```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df = pd.read_csv("uber.csv")
```

```
df.head()
```

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longit
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200000 entries, 0 to 199999
Data columns (total 9 columns):
 #   Column             Non-Null Count   Dtype  
---  ------             --------------   -----  
 0   Unnamed: 0         200000 non-null  int64  
 1   key                200000 non-null  object 
 2   fare_amount        200000 non-null  float64
 3   pickup_datetime    200000 non-null  object 
 4   pickup_longitude   200000 non-null  float64
 5   pickup_latitude    200000 non-null  float64
 6   dropoff_longitude  199999 non-null  float64
 7   dropoff_latitude   199999 non-null  float64
 8   passenger_count    200000 non-null  int64  
dtypes: float64(5), int64(2), object(2)
memory usage: 13.7+ MB
```

```
df=df.drop(['Unnamed: 0', 'key'], axis=1)
```

```
df.shape
```

```
(200000, 7)
```

```
df.dtypes
```

```
df.describe()
```

```
df.isnull().sum()
```

```
fare_amount          0
pickup_datetime      0
pickup_longitude     0
pickup_latitude      0
dropoff_longitude    1
dropoff_latitude     1
passenger_count      0
dtype: int64
```

```
df['dropoff_latitude'].fillna(value=df['dropoff_latitude'].mean(), inplace=True)
df['dropoff_longitude'].fillna(value=df['dropoff_longitude'].mean(), inplace=Tru
```

```
df.isnull().sum()
```

```
fare_amount          0
pickup_datetime      0
pickup_longitude     0
pickup_latitude      0
dropoff_longitude    0
dropoff_latitude     0
passenger_count      0
dtype: int64
```

```
corr=df.corr()
```

```
corr
```

```
<ipython-input-11-0a2117a8e592>:1: FutureWarning: The default value of nume
  corr=df.corr()
```

	fare_amount	**pickup_longitude**	**pickup_latitude**	**dropoff_l**
fare_amount	1.000000	0.010457	-0.008481	
pickup_longitude	0.010457	1.000000	-0.816461	
pickup_latitude	-0.008481	-0.816461	1.000000	-
dropoff_longitude	0.008986	0.833026	-0.774787	
dropoff_latitude	-0.011014	-0.846324	0.702367	-
passenger_count	0.010150	-0.000414	-0.001560	

```
x=df[['pickup_longitude', 'pickup_latitude', 'dropoff_longitude', 'dropoff_latit
y=df['fare_amount']
```

```
y=df['fare_amount']
```

```
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test=train_test_split(x,y,test_size=0.33)
```

```
from sklearn.linear_model import LinearRegression
regression=LinearRegression()
regression.fit(X_train, Y_train)
prediction=regression.predict(X_test)
print(prediction)
```

```
[11.36392632 11.29399232 11.2942145  ... 11.36430671 11.29528197
 11.29509535]
```

```
Y_test
```

```
145149     9.3
155648    11.0
124903     9.5
90823      4.5
174857     5.7
          ...
57547     14.0
57835      4.5
123562    13.0
76365     15.5
106005     9.3
Name: fare_amount, Length: 66000, dtype: float64
```

```
from sklearn.metrics import r2_score, mean_squared_error
print(r2_score(Y_test, prediction))
MSE=mean_squared_error(Y_test, prediction)
print(MSE)
print(np.sqrt(MSE))
```

```
1.203615155009885e-05
101.07593364758644
10.053652751492187
```

```
from sklearn.ensemble import RandomForestRegressor
rf=RandomForestRegressor(n_estimators=100)
rf.fit(X_train, Y_train)
y_pred=rf.predict(X_test)
print(y_pred)
```

```
[11.778 14.238  8.483 ... 10.907 15.34   7.024]
```

```
print(r2_score(Y_test, y_pred))
mean=mean_squared_error(Y_test, y_pred)
```

```
mser=mean_squared_error(Y_test, y_pred)
print(mser)
print(np.sqrt(mser))print(r2_score(Y_test, y_pred))
mser=mean_squared_error(Y_test, y_pred)
print(mser)
print(np.sqrt(mser))
```

```
0.7287275728764295
27.419443868943556
5.2363578820534755
```