In [2]:
```
import pandas as pd
```

In [3]:
```
data=pd.read_csv("/home/placement/Desktop/divyasri/fiat500.csv")#reading the file
```

In [4]:
```
data.describe()
```

Out[4]:

	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 [5]:
```
data.head(10)#shows top 10 columns
```

Out[5]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
0	1	lounge	51	882	25000	1	44.907242	8.611560	8900
1	2	pop	51	1186	32500	1	45.666359	12.241890	8800
2	3	sport	74	4658	142228	1	45.503300	11.417840	4200
3	4	lounge	51	2739	160000	1	40.633171	17.634609	6000
4	5	pop	73	3074	106880	1	41.903221	12.495650	5700
5	6	pop	74	3623	70225	1	45.000702	7.682270	7900
6	7	lounge	51	731	11600	1	44.907242	8.611560	10750
7	8	lounge	51	1521	49076	1	41.903221	12.495650	9190
8	9	sport	73	4049	76000	1	45.548000	11.549470	5600
9	10	sport	51	3653	89000	1	45.438301	10.991700	6000

In [6]:
```
data1=data.drop(['ID','lat','lon'],axis=1)#removing specific columns from dataframe
```

In [7]:
```
data1
```

Out[7]:

	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 [8]:
```
data1=pd.get_dummies(data1) #encodes the string into bits
```

In [9]:
```
data1
```

Out[9]:

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

1538 rows × 8 columns

In [10]:
```
data1.shape #rows and columns
```

Out[10]:
```
(1538, 8)
```

In [11]:
```
y=data1['price']#predicted value  removed from dataframe
x=data1.drop(['price'],axis=1)
```

In [12]:
```
y#prices only will displays
```

Out[12]:
```
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 [13]:
```
x #no prices
```

Out[13]:

	engine_power	age_in_days	km	previous_owners	model_lounge	model_pop	model_sport
0	51	882	25000	1	1	0	0
1	51	1186	32500	1	0	1	0
2	74	4658	142228	1	0	0	1
3	51	2739	160000	1	1	0	0
4	73	3074	106880	1	0	1	0
...	...	...	...	...	...	...	...
1533	51	3712	115280	1	0	0	1
1534	74	3835	112000	1	1	0	0
1535	51	2223	60457	1	0	1	0
1536	51	2557	80750	1	1	0	0
1537	51	1766	54276	1	0	1	0

1538 rows × 7 columns

In [14]:
```
!pip3 install scikit-learn#to install sklearn library
```

```
ERROR: Invalid requirement: 'scikit-learn#to'
```

In [15]:
```
from sklearn.model_selection import train_test_split
```

In [59]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.33,random_state=42)#spliting into training an
```

In [60]:
```
x_test.head(5)#shows the testing column values
```

Out[60]:

	engine_power	age_in_days	km	previous_owners	model_lounge	model_pop	model_sport
481	51	3197	120000	2	0	1	0
76	62	2101	103000	1	0	1	0
1502	51	670	32473	1	1	0	0
669	51	913	29000	1	1	0	0
1409	51	762	18800	1	1	0	0

In [61]:
```
x_train.head(5)
```

Out[61]:

	engine_power	age_in_days	km	previous_owners	model_lounge	model_pop	model_sport
527	51	425	13111	1	1	0	0
129	51	1127	21400	1	1	0	0
602	51	2039	57039	1	0	1	0
331	51	1155	40700	1	1	0	0
323	51	425	16783	1	1	0	0

In [62]:
```
y_train.head(5)
```

Out[62]:
```
527    9990
129    9500
602    7590
331    8750
323    9100
Name: price, dtype: int64
```

In [63]:
```
y_test.head(5)
```

Out[63]:
```
481     7900
76      7900
1502    9400
669     8500
1409    9700
Name: price, dtype: int64
```

In [64]:
```
from sklearn.linear_model import LinearRegression
reg=LinearRegression()#creating object of LinearRegression
reg.fit(x_train,y_train)#training and fitting LR object using training data
```

Out[64]:
```
LinearRegression()
```
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [65]:
```
ypred=reg.predict(x_test)#prediction of values(x_test*reg)
```

In [66]:
```
ypred
```

```
array([10354.61956534,  8552.21002673,  6726.65446676,  9381.22662706,
        6520.9999373 , 10352.85155564,  9063.7534579 , 10456.89121831,
        9127.72470241,  9952.37340054,  8376.6975881 ,  9220.36267675,
       10036.24981328,  8418.65456209,  4717.7579531 , 10076.86950203,
       10017.8490121 , 10590.33289679, 10161.75393066,  4927.49556508,
        7276.18410037,  9678.26477249,  9764.65653403,  5643.53722047,
       10062.84554534,  5163.04602382,  8307.60791348,  7441.80993846,
        7868.82460983,  9725.36143983,  8669.20982667, 10447.15719448,
        7124.58453563,  9718.32989102,  8059.66615638,  7430.65975056,
       10425.57075395, 10364.18738085,  5433.2724385 ,  9102.40298437,
        9629.06913727, 10532.3506032 , 10129.42684118,  9149.48843328,
        6158.13422239,  9721.03634157, 10419.02236947,  8838.50241314,
        8182.78836676, 10012.21373766,  9468.92324529,  9904.31954667,
       10475.66003551, 10475.0702782 ,  9609.27020577,  8115.22501265,
       10439.02404036, 10363.81936482,  8720.0683498 ,  8274.3579289 ,
        6889.7195761 , 10191.45963957,  4819.0674709 ,  8814.11814085,
        5737.62378403, 10051.06593609,  8840.87520652, 10054.31165256,
        9686.269121  , 10463.56977746, 10133.15815395,  9762.80613855,
        9793.03056946,  6796.69068198,  9599.3262671 ,  8488.31539047,
        6705.66818403, 10307.58651641, 10045.18332239, 10120.36242166
```

In [67]:
```
from sklearn.metrics import r2_score#efficiency
r2_score(y_test,ypred)#y_test is actual value #ypred is predicted value
```

Out[67]: 0.8415526986865394

In [70]:
```
from sklearn.metrics import mean_squared_error #to calculate rmse
mean_squared_error(ypred,y_test)
```

Out[70]: 581887.727391353

In [75]:
```
import math
```

In [76]:
```
n=581887.727391353 #squareroot value
math.sqrt(n)
```

Out[76]: 762.8156575420782

In [77]:
```
y_test.head(5)
```

Out[77]:
```
481     7900
76      7900
1502    9400
669     8500
1409    9700
Name: price, dtype: int64
```

In [79]:
```
ypred
```

Out[79]:
```
array([ 5867.6503378 ,  7133.70142341,  9866.35776216,  9723.28874535,
       10039.59101162,  9654.07582608,  9673.14563045, 10118.70728123,
        9903.85952664,  9351.55828437, 10434.34963575,  7732.26255693,
        7698.67240131,  6565.95240435,  9662.90103518, 10373.20344286,
        9599.94844451,  7699.34400418,  4941.33017994, 10455.2719478 ,
       10370.51555682, 10391.60424404,  7529.06622456,  9952.37340054,
        7006.13845729,  9000.1780961 ,  4798.36770637,  6953.10376491,
        7810.39767825,  9623.80497535,  7333.52158317,  5229.18705519,
        5398.21541073,  5157.65652129,  8948.63632836,  5666.62365159,
        9822.1231461 ,  8258.46551788,  6279.2040404 ,  8457.38443276,
        9773.86444066,  6767.04074749,  9182.99904787, 10210.05195479,
        8694.90545226, 10328.43369248,  9069.05761443,  8866.7826029 ,
        7058.39787506,  9073.33877162,  9412.68162121, 10293.69451263,
       10072.49011135,  6748.5794244 ,  9785.95841801,  9354.09969973,
        9507.9444386 , 10443.01608254,  9795.31884316,  7197.84932877,
       10108.31707235,  7009.6597206 ,  9853.90699412,  7146.87414965,
        6417.69133992,  9996.97382441,  9781.18795953,  8515.83255277,
        8456.30006203,  6499.76668237,  7768.57829985,  6832.86406122,
        8347.96113362, 10439.02404036,  7356.43463051,  8562.56562053,
```

In [96]:
```
Results=pd.DataFrame(columns=['Price','Predicted'])
Results['Price']=y_test #price column
Results['Predicted']=ypred #predicted  column
Results=Results.reset_index()
Results['ID']=Results.index
```

In [97]:
```
Results.head(15)
```

Out[97]:

	index	Price	Predicted	ID
0	481	7900	5867.650338	0
1	76	7900	7133.701423	1
2	1502	9400	9866.357762	2
3	669	8500	9723.288745	3
4	1409	9700	10039.591012	4
5	1414	9900	9654.075826	5
6	1089	9900	9673.145630	6
7	1507	9950	10118.707281	7
8	970	10700	9903.859527	8
9	1198	8999	9351.558284	9
10	1088	9890	10434.349636	10
11	576	7990	7732.262557	11
12	965	7380	7698.672401	12
13	1488	6800	6565.952404	13
14	1432	8900	9662.901035	14

In [99]:
```
Results['diff']=Results.apply(lambda row: row.Price-row.Predicted,axis=1) #difference value
```

In [100]:
```
Results
```

Out[100]:

	index	Price	Predicted	ID	diff
0	481	7900	5867.650338	0	2032.349662
1	76	7900	7133.701423	1	766.298577
2	1502	9400	9866.357762	2	-466.357762
3	669	8500	9723.288745	3	-1223.288745
4	1409	9700	10039.591012	4	-339.591012
...	...	...	...	...	...
503	291	10900	10032.665135	503	867.334865
504	596	5699	6281.536277	504	-582.536277
505	1489	9500	9986.327508	505	-486.327508
506	1436	6990	8381.517020	506	-1391.517020
507	575	10900	10371.142553	507	528.857447

508 rows × 5 columns

In [ ]:
```

```