mean	2.779091e+07	10.620050	-71.388553	39.327046	-71.387016	39.32
std	1.578378e+07	8.023976	13.629815	7.508297	13.629487	7.50
min	2.268700e+05	2.500000	-74.015122	0.000000	-74.016152	0.00
25%	1.418957e+07	6.000000	-73.992744	40.736897	-73.989371	40.73
50%	2.799295e+07	8.100000	-73.982225	40.753583	-73.979274	40.75
75%	4.126453e+07	12.125000	-73.968338	40.766672	-73.962785	40.77
max	5.519870e+07	56.800000	0.001782	40.850558	0.000875	40.89

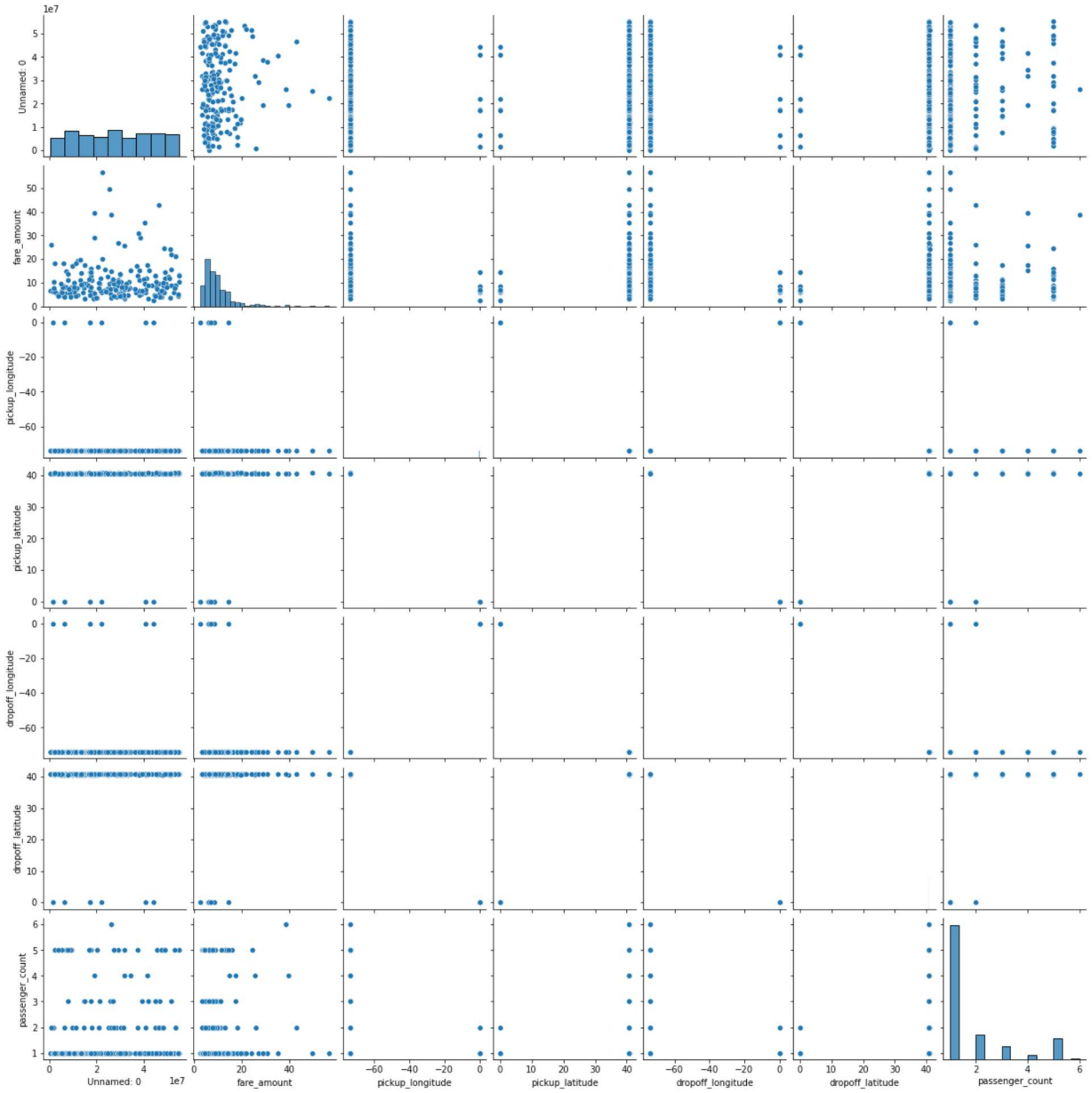
In [9]:
```
#to display the column heading
data.columns
```

Out[9]:
```
Index(['Unnamed: 0', 'key', 'fare_amount', 'pickup_datetime',
       'pickup_longitude', 'pickup_latitude', 'dropoff_longitude',
       'dropoff_latitude', 'passenger_count'],
      dtype='object')
```

EDA and DATA VISUALIZATION

In [10]:
```
sns.pairplot(data)
```

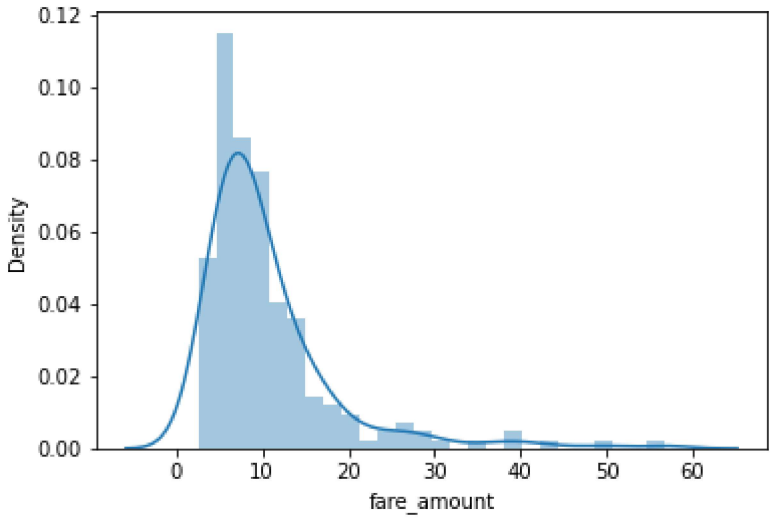
Out[10]:
```
<seaborn.axisgrid.PairGrid at 0x2298b677340>
```



In [12]:
```
sns.distplot(data['fare_amount'])
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: Futur
eWarning: `distplot` is a deprecated function and will be removed in a future v
ersion. Please adapt your code to use either `displot` (a figure-level function
with similar flexibility) or `histplot` (an axes-level function for histogram
s).
  warnings.warn(msg, FutureWarning)
```

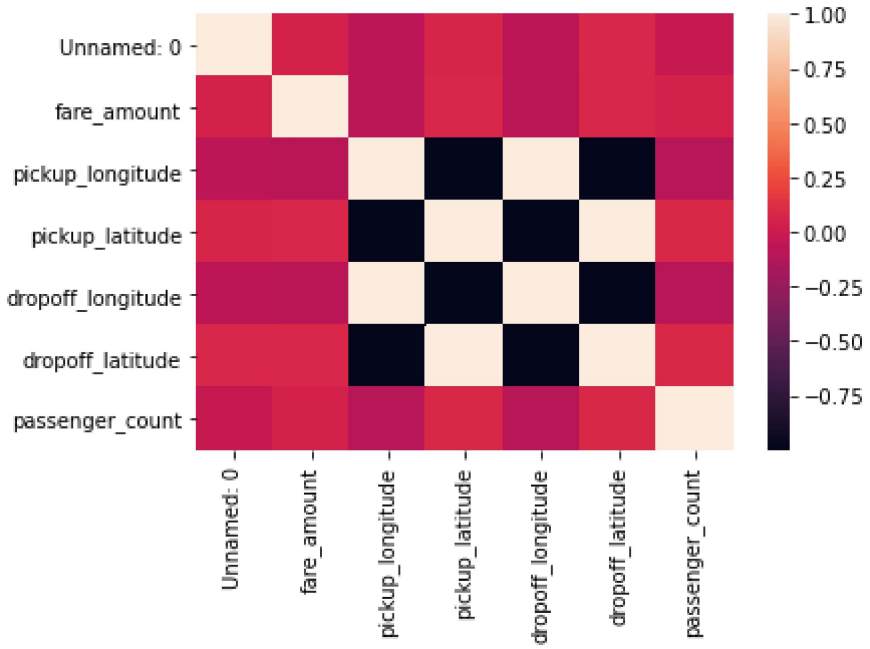
Out[12]: <AxesSubplot:xlabel='fare_amount', ylabel='Density'>



In [13]:
```
df=data[['Unnamed: 0', 'key', 'fare_amount', 'pickup_datetime',
       'pickup_longitude', 'pickup_latitude', 'dropoff_longitude',
       'dropoff_latitude', 'passenger_count']]
```

In [14]:
```
sns.heatmap(df.corr())
```

Out[14]: <AxesSubplot:>



TRAINING MODEL

In [24]:
```
df[['pickup_longitude', 'pickup_latitude', 'dropoff_longitude','dropoff_latitude'
df[['fare_amount']]
```

In [25]:
```
#to split my dataset into trainning and test
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 [26]:
```
from sklearn.linear_model import LinearRegression

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

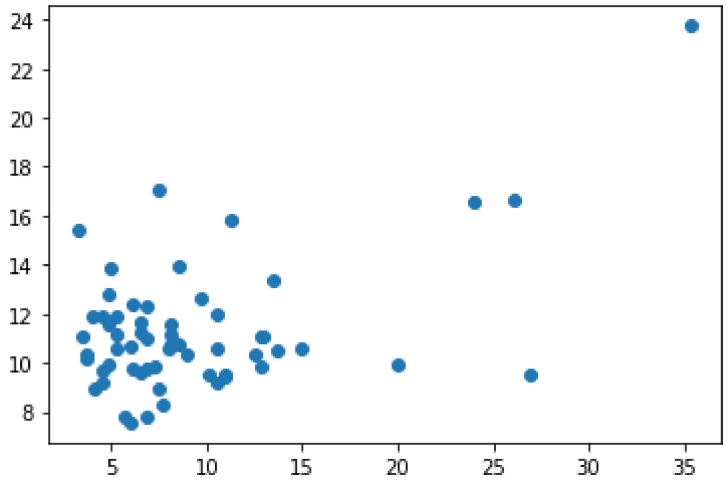
Out[26]:
```
LinearRegression()
```

In [27]:
```
#to find intercept
print(lr.intercept_)
```

```
[7.32346528]
```

In [28]:
```
prediction = lr.predict(x_test)
plt.scatter(y_test,prediction)
```

Out[28]:
```
<matplotlib.collections.PathCollection at 0x229a6a28fa0>
```



In [29]:
```
print(lr.score(x_test,y_test))
```

```
0.13177604652096764
```

RIDGE AND LASSO REGRESSION

In [30]:
```
from sklearn.linear_model import Ridge,Lasso
```

In [31]:
```
rr=Ridge(alpha=10)
rr.fit(x_train,y_train)
```

Out[31]:
```
Ridge(alpha=10)
```

In [32]:
```
rr.score(x_test,y_test)
```

Out[32]:
```
-0.10626904689186656
```

In [33]:
```
la=Lasso(alpha=10)
la.fit(x_train,y_train)
```

Out[33]:
```
Lasso(alpha=10)
```

In [34]:
```
la.score(x_test,y_test)
```

Out[34]:
```
-0.10198359066872098
```