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(r'C:\Users\Admin\Downloads\Data thecnogeeks\Data_Engineering\DataScience\1. Linear-Regression\USA_Housing.csv')
```

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

Out[3]:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Address
0	79545.45857	5.682861	7.009188	4.09	23086.80050	1.059034e+06	208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
1	79248.64245	6.002900	6.730821	3.09	40173.07217	1.505891e+06	188 Johnson Views Suite 079\nLake Kathleen, CA...
2	61287.06718	5.865890	8.512727	5.13	36882.15940	1.058988e+06	9127 Elizabeth Stravenue\nDanielton, WI 06482...
3	63345.24005	7.188236	5.586729	3.26	34310.24283	1.260617e+06	USS Barnett\nFPO AP 44820
4	59982.19723	5.040555	7.839388	4.23	26354.10947	6.309435e+05	USNS Raymond\nFPO AE 09386

In [4]:
```
data.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
```

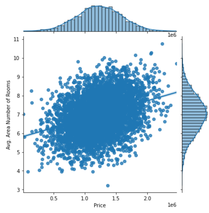
In [5]:
```
data.corr()
```

Out[5]:

	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
Avg. Area Number of Bedrooms	0.019788	0.006149	0.462695	1.000000	-0.022168	0.171071
Area Population	-0.016234	-0.018743	0.002040	-0.022168	1.000000	0.408556
Price	0.639734	0.452543	0.335664	0.171071	0.408556	1.000000

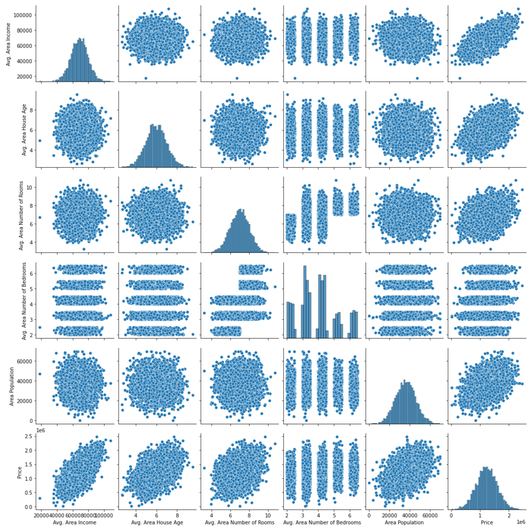
In [6]:
```
sns.jointplot(x='Price',y='Avg. Area Number of Rooms',data=data,kind='reg')
```

Out[6]:
```
<seaborn.axisgrid.JointGrid at 0x2ad00312490>
```



In [7]:
```
sns.pairplot(data)
```

Out[7]:
```
<seaborn.axisgrid.PairGrid at 0x2ad01d28610>
```



In [9]:
```
data.columns
```

Out[9]:
```
Index(['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',
       'Avg. Area Number of Bedrooms', 'Area Population', 'Price', 'Address'],
      dtype='object')
```

In [10]:
```
from sklearn.model_selection import train_test_split
```

In [11]:
```
x = data[['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',
          'Avg. Area Number of Bedrooms', 'Area Population']]
y = data['Price']
```

In [12]:
```
x_train,x_test,y_train,y_test = train_test_split(x,y,train_size=0.7)
```

In [13]:
```
x_train.shape
```

Out[13]:
```
(3500, 5)
```

In [14]:
```
y_train.shape
```

Out[14]:
```
(3500,)
```

In [15]:
```
from sklearn.linear_model import LinearRegression
```

In [16]:
```
lgr = LinearRegression()
```

In [17]:
```
lgr.fit(x_train,y_train)
```

Out[17]:
```
LinearRegression()
```

In [18]:
```
y_pred=lgr.predict(x_test)
```

In [20]:
```
lgr.coef_
```

Out[20]:
```
array([2.15693274e+01, 1.65867350e+05, 1.21998615e+05, 2.15709835e+03,
       1.51406875e+01])
```

In [21]:
```
y_pred[102]
```

Out[21]:
```
1591744.499872785
```

In [24]:
```
coeff_df = pd.DataFrame(lgr.coef_,x.columns,columns=['Coefficient'])
coeff_df
```

Out[24]:

	Coefficient
Avg. Area Income	21.569327
Avg. Area House Age	165867.349727
Avg. Area Number of Rooms	121998.614804
Avg. Area Number of Bedrooms	2157.098355
Area Population	15.140687

In [26]:
```
x.mean()
```

Out[26]:
```
Avg. Area Income                68583.108984
Avg. Area House Age                 5.977222
Avg. Area Number of Rooms           6.987792
Avg. Area Number of Bedrooms        3.981330
Area Population                 36163.516039
dtype: float64
```

y_test.iloc[102]

In [27]:
```
y_pred[100]
```

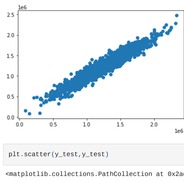
Out[27]:
```
1294130.1849425738
```

In [28]:
```
y_test.iloc[100]
```

Out[28]:
```
1339636.187
```

In [29]:
```
plt.scatter(y_test,y_pred)
```

Out[29]:
```
<matplotlib.collections.PathCollection at 0x2ad04d9a7c0>
```



In [31]:
```
plt.scatter(y_test,y_test)
```

Out[31]:
```
<matplotlib.collections.PathCollection at 0x2ad04def9d0>
```

In [34]:
```
from sklearn import metrics
```

In [35]:
```
print('MAE:',metrics.mean_absolute_error(y_test,y_pred))
```

```
MAE: 82633.91684021002
```

In [36]:
```
print('MSE:',metrics.mean_squared_error(y_test,y_pred))
```

```
MSE: 10545750096.667534
```

In [37]:
```
print('RMSE:',np.sqrt(metrics.mean_squared_error(y_test,y_pred)))
```

```
RMSE: 102692.50263124146
```

In [ ]: