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

In [3]:
```
data=pd.read_csv("/home/placement/Downloads/fiat500.csv")
```

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

Out[4]:

	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

In [5]:
```
data1=data.drop(['lat','lon','ID'],axis=1)
```

In [6]:
```
data1.head()
```

Out[6]:

	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

In [7]:
```
data=pd.get_dummies(data)
```

In [8]:
```
data.shape
```

Out[8]:
```
(1538, 11)
```

In [ ]:
```

```

In [9]:
```
y=data['price']
```

In [10]:
```
x=data.drop('price',axis=1)
```

In [ ]:
```

```

In [11]:
```
y
```

Out[11]:
```
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 [12]:
```
from sklearn.model_selection import train_test_split
```

In [13]:
```
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.shape
```

Out[17]: (1030,)

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

Out[18]:
```
527    9990
129    9500
602    7590
331    8750
323    9100
Name: price, dtype: int64
```

In [19]:
```
from sklearn.linear_model import LinearRegression
reg=LinearRegression()
reg.fit(x_train,y_train)
```

Out[19]:
```
▾ LinearRegression

LinearRegression()
```

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

In [21]:
```
ypred
```

```
array([10420.91991521,  8559.36348376,  6802.41765474,  9463.60914351,
        6595.32443302, 10452.79948148,  8985.36620712, 10381.62641587,
        9044.61727199,  9858.25635967,  8431.44396488,  9332.03262887,
        9968.53895502,  8373.53182638,  4730.06355541, 10128.50049952,
        9951.43993841, 10514.34035875,  9980.65273426,  4951.73872798,
        7295.05742766,  9562.10832139,  9907.19829915,  5552.67362672,
        9997.10882667,  5143.10046698,  8426.58953237,  7496.77298917,
        7788.58793224,  9652.49535927,  8621.5141248 , 10372.95686992,
        7121.85593538,  9807.45515506,  8055.56244116,  7460.56572873,
       10280.14093316, 10476.37888795,  5412.86365493,  9159.69570754,
        9570.24451488, 10631.67197425, 10191.16722391,  9246.40215298,
        6018.78924109,  9615.95984878, 10520.04528994,  8915.29037502,
        8227.95081736,  9888.28916316,  9469.40296538,  9984.09733556,
       10405.22999333, 10510.19490347,  9686.14366923,  8096.65870148,
       10493.03817153, 10386.55525134,  8807.23351267,  8280.41931335,
        6855.39153451, 10228.17167191,  4741.4113269 ,  8847.08211454,
        5777.60377085, 10140.98307313,  8867.00881496, 10115.19879685,
        9581.17073619, 10485.09114495, 10213.14966572,  9750.35350756,
        9670.90795615,  6722.33834114,  9564.21623605,  8616.86940955,
        6677.30400889, 10373.25962634, 10171.86282428, 10201.5589761
```

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]:
```
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[25]:

	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
11	576	7990	7807.300526	11
12	965	7380	7705.158738	12
13	1488	6800	6295.632449	13
14	1432	8900	9545.404863	14

In [26]:
```
Results['diff_price']=Results.apply(lambda row:row.price-row.predicate,axis=1)
```

In [27]:
```
Results.head(15)
```

Out[27]:

	index	price	predicate	Id	diff_price
0	481	7900	5819.193088	0	2080.806912
1	76	7900	7248.829142	1	651.170858
2	1502	9400	9741.893697	2	-341.893697
3	669	8500	9798.980331	3	-1298.980331
4	1409	9700	10055.006246	4	-355.006246
5	1414	9900	9551.495568	5	348.504432
6	1089	9900	9758.017439	6	141.982561
7	1507	9950	10122.977837	7	-172.977837
8	970	10700	9654.966181	8	1045.033819
9	1198	8999	9251.140326	9	-252.140326
10	1088	9890	10478.095123	10	-588.095123
11	576	7990	7807.300526	11	182.699474
12	965	7380	7705.158738	12	-325.158738
13	1488	6800	6295.632449	13	504.367551
14	1432	8900	9545.404863	14	-645.404863

In [ ]:
```

```