In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```
data=pd.read_csv('/home/placement/Downloads/fiat500.csv')
```

In [3]:
```
data.head(10)
```

Out[3]:

	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 [4]:
```
data.isnull().sum()
```

Out[4]:
```
ID                0
model             0
engine_power      0
age_in_days       0
km                0
previous_owners   0
lat               0
lon               0
price             0
dtype: int64
```

In [5]:
```
data.describe()
```

Out[5]:

	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 [6]:
```
data=data.drop(['lat','ID','lon'],axis=1)
```

In [7]:
```
data.describe()
```

Out[7]:

	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 [8]:
```
data=pd.get_dummies(data)
```

In [9]:
```
data
```

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]:
```
!pip3 install scikit-learn
```

```
Requirement already satisfied: scikit-learn in ./anaconda3/lib/python3.10/site-packages (1.2.1)
Requirement already satisfied: numpy>=1.17.3 in ./anaconda3/lib/python3.10/site-packages (from scikit-lear
n) (1.23.5)
Requirement already satisfied: joblib>=1.1.1 in ./anaconda3/lib/python3.10/site-packages (from scikit-lear
n) (1.1.1)
Requirement already satisfied: scipy>=1.3.2 in ./anaconda3/lib/python3.10/site-packages (from scikit-learn)
(1.10.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in ./anaconda3/lib/python3.10/site-packages (from sciki
t-learn) (2.2.0)
```

In [11]:
```
data.shape
```

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

In [12]:
```
y=data['price']
x=data.drop(['price'],axis=1)
```

In [13]:
```
y
```

Out[13]:
```
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 [14]:
```
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 [15]:
```
x_test.head(5)
```

Out[15]:

	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 [16]:
```
x_test.shape
```

Out[16]:
```
(508, 7)
```

In [17]:
```
y_test.head(5)
```

Out[17]:
```
481     7900
76      7900
1502    9400
669     8500
1409    9700
Name: price, dtype: int64
```

Linear Regression

In [18]:
```
from sklearn.linear_model import LinearRegression
reg=LinearRegression()
reg.fit(x_train,y_train)
```

Out[18]:
```
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 [19]:
```
ypred=reg.predict(x_test)
```

In [20]:
```
ypred
```

Out[20]:
```
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 [21]:
```
from sklearn.metrics import r2_score
r2_score(y_test,ypred)
```

Out[21]:
```
0.8415526986865394
```

In [22]:
```
from sklearn.metrics import mean_squared_error
b=mean_squared_error(ypred,y_test)
```

In [23]:
```
srt=b**(1/2)
srt
```

Out[23]:
```
762.8156575420782
```

In [24]:
```
Results=pd.DataFrame(columns=['price'])
```

In [25]:
```
Results=pd.DataFrame(columns=['price','predicted'])
Results['price']=y_test
Results['predicted']=ypred
#result=results.reset_index()
#results['Id']=results.index()
Results.head(15)
```

Out[25]:

	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 [ ]:

In [ ]:

Ridge Regression

In [26]:
```
import warnings
warnings.filterwarnings('ignore')
```

In [27]:
```
from sklearn.model_selection import GridSearchCV
from sklearn.linear_model import Ridge

alpha = [1e-15,1e-10,1e-8,1e-4,1e-3,1e-2,1,5,10,20,30]

ridge=Ridge()

parameters={'alpha':alpha}
ridge_regressor=GridSearchCV(ridge,parameters)
ridge_regressor.fit(x_train,y_train)
```

Out[27]:
```
GridSearchCV(estimator=Ridge(),
             param_grid={'alpha': [1e-15, 1e-10, 1e-08, 0.0001, 0.001, 0.01, 1,
                                   5, 10, 20, 30]})
```

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 [28]:
```
ridge_regressor.best_params_
```

Out[28]:
```
{'alpha': 30}
```

In [29]:
```
ridge=Ridge(alpha=30)
ridge.fit(x_train,y_train)
y_pred_ridge=ridge.predict(x_train)
```

In [30]:
```
Ridge_Error=mean_squared_error(y_pred_ridge,y_train)
Ridge_Error
```

Out[30]:
```
584715.8955379989
```

In [31]:
```
from sklearn.metrics import r2_score
r2_score(y_train,y_pred_ridge)
```

Out[31]:
```
0.8462648330826763
```

In [32]:
```
Results=pd.DataFrame(columns=['Actual','Predicted'])
Results['Actual']=y_train
Results['Predicted']=y_pred_ridge
#Results['km']=x_train['km']
Result=Results.reset_index()
Results['Id']=Results.index
Results.head(10)
```

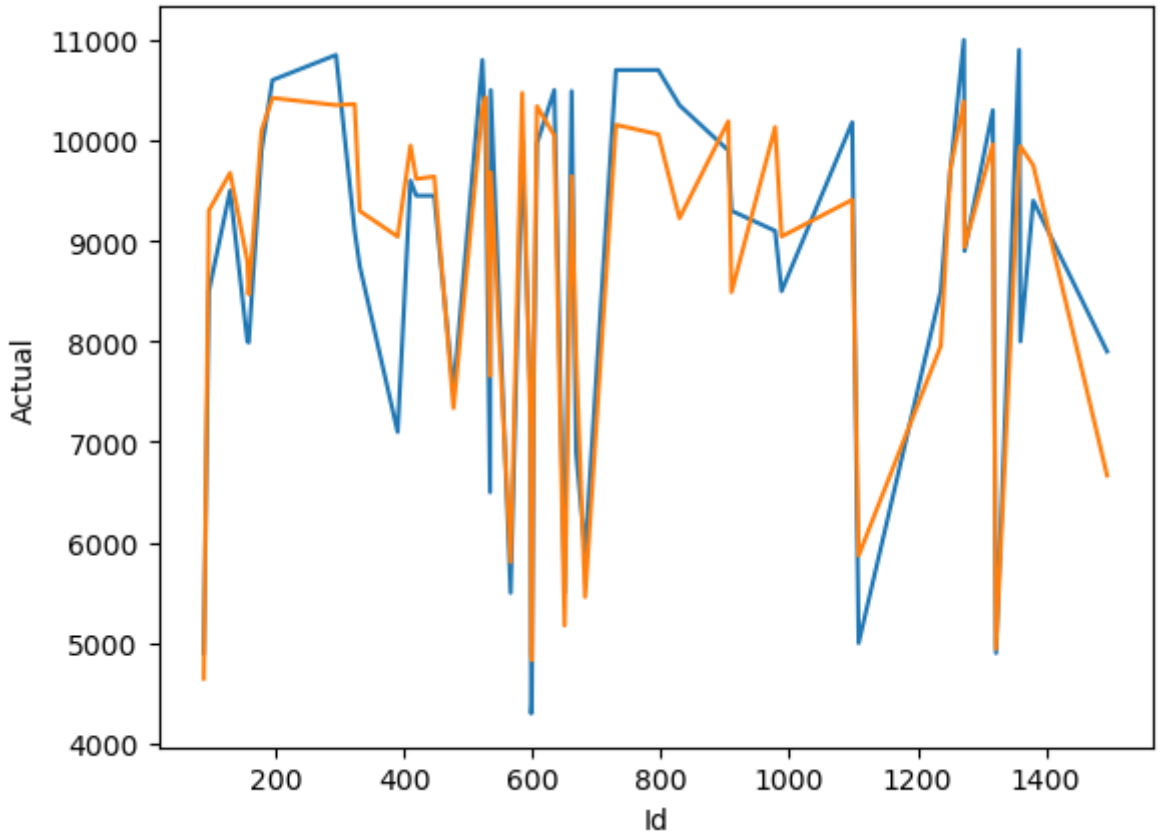
Out[32]:

	Actual	Predicted	Id
527	9990	10428.252278	527
129	9500	9676.489234	129
602	7590	7938.112518	602
331	8750	9297.389664	331
323	9100	10360.672371	323
1358	10900	9841.216966	1358
522	10800	10330.121542	522
584	9999	10472.654629	584
1236	8500	7955.649421	1236
535	10500	9682.904429	535

In [33]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

In [34]:
```
sns.lineplot(x='Id',y='Actual',data=Results.head(50))
sns.lineplot(x='Id',y='Predicted',data=Results.head(50))
plt.plot()
```

Out[34]: []



Elastic Regressor

In [35]:
```
from sklearn.linear_model import ElasticNet

elastic=ElasticNet()

parameters={'alpha':[1e-15,1e-10,1e-8,1e-4,1e-3,1e-2,1,5,10,20]}

elastic_regressor=GridSearchCV(elastic,parameters)

elastic_regressor.fit(x_train,y_train)
```

Out[35]:
```
GridSearchCV(estimator=ElasticNet(),
             param_grid={'alpha': [1e-15, 1e-10, 1e-08, 0.0001, 0.001, 0.01, 1,
                                   5, 10, 20]})
```
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 [36]:
```
elastic_regressor.best_params_
```

Out[36]:
```
{'alpha': 0.01}
```

In [37]:
```
elastic=ElasticNet(alpha=0.1)
elastic.fit(x_train,y_train)
y_pred_elastic=elastic.predict(x_test)
```

In [38]:
```
from sklearn.metrics import r2_score
r2_score(y_test,y_pred_elastic)
```

Out[38]:
```
0.8425222843073693
```

In [39]:
```
elastic_Error=mean_squared_error(y_pred_elastic,y_test)
elastic_Error
```

Out[39]:
```
578326.9853103004
```

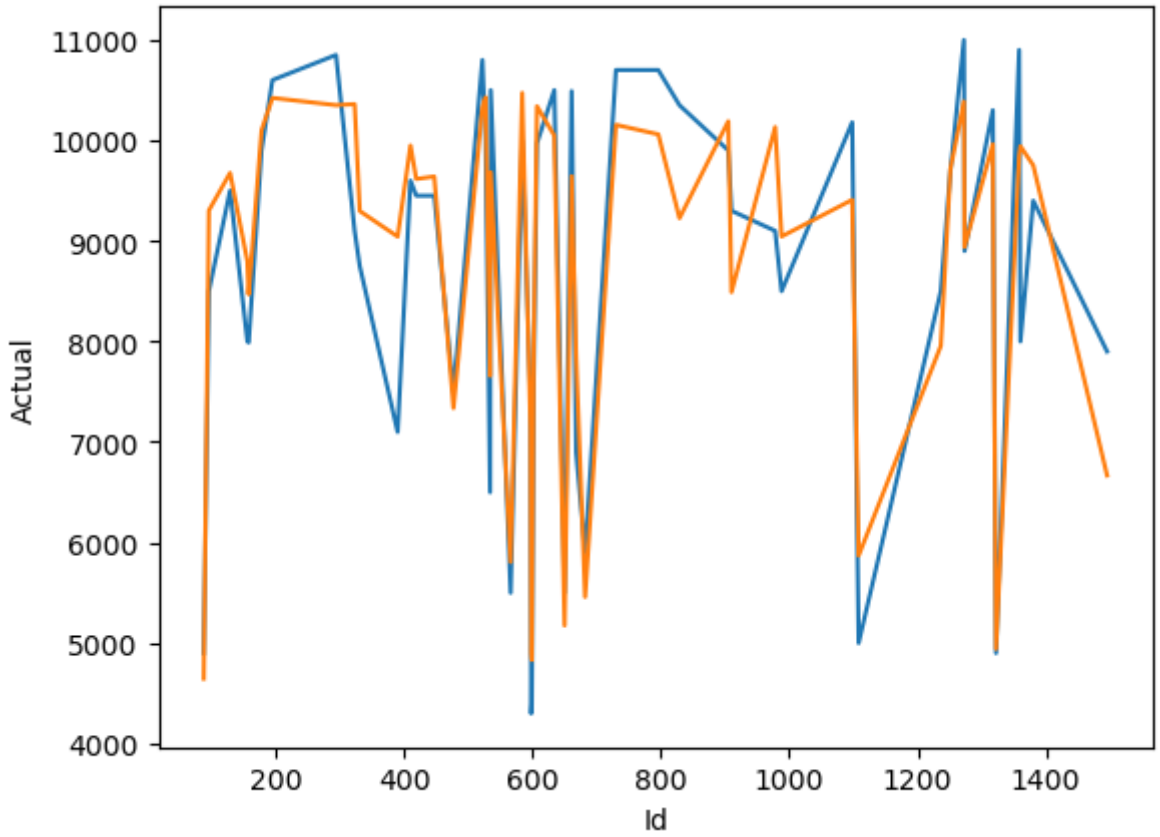
In [40]:
```
Results=pd.DataFrame(columns=['Actual','Predicted'])
Results['Actual']=y_train
Results['Predicted']=y_pred_ridge
#Results['km']=x_train['km']
Result=Results.reset_index()
Results['Id']=Results.index
Results.head(10)
```

Out[40]:

	Actual	Predicted	Id
527	9990	10428.252278	527
129	9500	9676.489234	129
602	7590	7938.112518	602
331	8750	9297.389664	331
323	9100	10360.672371	323
1358	10900	9841.216966	1358
522	10800	10330.121542	522
584	9999	10472.654629	584
1236	8500	7955.649421	1236
535	10500	9682.904429	535

In [41]:
```
sns.lineplot(x='Id',y='Actual',data=Results.head(50))
sns.lineplot(x='Id',y='Predicted',data=Results.head(50))
plt.plot()
```

Out[41]: []



In [ ]:

In [ ]:

In [ ]: