

```
In [126]: import pandas as pd
```

```
In [127]: data=pd.read_csv("fiat500.csv")
```

```
In [128]: data
```

```
Out[128]:
```

	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
...	...	...	...	...	...	...	...	...	...
1533	1534	sport	51	3712	115280	1	45.069679	7.704920	5200
1534	1535	lounge	74	3835	112000	1	45.845692	8.666870	4600
1535	1536	pop	51	2223	60457	1	45.481541	9.413480	7500
1536	1537	lounge	51	2557	80750	1	45.000702	7.682270	5990
1537	1538	pop	51	1766	54276	1	40.323410	17.568270	7900

1538 rows × 9 columns

```
In [129]: data=data.drop(['lat','lon','ID'],axis=1)
```

```
In [130]: data
```

```
Out[130]:
```

	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 [131]: data2=pd.get_dummies(data)
```

In [132]: data

Out[132]:

	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 [133]: data
```

```
Out[133]:
```

	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 [134]: data2.shape  
#data['model']=data['model']
```

```
Out[134]: (1538, 8)
```

```
In [135]: data2
```

```
Out[135]:
```

	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 [136]: y=data2['price']  
x=data2.drop('price',axis=1)
```

```
In [137]: y
```

```
Out[137]: 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 [138]: 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 [139]: x_test.head(5)
```

```
Out[139]:
```

	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 [140]: x_train.shape
```

```
Out[140]: (1030, 7)
```

```
In [141]: y_train
```

```
Out[141]: 527      9990
          129      9500
          602      7590
          331      8750
          323      9100
          ...
          1130     10990
          1294      9800
          860      5500
          1459      9990
          1126      8900
          Name: price, Length: 1030, dtype: int64
```

```
In [142]: 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[142]: 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 [143]: ypred=reg.predict(x_test)
```

In [144]: ypred

```
10582.50828537, 7568.91955482, 6804.97705225, 8065.01292384,  
10310.29143419, 8836.34894739, 8390.05091229, 9582.13932508,  
  
9745.34784981, 10045.45021387, 10294.09872915, 7145.15315349,  
9727.85493167, 6281.78952194, 7901.36245623, 9387.9203723 ,  
5039.55649797, 9351.49777725, 9980.70844784, 10094.79341516,  
6359.24321991, 9856.10227211, 9099.07023804, 5234.05388382,  
5534.45288323, 4495.02309231, 10199.78432943, 10024.87037067,  
5465.58034188, 8520.72057674, 7034.71038647, 10054.65061446,  
10191.12067767, 6008.34860428, 9748.18097947, 9669.4333196 ,  
9145.3756075 , 9175.66562699, 10087.86753845, 9825.02990067,  
7340.29803785, 5083.8487301 , 9441.50914802, 10243.05490667,  
5556.42300245, 10676.01945733, 6126.99295838, 9845.16661356,  
9850.77978959, 7840.83596305, 6552.05146566, 9938.82104889,  
8327.79232274, 9119.62204137, 6111.83787367, 10410.00504522,  
6360.97695249, 8601.59209793, 8377.80258216, 9803.81343895,  
8285.09831762, 10091.75635129, 10003.86694939, 10028.60283146,  
10354.61956534, 8552.21002673, 6726.65446676, 9381.22662706,  
6520.9999373 , 10352.85155564, 9063.7534579 , 10456.89121831,  
9127.72470241, 9952.37340054, 8376.6975881 , 9220.36267675,
```

```
In [145]: #savedmodel=pickle.load(open(filename,'rb'))  
#x_test=[[1,75,1062,8000,1]]  
#savedmodel.predict(x_test)
```

```
In [146]: from sklearn.metrics import r2_score  
r2_score(y_test,ypred)
```

Out[146]: 0.8415526986865394

In [ ]:

```
In [147]: from sklearn.metrics import mean_squared_error  
mean_squared_error(ypred,y_test)
```

Out[147]: 581887.727391353



```
In [148]: #from sklearn.metrics import accuracy_score  
#accuracy_score(y_test,ypred)
```

```
In [151]: #results=pd.dataframe(coloumns=['actual','predicted'])  
#Results['Acutal']=y_test  
Results=pd.DataFrame(columns=['Price','Predicted'])  
Results['Price']=y_test  
Results['Predicted']=ypred  
#result['km']=x_test['km']  
Results=Results.reset_index()  
Results['Id']=Results.index  
Results.head(15)
```

Out[151]:

	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 [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: