

```
In [ ]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

%matplotlib inline
```

```
In [15]: HouseDF= pd.read_csv('data.csv')
```

```
In [16]: HouseDF.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4600 entries, 0 to 4599
Data columns (total 18 columns):
#   Column              Non-Null Count  Dtype  
---  --
0    date                4600 non-null  object  
1    price               4600 non-null  float64 
2    bedrooms            4600 non-null  float64 
3    bathrooms            4600 non-null  float64 
4    sqft_living          4600 non-null  int64   
5    sqft_lot             4600 non-null  int64   
6    floors              4600 non-null  float64 
7    waterfront          4600 non-null  int64   
8    view                4600 non-null  int64   
9    condition           4600 non-null  int64   
10   sqft_above           4600 non-null  int64   
11   sqft_basement        4600 non-null  int64   
12   yr_built             4600 non-null  int64   
13   yr_renovated         4600 non-null  int64   
14   street              4600 non-null  object  
15   city                4600 non-null  object  
16   statezip            4600 non-null  object  
17   country             4600 non-null  object  
dtypes: float64(4), int64(9), object(5)
memory usage: 647.0+ KB
```

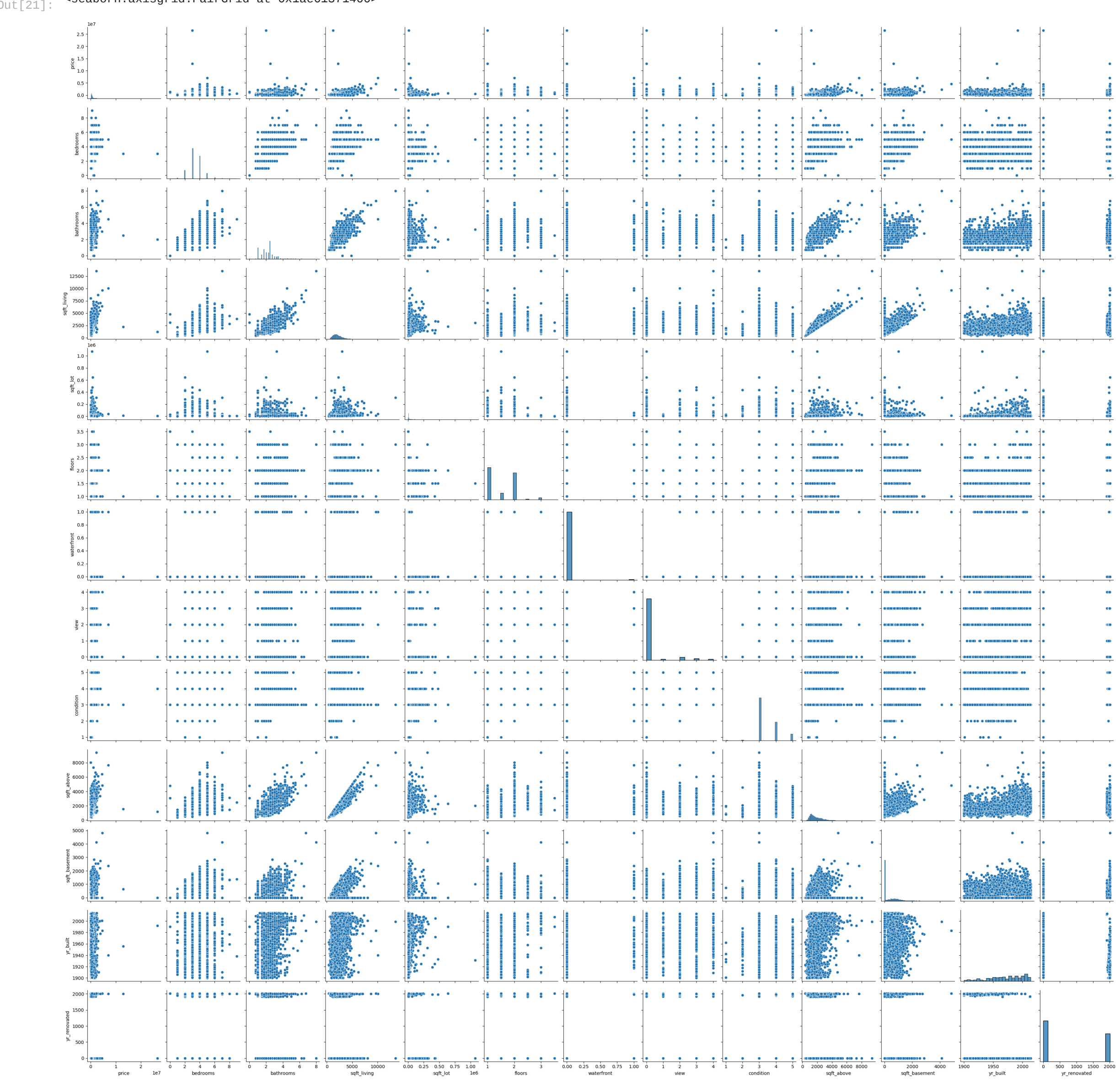
```
In [17]: HouseDF.describe()
```

	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	sqft_above	sqft_basement	yr_built	yr_renovated
count	4.600000e+03	4600.000000	4600.000000	4600.000000	4.600000e+03	4600.000000	4600.000000	4600.000000	4600.000000	4600.000000	4600.000000	4600.000000	4600.000000
mean	5.519630e+05	3.400870	2.160815	2139.346957	1.485252e+04	1.512065	0.007174	0.240652	3.451739	1827.265435	312.081522	1970.786304	808.608261
std	5.638347e+05	0.908848	0.783781	963.206916	3.588444e+04	0.538288	0.084404	0.778405	0.677230	862.168977	464.137228	29.731848	979.414536
min	0.000000e+00	0.000000	0.000000	370.000000	6.380000e+02	1.000000	0.000000	0.000000	1.000000	370.000000	0.000000	1900.000000	0.000000
25%	3.228750e+05	3.000000	1.750000	1460.000000	5.000750e+03	1.000000	0.000000	0.000000	3.000000	1190.000000	0.000000	1951.000000	0.000000
50%	4.609435e+05	3.000000	2.250000	1980.000000	7.683000e+03	1.500000	0.000000	0.000000	3.000000	1590.000000	0.000000	1976.000000	0.000000
75%	6.549625e+05	4.000000	2.500000	2620.000000	1.100125e+04	2.000000	0.000000	0.000000	4.000000	2300.000000	610.000000	1997.000000	1999.000000
max	2.659000e+07	9.000000	8.000000	13540.000000	1.074218e+06	3.500000	1.000000	4.000000	5.000000	9410.000000	4820.000000	2014.000000	2014.000000

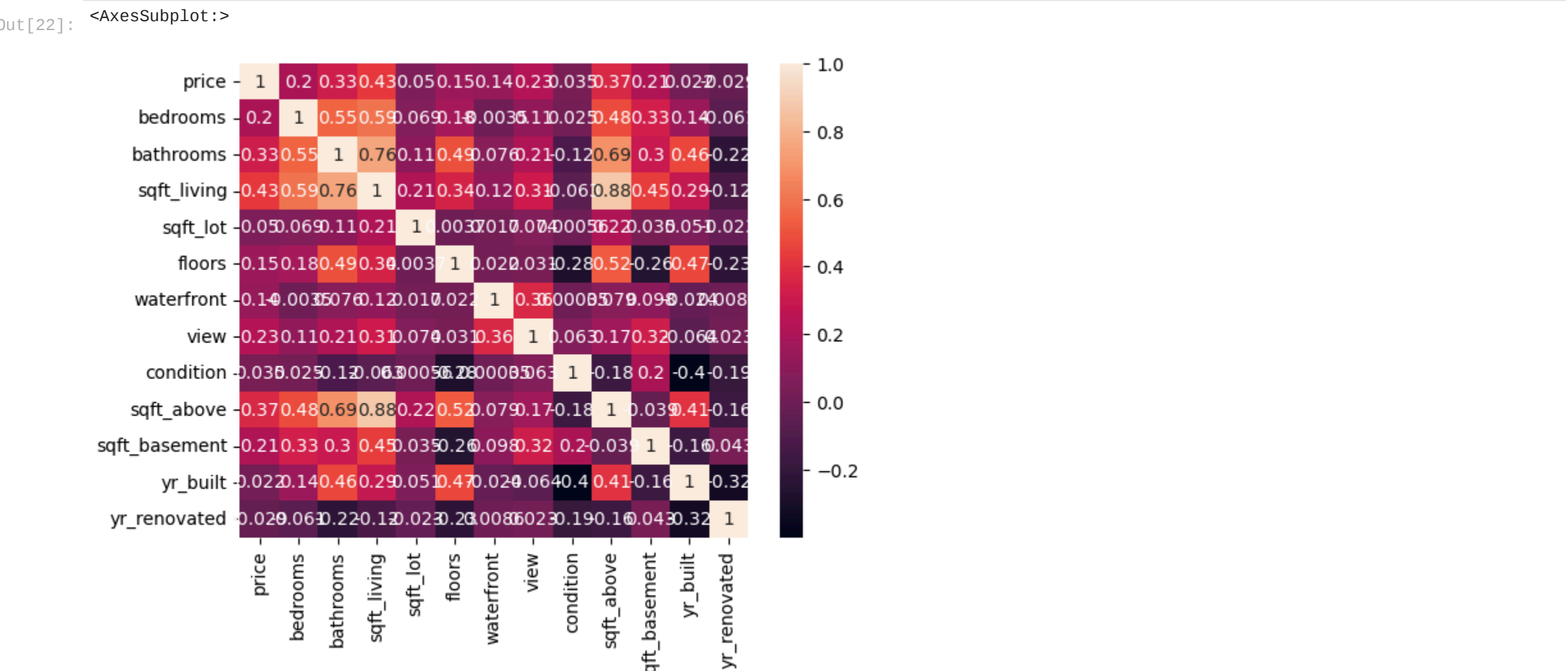
```
In [18]: HouseDF.columns
```

```
Out[18]: Index(['date', 'price', 'bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot',
      'floors', 'waterfront', 'view', 'condition', 'sqft_above', 'sqft_basement',
      'yr_built', 'yr_renovated', 'street', 'city', 'statezip', 'country'],
      dtype='object')
```

```
In [21]: sns.pairplot(HouseDF)
```



```
In [22]: sns.heatmap(HouseDF.corr(), annot=True)
```



```
In [31]: x = HouseDF[['floors', 'waterfront', 'view', 'condition', 'sqft_above','sqft_basement', 'yr_built', 'yr_renovated']]
y =HouseDF['price']
```

```
In [32]: from sklearn.model_selection import train_test_split
```

```
In [34]: x_train,x_test,y_train,y_test= train_test_split(x,y, test_size=0.40, random_state=101)
```

```
In [35]: x_train
```

Out[35]:

	floors	waterfront	view	condition	sqft_above	sqft_basement	yr_built	yr_renovated
695	2.0	0	0	3	3200	0	2004	2003
1170	1.0	0	0	4	1430	920	1977	0
684	1.0	0	0	4	700	0	1949	1985
2490	1.0	0	0	3	1150	0	1950	2005
2882	2.0	0	0	4	1290	0	1906	1990
...
4079	2.0	0	0	3	1810	0	1992	0
4171	1.0	0	0	3	2210	700	1957	1995
599	1.0	0	0	4	2450	0	1978	2000
1361	2.0	0	0	3	1982	0	2004	2003
1547	1.0	0	0	4	1240	1210	1900	1955

2760 rows × 8 columns

```
In [36]: from sklearn.linear_model import LinearRegression
```

```
In [37]: lm = LinearRegression()
```

```
In [38]: lm.fit(x_train, y_train)
```

```
Out[38]: LinearRegression()
```

```
In [40]: coeff_df = pd.DataFrame(lm.coef_,x.columns,columns=['Coefficient'])
```

```
In [41]: coeff_df
```

Out[41]:

	Coefficient
floors	82970.677923
waterfront	250039.925430
view	53189.095601
condition	23005.435232
sqft_above	258.970698
sqft_basement	234.135846
yr_built	-2549.855688
yr_renovated	3.319011

```
In [42]: predictions = lm.predict(x_test)
```

```
In [44]: plt.scatter(y_test, predictions)
```

Out[44]: <matplotlib.collections.PathCollection at 0x1ae8b7ec760>

```
In [ ]:
```