

```
In [176]: import pandas as pd  
import numpy as np
```

```
In [177]: data=pd.read_csv('/home/placement/Downloads/fiat500.csv')
```

```
In [178]: data.head(10)
```

Out[178]:

| | 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 [179]: data.describe()
```

```
Out[179]:
```

| | 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 [180]: data=data.drop(['lat','ID','lon'],axis=1)
```

```
In [181]: data.describe()
```

```
Out[181]:
```

| | engine_power | age_in_days | km | previous_owners | price |
|--------------|--------------|-------------|---------------|-----------------|--------------|
| count | 1538.000000 | 1538.000000 | 1538.000000 | 1538.000000 | 1538.000000 |
| mean | 51.904421 | 1650.980494 | 53396.011704 | 1.123537 | 8576.003901 |
| std | 3.988023 | 1289.522278 | 40046.830723 | 0.416423 | 1939.958641 |
| min | 51.000000 | 366.000000 | 1232.000000 | 1.000000 | 2500.000000 |
| 25% | 51.000000 | 670.000000 | 20006.250000 | 1.000000 | 7122.500000 |
| 50% | 51.000000 | 1035.000000 | 39031.000000 | 1.000000 | 9000.000000 |
| 75% | 51.000000 | 2616.000000 | 79667.750000 | 1.000000 | 10000.000000 |
| max | 77.000000 | 4658.000000 | 235000.000000 | 4.000000 | 11100.000000 |

```
In [182]: data=pd.get_dummies(data)
```

```
In [183]: data
```

| | engine | price | horsepower | weight | displacement | price | horsepower | weight | displacement |
|------|--------|-------|------------|--------|--------------|-------|------------|--------|--------------|
| 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 [184]: data.shape
```

```
Out[184]: (1538, 8)
```

```
In [185]: y=data['price']  
x=data.drop(['price'],axis=1)
```

In [186]:

y

Out[186]:

| | |
|------|------|
| 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 [187]:

```
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 [188]:

x_test.head(5)

Out[188]:

| | 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 [189]:

x_test.shape

Out[189]: (508, 7)

```
In [190]: y_test.head(5)
```

```
Out[190]: 481      7900  
          76      7900  
          1502    9400  
          669    8500  
          1409    9700  
          Name: price, dtype: int64
```

```
In [191]: from sklearn.linear_model import LinearRegression  
          reg=LinearRegression()  
          reg.fit(x_train,y_train)
```

```
Out[191]: ▾ LinearRegression  
          LinearRegression()
```

```
In [192]: ypred=reg.predict(x_test)
```

In [193]: ypred

```
9791.34520178, 9779.16293972, 6753.27416058, 7354.16762745,
8760.24542762, 9923.66596418, 9812.92276721, 10466.90125415,
8163.46726237, 6659.46839415, 9987.65677522, 8866.7826029 ,
9952.37340054, 10187.72427693, 10231.39378767, 10091.11325493,
9365.98570732, 10009.10088406, 9141.00566394, 10099.11667176,
7803.77049829, 6009.84398185, 8800.33824151, 10237.60733785,
5609.98366311, 10097.61555355, 9684.99946572, 7644.67379732,
9276.37891542, 7371.5492091 , 10287.98873148, 10067.26428381,
10552.64805598, 9966.72383894, 10068.46126756, 6232.53552963,
10584.55044373, 9965.98687522, 10529.44404458, 9602.67646085,
9665.77720284, 6186.06948587, 8073.87436253, 10345.58323918,
6344.74803956, 7361.62678204, 10058.57116223, 6792.219309 ,
7897.72464823, 5261.45936067, 4540.24137423, 8709.36468047,
6882.0117409 , 7406.73353952, 6795.61189392, 7047.27998963,
9945.33400083, 8856.93910595, 9378.02074127, 10389.561154 ,
10092.46332921, 10381.52000388, 9723.92466625, 5996.3331428 ,
9786.14866981, 7708.49649098, 5583.48163469, 4932.92788329,
9856.66053994, 9236.22981005, 10092.64052142, 6256.43516278,
8592.63841379, 10341.5365957 , 5177.96595576, 10032.66513491,
6781.53627686 0086.327508 8281.51701051 10371.142553131\
```

In [194]: `from sklearn.metrics import r2_score`
`r2_score(y_test,ypred)`

Out[194]: 0.8415526986865394

In [195]: `from sklearn.metrics import mean_squared_error`
`b=mean_squared_error(ypred,y_test)`

In [196]: `srt=b**(1/2)`
`srt`

Out[196]: 762.8156575420782

```
In [197]: results=pd.DataFrame(columns=['price','predicted'])
results['price']=y_test
results['predicted']=ypred
#results=results.reset_index()
#results['Id']=results.index()
results.head(15)
```

Out[197]:

| | price | predicted |
|------|-------|--------------|
| 481 | 7900 | 5867.650338 |
| 76 | 7900 | 7133.701423 |
| 1502 | 9400 | 9866.357762 |
| 669 | 8500 | 9723.288745 |
| 1409 | 9700 | 10039.591012 |
| 1414 | 9900 | 9654.075826 |
| 1089 | 9900 | 9673.145630 |
| 1507 | 9950 | 10118.707281 |
| 970 | 10700 | 9903.859527 |
| 1198 | 8999 | 9351.558284 |
| 1088 | 9890 | 10434.349636 |
| 576 | 7990 | 7732.262557 |
| 965 | 7380 | 7698.672401 |
| 1488 | 6800 | 6565.952404 |
| 1432 | 8900 | 9662.901035 |

```
In [198]: import warnings
warnings.filterwarnings('ignore')
```