COLLEGE CODE: 9133

COURSE:Artificial intelligence

PHASE 3:Development Part 1

PROJECT TITLE: House price predictor

Team Members:

maalinivcse2021@gmail.com - Maalini.V

pavithrancse2021@gmail.com - Pavithra.N

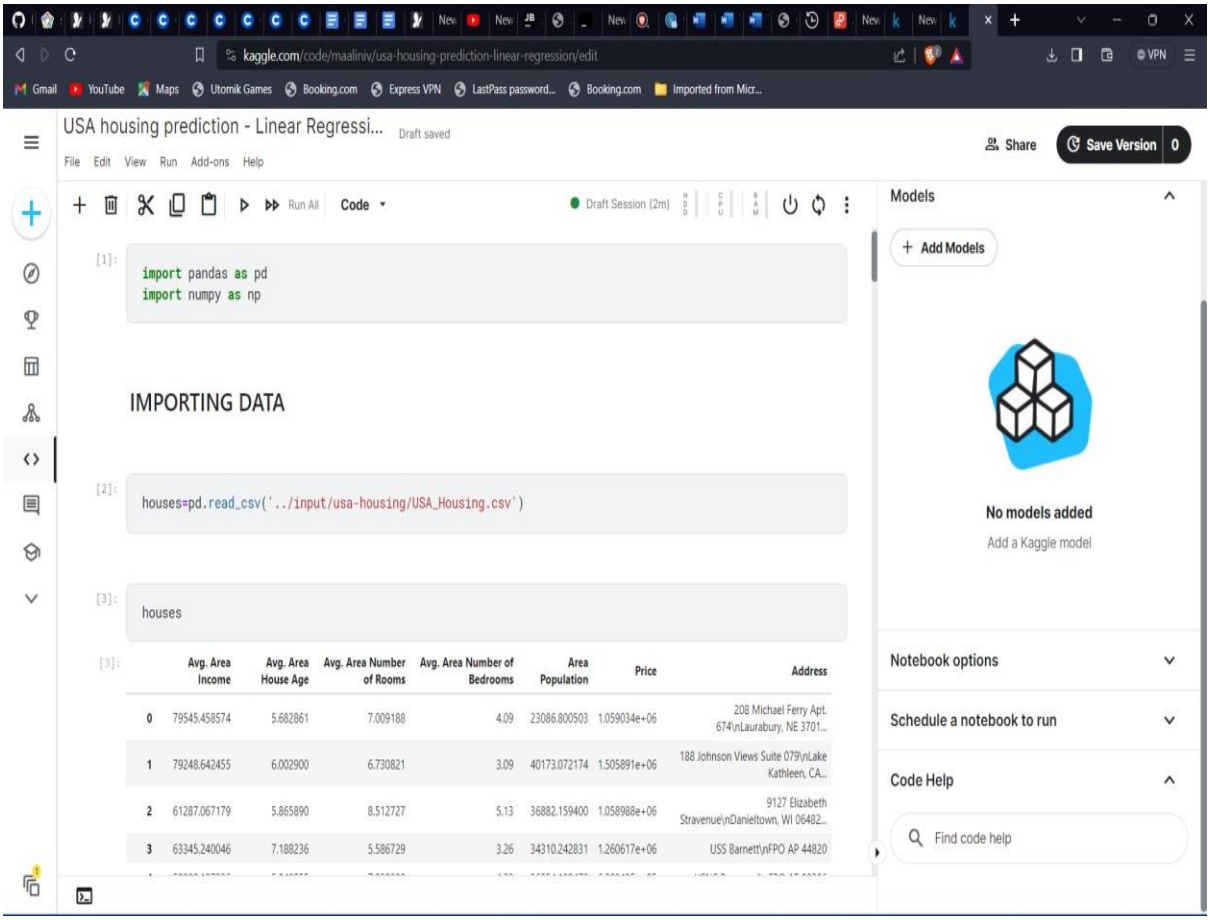
divyadharshiniscse2021@gmail.com - Divyadharshini.S

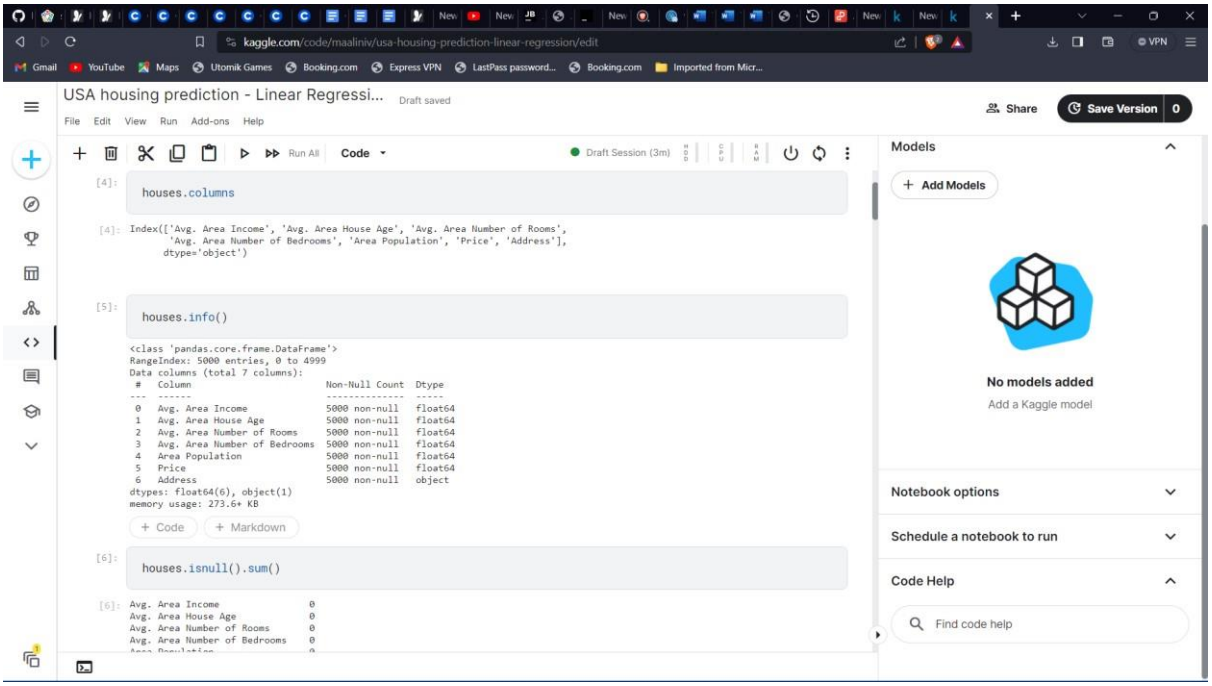
manoranjithampcse2021@gmail.com - Manoranjitham.P

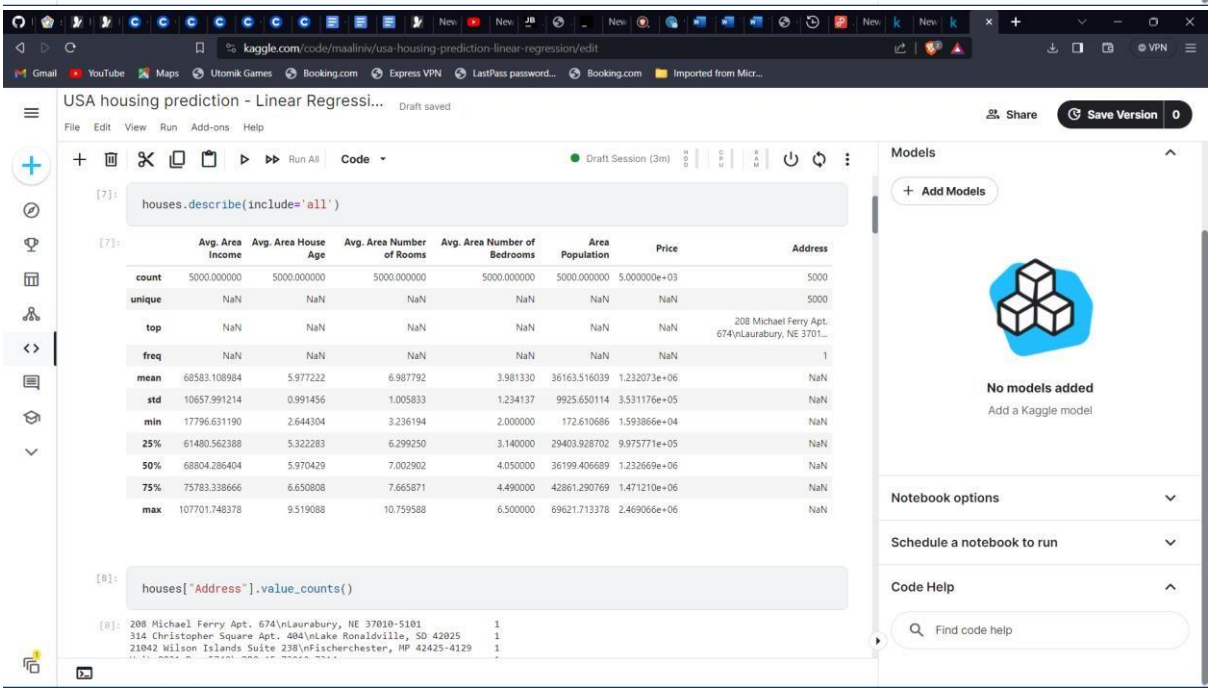
priyadharshiniscse2021@gmail.com - Priyadharshini.S

DATASET: https://www.kaggle.com/datasets/vedavyasv/usa-housin

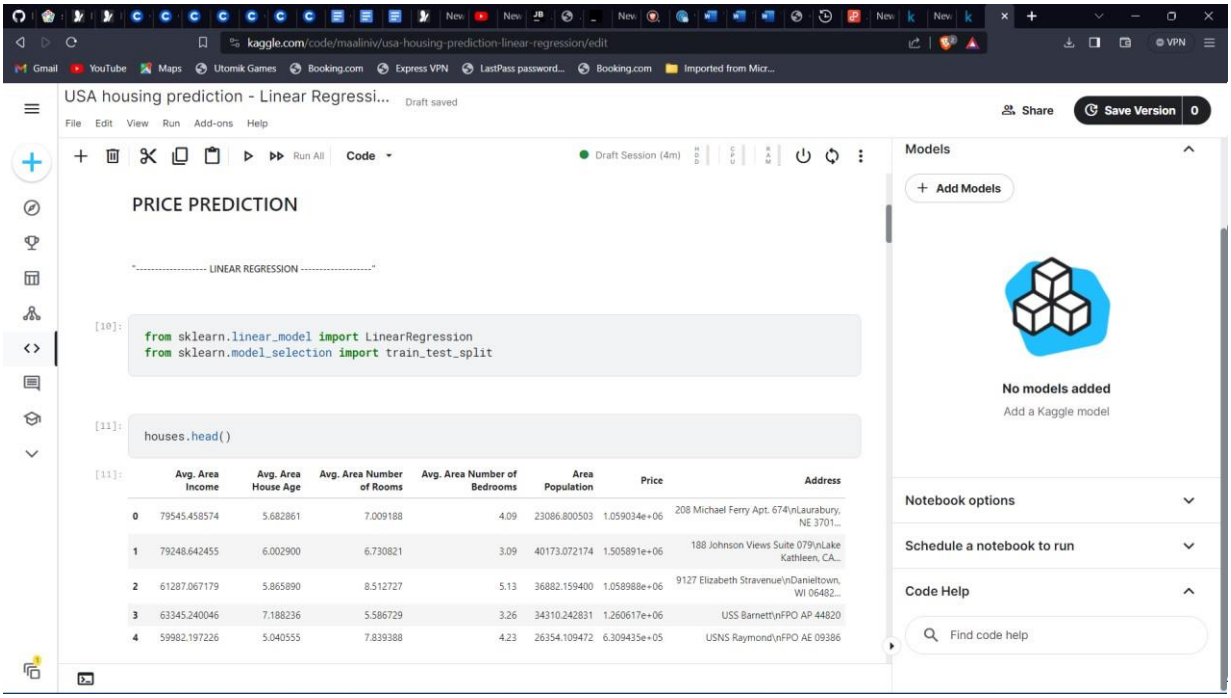
IMPORTING DATA:

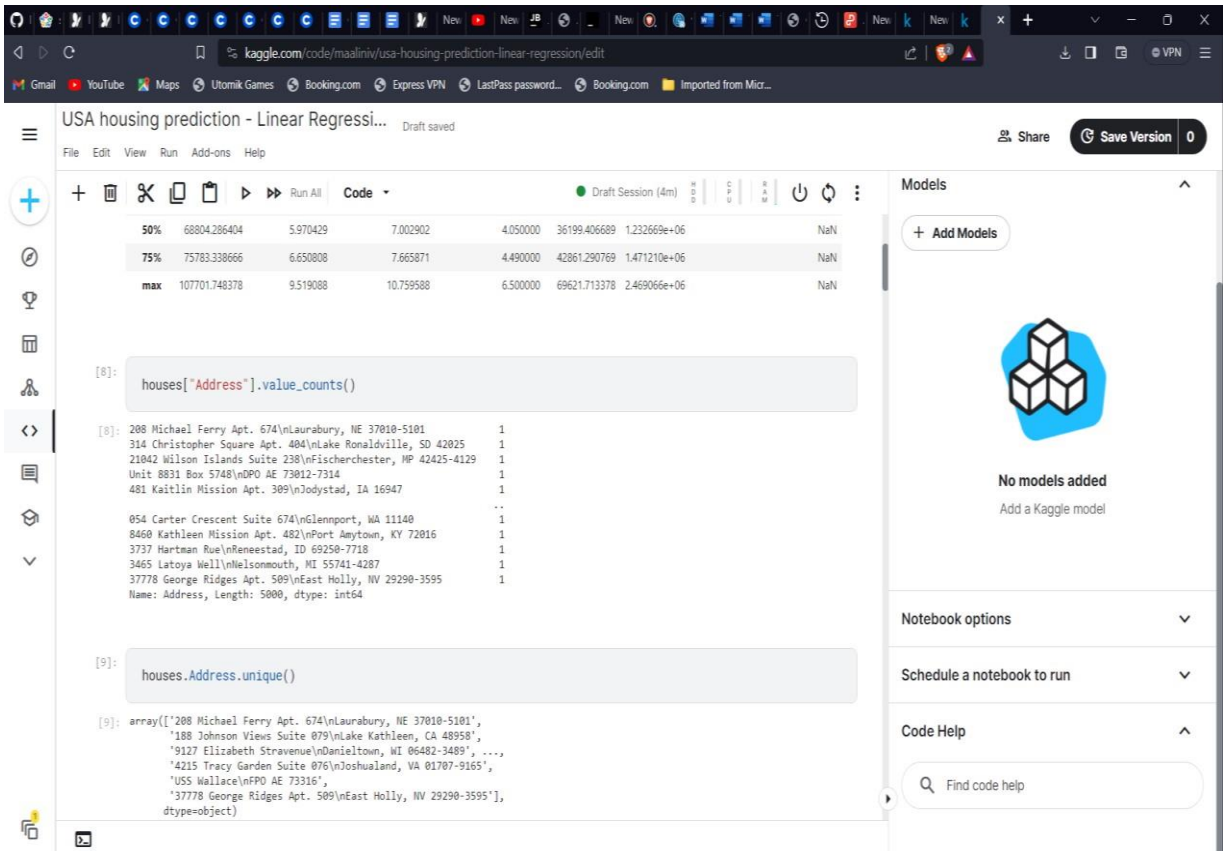


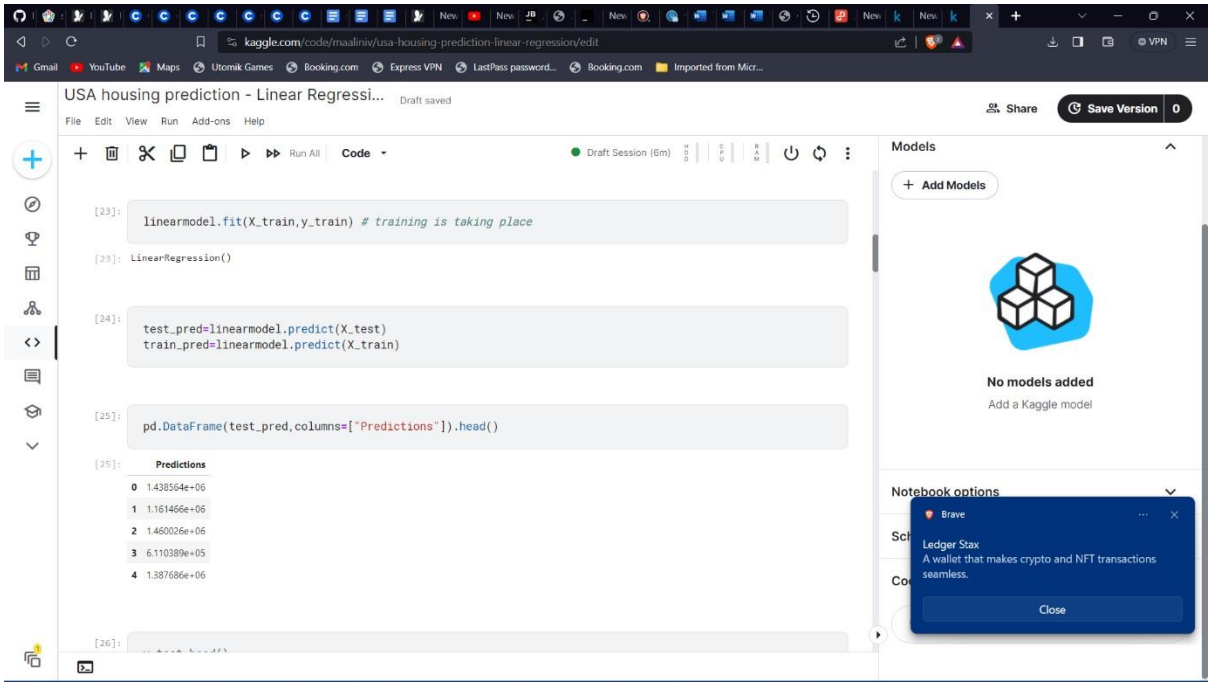


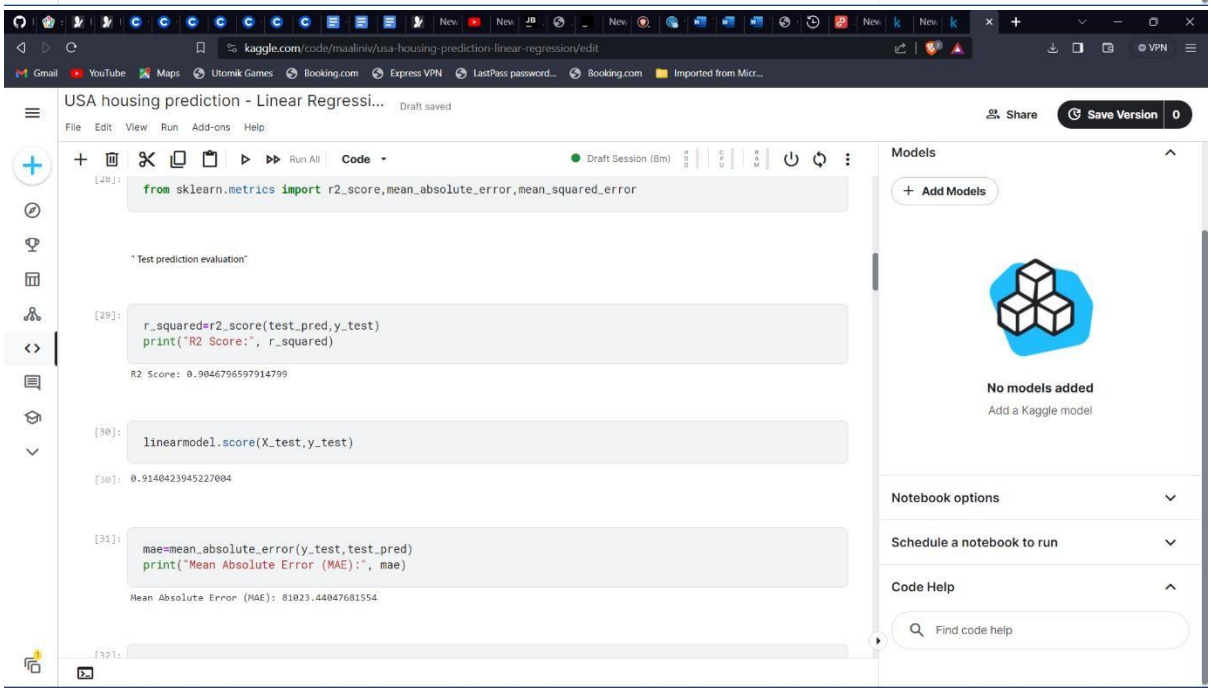


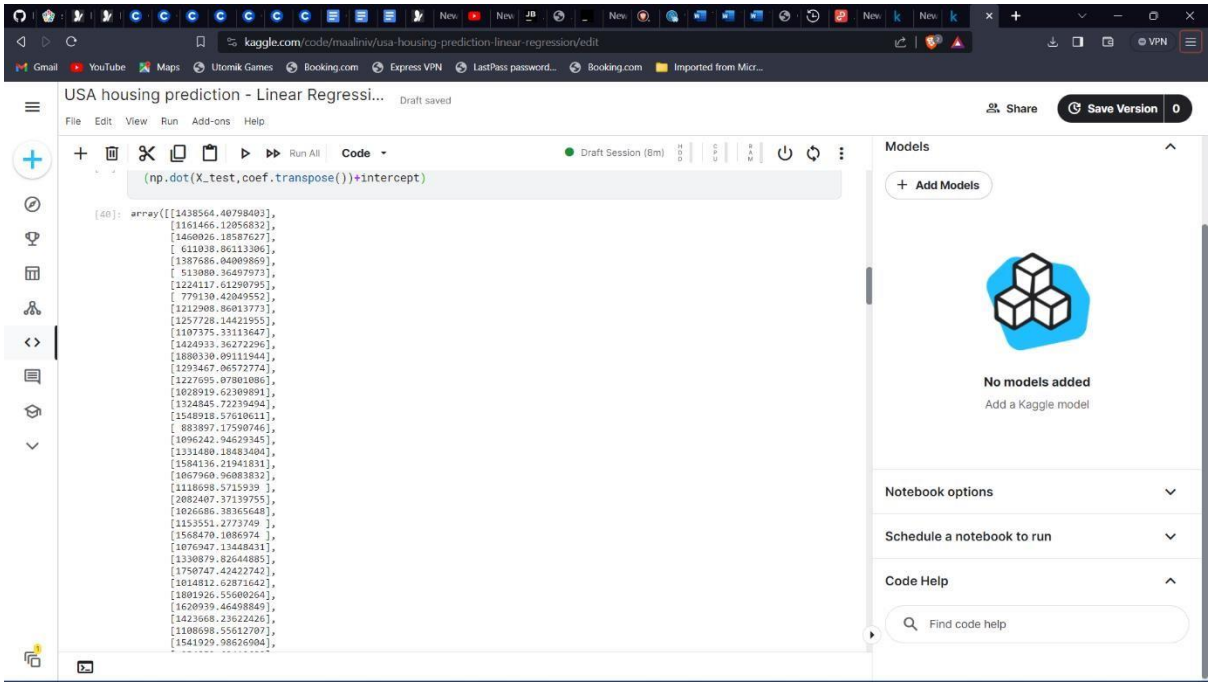
PRICE PREDICTION:

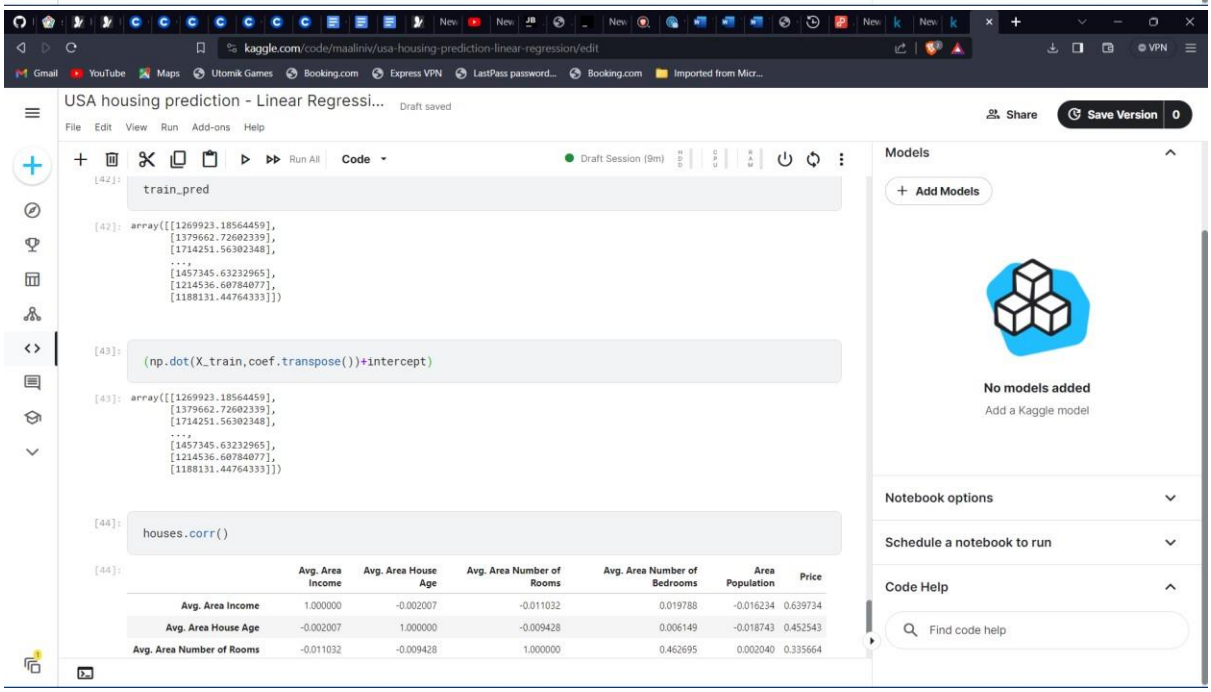

kaggle.com/code/maaliniv/usa-housing-prediction-linear-regression/edit
USA housing prediction - Linear Regressi...
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (4m)
PRICE PREDICTION
"---------- LINEAR REGRESSION ----------"
[10]:
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
[11]:
houses.head()
[11]:
Avg. Area Income
Avg. Area House Age
Avg. Area Number of Rooms
Avg. Area Number of Bedrooms
Area Population
Price
Address
0 79545.458574 5.682861 7.009188 4.09 23086.800503 1.059034e+06 208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
1 79248.642455 6.002900 6.730821 3.09 40173.072174 1.505891e+06 188 Johnson Views Suite 079\nLake Kathleen, CA...
2 61287.067179 5.865890 8.512727 5.13 36882.159400 1.058988e+06 9127 Elizabeth Stravenue\nDanieltown, WI 06482...
3 63345.240046 7.188236 5.586729 3.26 34310.242831 1.260617e+06 USS Barnett\nFPO AP 44820
4 59982.197226 5.040555 7.839388 4.23 26354.109472 6.309435e+05 USNS Raymond\nFPO AE 09386
Models
Add Models
No models added
Add a Kaggle model
Notebook options
Schedule a notebook to run
Code Help
Find code help

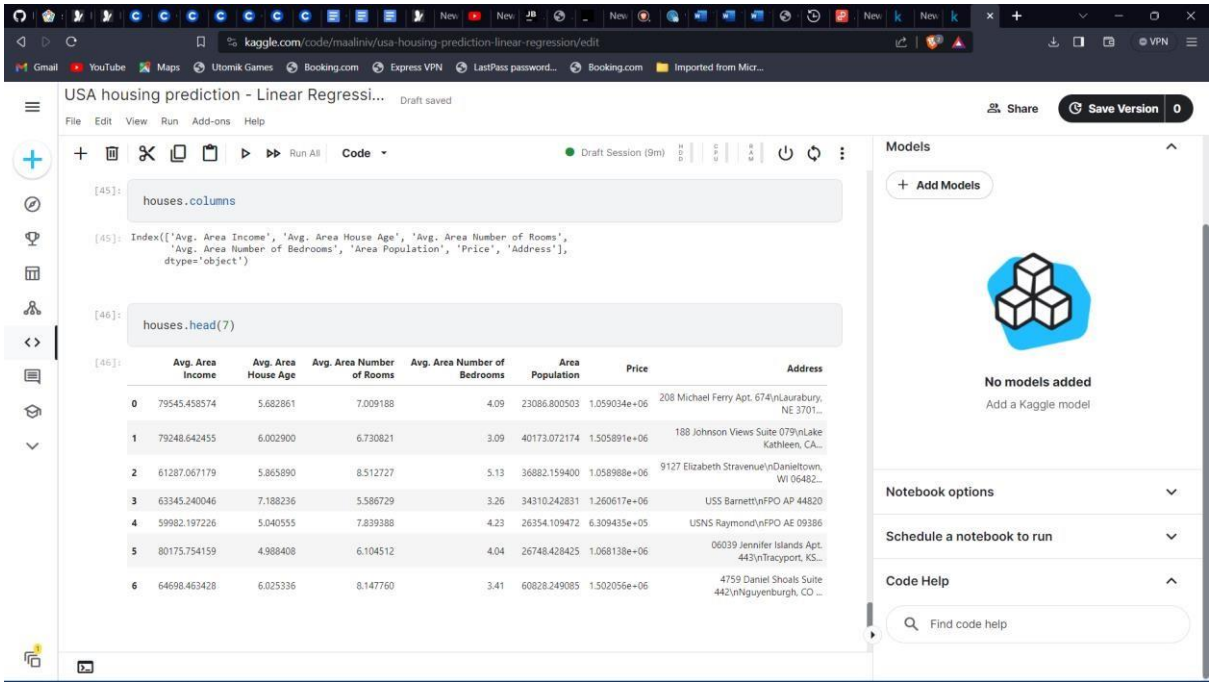

kaggle.com/code/maaliniv/usa-housing-prediction-linear-regression/edit
USA housing prediction - Linear Regressi...
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (4m)
50% 68804.286404 5.970429 7.002902 4.050000 36199.406689 1.232669e+06 NaN
75% 75783.338666 6.650808 7.665871 4.490000 42861.290769 1.471210e+06 NaN
max 107701.748378 9.519088 10.759588 6.500000 69621.713378 2.469066e+06 NaN
[8]:
houses["Address"].value_counts()
[8]:
208 Michael Ferry Apt. 674\nLaurabury, NE 37010-5101 1
314 Christopher Square Apt. 404\nLake Ronaldville, SD 42025 1
21042 Wilson Islands Suite 238\nFischerchester, MP 42425-4129 1
Unit 8831 Box 5748\nDPO AE 73012-7314 1
481 Kaitlin Mission Apt. 309\nJodystad, IA 16947 1
..
054 Carter Crescent Suite 674\nGlennport, WA 11140 1
8460 Kathleen Mission Apt. 482\nPort Amytown, KY 72016 1
3737 Hartman Rue\nReneestad, ID 69250-7718 1
3465 Latoya Well\nNelsonmouth, MI 55741-4287 1
37778 George Ridges Apt. 509\nEast Holly, NV 29290-3595 1
Name: Address, Length: 5000, dtype: int64
[9]:
houses.Address.unique()
[9]:
array(['208 Michael Ferry Apt. 674\nLaurabury, NE 37010-5101',
'188 Johnson Views Suite 079\nLake Kathleen, CA 48958',
'9127 Elizabeth Stravenue\nDanieltown, WI 06482-3489', ...,
'4215 Tracy Garden Suite 076\nJoshualand, VA 01707-9165',
'USS Wallace\nFPO AE 73316',
'37778 George Ridges Apt. 509\nEast Holly, NV 29290-3595'],
dtype=object)
Models
Add Models
No models added
Add a Kaggle model
Notebook options
Schedule a notebook to run
Code Help
Find code help

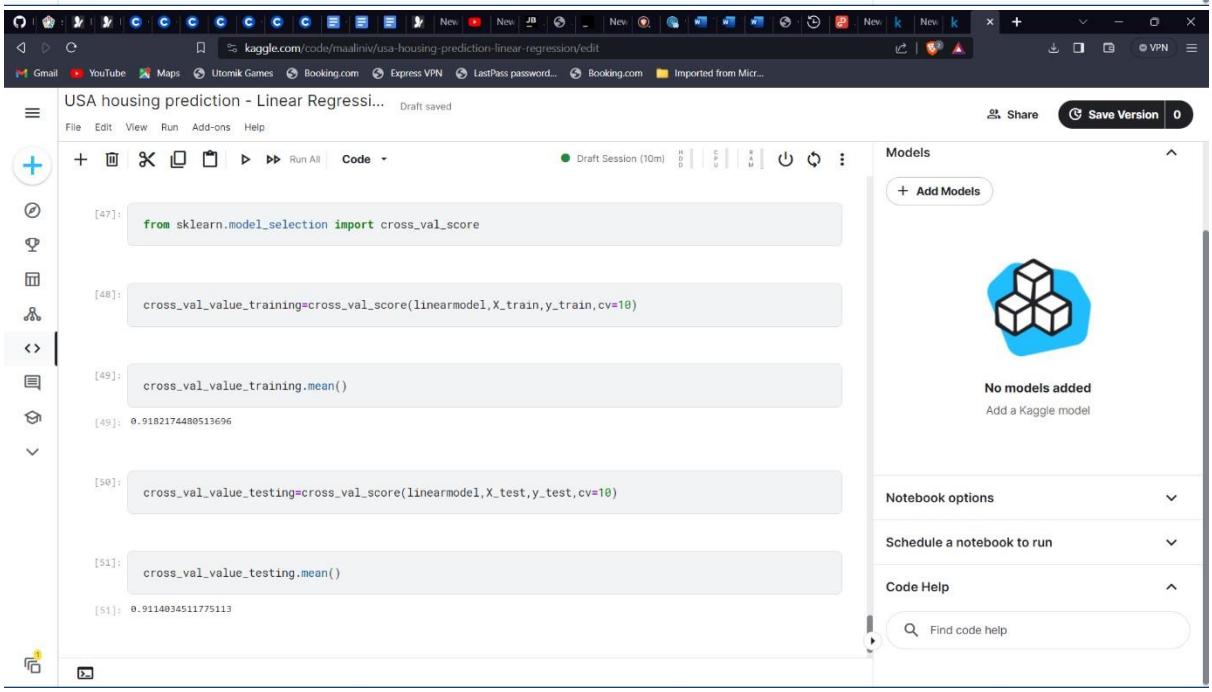

kaggle.com/code/maaliniv/usa-housing-prediction-linear-regression/edit
USA housing prediction - Linear Regressi...
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (6m)
[23]: linearmodel.fit(X_train,y_train) # training is taking place
[23]: LinearRegression()
[24]: test_pred=linearmodel.predict(X_test)
train_pred=linearmodel.predict(X_train)
[25]: pd.DataFrame(test_pred,columns=["Predictions"]).head()
[25]:
Predictions
0 1.438564e+06
1 1.161466e+06
2 1.460026e+06
3 6.110389e+05
4 1.387686e+06
[26]:
Models
Add Models
No models added
Add a Kaggle model
Notebook options
Brave
Ledger Stax
A wallet that makes crypto and NFT transactions seamless.
Close

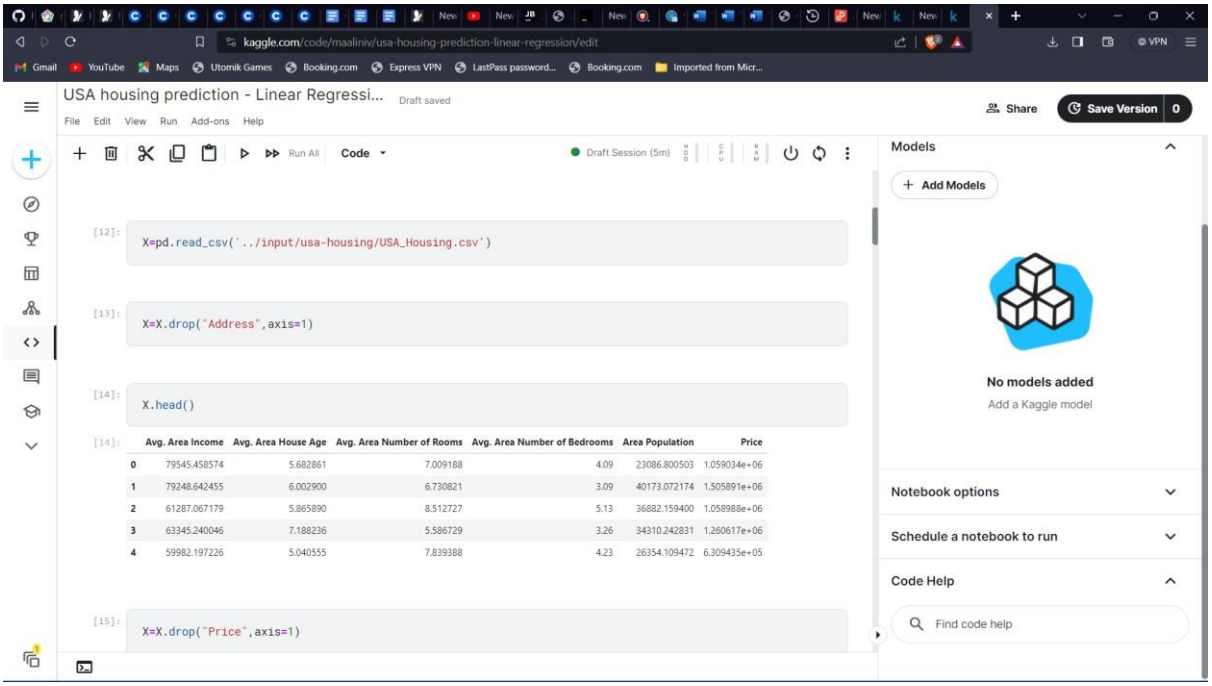

kaggle.com/code/maaliniv/usa-housing-prediction-linear-regression/edit
USA housing prediction - Linear Regressi...
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (8m)
[28]: from sklearn.metrics import r2_score,mean_absolute_error,mean_squared_error
" Test prediction evaluation"
[29]: r_squared=r2_score(test_pred,y_test)
print("R2 Score:", r_squared)
R2 Score: 0.9046796597914799
[30]: linearmodel.score(X_test,y_test)
[30]: 0.9140423945227004
[31]: mae=mean_absolute_error(y_test,test_pred)
print("Mean Absolute Error (MAE):", mae)
Mean Absolute Error (MAE): 81023.44047681554
Models
Add Models
No models added
Add a Kaggle model
Notebook options
Schedule a notebook to run
Code Help
Find code help


kaggle.com/code/maaliniv/usa-housing-prediction-linear-regression/edit
USA housing prediction - Linear Regressi...
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (8m)
(np.dot(X_test,coef.transpose())+intercept)
[40]:
array([[1438564.40798403],
[1161466.12056832],
[1460026.18587627],
[611038.86113306],
[1387686.04009869],
[513080.36497973],
[1224117.61290795],
[779130.42049552],
[1212908.86013773],
[1257728.14421955],
[1107375.33113647],
[1424933.36272296],
[1880330.09111944],
[1293467.06572774],
[1227695.07801086],
[1028919.62309891],
[1324845.72239494],
[1548918.57610611],
[883897.17590746],
[1096242.94629345],
[1331480.18483404],
[1584136.21941831],
[1067960.96083832],
[1118698.5715939],
[2082407.37139755],
[1026686.38365648],
[1153551.2773749],
[1568470.1086974],
[1076947.13448431],
[1330879.82644885],
[1750747.42422742],
[1014812.62871642],
[1801926.55600264],
[1620939.46498849],
[1423668.23622426],
[1108698.55612707],
[1541929.98626904],
Models
Add Models
No models added
Add a Kaggle model
Notebook options
Schedule a notebook to run
Code Help
Find code help

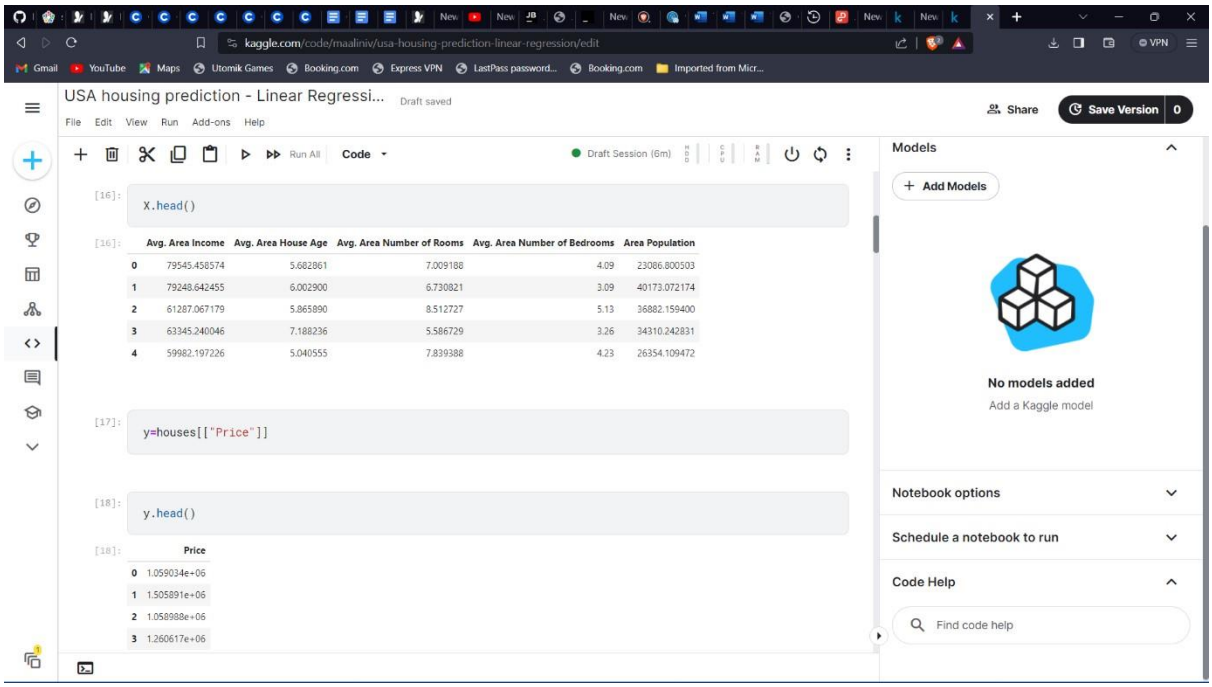

kaggle.com/code/maaliniv/usa-housing-prediction-linear-regression/edit
USA housing prediction - Linear Regressi...
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (9m)
[42]:
train_pred
[42]:
array([[1269923.18564459],
[1379662.72602339],
[1714251.56302348],
...,
[1457345.63232965],
[1214536.60784077],
[1188131.44764333]])
[43]:
(np.dot(X_train,coef.transpose())+intercept)
[43]:
array([[1269923.18564459],
[1379662.72602339],
[1714251.56302348],
...,
[1457345.63232965],
[1214536.60784077],
[1188131.44764333]])
[44]:
houses.corr()
[44]:
Avg. Area Income
Avg. Area House Age
Avg. Area Number of Rooms
Avg. Area Number of Bedrooms
Area Population
Price
Avg. Area Income 1.000000 -0.002007 -0.011032 0.019788 -0.016234 0.639734
Avg. Area House Age -0.002007 1.000000 -0.009428 0.006149 -0.018743 0.452543
Avg. Area Number of Rooms -0.011032 -0.009428 1.000000 0.462695 0.002040 0.335664
Models
Add Models
No models added
Add a Kaggle model
Notebook options
Schedule a notebook to run
Code Help
Find code help


kaggle.com/code/maaliniv/usa-housing-prediction-linear-regression/edit
USA housing prediction - Linear Regressi...
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (9m)
[45]: houses.columns
[45]: Index(['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',
'Avg. Area Number of Bedrooms', 'Area Population', 'Price', 'Address'],
dtype='object')
[46]: houses.head(7)
[46]:
Avg. Area Income | Avg. Area House Age | Avg. Area Number of Rooms | Avg. Area Number of Bedrooms | Area Population | Price | Address
0 | 79545.458574 | 5.682861 | 7.009188 | 4.09 | 23086.800503 | 1.059034e+06 | 208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
1 | 79248.642455 | 6.002900 | 6.730821 | 3.09 | 40173.072174 | 1.505891e+06 | 188 Johnson Views Suite 079\nLake Kathleen, CA...
2 | 61287.067179 | 5.865890 | 8.512727 | 5.13 | 36882.159400 | 1.058988e+06 | 9127 Elizabeth Stravenue\nDanieltown, WI 06482...
3 | 63345.240046 | 7.188236 | 5.586729 | 3.26 | 34310.242831 | 1.260617e+06 | USS Barnett\nFPO AP 44820
4 | 59982.197226 | 5.040555 | 7.839388 | 4.23 | 26354.109472 | 6.309435e+05 | USNS Raymond\nFPO AE 09386
5 | 80175.754159 | 4.988408 | 6.104512 | 4.04 | 26748.428425 | 1.068138e+06 | 06039 Jennifer Islands Apt. 443\nTracyport, KS...
6 | 64698.463428 | 6.025336 | 8.147760 | 3.41 | 60828.249085 | 1.502056e+06 | 4759 Daniel Shoals Suite 442\nNguyenburgh, CO ...
Models
Add Models
No models added
Add a Kaggle model
Notebook options
Schedule a notebook to run
Code Help
Find code help


kaggle.com/code/maaliniv/usa-housing-prediction-linear-regression/edit
USA housing prediction - Linear Regressi...
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (10m)
[47]: from sklearn.model_selection import cross_val_score
[48]: cross_val_value_training=cross_val_score(linearmodel,X_train,y_train,cv=10)
[49]: cross_val_value_training.mean()
[49]: 0.9182174480513696
[50]: cross_val_value_testing=cross_val_score(linearmodel,X_test,y_test,cv=10)
[51]: cross_val_value_testing.mean()
[51]: 0.9114034511775113
Models
Add Models
No models added
Add a Kaggle model
Notebook options
Schedule a notebook to run
Code Help
Find code help



	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price
0	79545.458574	5.682861	7.009188	4.09	23086.800503	1.059034e+06
1	79248.642455	6.002900	6.730821	3.09	40173.072174	1.505891e+06
2	61287.067179	5.865890	8.512727	5.13	36882.159400	1.058988e+06
3	63345.240046	7.188236	5.586729	3.26	34310.242831	1.260617e+06
4	59982.197226	5.040555	7.839388	4.23	26354.109472	6.309435e+05



	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population
0	79545.458574	5.682861	7.009188	4.09	23086.800503
1	79248.642455	6.002900	6.730821	3.09	40173.072174
2	61287.067179	5.865890	8.512727	5.13	36882.159400
3	63345.240046	7.188236	5.586729	3.26	34310.242831
4	59982.197226	5.040555	7.839388	4.23	26354.109472

	Price
0	1.059034e+06
1	1.505891e+06
2	1.058988e+06
3	1.260617e+06

RESULT:

Thus the dataset has been loaded and preprocessed successfully.