ai-phase3

October 18, 2023

[1]:
```
#IMPORTING THE LIBRARIES

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score, mean_absolute_error,mean_squared_error
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Lasso
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import SVR
import xgboost as xg

#IMPORTING & LOADING THE DATASET
print("_____________")
print("THE DATASET IS:")
print("_____________")
print()
dataset = pd.read_csv("C:/Users/amald/OneDrive/Desktop/jupyter/USA HOUSING.csv")
print(dataset)
print()
print()

#PREPROCESSING THE DATASET
print("_____________________________")
print("THE INFORMATION ABOUT DATASET:")
print("_____________________________")
print()
print(dataset.info())
print()
print("___________________")
print("DESCRIBING DATASET:")
print("___________________")
print()
print(dataset.describe())
```

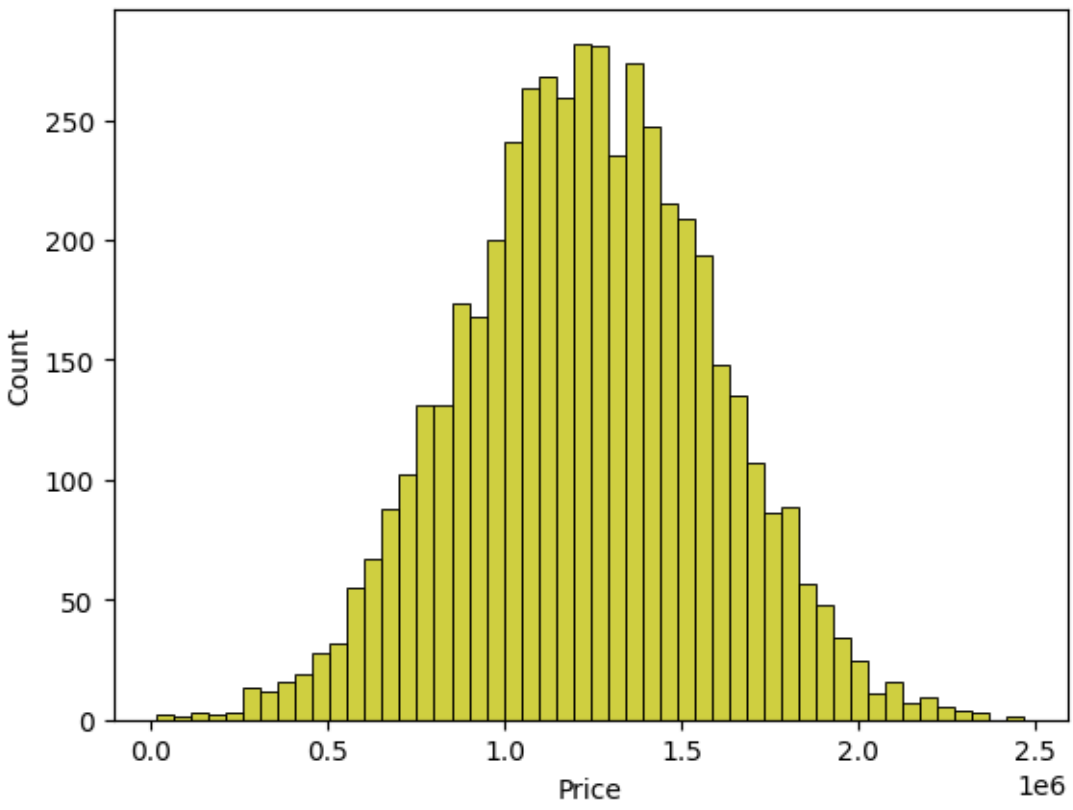
```
print()
print("_________________")
print("THE COLUMNS ARE:")
print("_________________")
print()
print(dataset.columns)
print()
print("_________")
print("HISTPLOT:")
print("_________")
print()
m=sns.histplot(dataset, x='Price', bins=50, color='y')
plt.show()
print()
print("________")
print("JOINTPLOT:")
print("________")
sns.jointplot(dataset, x='Avg. Area House Age', y='Price', kind='hex')
sns.jointplot(dataset, x='Avg. Area Income', y='Price')
plt.show()
print()
print("_________")
print("HISTOGRAM")
print("_________")
dataset.hist(figsize=(10,8))
plt.show()
print("__________________")
print("CORRELATING DATA:")
print("__________________")
print(dataset.corr(numeric_only=True))
print()
print("________")
print("HEATMAP:")
print("________")
sns.heatmap(dataset.corr(numeric_only = True), annot=True)
plt.show()
print()

#DIVIDING DATASET INTO FEATURES AND TARGET
x = dataset[['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of␣
 ↪Rooms',
       'Avg. Area Number of Bedrooms', 'Area Population']]
y = dataset['Price']
print()
print("THE VALUE OF X IS:")
print()
print(x)
```

```
print()
print("THE VALUE OF Y IS:")
print()
print(y)
```

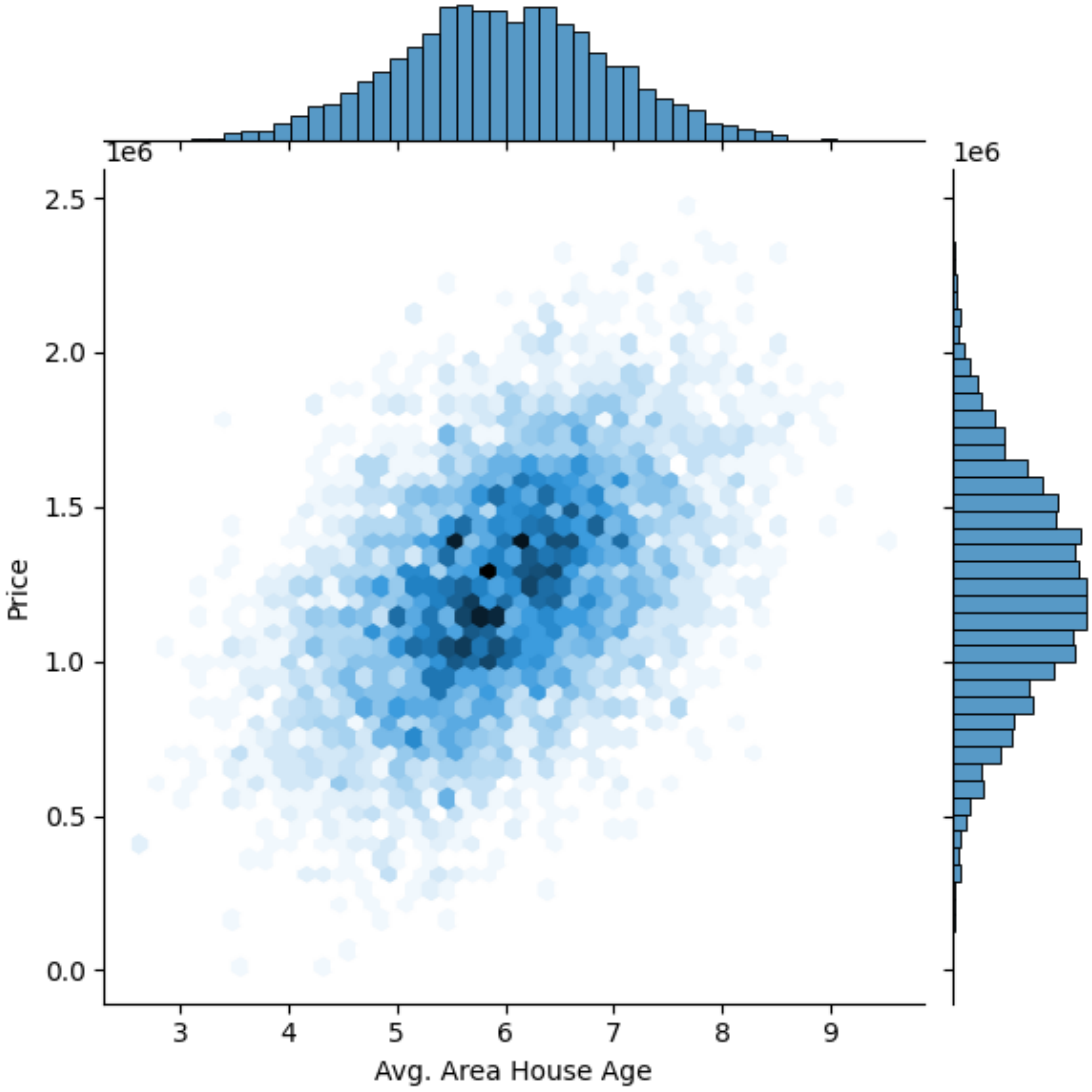
```
---------------
THE DATASET IS:
---------------

      Avg. Area Income  Avg. Area House Age  Avg. Area Number of Rooms  \
0          79545.45857             5.682861                   7.009188
1          79248.64245             6.002900                   6.730821
2          61287.06718             5.865890                   8.512727
3          63345.24005             7.188236                   5.586729
4          59982.19723             5.040555                   7.839388
…                  …                    …                          …
4995       60567.94414             7.830362                   6.137356
4996       78491.27543             6.999135                   6.576763
4997       63390.68689             7.250591                   4.805081
4998       68001.33124             5.534388                   7.130144
4999       65510.58180             5.992305                   6.792336

      Avg. Area Number of Bedrooms  Area Population         Price  \
0                             4.09      23086.80050  1.059034e+06
1                             3.09      40173.07217  1.505891e+06
2                             5.13      36882.15940  1.058988e+06
3                             3.26      34310.24283  1.260617e+06
4                             4.23      26354.10947  6.309435e+05
…                             …              …           …
4995                          3.46      22837.36103  1.060194e+06
4996                          4.02      25616.11549  1.482618e+06
4997                          2.13      33266.14549  1.030730e+06
4998                          5.44      42625.62016  1.198657e+06
4999                          4.07      46501.28380  1.298950e+06

                                          Address
0     208 Michael Ferry Apt. 674\nLaurabury, NE 3701…
1     188 Johnson Views Suite 079\nLake Kathleen, CA…
2     9127 Elizabeth Stravenue\nDanieltown, WI 06482…
3                           USS Barnett\nFPO AP 44820
4                          USNS Raymond\nFPO AE 09386
…                                                 …
4995                 USNS Williams\nFPO AP 30153-7653
4996             PSC 9258, Box 8489\nAPO AA 42991-3352
4997  4215 Tracy Garden Suite 076\nJoshualand, VA 01…
4998                          USS Wallace\nFPO AE 73316
4999  37778 George Ridges Apt. 509\nEast Holly, NV 2…
```

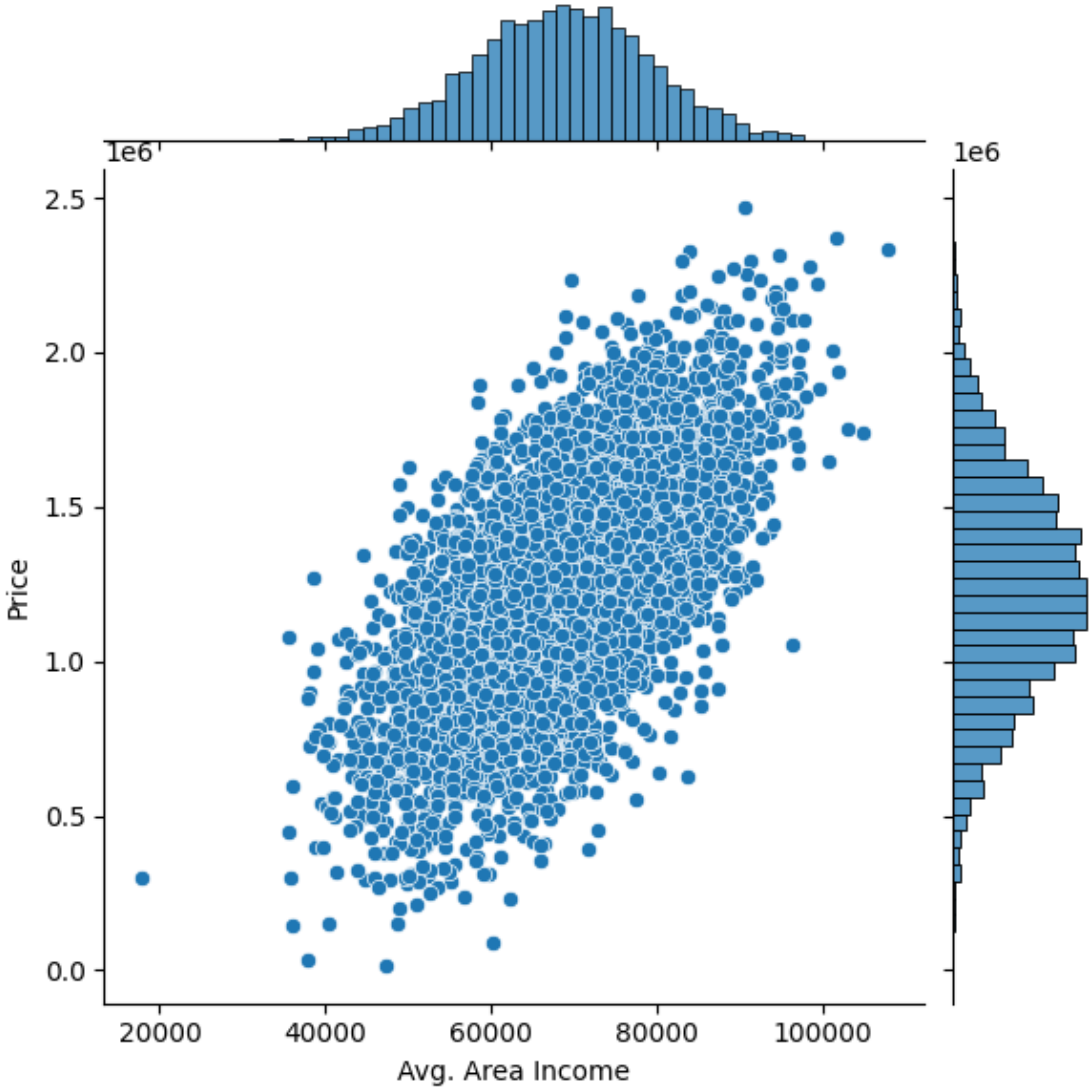
```
[5000 rows x 7 columns]


----------------------------
THE INFORMATION ABOUT DATASET:
----------------------------

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 4999
Data columns (total 7 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Avg. Area Income              5000 non-null   float64
 1   Avg. Area House Age           5000 non-null   float64
 2   Avg. Area Number of Rooms     5000 non-null   float64
 3   Avg. Area Number of Bedrooms  5000 non-null   float64
 4   Area Population               5000 non-null   float64
 5   Price                         5000 non-null   float64
 6   Address                       5000 non-null   object
dtypes: float64(6), object(1)
memory usage: 273.6+ KB
None


-------------------
DESCRIBING DATASET:
-------------------

       Avg. Area Income  Avg. Area House Age  Avg. Area Number of Rooms  \
count       5000.000000          5000.000000                5000.000000
mean       68583.108984             5.977222                   6.987792
std        10657.991214             0.991456                   1.005833
min        17796.631190             2.644304                   3.236194
25%        61480.562390             5.322283                   6.299250
50%        68804.286405             5.970429                   7.002902
75%        75783.338665             6.650808                   7.665871
max       107701.748400             9.519088                  10.759588

       Avg. Area Number of Bedrooms  Area Population         Price
count                   5000.000000      5000.000000  5.000000e+03
mean                       3.981330     36163.516039  1.232073e+06
std                        1.234137      9925.650114  3.531176e+05
min                        2.000000       172.610686  1.593866e+04
25%                        3.140000     29403.928700  9.975771e+05
50%                        4.050000     36199.406690  1.232669e+06
75%                        4.490000     42861.290770  1.471210e+06
max                        6.500000     69621.713380  2.469066e+06
```

```
----------------
THE COLUMNS ARE:
----------------

Index(['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',
       'Avg. Area Number of Bedrooms', 'Area Population', 'Price', 'Address'],
      dtype='object')


---------
HISTPLOT:
---------
```

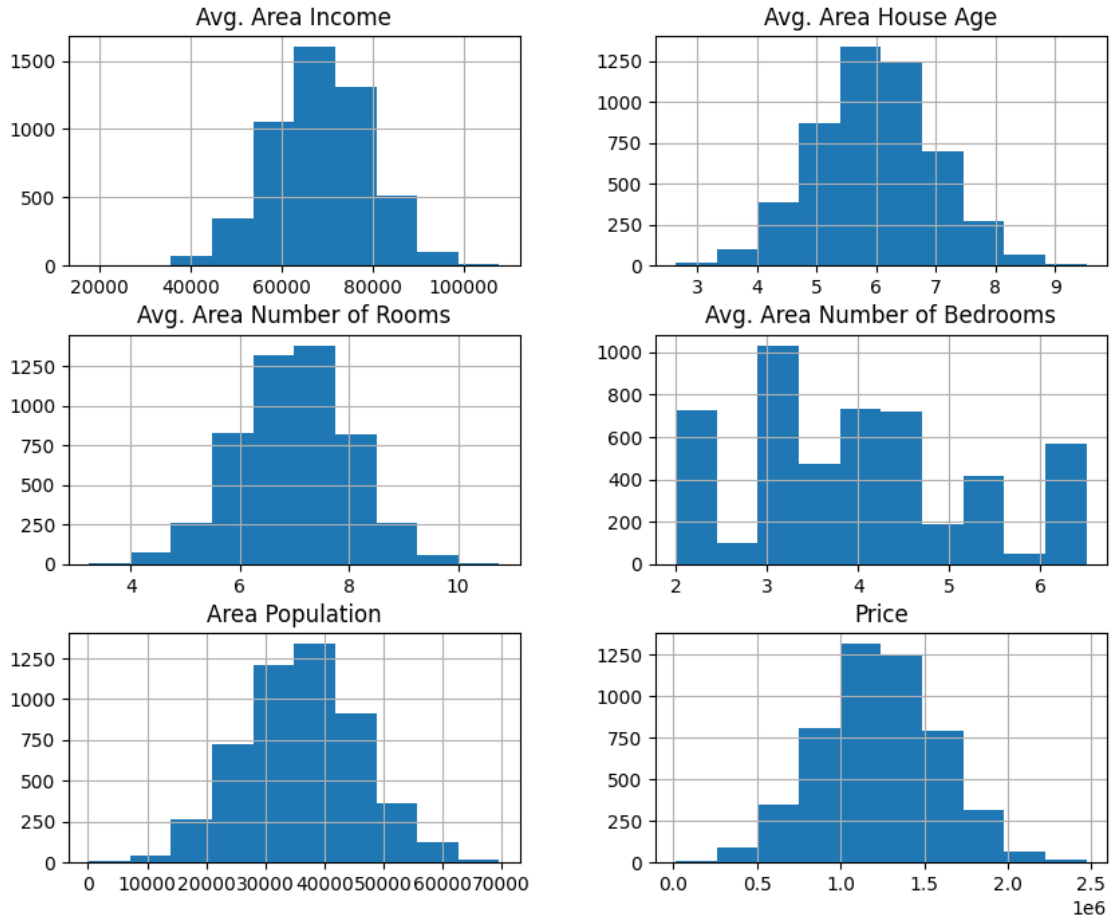



```
--------
JOINTPLOT:
--------
```


1e6
1e6
2.5
2.0
1.5
1.0
0.5
0.0
Price
3
4
5
6
7
8
9
Avg. Area House Age

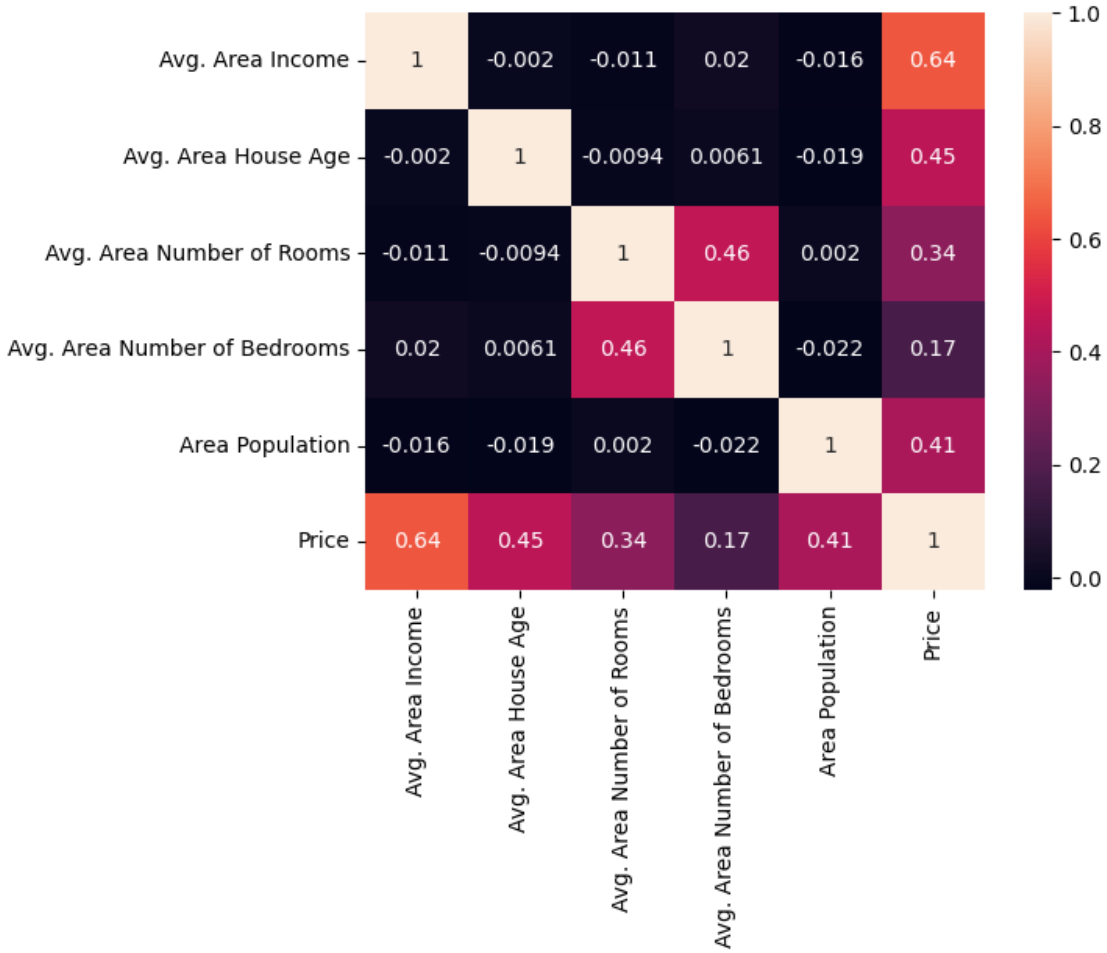

1e6
1e6
2.5
2.0
1.5
1.0
0.5
0.0
Price
20000
40000
60000
80000
100000
Avg. Area Income

HISTOGRAM




```
-----------------
CORRELATING DATA:

-----------------
                            Avg. Area Income  Avg. Area House Age  \
Avg. Area Income                    1.000000            -0.002007
Avg. Area House Age                -0.002007             1.000000
Avg. Area Number of Rooms          -0.011032            -0.009428
Avg. Area Number of Bedrooms        0.019788             0.006149
Area Population                    -0.016234            -0.018743
Price                               0.639734             0.452543

                            Avg. Area Number of Rooms  \
Avg. Area Income                            -0.011032
Avg. Area House Age                         -0.009428
Avg. Area Number of Rooms                    1.000000
Avg. Area Number of Bedrooms                 0.462695
Area Population                              0.002040
Price                                        0.335664
```

```
                             Avg. Area Number of Bedrooms  Area Population  \
Avg. Area Income                                 0.019788        -0.016234
Avg. Area House Age                              0.006149        -0.018743
Avg. Area Number of Rooms                        0.462695         0.002040
Avg. Area Number of Bedrooms                     1.000000        -0.022168
Area Population                                 -0.022168         1.000000
Price                                            0.171071         0.408556

                                Price
Avg. Area Income             0.639734
Avg. Area House Age          0.452543
Avg. Area Number of Rooms    0.335664
Avg. Area Number of Bedrooms 0.171071
Area Population              0.408556
Price                        1.000000

--------
HEATMAP:
--------
```



```
THE VALUE OF X IS:

      Avg. Area Income  Avg. Area House Age  Avg. Area Number of Rooms  \
0          79545.45857             5.682861                   7.009188
1          79248.64245             6.002900                   6.730821
2          61287.06718             5.865890                   8.512727
3          63345.24005             7.188236                   5.586729
4          59982.19723             5.040555                   7.839388
…                  …                    …                          …
4995       60567.94414             7.830362                   6.137356
4996       78491.27543             6.999135                   6.576763
4997       63390.68689             7.250591                   4.805081
4998       68001.33124             5.534388                   7.130144
4999       65510.58180             5.992305                   6.792336

      Avg. Area Number of Bedrooms  Area Population
0                             4.09      23086.80050
1                             3.09      40173.07217
2                             5.13      36882.15940
3                             3.26      34310.24283
4                             4.23      26354.10947
…                             …             …
4995                          3.46      22837.36103
4996                          4.02      25616.11549
4997                          2.13      33266.14549
4998                          5.44      42625.62016
4999                          4.07      46501.28380

[5000 rows x 5 columns]

THE VALUE OF Y IS:

0       1.059034e+06
1       1.505891e+06
2       1.058988e+06
3       1.260617e+06
4       6.309435e+05
           …
4995    1.060194e+06
4996    1.482618e+06
4997    1.030730e+06
4998    1.198657e+06
4999    1.298950e+06
```

```
Name: Price, Length: 5000, dtype: float64
```