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

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

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

In [6]:
```
data1
```

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
...      ...           ...          ...     ...              ...    ...
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
```

Type *Markdown* and LaTeX: $\alpha^2$

In [7]:
```
data1['model']=data1['model'].map({'lounge':1,'pop':2,'sport':3})
```

In [8]:

```
data1
```

Out[8]:

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

1538 rows × 6 columns

In [9]:

```
y=data1['price']
x=data1.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]:

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

1538 rows × 5 columns

In [12]:
```
#!pip3 install scikit-learn
```

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]:

| | model | engine_power | age_in_days | km | previous_owners |
|---|---|---|---|---|---|
| **481** | 2 | 51 | 3197 | 120000 | 2 |
| **76** | 2 | 62 | 2101 | 103000 | 1 |
| **1502** | 1 | 51 | 670 | 32473 | 1 |
| **669** | 1 | 51 | 913 | 29000 | 1 |
| **1409** | 1 | 51 | 762 | 18800 | 1 |

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

Out[15]:
```
(1030, 5)
```

In [16]:
```
y_train.shape
```

Out[16]:
```
(1030,)
```

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

Out[17]:

| | model | engine_power | age_in_days | km | previous_owners |
|---|---|---|---|---|---|
| **527** | 1 | 51 | 425 | 13111 | 1 |
| **129** | 1 | 51 | 1127 | 21400 | 1 |
| **602** | 2 | 51 | 2039 | 57039 | 1 |
| **331** | 1 | 51 | 1155 | 40700 | 1 |
| **323** | 1 | 51 | 425 | 16783 | 1 |

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

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

In [19]:
```
x_test.head()
```

Out[19]:

| | model | engine_power | age_in_days | km | previous_owners |
|---|---|---|---|---|---|
| **481** | 2 | 51 | 3197 | 120000 | 2 |
| **76** | 2 | 62 | 2101 | 103000 | 1 |
| **1502** | 1 | 51 | 670 | 32473 | 1 |
| **669** | 1 | 51 | 913 | 29000 | 1 |
| **1409** | 1 | 51 | 762 | 18800 | 1 |

In [20]:
```
y_test.head()
```

Out[20]:
```
481     7900
76      7900
1502    9400
669     8500
1409    9700
Name: price, dtype: int64
```

In [21]:
```
from sklearn.linear_model import LinearRegression
reg=LinearRegression()
reg.fit(x_train,y_train)
```

Out[21]:
```
▾ LinearRegression

LinearRegression()
```

In [22]:
```
ypred=reg.predict(x_test)
```

In [23]:
```
ypred
```

```
         9767.12327701,  9755.03033027,  6859.84033207,  7339.68592914,
         8740.34003982,  9898.84623968,  9788.7072129 , 10439.74281794,
         8145.90808395,  6767.15633519,  9962.57850061,  8846.92420399,
         9927.58506055, 10279.88133318, 10205.11210182, 10065.46678709,
         9343.97683092,  9983.85933876,  9237.93178546, 10073.45985579,
         7906.63849672,  6017.75726035,  8780.77873324, 10211.55465771,
         5737.35007744, 10190.21750673,  9661.444679  ,  7747.41088806,
         9396.65945773,  7357.03908605, 10261.68730153, 10041.70922157,
        10525.09542651,  9941.6915233 , 10042.87112799,  6342.10368715,
        10588.92756092,  9940.98736563, 10501.95046891,  9697.00608104,
         9642.20441674,  6177.49903451,  8056.81304643, 10318.99744586,
         6334.90676093,  7347.76781534, 10049.18638926,  6780.85650138,
         7897.31981053,  5062.64376289,  4656.55980585,  8690.25433913,
         6988.39956167,  7416.44791638,  6784.57575877,  7034.60046808,
         9920.49385534,  8837.10014026,  9473.86037871, 10362.758227  ,
        10185.09072066, 10354.75677873,  9700.03561924,  6004.14282877,
         9762.0944756 ,  7693.58259057,  5694.99266265,  5047.53800681,
         9832.25820502,  9332.68577808, 10066.98651296,  6061.66190757,
         8691.8702504 , 10314.99712678,  5011.82707762, 10007.36463896,
         6390.17471479, 10079.47892753,  8363.33758536, 10344.4860771 ])
```

In [24]:
```
from sklearn.metrics import r2_score
r2_score(y_test,ypred)
```

Out[24]:
```
0.8383895235218546
```

In [25]:
```
from sklearn.metrics import mean_squared_error as ns
o=ns(y_test,ypred)
o
```

Out[25]:
```
593504.2888137395
```

In [26]:
```
import math
math.sqrt(o)
```

Out[26]:
```
770.3922954013361
```

```
In [27]: ypred
```

```
         9915.79926869,  8255.93615893,  6270.40332834,  8556.73835062,
         9749.72882426,  6873.76758364,  8951.72659758, 10301.95669828,
         8674.89268564, 10301.93257222,  9165.73586068,  8846.92420399,
         7044.68964545,  9052.4031418 ,  9390.75738772, 10267.3912561 ,
        10046.90924744,  6855.71260655,  9761.93338967,  9450.05744337,
         9274.98388541, 10416.00474283,  9771.10646661,  7302.96566423,
        10082.61483093,  6996.96553454,  9829.40534825,  7134.21944391,
         6407.26222178,  9971.82132188,  9757.01618446,  8614.84049875,
         8437.92452169,  6489.24658616,  7752.65456507,  6626.60510856,
         8329.88998217, 10412.00324329,  7342.77348105,  8543.63624413,
         9796.44743777, 10010.43582651,  7356.86786062,  9523.1488851 ,
        10326.22987622,  8076.40060487, 10419.77659259,  3528.47185472,
        10322.06476759, 10425.03337221,  6183.26049817, 10363.3075255 ,
         6517.32275335,  9095.0117622 , 10457.31199063,  9431.7474931 ,
         6698.67676865,  3459.16038547, 10198.89642316,  9768.32453463,
         6349.19977644,  4992.62377497,  9157.57378424,  9843.66900765,
         5300.60403791,  5481.29193836, 10061.18107826,  8216.74437963,
        10365.58751936,  6942.04538818,  6646.91921519,  5747.75873375,
         8993.90183905,  9927.58506055, 10363.24676269,  9396.86589768,
         9076.99796116, 10215.3021386 , 10383.15669206, 10126.03200513,
```

In [28]:

```
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[28]:

| | index | price | predicted | ID |
|---|---|---|---|---|
| **0** | 481 | 7900 | 5994.517032 | 0 |
| **1** | 76 | 7900 | 7263.587267 | 1 |
| **2** | 1502 | 9400 | 9841.907549 | 2 |
| **3** | 669 | 8500 | 9699.316277 | 3 |
| **4** | 1409 | 9700 | 10014.198926 | 4 |
| **5** | 1414 | 9900 | 9630.587158 | 5 |
| **6** | 1089 | 9900 | 9649.449903 | 6 |
| **7** | 1507 | 9950 | 10092.981966 | 7 |
| **8** | 970 | 10700 | 9879.194987 | 8 |
| **9** | 1198 | 8999 | 9329.193479 | 9 |
| **10** | 1088 | 9890 | 10407.296406 | 10 |
| **11** | 576 | 7990 | 7716.917060 | 11 |
| **12** | 965 | 7380 | 7682.891525 | 12 |
| **13** | 1488 | 6800 | 6673.958110 | 13 |
| **14** | 1432 | 8900 | 9639.426188 | 14 |

In [29]:

```
Results['price_diff']=Results.apply(lambda row: row.price - row.predicted,axis=1)
```

```
In [30]: Results

Out[30]:
       index  price     predicted   ID    price_diff
  0      481   7900   5994.517032    0   1905.482968
  1       76   7900   7263.587267    1    636.412733
  2     1502   9400   9841.907549    2   -441.907549
  3      669   8500   9699.316277    3  -1199.316277
  4     1409   9700  10014.198926    4   -314.198926
...      ...    ...           ...  ...           ...
503      291  10900  10007.364639  503    892.635361
504      596   5699   6390.174715  504   -691.174715
505     1489   9500  10079.478928  505   -579.478928
506     1436   6990   8363.337585  506  -1373.337585
507      575  10900  10344.486077  507    555.513923

508 rows × 5 columns

In [ ]:
```