

```
In [1]: import pandas as pd
```

```
In [2]: data=pd.read_csv("/home/placement/Downloads/fiat500.csv")
```

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

Out[3]:

| | ID | engine_power | age_in_days | km | previous_owners | lat | lon | price |
|--------------|-------------|--------------|-------------|---------------|-----------------|-------------|-------------|--------------|
| count | 1538.000000 | 1538.000000 | 1538.000000 | 1538.000000 | 1538.000000 | 1538.000000 | 1538.000000 | 1538.000000 |
| mean | 769.500000 | 51.904421 | 1650.980494 | 53396.011704 | 1.123537 | 43.541361 | 11.563428 | 8576.003901 |
| std | 444.126671 | 3.988023 | 1289.522278 | 40046.830723 | 0.416423 | 2.133518 | 2.328190 | 1939.958641 |
| min | 1.000000 | 51.000000 | 366.000000 | 1232.000000 | 1.000000 | 36.855839 | 7.245400 | 2500.000000 |
| 25% | 385.250000 | 51.000000 | 670.000000 | 20006.250000 | 1.000000 | 41.802990 | 9.505090 | 7122.500000 |
| 50% | 769.500000 | 51.000000 | 1035.000000 | 39031.000000 | 1.000000 | 44.394096 | 11.869260 | 9000.000000 |
| 75% | 1153.750000 | 51.000000 | 2616.000000 | 79667.750000 | 1.000000 | 45.467960 | 12.769040 | 10000.000000 |
| max | 1538.000000 | 77.000000 | 4658.000000 | 235000.000000 | 4.000000 | 46.795612 | 18.365520 | 11100.000000 |

```
In [4]: data1=data.drop(['ID','lat','lon'],axis=1) #unwanted columns removed
```

```
In [5]: data1
```

```
Out[5]:
```

| | model | engine_power | age_in_days | km | previous_owners | price |
|------|--------|--------------|-------------|--------|-----------------|-------|
| 0 | lounge | 51 | 882 | 25000 | 1 | 8900 |
| 1 | pop | 51 | 1186 | 32500 | 1 | 8800 |
| 2 | sport | 74 | 4658 | 142228 | 1 | 4200 |
| 3 | lounge | 51 | 2739 | 160000 | 1 | 6000 |
| 4 | pop | 73 | 3074 | 106880 | 1 | 5700 |
| ... | ... | ... | ... | ... | ... | ... |
| 1533 | sport | 51 | 3712 | 115280 | 1 | 5200 |
| 1534 | lounge | 74 | 3835 | 112000 | 1 | 4600 |
| 1535 | pop | 51 | 2223 | 60457 | 1 | 7500 |
| 1536 | lounge | 51 | 2557 | 80750 | 1 | 5990 |
| 1537 | pop | 51 | 1766 | 54276 | 1 | 7900 |

1538 rows × 6 columns

```
In [6]: data=pd.get_dummies(data)
```

In [7]: data

Out[7]:

| | ID | engine_power | age_in_days | km | previous_owners | lat | lon | price | model_lounge | model_pop | model_sport |
|------|------|--------------|-------------|--------|-----------------|-----------|-----------|-------|--------------|-----------|-------------|
| 0 | 1 | 51 | 882 | 25000 | 1 | 44.907242 | 8.611560 | 8900 | 1 | 0 | 0 |
| 1 | 2 | 51 | 1186 | 32500 | 1 | 45.666359 | 12.241890 | 8800 | 0 | 1 | 0 |
| 2 | 3 | 74 | 4658 | 142228 | 1 | 45.503300 | 11.417840 | 4200 | 0 | 0 | 1 |
| 3 | 4 | 51 | 2739 | 160000 | 1 | 40.633171 | 17.634609 | 6000 | 1 | 0 | 0 |
| 4 | 5 | 73 | 3074 | 106880 | 1 | 41.903221 | 12.495650 | 5700 | 0 | 1 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1533 | 1534 | 51 | 3712 | 115280 | 1 | 45.069679 | 7.704920 | 5200 | 0 | 0 | 1 |
| 1534 | 1535 | 74 | 3835 | 112000 | 1 | 45.845692 | 8.666870 | 4600 | 1 | 0 | 0 |
| 1535 | 1536 | 51 | 2223 | 60457 | 1 | 45.481541 | 9.413480 | 7500 | 0 | 1 | 0 |
| 1536 | 1537 | 51 | 2557 | 80750 | 1 | 45.000702 | 7.682270 | 5990 | 1 | 0 | 0 |
| 1537 | 1538 | 51 | 1766 | 54276 | 1 | 40.323410 | 17.568270 | 7900 | 0 | 1 | 0 |

1538 rows × 11 columns

In [8]: data.shape

Out[8]: (1538, 11)

In [9]: y=data['price']
x=data.drop('price',axis=1)

In [10]:

y

Out[10]:

```
0      8900
1      8800
2      4200
3      6000
4      5700
```

```
...
1533   5200
1534   4600
1535   7500
1536   5990
1537   7900
```

Name: price, Length: 1538, dtype: int64

In [11]:

x

Out[11]:

| | ID | engine_power | age_in_days | km | previous_owners | lat | lon | model_lounge | model_pop | model_sport |
|------|------|--------------|-------------|--------|-----------------|-----------|-----------|--------------|-----------|-------------|
| 0 | 1 | 51 | 882 | 25000 | 1 | 44.907242 | 8.611560 | 1 | 0 | 0 |
| 1 | 2 | 51 | 1186 | 32500 | 1 | 45.666359 | 12.241890 | 0 | 1 | 0 |
| 2 | 3 | 74 | 4658 | 142228 | 1 | 45.503300 | 11.417840 | 0 | 0 | 1 |
| 3 | 4 | 51 | 2739 | 160000 | 1 | 40.633171 | 17.634609 | 1 | 0 | 0 |
| 4 | 5 | 73 | 3074 | 106880 | 1 | 41.903221 | 12.495650 | 0 | 1 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1533 | 1534 | 51 | 3712 | 115280 | 1 | 45.069679 | 7.704920 | 0 | 0 | 1 |
| 1534 | 1535 | 74 | 3835 | 112000 | 1 | 45.845692 | 8.666870 | 1 | 0 | 0 |
| 1535 | 1536 | 51 | 2223 | 60457 | 1 | 45.481541 | 9.413480 | 0 | 1 | 0 |
| 1536 | 1537 | 51 | 2557 | 80750 | 1 | 45.000702 | 7.682270 | 1 | 0 | 0 |
| 1537 | 1538 | 51 | 1766 | 54276 | 1 | 40.323410 | 17.568270 | 0 | 1 | 0 |

1538 rows × 10 columns

```
In [12]: !pip install scikit-learn
```

```
Requirement already satisfied: scikit-learn in ./anaconda3/lib/python3.10/site-packages (1.2.1)  
Requirement already satisfied: threadpoolctl>=2.0.0 in ./anaconda3/lib/python3.10/site-packages (from scikit-learn) (2.2.0)  
Requirement already satisfied: joblib>=1.1.1 in ./anaconda3/lib/python3.10/site-packages (from scikit-learn) (1.1.1)  
Requirement already satisfied: numpy>=1.17.3 in ./anaconda3/lib/python3.10/site-packages (from scikit-learn) (1.23.5)  
Requirement already satisfied: scipy>=1.3.2 in ./anaconda3/lib/python3.10/site-packages (from scikit-learn) (1.10.0)
```

```
In [13]: 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,random_state=42)
```

```
In [14]: x_test.head(5)
```

Out[14]:

| | ID | engine_power | age_in_days | km | previous_owners | lat | lon | model_lounge | model_pop | model_sport |
|------|------|--------------|-------------|--------|-----------------|-----------|-----------|--------------|-----------|-------------|
| 481 | 482 | 51 | 3197 | 120000 | 2 | 40.174702 | 18.167629 | 0 | 1 | 0 |
| 76 | 77 | 62 | 2101 | 103000 | 1 | 45.797859 | 8.644440 | 0 | 1 | 0 |
| 1502 | 1503 | 51 | 670 | 32473 | 1 | 41.107880 | 14.208810 | 1 | 0 | 0 |
| 669 | 670 | 51 | 913 | 29000 | 1 | 45.778591 | 8.946250 | 1 | 0 | 0 |
| 1409 | 1410 | 51 | 762 | 18800 | 1 | 45.538689 | 9.928310 | 1 | 0 | 0 |

```
In [15]: x_train.shape
```

Out[15]: (1030, 10)

```
In [16]: x_train.head()
```

```
Out[16]:
```

| | ID | engine_power | age_in_days | km | previous_owners | lat | lon | model_lounge | model_pop | model_sport |
|-----|-----|--------------|-------------|-------|-----------------|-----------|----------|--------------|-----------|-------------|
| 527 | 528 | 51 | 425 | 13111 | 1 | 45.022388 | 7.58602 | 1 | 0 | 0 |
| 129 | 130 | 51 | 1127 | 21400 | 1 | 44.332531 | 7.54592 | 1 | 0 | 0 |
| 602 | 603 | 51 | 2039 | 57039 | 1 | 40.748241 | 14.52835 | 0 | 1 | 0 |
| 331 | 332 | 51 | 1155 | 40700 | 1 | 42.143860 | 12.54016 | 1 | 0 | 0 |
| 323 | 324 | 51 | 425 | 16783 | 1 | 41.903221 | 12.49565 | 1 | 0 | 0 |

```
In [17]: y_train.head()
```

```
Out[17]: 527    9990
129     9500
602     7590
331     8750
323     9100
Name: price, dtype: int64
```

```
In [18]: y_test.head()
```

```
Out[18]: 481     7900
76      7900
1502    9400
669     8500
1409    9700
Name: price, dtype: int64
```

```
In [19]: from sklearn.linear_model import LinearRegression
reg=LinearRegression() #creating object of LinearRegression
reg.fit(x_train,y_train) #training are fitting LR object using training data
```

```
Out[19]: ▼ LinearRegression
LinearRegression()
```

```
In [20]: ypred=reg.predict(x_test)
```

```
In [21]: ypred
9775.32352003, 10408.91107298, 5697.84384985, 10227.82627464,
9499.28759155, 8063.91598111, 5558.91344769, 9800.12594902,
10534.65340782, 6255.9526295, 9623.6837517, 9474.68837969,
10386.79158214, 9590.83795102, 9586.78339, 9678.64671073,
6877.78966053, 7815.97344878, 10292.13370401, 10416.92788043,
7338.21169569, 9995.55545352, 10491.24268381, 10629.34614975,
10375.07768021, 10067.07191604, 9568.3301962, 7811.49275809,
9221.47869229, 9935.53077254, 9888.11152759, 9905.19892591,
9447.64222531, 9682.52138303, 9634.69968187, 9811.31864154,
8756.03555754, 6193.65129714, 6354.87146951, 8207.3621036,
8598.2667194, 6487.89489493, 6832.90460176, 5301.78688917,
8187.42493969, 9828.98271296, 7806.33327627, 9782.73212406,
10085.78321491, 5680.95952529, 9926.93972284, 9950.95892668,
8098.05368723, 4565.24396633, 10709.71543595, 3901.21685234,
9783.97388187, 10514.70878639, 5702.77097363, 5515.28972165,
10534.65884212, 6791.82026062, 9022.87088921, 10542.0619477,
9388.22098306, 9969.20840168, 8529.3315146, 7909.48052462,
10371.21600688, 5343.40955708, 9793.46202119, 10248.73590685,
10350.88894338, 9418.6649538, 9246.75432375, 9726.77158038,
5646.60360194, 4954.59993355, 4854.00609399, 9667.14070802,
```

```
In [22]: from sklearn.metrics import r2_score
r2_score(y_test,ypred)
```

```
Out[22]: 0.8428319728488683
```

```
In [23]: from sklearn.metrics import mean_squared_error
mean_squared_error(ypred,y_test)
```

```
Out[23]: 577189.6736608233
```

```
In [24]: import math
a=577189.6736608233
print(math.sqrt(a))
```

```
759.7300005007195
```

```
In [25]: ypred
```

```
Out[25]: array([ 5819.19308764,  7248.82914161,  9741.8936974 ,  9798.98033074,
 10055.00624601,  9551.4955679 ,  9758.01743879, 10122.9778365 ,
  9654.9661814 ,  9251.1403257 , 10478.09512253,  7807.3005255 ,
  7705.15873781,  6295.63244894,  9545.40486313, 10422.92177704,
  9616.90811615,  7756.9171161 ,  4893.88454414, 10581.46142719,
 10465.24078346, 10443.29318231,  7518.43696046, 10028.21911459,
  6990.73118896,  8989.86900819,  4823.51364349,  6989.03118684,
  7822.83203734,  9683.17944083,  7344.21343132,  5341.43860798,
  5420.78405336,  5092.38401339,  8971.44357515,  5702.81242412,
  9920.16285466,  8334.58448277,  6220.93323723,  8389.23958511,
  9695.84208061,  6859.59630725,  9101.22635456, 10063.22592995,
  8621.83915759, 10175.06753933,  9063.21918346,  8867.24865352,
  7094.44228184,  9058.37693565,  9474.82390731, 10406.09102832,
 10112.65006224,  6820.90463865,  9700.36507783,  9382.18149429,
  9632.57617775, 10553.81356008,  9847.21129432,  7247.16814789,
  9990.23331336,  7084.23300123,  9977.34233656,  7245.01115798,
  6490.89305576,  9737.86785115,  9853.54349825,  8568.7125607 ,
  8506.81438703,  6484.69051659,  7883.1895563 ,  6870.28308427,
  8263.36833348, 10551.03496347,  7434.71134313,  8637.85174602,
  8762.87817027, 10010.47000077,  7324.60000000,  8527.72426022])
```



```
In [26]: Results=pd.DataFrame(columns=['price','predicate'])
Results['price']=y_test
Results['predicate']=ypred
Results=Results.reset_index()
Results['Id']=Results.index
Results.head(15)
```

Out[26]:

| | index | price | predicate | Id |
|----|-------|-------|--------------|----|
| 0 | 481 | 7900 | 5819.193088 | 0 |
| 1 | 76 | 7900 | 7248.829142 | 1 |
| 2 | 1502 | 9400 | 9741.893697 | 2 |
| 3 | 669 | 8500 | 9798.980331 | 3 |
| 4 | 1409 | 9700 | 10055.006246 | 4 |
| 5 | 1414 | 9900 | 9551.495568 | 5 |
| 6 | 1089 | 9900 | 9758.017439 | 6 |
| 7 | 1507 | 9950 | 10122.977837 | 7 |
| 8 | 970 | 10700 | 9654.966181 | 8 |
| 9 | 1198 | 8999 | 9251.140326 | 9 |
| 10 | 1088 | 9890 | 10478.095123 | 10 |

In []: