```python
import pandas as pd

dataset = pd.read_csv(r'./USA_Housing.csv')

print(len(dataset['Address'].unique()))
```

```
5000
```

```python
dataset.head()
```

```
   Avg. Area Income  Avg. Area House Age  Avg. Area Number of Rooms  \
0      79545.458574             5.682861                   7.009188
1      79248.642455             6.002900                   6.730821
2      61287.067179             5.865890                   8.512727
3      63345.240046             7.188236                   5.586729
4      59982.197226             5.040555                   7.839388

   Avg. Area Number of Bedrooms  Area Population         Price  \
0                          4.09     23086.800503  1.059034e+06
1                          3.09     40173.072174  1.505891e+06
2                          5.13     36882.159400  1.058988e+06
3                          3.26     34310.242831  1.260617e+06
4                          4.23     26354.109472  6.309435e+05

                                             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
```

```python
dataset = dataset.drop('Address', axis=1)

x = dataset.drop('Price', axis=1)
y = dataset['Price']

x.head()
```

```
   Avg. Area Income  Avg. Area House Age  Avg. Area Number of Rooms  \
0      79545.458574             5.682861                   7.009188
1      79248.642455             6.002900                   6.730821
2      61287.067179             5.865890                   8.512727
3      63345.240046             7.188236                   5.586729
4      59982.197226             5.040555                   7.839388

   Avg. Area Number of Bedrooms  Area Population
0                          4.09     23086.800503
1                          3.09     40173.072174
2                          5.13     36882.159400
3                          3.26     34310.242831
4                          4.23     26354.109472
```
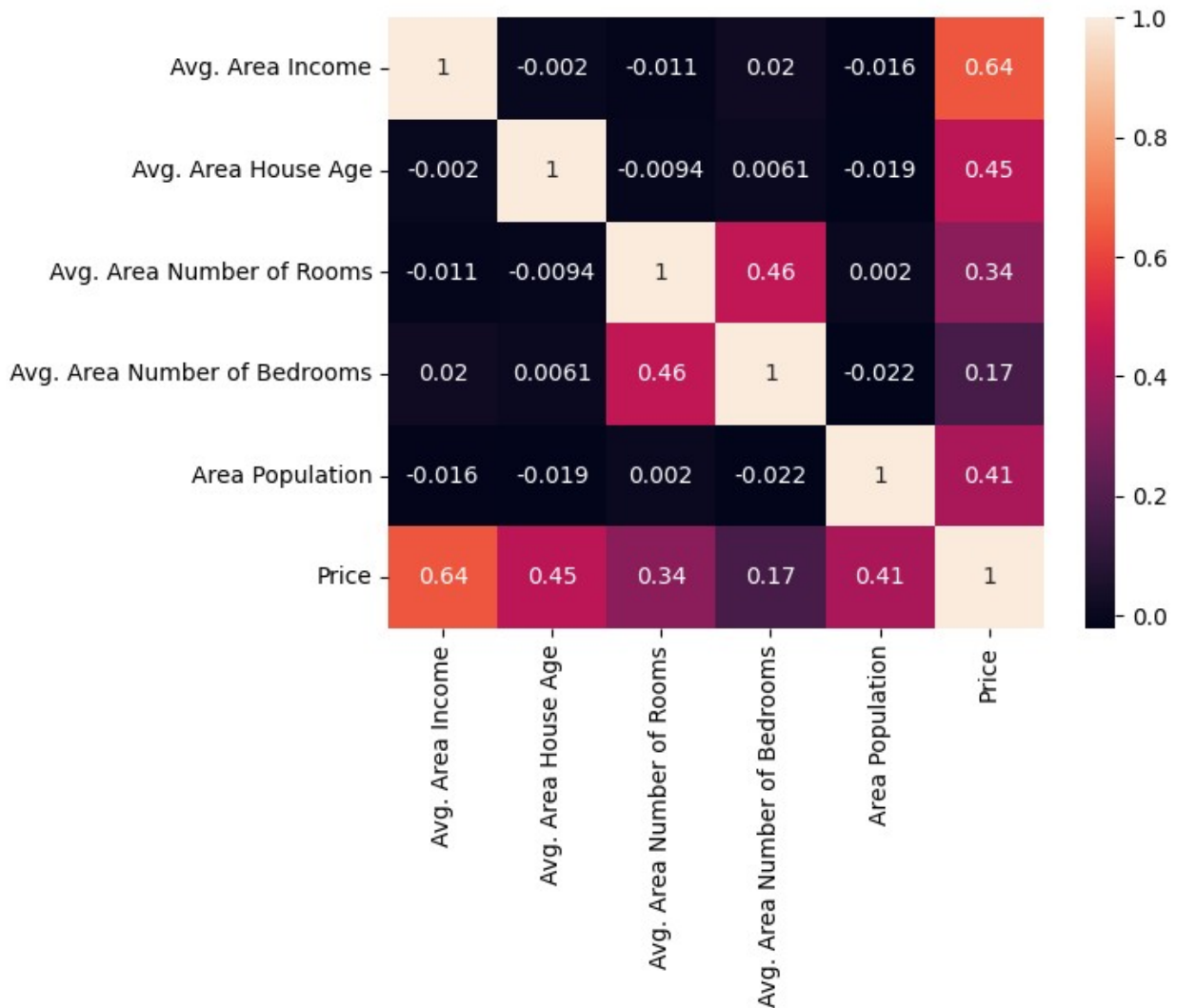
```python
import seaborn as sns
import matplotlib as plt
%matplotlib inline

sns.heatmap(dataset.corr(), annot=True)
# plt.show/
```

```
<Axes: >
```



```python
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(x, y,
random_state=40)

scaler = StandardScaler()
```

```
x_train_scaled = scaler.fit_transform(X_train)
x_test_scaled = scaler.fit_transform(X_test)

x_train_scaled

array([[-0.66041094, -0.4484942 , -0.77296479, -1.44109405,
0.58868104],
       [ 0.22936865,  0.5795949 ,  0.03704415,  1.86893391,
0.36165789],
       [ 0.73798528, -0.39761892, -0.14812542, -1.4491281 ,  0.0659088
],
       ...,
       [ 1.1513024 ,  0.30975303, -1.1728203 , -1.23220879,
0.62627246],
       [ 2.40776591,  1.71210444,  1.07286112,  1.83679772,
0.29735505],
       [-1.56209424,  0.26164926,  0.1843303 ,  1.9492744 , -
0.50432325]])

x_test_scaled

array([[ 0.62134846,  0.55705943, -1.05371838, -0.6206854 ,
0.73334946],
       [ 1.61315092, -0.33867987,  0.37126806,  1.28110827, -
0.29210403],
       [ 0.63520205,  1.97554275,  0.58128115,  1.99011029, -
0.33471659],
       ...,
       [-0.66021978, -0.04985402, -0.80781138,  0.22177583, -
0.02631411],
       [ 0.89888266, -0.46882496, -1.11162927, -1.55489983,
0.55269839],
       [ 0.67036641, -2.03100441,  0.08189237,  1.04755466, -
0.10636859]])
```

```python
from sklearn.linear_model import LinearRegression, LogisticRegression
```

```python
lr_model= LinearRegression()
lr_model.fit(x_train_scaled, y_train)
```

```
0.9142264116316149
```

```python
lr_y_prei = lr_model.predict(x_test_scaled)
lr_score = lr_model.score(x_test_scaled,y_test)
print("accuracy = ", lr_score)
```

```
accuracy =  0.9142264116316149
```