In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
import cufflinks as cf
from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
init_notebook_mode(connected = True)
cf.go_offline()
```

In [ ]:

In [ ]:

In [3]:
```
us_housing = pd.read_csv("USAHousing.csv")
```

In [10]:
```
us_housing
```

Out[10]:

	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\nDanieltown, 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
...	...	...	...	...	...	...	...
4995	60567.94414	7.830362	6.137356	3.46	22837.36103	1.060194e+06	USNS Williams\nFPO AP 30153-7653
4996	78491.27543	6.999135	6.576763	4.02	25616.11549	1.482618e+06	PSC 9258, Box 8489\nAPO AA 42991-3352
4997	63390.68689	7.250591	4.805081	2.13	33266.14549	1.030730e+06	4215 Tracy Garden Suite 076\nJoshualand, VA 01...
4998	68001.33124	5.534388	7.130144	5.44	42625.62016	1.198657e+06	USS Wallace\nFPO AE 73316
4999	65510.58180	5.992305	6.792336	4.07	46501.28380	1.298950e+06	37778 George Ridges Apt. 509\nEast Holly, NV 2...

5000 rows × 7 columns

In [11]:
```
us_housing.count()
```

Out[11]:
```
Avg. Area Income                5000
Avg. Area House Age             5000
Avg. Area Number of Rooms       5000
Avg. Area Number of Bedrooms    5000
Area Population                 5000
Price                           5000
Address                         5000
dtype: int64
```

In [67]:
```
us_housing.tail()
```

Out[67]:

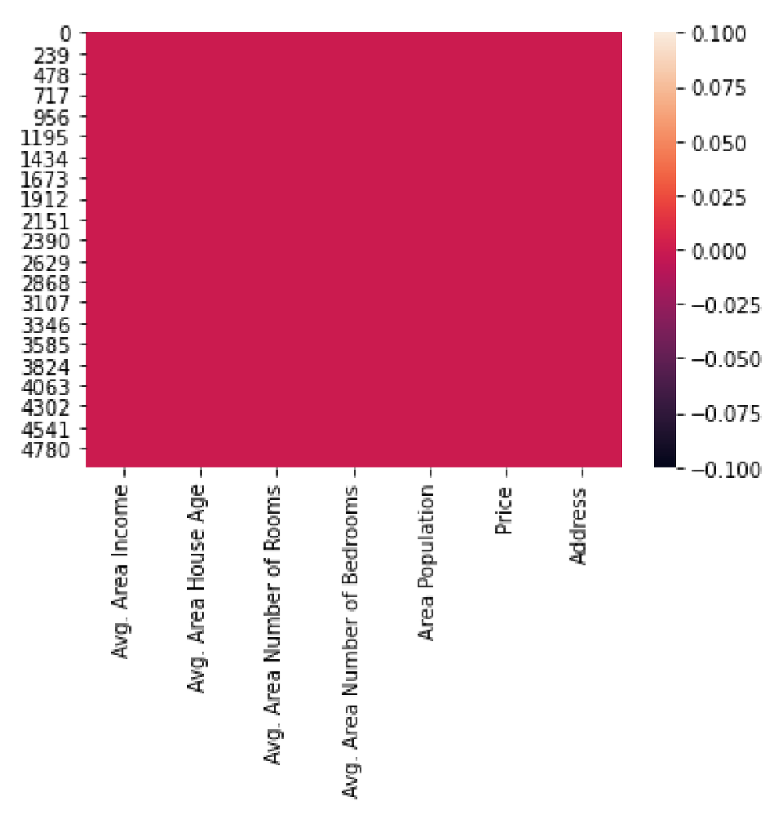
	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Address
4995	60567.94414	7.830362	6.137356	3.46	22837.36103	1060193.786	USNS Williams\nFPO AP 30153-7653
4996	78491.27543	6.999135	6.576763	4.02	25616.11549	1482617.729	PSC 9258, Box 8489\nAPO AA 42991-3352
4997	63390.68689	7.250591	4.805081	2.13	33266.14549	1030729.583	4215 Tracy Garden Suite 076\nJoshualand, VA 01...
4998	68001.33124	5.534388	7.130144	5.44	42625.62016	1198656.872	USS Wallace\nFPO AE 73316
4999	65510.58180	5.992305	6.792336	4.07	46501.28380	1298950.480	37778 George Ridges Apt. 509\nEast Holly, NV 2...

In [73]:
```
us_housing['Avg. Area Number of Bedrooms'].nunique()
```

Out[73]: 255

In [6]:
```
sns.heatmap(us_housing.isnull())
```

Out[6]: <matplotlib.axes._subplots.AxesSubplot at 0x1b715d47070>



In [8]:
```
us_housing.columns
```

Out[8]:
```
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 [44]:
```
us_housing.iloc[465]
```

Out[44]:
```
Avg. Area Income                                    90592.5
Avg. Area House Age                                 7.70013
Avg. Area Number of Rooms                            9.7088
Avg. Area Number of Bedrooms                           5.19
Area Population                                     37223.9
Price                                           2.46907e+06
Address                         USNS Vargas\nFPO AE 56319-6904
Name: 465, dtype: object
```

In [46]:
```
print(us_housing['Price'].max())
```

```
2469065.594
```

In [50]:
```
print(us_housing[us_housing.Price ==us_housing.Price.min(axis =0)])
```

```
      Avg. Area Income  Avg. Area House Age  Avg. Area Number of Rooms  \
3212       47320.65721             3.558054                   7.006987

      Avg. Area Number of Bedrooms  Area Population        Price  \
3212                          3.16       15776.6186  15938.65792

                                          Address
3212  91410 Megan Camp Suite 360\nLaurafort, OH 15735
```

In [51]:
```
print(us_housing[us_housing.Price == us_housing.Price.max()])
```

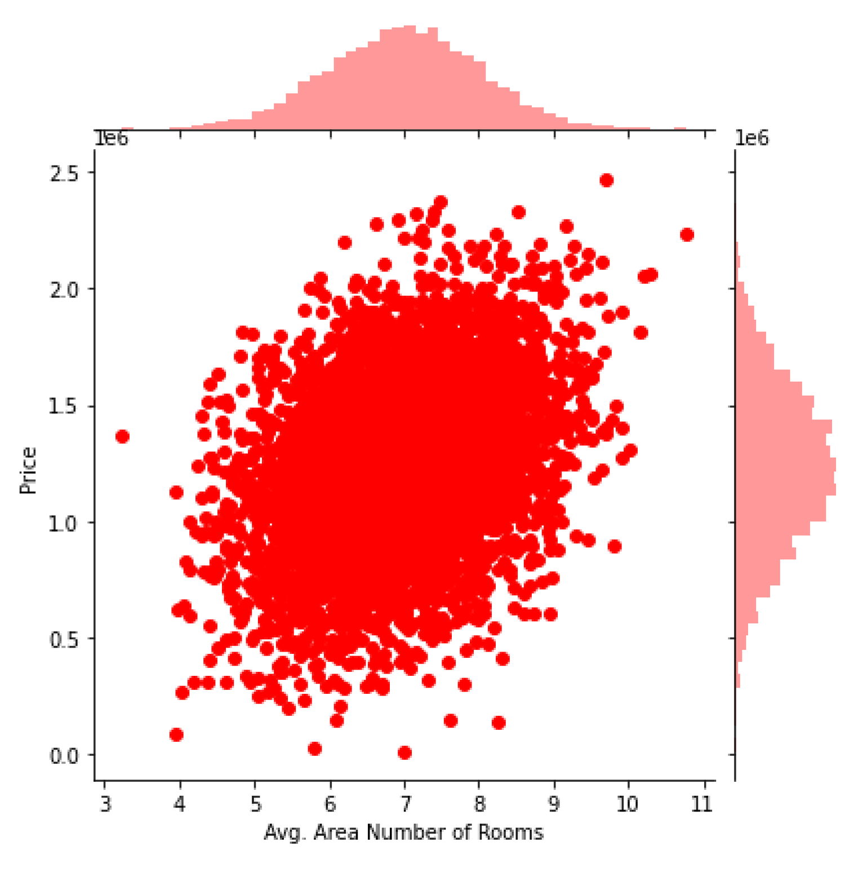
```
     Avg. Area Income  Avg. Area House Age  Avg. Area Number of Rooms  \
465       90592.46961             7.700132                   9.708803

     Avg. Area Number of Bedrooms  Area Population        Price  \
465                          5.19      37223.87617  2469065.594

                           Address
465  USNS Vargas\nFPO AE 56319-6904
```

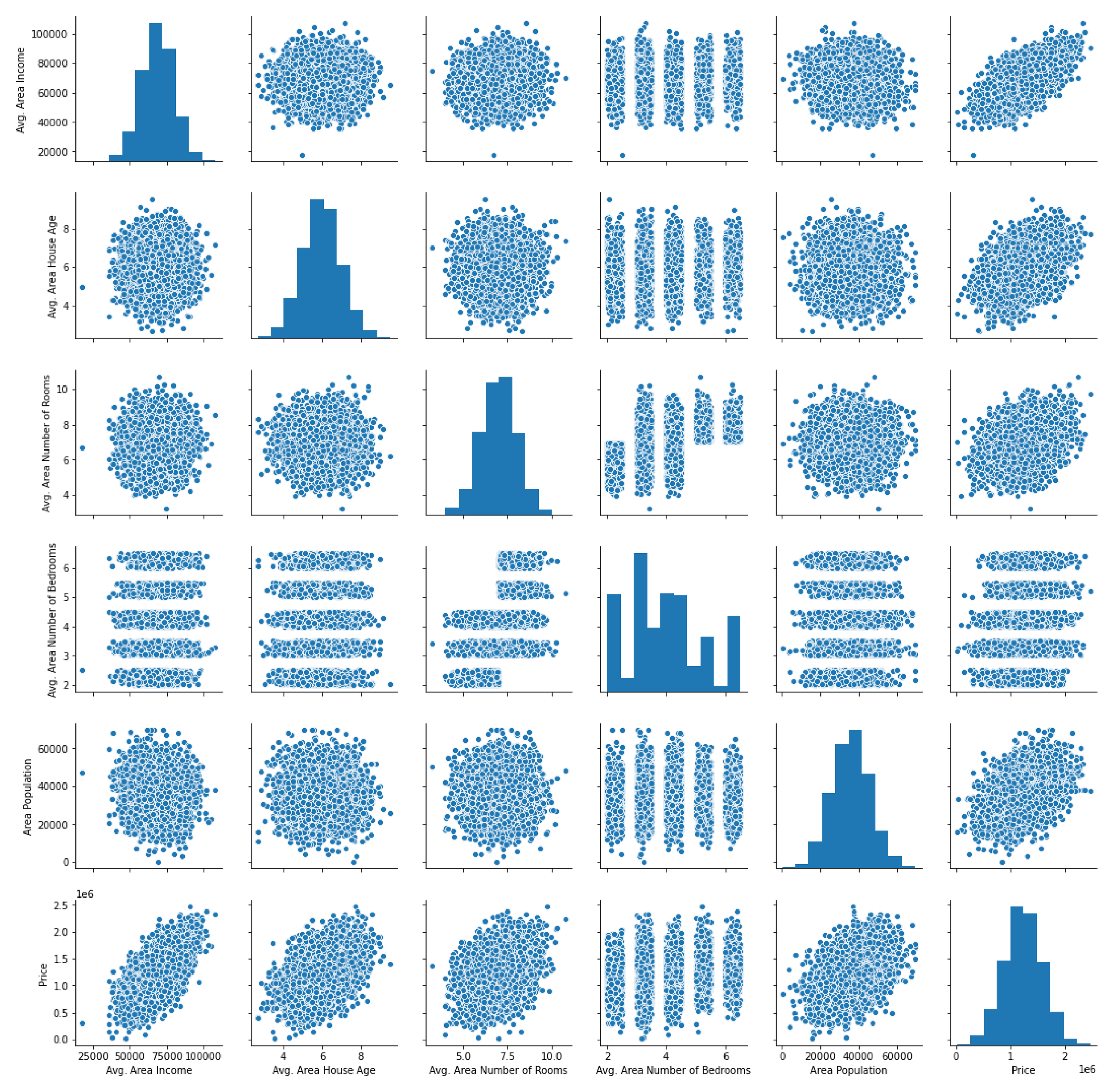
In [60]:
```
sns.jointplot(us_housing['Avg. Area Number of Rooms'],us_housing['Price'],kind='scatter',color ='red')
```

Out[60]: <seaborn.axisgrid.JointGrid at 0x1b7192113a0>



In [62]:
```
sns.pairplot(us_housing)
```

Out[62]: <seaborn.axisgrid.PairGrid at 0x1b71b336d60>



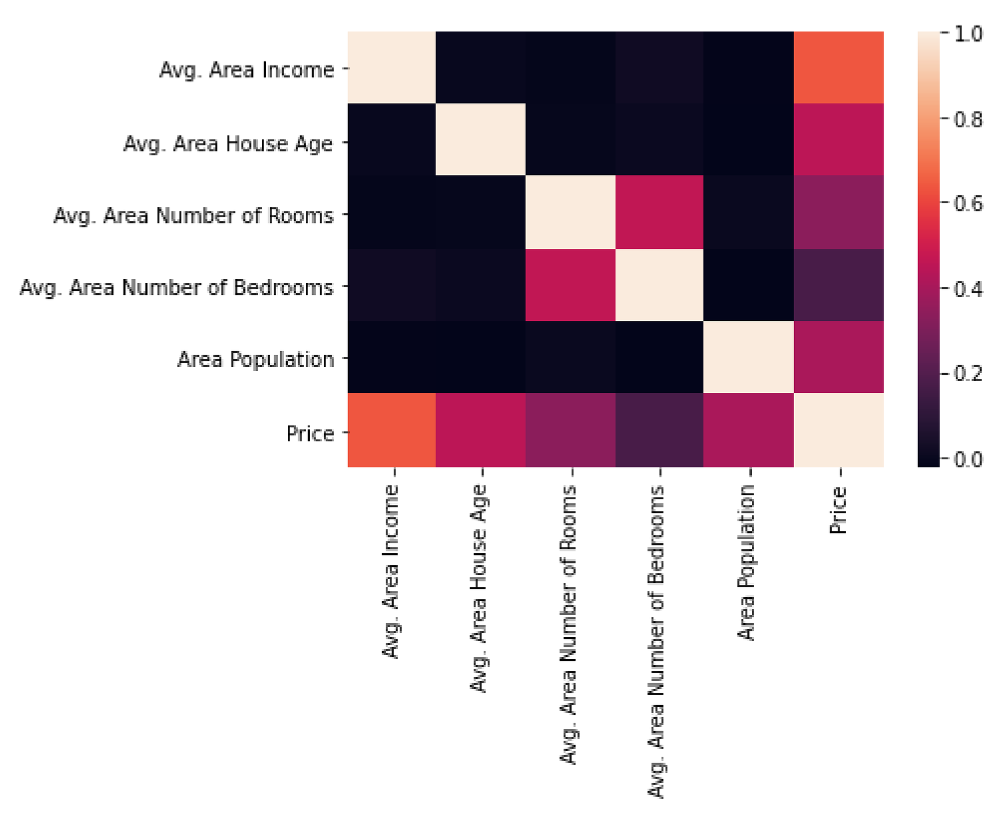
In [64]:
```
us_housing.corr(method ='pearson')
```

Out[64]:

	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 [66]:
```
sns.heatmap(us_housing.corr(method ='pearson'))
```

Out[66]: <matplotlib.axes._subplots.AxesSubplot at 0x1b71dd049a0>



In [90]:
```
us_housing(us_housing.)
```

Out[90]:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price
0	79545.45857	5.682861	7.009188	4.09	23086.80050	1.059034e+06
1	79248.64245	6.002900	6.730821	3.09	40173.07217	1.505891e+06
2	61287.06718	5.865890	8.512727	5.13	36882.15940	1.058988e+06
3	63345.24005	7.188236	5.586729	3.26	34310.24283	1.260617e+06
4	59982.19723	5.040555	7.839388	4.23	26354.10947	6.309435e+05
...	...	...	...	...	...	...
4995	60567.94414	7.830362	6.137356	3.46	22837.36103	1.060194e+06
4996	78491.27543	6.999135	6.576763	4.02	25616.11549	1.482618e+06
4997	63390.68689	7.250591	4.805081	2.13	33266.14549	1.030730e+06
4998	68001.33124	5.534388	7.130144	5.44	42625.62016	1.198657e+06
4999	65510.58180	5.992305	6.792336	4.07	46501.28380	1.298950e+06

5000 rows × 6 columns

In [124]:
```
X = us_housing.drop('Price',axis= 1).columns
```

In [127]:
```
x.columns
```

Out[127]:
```
Index(['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',
       'Avg. Area Number of Bedrooms', 'Area Population'],
      dtype='object')
```

In [ ]:

In [95]:
```
y = us_housing['Price']
```

In [86]:
```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

In [89]:
```
reg = LinearRegression()
```

In [133]:
```
reg.fit(us_housing[['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',
       'Avg. Area Number of Bedrooms', 'Area Population']],us_housing.Price)
```

Out[133]: LinearRegression()

In [135]:
```
us_housing.iloc[46]
```

Out[135]:
```
Avg. Area Income                7.394685e+04
Avg. Area House Age             4.863154e+00
Avg. Area Number of Rooms       7.537182e+00
Avg. Area Number of Bedrooms    6.350000e+00
Area Population                 3.526113e+04
Price                           1.109588e+06
Name: 46, dtype: float64
```

In [136]:
```
reg.predict([[50000,5,3,5,73734]])
```

Out[136]: array([760835.76798376])

In [140]:
```
reg.coef_
```

Out[140]:
```
array([2.15780494e+01, 1.65637027e+05, 1.20659949e+05, 1.65113905e+03,
       1.52007439e+01])
```

In [142]:
```
reg.intercept_
```

Out[142]: -2637299.0333314524

In [144]:
```
r_score =reg.score(X_test,y_test)
```

In [146]:
```
r_score
```

Out[146]: 0.9165644843776913

In [ ]: