In [4]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [5]:

```
df=pd.read_csv(r"C:\Users\Arshad Shaik\Downloads\archive.zip")
```

In [4]:

```
df.head(10)
```

Out[4]:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	A
0	79545.458574	5.682861	7.009188	4.09	23086.800503	1.059034e+06	208 Michael Fe 674\nLaurab
1	79248.642455	6.002900	6.730821	3.09	40173.072174	1.505891e+06	188 Johnsor Suite 079 Kathlee
2	61287.067179	5.865890	8.512727	5.13	36882.159400	1.058988e+06	9127 El Stravenue\nDani WI (
3	63345.240046	7.188236	5.586729	3.26	34310.242831	1.260617e+06	USS Barnett\nF
4	59982.197226	5.040555	7.839388	4.23	26354.109472	6.309435e+05	USNS Raymonc AE
5	80175.754159	4.988408	6.104512	4.04	26748.428425	1.068138e+06	06039 Jennifer Apt. 443\nTra
6	64698.463428	6.025336	8.147760	3.41	60828.249085	1.502056e+06	4759 Daniel 442\nNguyenbur
7	78394.339278	6.989780	6.620478	2.42	36516.358972	1.573937e+06	97 Viaduct\nLake V TN 1777
8	59927.660813	5.362126	6.393121	2.30	29387.396003	7.988695e+05	USS Gilbert\nF
9	81885.927184	4.423672	8.167688	6.10	40149.965749	1.545155e+06	Unit 94 0958\nDPO AE

In [5]:

```
df.describe()
```

Out[5]:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5.000000e+03
mean	68583.108984	5.977222	6.987792	3.981330	36163.516039	1.232073e+06
std	10657.991214	0.991456	1.005833	1.234137	9925.650114	3.531176e+05
min	17796.631190	2.644304	3.236194	2.000000	172.610686	1.593866e+04
25%	61480.562388	5.322283	6.299250	3.140000	29403.928702	9.975771e+05
50%	68804.286404	5.970429	7.002902	4.050000	36199.406689	1.232669e+06
75%	75783.338666	6.650808	7.665871	4.490000	42861.290769	1.471210e+06
max	107701.748378	9.519088	10.759588	6.500000	69621.713378	2.469066e+06

In [6]:

```
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
```

In [7]:

```
df.columns
```

Out[7]:

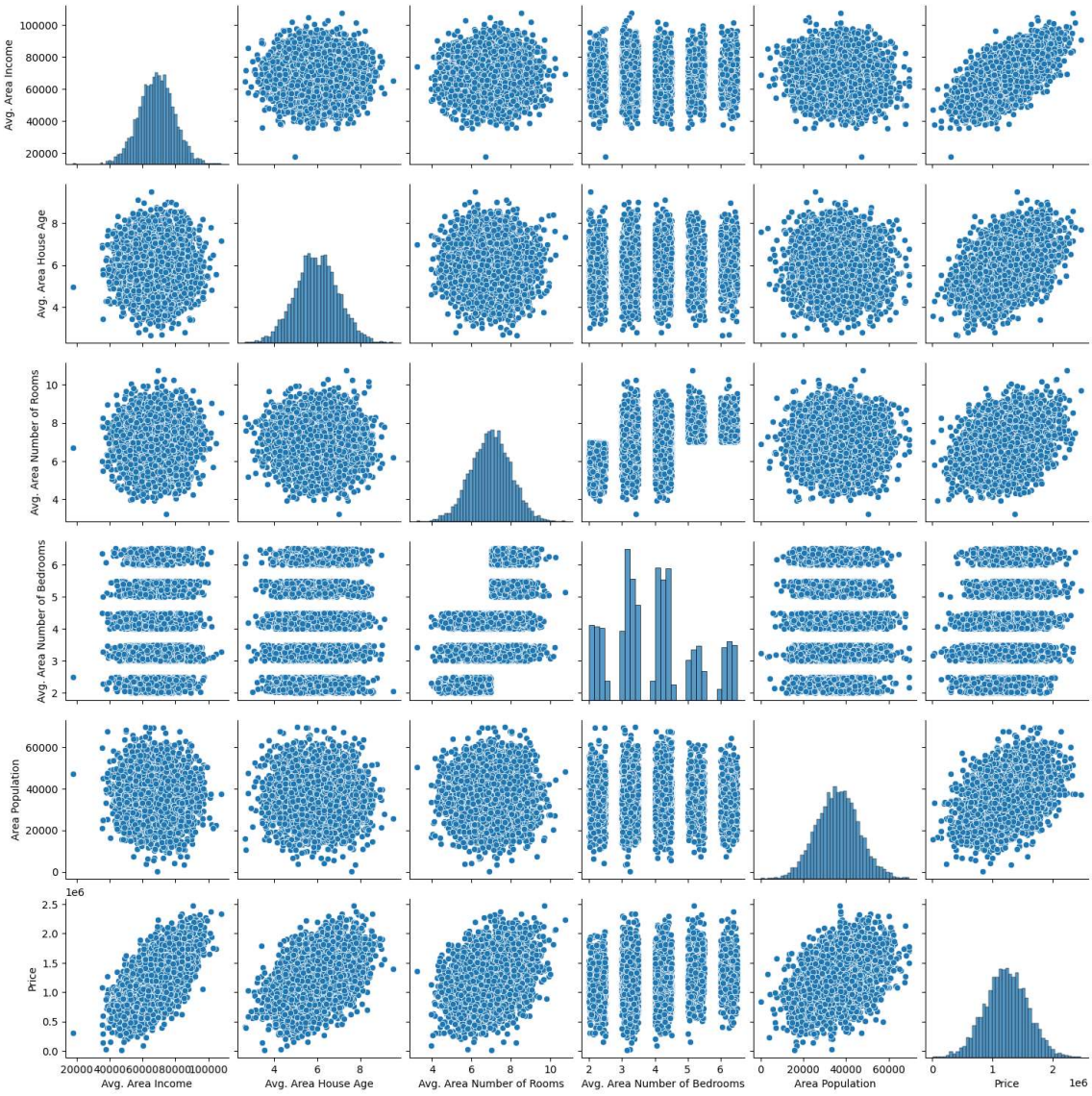
```
Index(['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Ro
oms',
       'Avg. Area Number of Bedrooms', 'Area Population', 'Price', 'Addre
ss'],
      dtype='object')
```

In [8]:

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

Out[8]:

```
<seaborn.axisgrid.PairGrid at 0x1d518449d20>
```

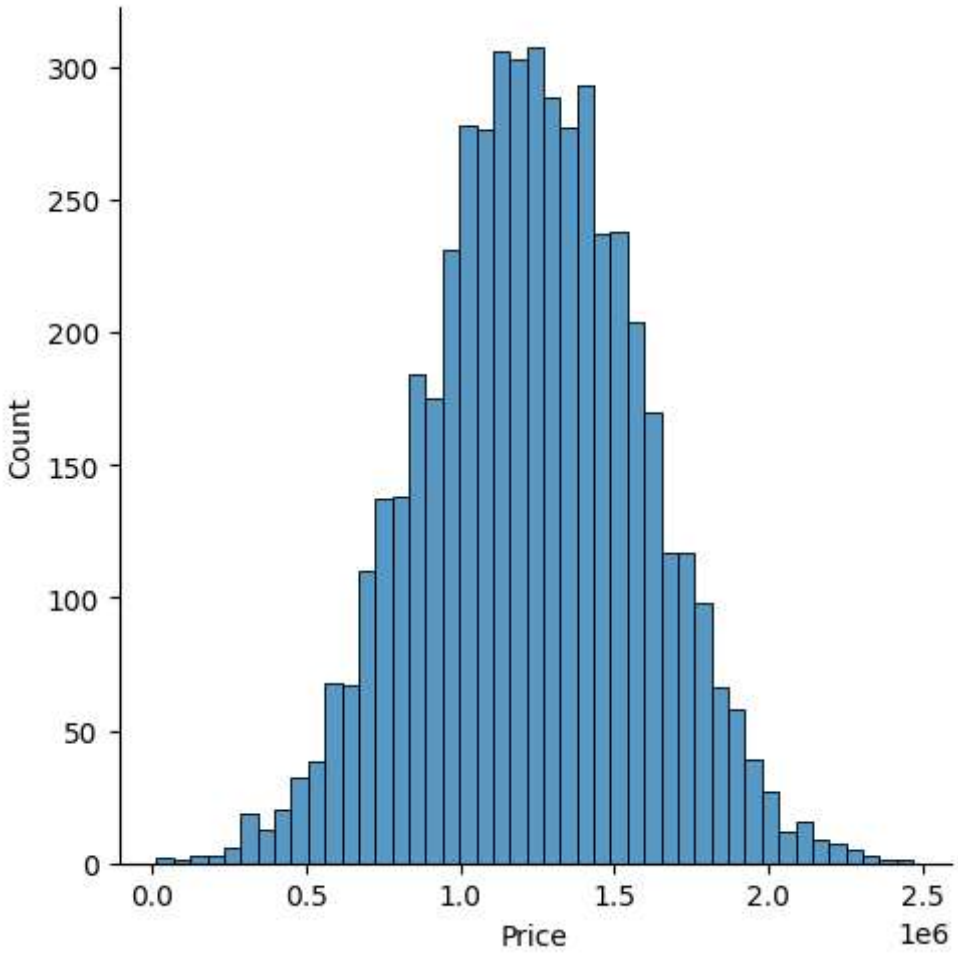


In [9]:

```
sns.displot(df['Price'])
```

Out[9]:

```
<seaborn.axisgrid.FacetGrid at 0x1d52ddab640>
```

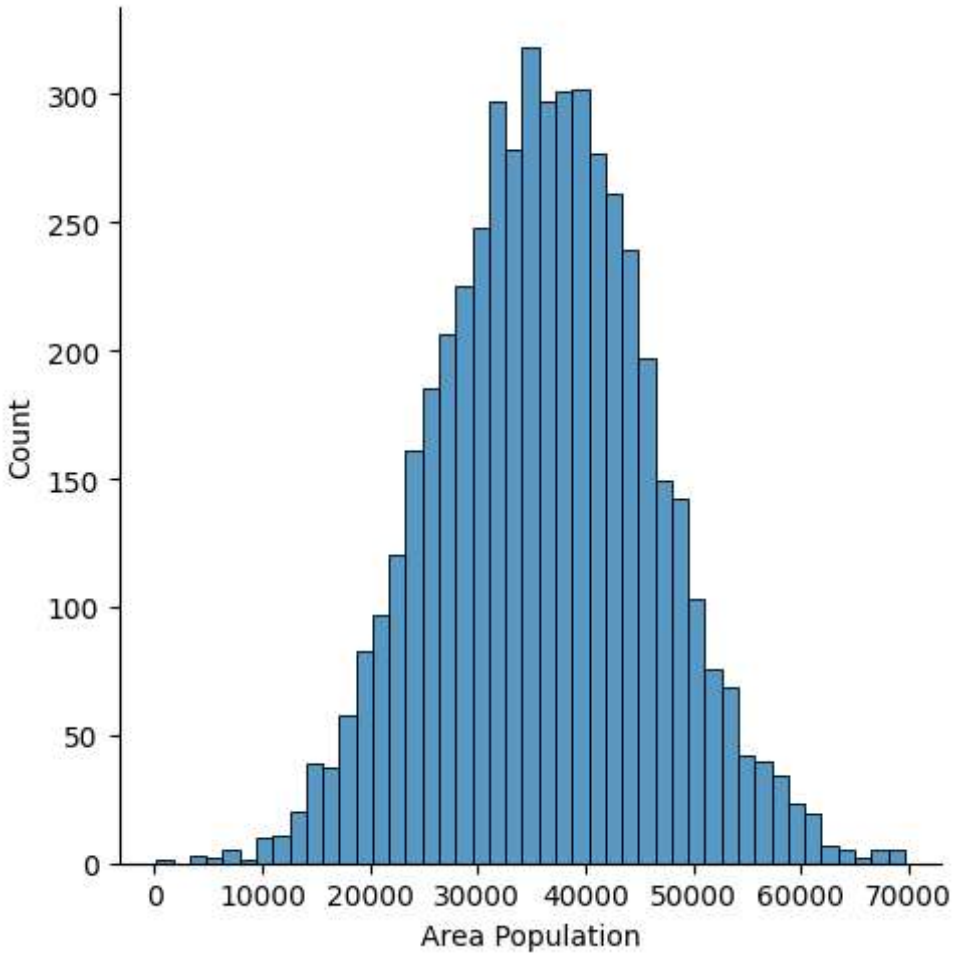


In [10]:

```
sns.displot(df['Area Population'])
```

Out[10]:

```
<seaborn.axisgrid.FacetGrid at 0x1d52dede830>
```



In [11]:

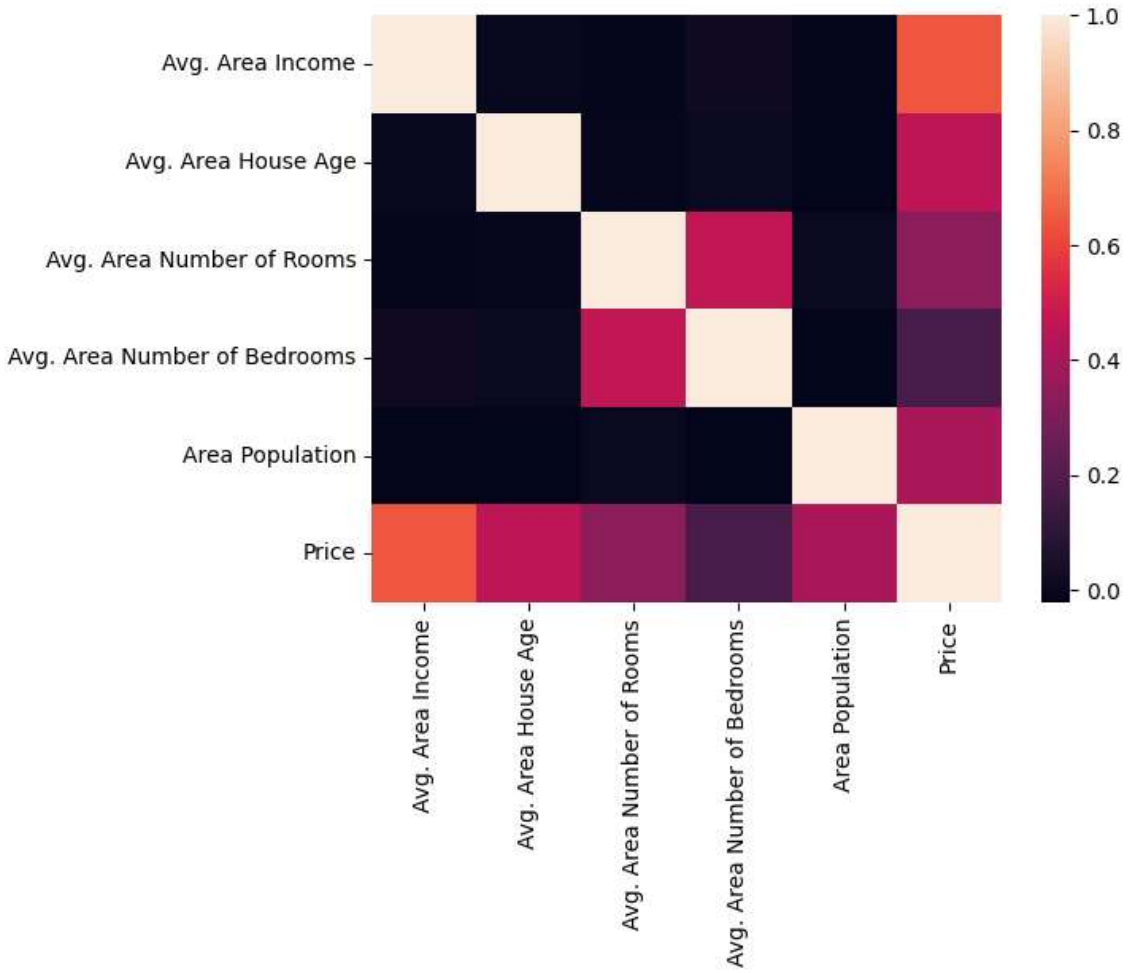
```
Housedf=df[['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',
            'Avg. Area Number of Bedrooms', 'Area Population', 'Price']]
```

In [12]:

```
sns.heatmap(Housedf.corr())
```

Out[12]:

```
<Axes: >
```



In [6]:

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

In [7]:

```
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 [8]:

```
from sklearn.linear_model import LinearRegression
lm=LinearRegression()
lm.fit(x_train,y_train)
```

Out[8]:

LinearRegression

LinearRegression()

In [9]:

```
print(lm.intercept_)
```

```
-2641372.6673006266
```

In [10]:

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

Out[10]:

	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 [11]:

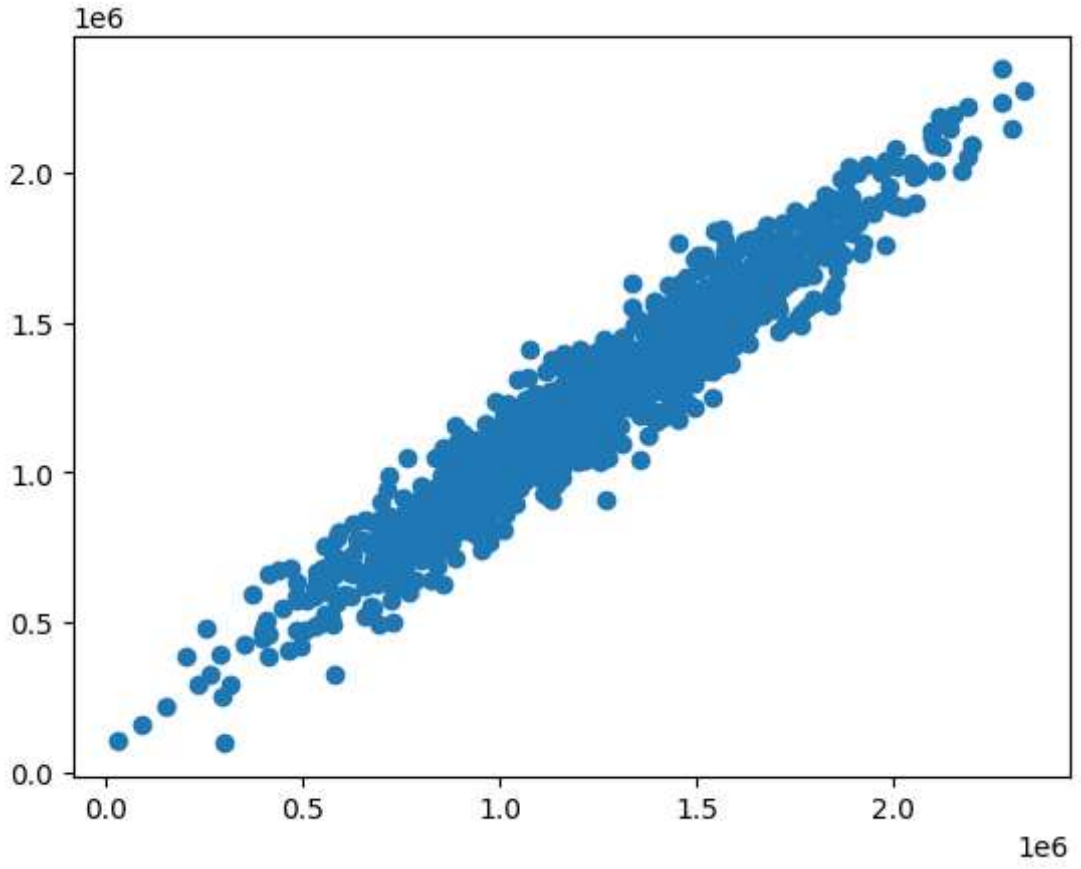
```
predictions=lm.predict(x_test)
```

In [12]:

```
plt.scatter(y_test,predictions)
```

Out[12]:

```
<matplotlib.collections.PathCollection at 0x1a602ebe020>
```

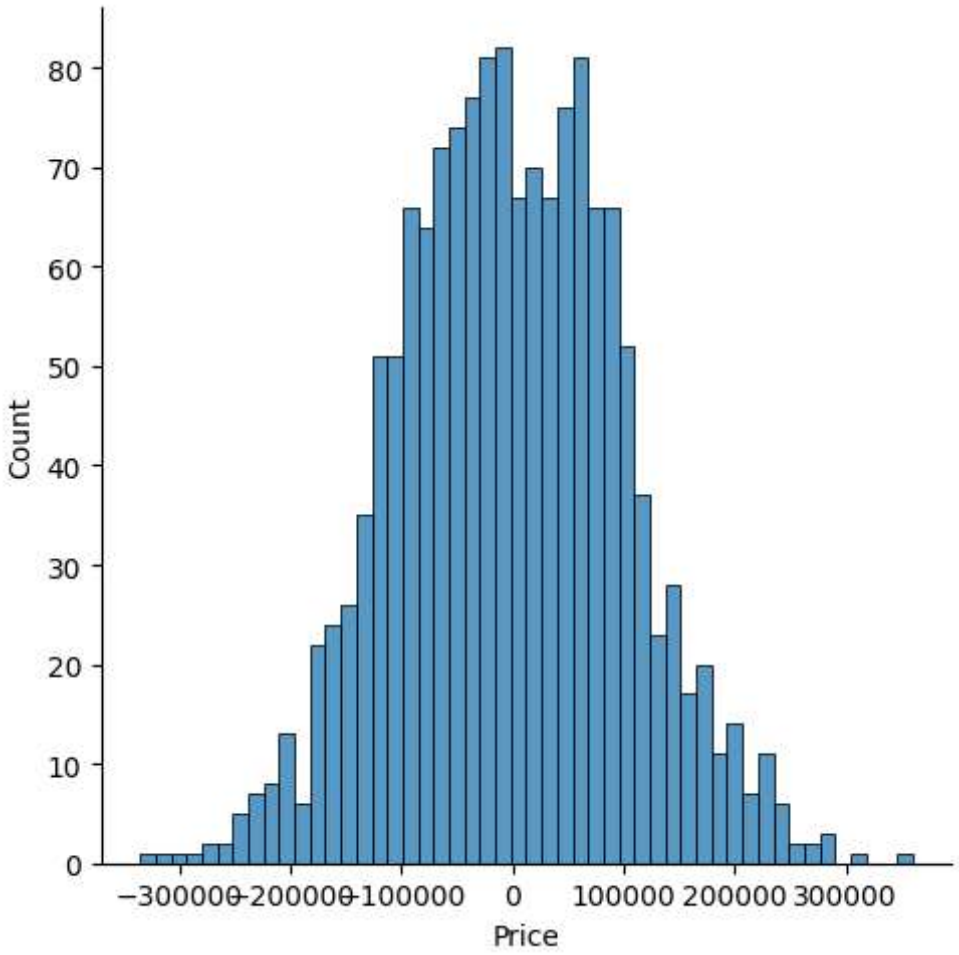


In [13]:

```
sns.displot((y_test-predictions),bins=50)
```

Out[13]:

```
<seaborn.axisgrid.FacetGrid at 0x1a6022e7010>
```



In [14]:

```
from sklearn import metrics
```

In [15]:

```
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: 81257.5579585557
MSE: 10169125565.89724
RMSE: 100842.08231634866
```

In [ ]:

```

```