

```
In [1]: import numpy as np  
import pandas as pd
```

```
In [2]: import numpy as np  
import pandas as pd  
import seaborn as sns  
import matplotlib.pyplot as plt
```

In [7]: df.head()

Out[7]:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Address
0	79545.458574	5.682861	7.009188	4.09	23086.800503	1.059034e+06	208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
1	79248.642455	6.002900	6.730821	3.09	40173.072174	1.505891e+06	188 Johnson Views Suite 079\nLake Kathleen, CA...
2	61287.067179	5.865890	8.512727	5.13	36882.159400	1.058988e+06	9127 Elizabeth Stravenue\nDanielstown, WI 06482...
3	63345.240046	7.188236	5.586729	3.26	34310.242831	1.260617e+06	USS Barnett\nFPO AP 44820
4	59982.197226	5.040555	7.839388	4.23	26354.109472	6.309435e+05	USNS Raymond\nFPO AE 09386

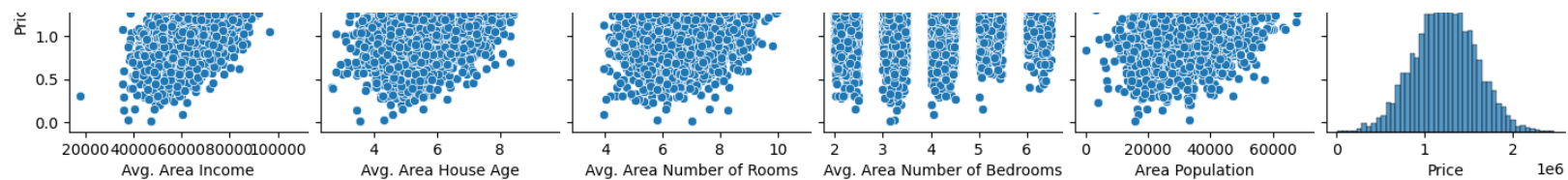
In [9]: df.tail()

Out[9]:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Address
4995	60567.944140	7.830362	6.137356	3.46	22837.361035	1.060194e+06	USNS Williams\nFPO AP 30153-7653
4996	78491.275435	6.999135	6.576763	4.02	25616.115489	1.482618e+06	PSC 9258, Box 8489\nAPO AA 42991-3352
4997	63390.686886	7.250591	4.805081	2.13	33266.145490	1.030730e+06	4215 Tracy Garden Suite 076\nJoshualand, VA 01...
4998	68001.331235	5.534388	7.130144	5.44	42625.620156	1.198657e+06	USS Wallace\nFPO AE 73316
4999	65510.581804	5.992305	6.792336	4.07	46501.283803	1.298950e+06	37778 George Ridges Apt. 509\nEast Holly, NV 2...

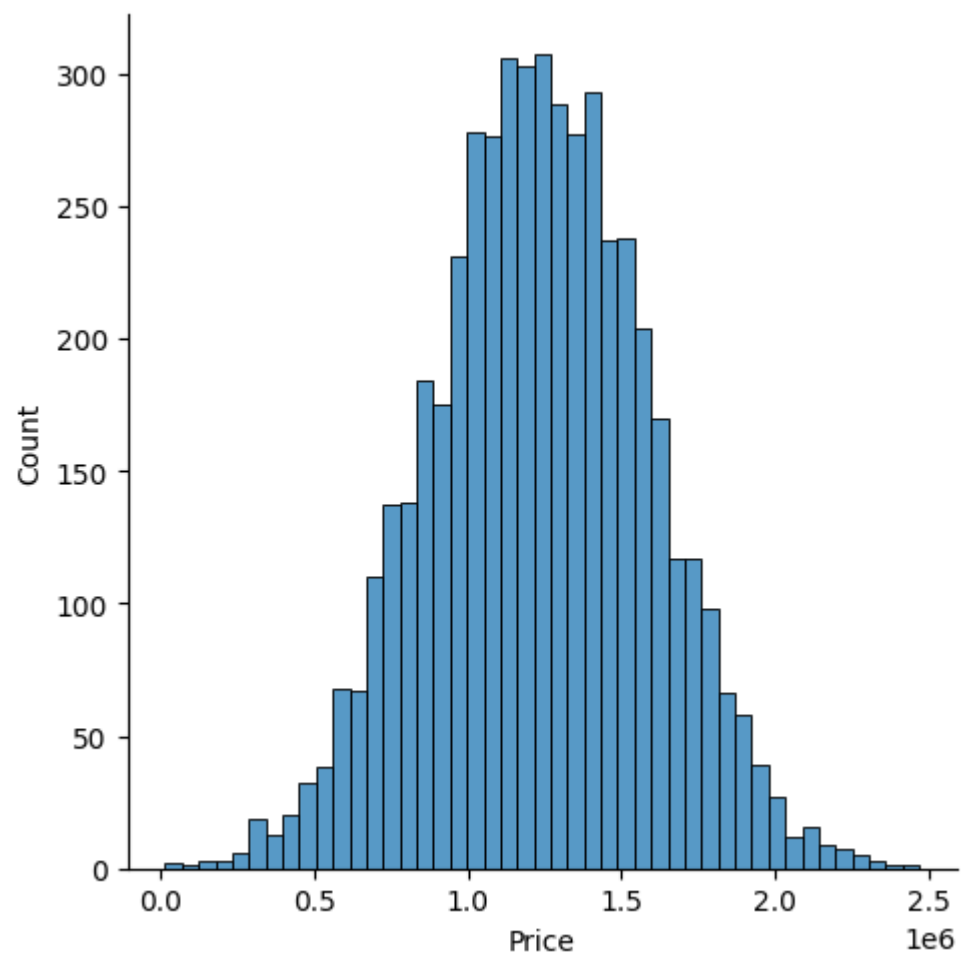
```
In [12]: df.columns
```

```
Out[12]: Index(['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',  
               'Avg. Area Number of Bedrooms', 'Area Population', 'Price', 'Address'],  
              dtype='object')
```

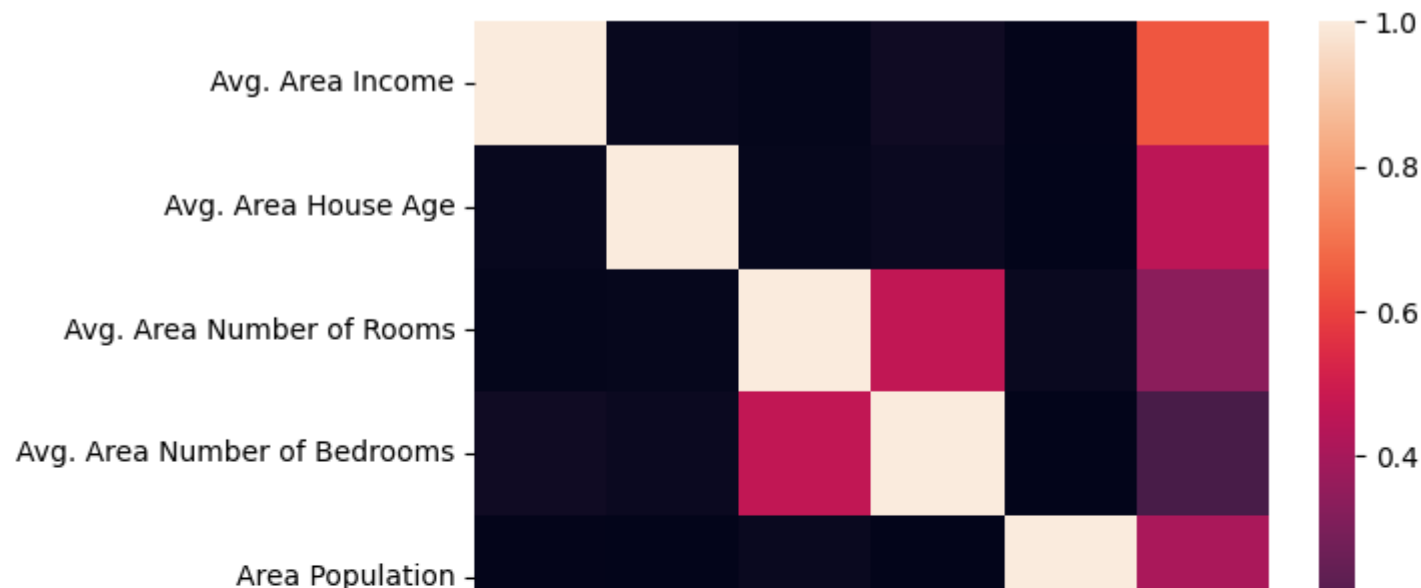
```
In [14]: sns.displot(df['Price'])
```

```
Out[14]: <seaborn.axisgrid.FacetGrid at 0x237ff43ecd0>
```



```
In [20]: sns.heatmap(Housedf.corr())
```

```
Out[20]: <Axes: >
```



```
In [21]: x=Housedf[['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',  
                  'Avg. Area Number of Bedrooms', 'Area Population']]  
y=df['Price']
```

```
In [22]: from sklearn.model_selection import train_test_split  
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=101)
```

```
In [23]: pip install scikitlearn
```

Note: you may need to restart the kernel to use updated packages.

ERROR: Could not find a version that satisfies the requirement scikitlearn (from versions: none)
ERROR: No matching distribution found for scikitlearn

```
In [30]: coeff_df=pd.DataFrame(lm.coef_,x.columns,columns=['coefficient'])  
coeff_df
```

Out[30]:

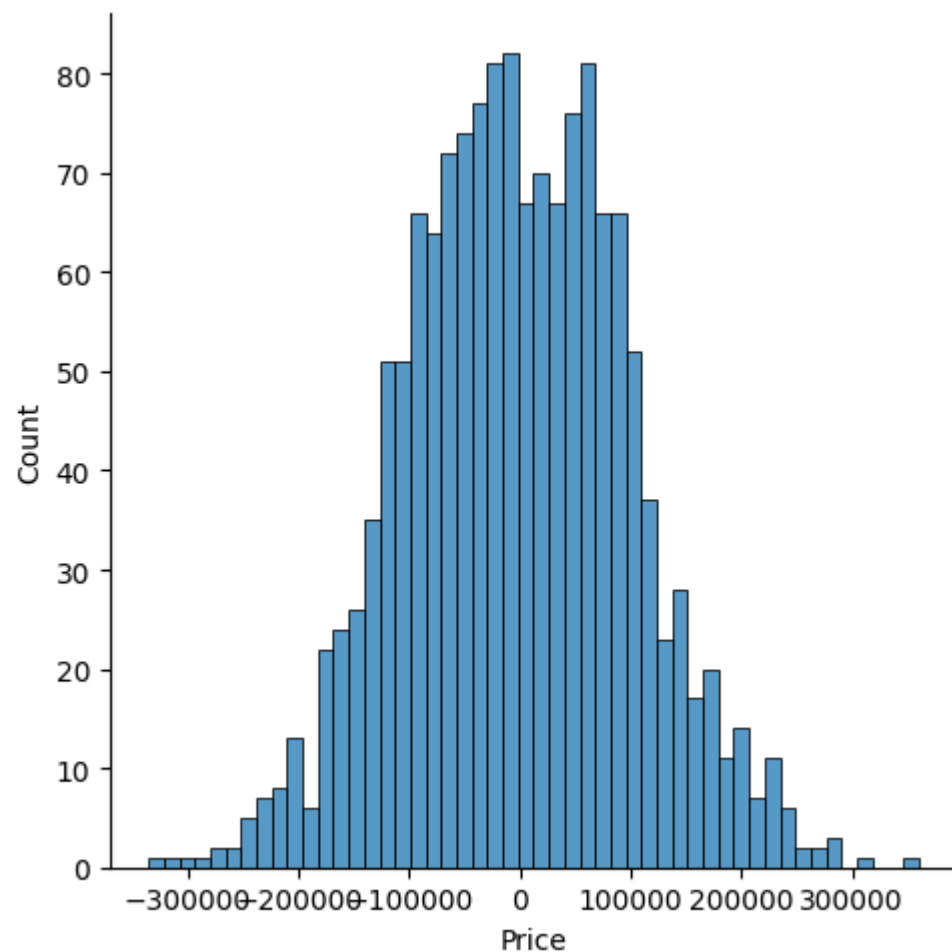
	coefficient
Avg. Area Income	21.617635
Avg. Area House Age	165221.119872
Avg. Area Number of Rooms	121405.376596
Avg. Area Number of Bedrooms	1318.718783
Area Population	15.225196

```
In [31]: coeff_df=pd.DataFrame(lm.coef_,x.columns,columns=['coefficient'])  
coeff_df
```

Out[31]:

	coefficient
Avg. Area Income	21.617635
Avg. Area House Age	165221.119872
Avg. Area Number of Rooms	121405.376596
Avg. Area Number of Bedrooms	1318.718783
Area Population	15.225196

```
In [33]: sns.displot((y_test-predictions),bins=50);
```



```
In [34]: from sklearn import metrics
print('MAE:',metrics.mean_absolute_error(y_test,predictions))
print('MSE:',metrics.mean_squared_error(y_test,predictions))
print('RMSE:',np.sqrt(metrics.mean_squared_error(y_test,predictions)))
```

```
MAE: 81257.55795855941
MSE: 10169125565.897606
RMSE: 100842.08231635048
```