```
import pandas as pd
import numpy as np


import seaborn as sns
import matplotlib.pyplot as plt


from sklearn.preprocessing import StandardScaler

from sklearn.metrics import mean_absolute_error , mean_squared_error ,
median_absolute_error , r2_score

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression ,
SGDRegressor ,Ridge , Lasso
```

Loading Dataset

```
df=pd.read_csv(r'USA_Housing.csv')
df.head(10)
```

```
   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
5      80175.754159             4.988408                   6.104512
6      64698.463428             6.025336                   8.147760
7      78394.339278             6.989780                   6.620478
8      59927.660813             5.362126                   6.393121
9      81885.927184             4.423672                   8.167688

   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
5                          4.04     26748.428425  1.068138e+06
6                          3.41     60828.249085  1.502056e+06
7                          2.42     36516.358972  1.573937e+06
8                          2.30     29387.396003  7.988695e+05
9                          6.10     40149.965749  1.545155e+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
5  06039 Jennifer Islands Apt. 443\nTracyport, KS...
6  4759 Daniel Shoals Suite 442\nNguyenburgh, CO ...
7     972 Joyce Viaduct\nLake William, TN 17778-6483
8                          USS Gilbert\nFPO AA 20957
9                   Unit 9446 Box 0958\nDPO AE 97025
```

Preprocessing the dataset

```
df.shape
```

```
(5000, 7)
```

```
df.info()
```

```
<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
```

```
df.describe()
```

```
       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%        61480.562388             5.322283
6.299250
50%        68804.286404             5.970429
7.002902
```

```
75%        75783.338666                6.650808
7.665871
max       107701.748378                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     29403.928702  9.975771e+05
50%                        4.050000     36199.406689  1.232669e+06
75%                        4.490000     42861.290769  1.471210e+06
max                        6.500000     69621.713378  2.469066e+06
```

```
df.isnull().sum()
```

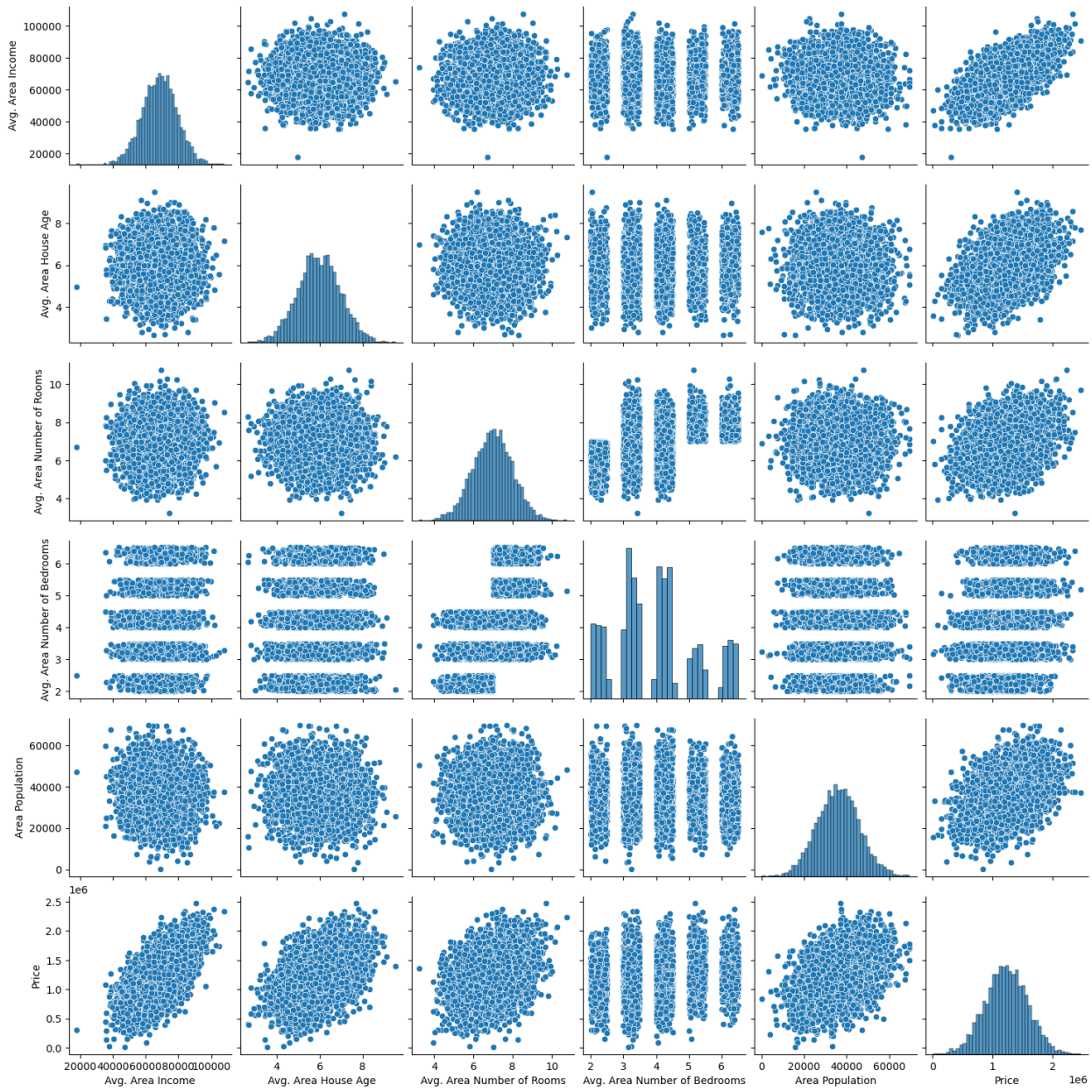
```
Avg. Area Income                0
Avg. Area House Age             0
Avg. Area Number of Rooms       0
Avg. Area Number of Bedrooms    0
Area Population                 0
Price                           0
Address                         0
dtype: int64
```

```
df.duplicated()
```

```
0       False
1       False
2       False
3       False
4       False
        ...
4995    False
4996    False
4997    False
4998    False
4999    False
Length: 5000, dtype: bool
```

```
sns.pairplot(df)
```

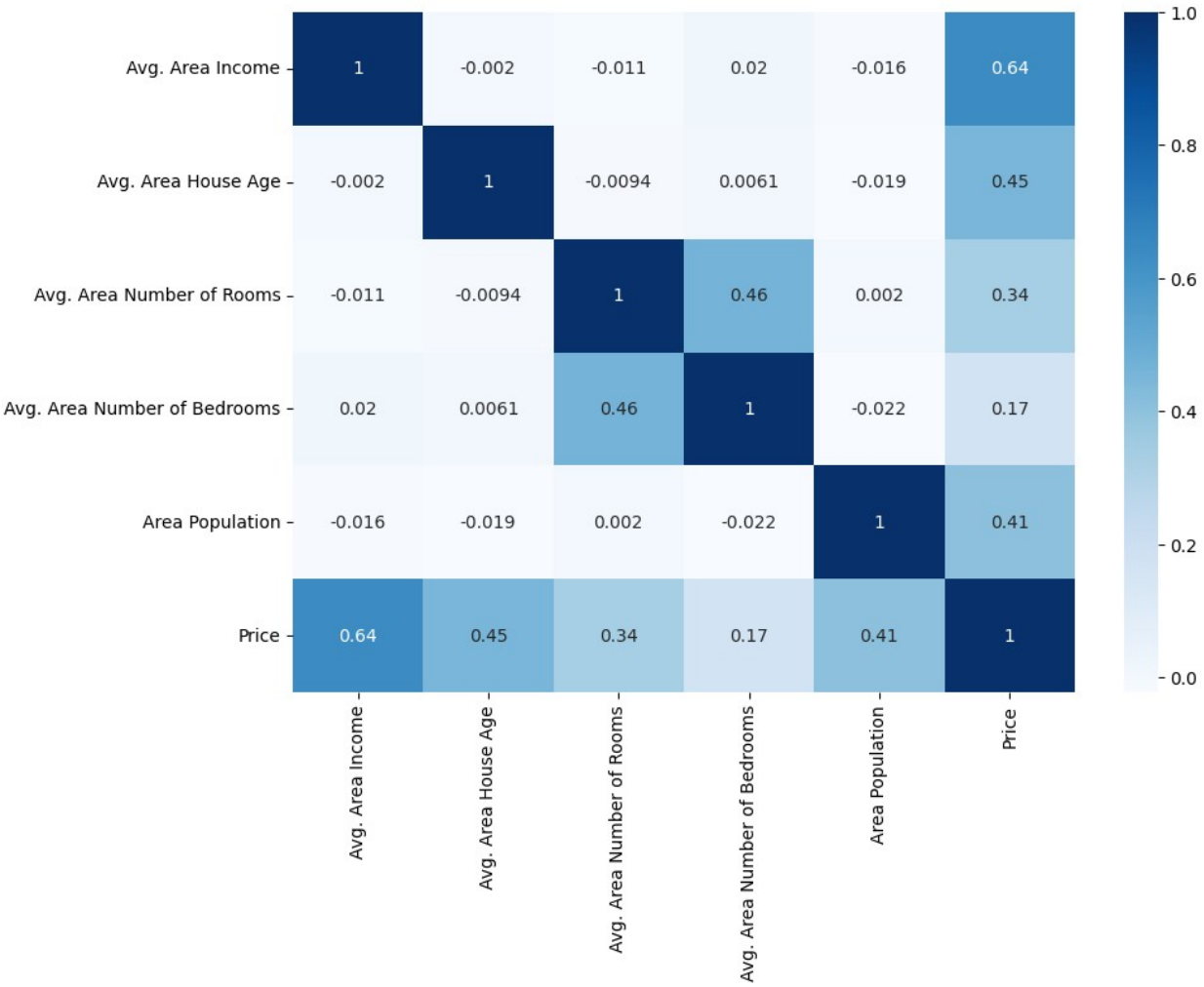
```
<seaborn.axisgrid.PairGrid at 0x7f4d11732710>
```




```
plt.figure(figsize=(10,7))
sns.heatmap(df.corr() , annot=True ,cmap='Blues')
```

```
/tmp/ipykernel_12463/850008773.py:2: FutureWarning: The default value
of numeric_only in DataFrame.corr is deprecated. In a future version,
it will default to False. Select only valid columns or specify the
value of numeric_only to silence this warning.
  sns.heatmap(df.corr() , annot=True ,cmap='Blues')
```

```
<Axes: >
```




```
y=df['Price'].values
X=df.iloc[:,:-2].values

X_train, X_test, y_train, y_test = train_test_split( X, y,
test_size=0.3, random_state=42)

scaler=StandardScaler()

X_train=scaler.fit_transform(X_train)
X_test=scaler.transform(X_test)
```