In [1]:

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

In [2]:

```
df=pd.read_csv(r"C:\Users\MSI\Downloads\USA_Housing.csv")
df
```

Out[2]:

	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\nDanielltown, WI 06482...
3	63345.240046	7.188236	5.586729	3.26	34310.242831	1.260617e+06	USS Barnett\nFPO AP 44820

In [3]:

```
df.head()
```

Out[3]:

	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\nLaurabı
1	79248.642455	6.002900	6.730821	3.09	40173.072174	1.505891e+06	188 Johnsor Suite 079 Kathleer
2	61287.067179	5.865890	8.512727	5.13	36882.159400	1.058988e+06	9127 Eli Stravenue\nDani WI 0
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

In [4]:

```
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 [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 [7]:

```
df.columns
```

Out[7]:

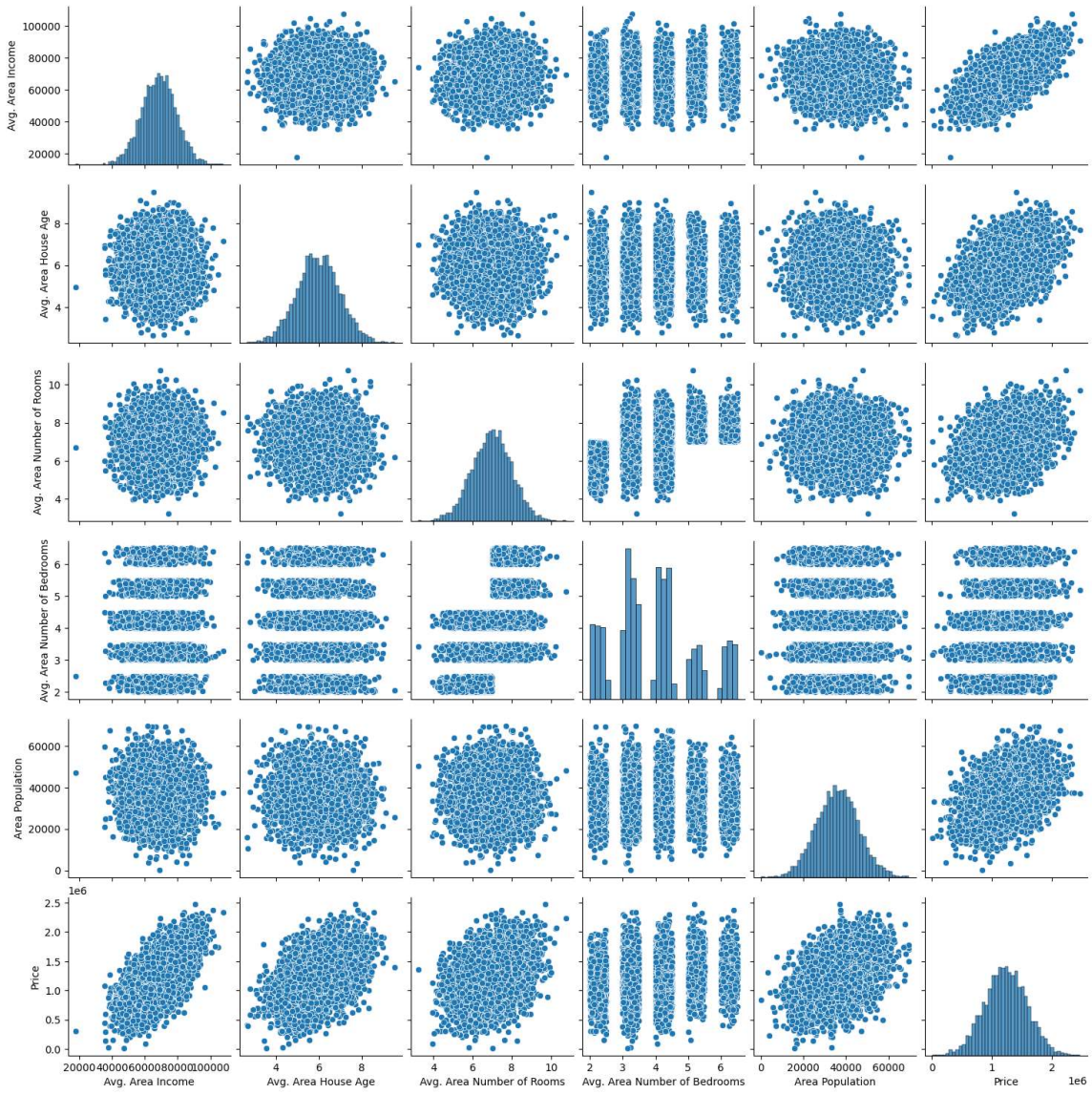
```
Index(['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Roo
ms',
       'Avg. Area Number of Bedrooms', 'Area Population', 'Price', 'Addres
s'],
      dtype='object')
```

In [8]:

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

Out[8]:

```
<seaborn.axisgrid.PairGrid at 0x24de843f210>
```

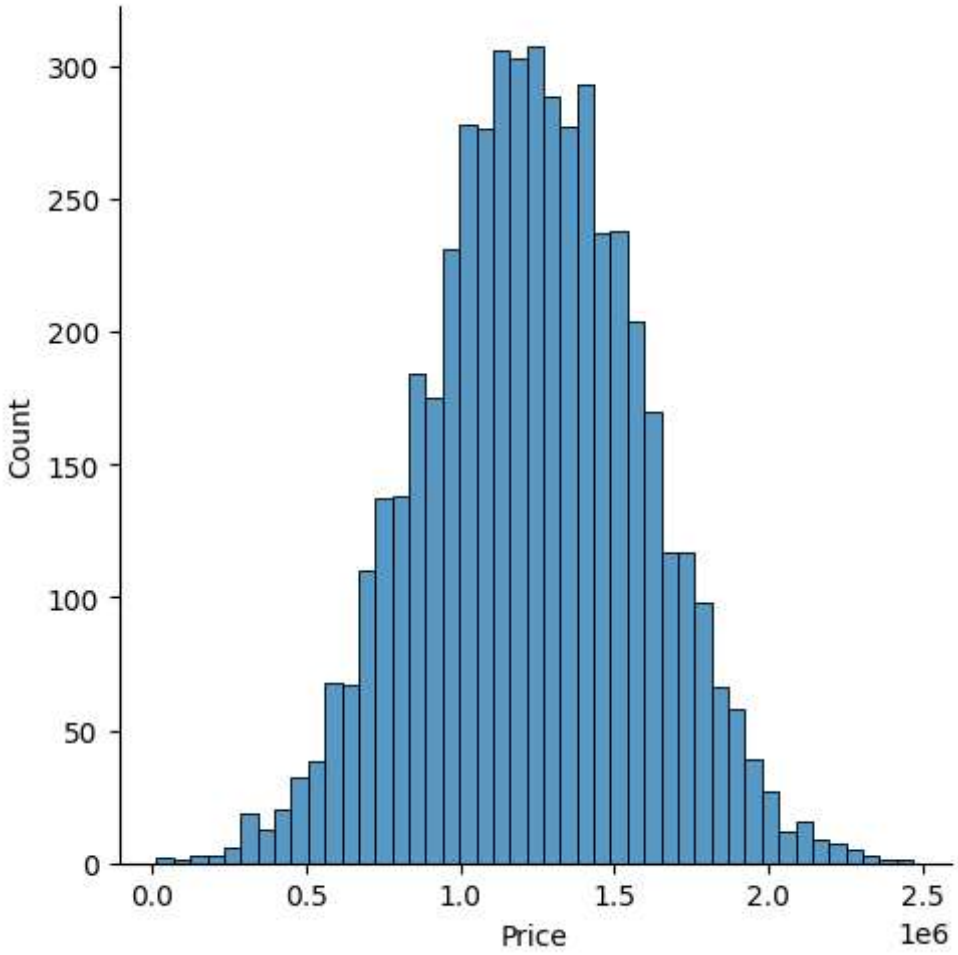


In [10]:

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

Out[10]:

```
<seaborn.axisgrid.FacetGrid at 0x24ded344ad0>
```

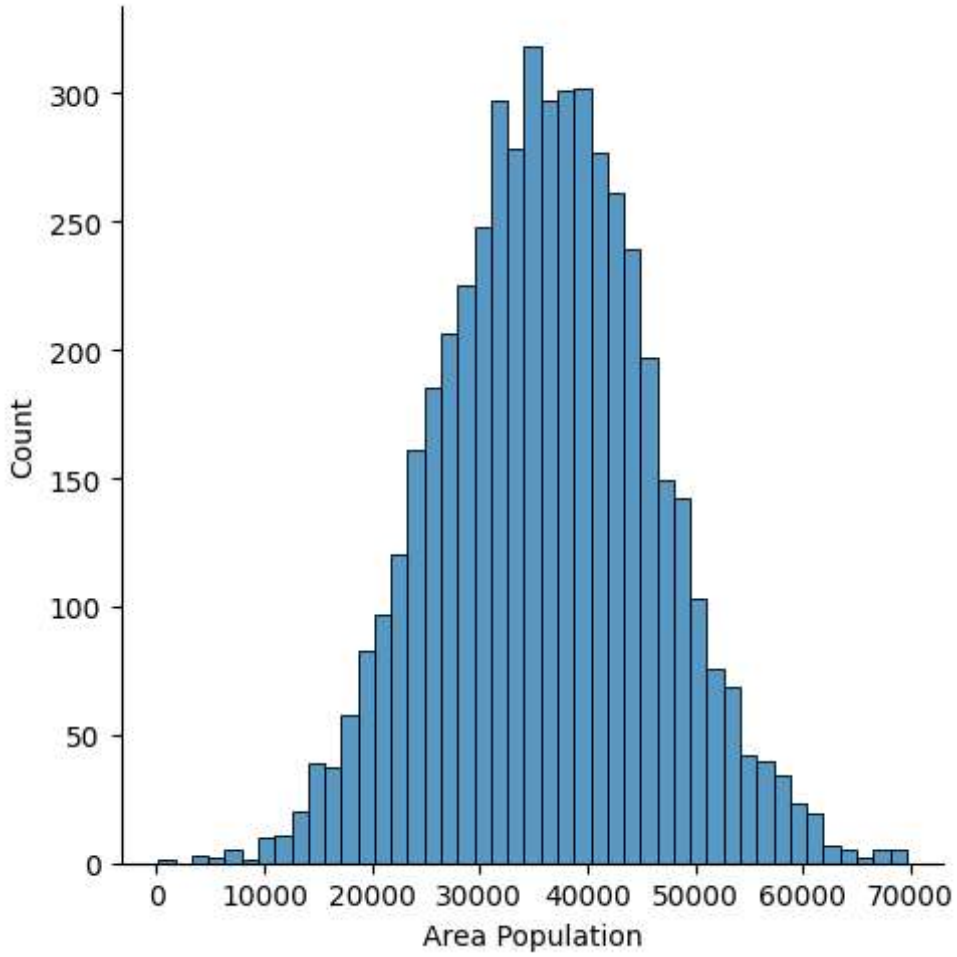


In [12]:

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

Out[12]:

```
<seaborn.axisgrid.FacetGrid at 0x24dec204d50>
```



In [19]:

```
Housedf=df[['Avg.Area Income','Avg.Area House Age','Avg.Area Number Of Rooms','Avg.Area
sns.heatmap(Housedf.corr())
```

```
---------------------------------------------------------------------------
-
KeyError                                  Traceback (most recent call las
t)
Cell In[19], line 1
----> 1 Housedf=df[['Avg.Area Income','Avg.Area House Age','Avg.Area Numbe
r Of Rooms','Avg.Area Of Number Of Bedrooms','Area Population','Price']]
      2 sns.heatmap(Housedf.corr())

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\co
re\frame.py:3767, in DataFrame.__getitem__(self, key)
   3765     if is_iterator(key):
   3766         key = list(key)
-> 3767     indexer = self.columns._get_indexer_strict(key, "columns")[1]
   3769 # take() does not accept boolean indexers
   3770 if getattr(indexer, "dtype", None) == bool:

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\co
re\indexes\base.py:5876, in Index._get_indexer_strict(self, key, axis_nam
e)
   5873 else:
   5874     keyarr, indexer, new_indexer = self._reindex_non_unique(keyar
r)
-> 5876 self._raise_if_missing(keyarr, indexer, axis_name)
   5878 keyarr = self.take(indexer)
   5879 if isinstance(key, Index):
   5880     # GH 42790 - Preserve name from an Index

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\pandas\co
re\indexes\base.py:5938, in Index._raise_if_missing(self, key, indexer, ax
is_name)
   5935     raise KeyError(f"None of [{key}] are in the [{axis_name}]")
   5937 not_found = list(ensure_index(key)[missing_mask.nonzero()[0]].uniq
ue())
-> 5938 raise KeyError(f"{not_found} not in index")

KeyError: "['Avg.Area Income', 'Avg.Area House Age', 'Avg.Area Number Of R
ooms', 'Avg.Area Of Number Of Bedrooms'] not in index"
```

In [22]:

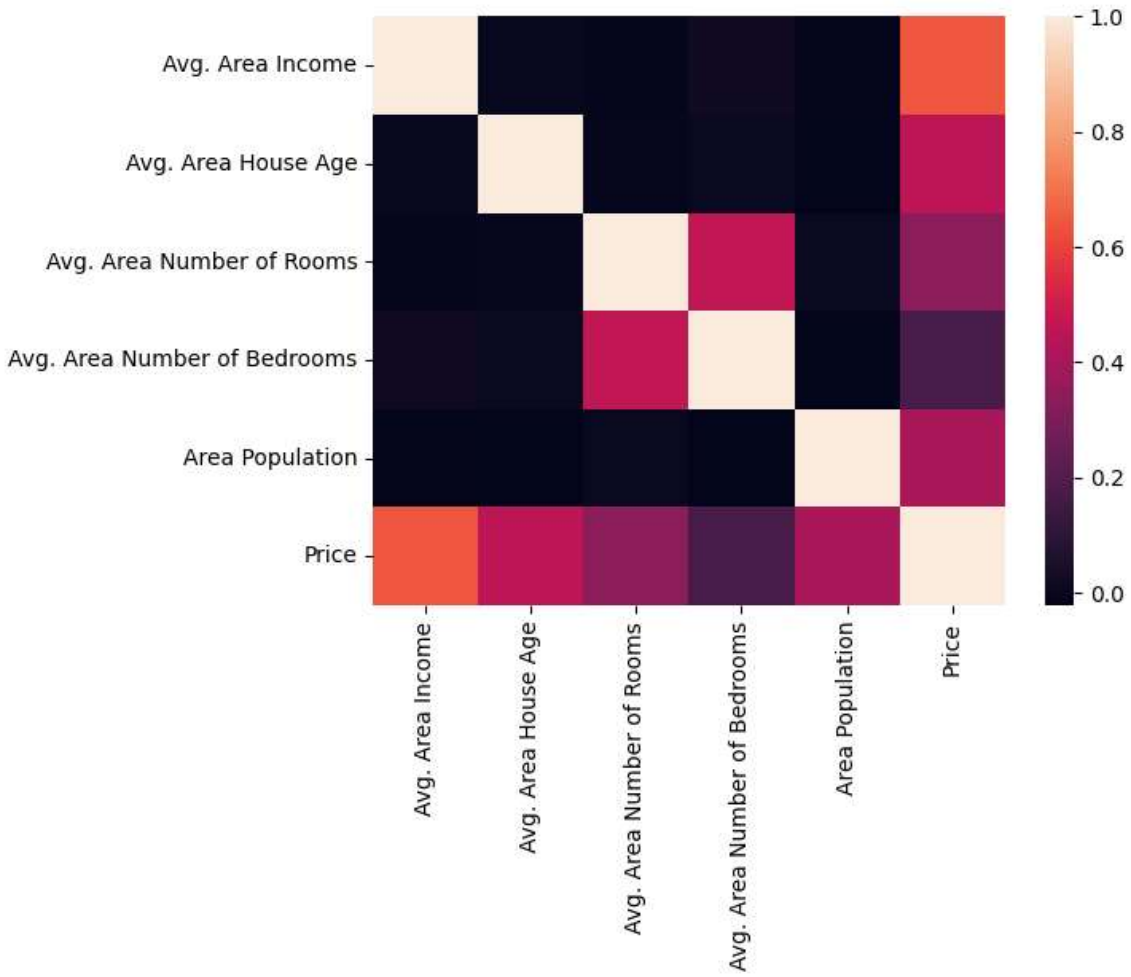
```
Housedf=df[['Avg. Area Income','Avg. Area House Age','Avg. Area Number of Rooms',
       'Avg. Area Number of Bedrooms','Area Population','Price']]
```

In [23]:

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

Out[23]:

```
<Axes: >
```



In [24]:

```
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 [25]:

```
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 [27]:

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

Out[27]:

LinearRegression

LinearRegression()

In [28]:

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

```
-2641372.6673014304
```

In [29]:

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

Out[29]:

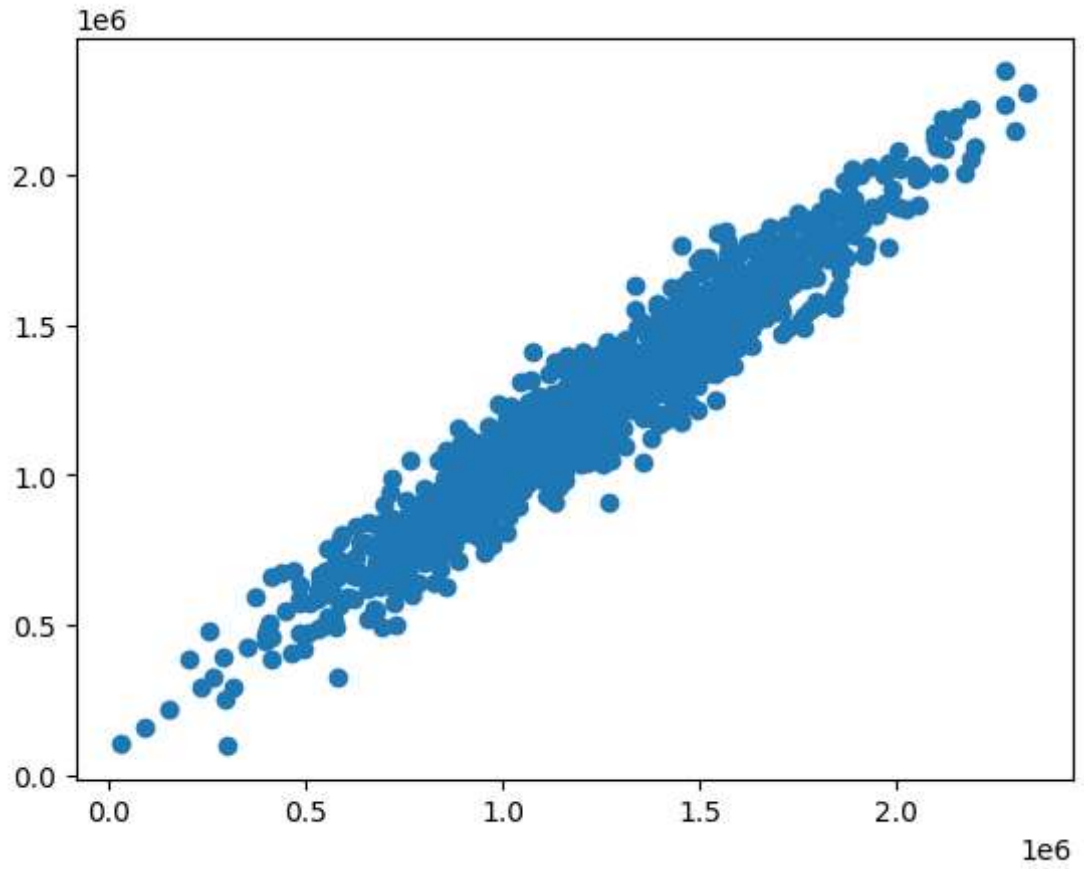
	coefficient
Avg. Area Income	21.617635
Avg. Area House Age	165221.119872
Avg. Area Number of Rooms	121405.376596
Avg. Area Number of Bedrooms	1318.718783
Area Population	15.225196

In [31]:

```
predictions=lm.predict(x_test)
plt.scatter(y_test,predictions)
```

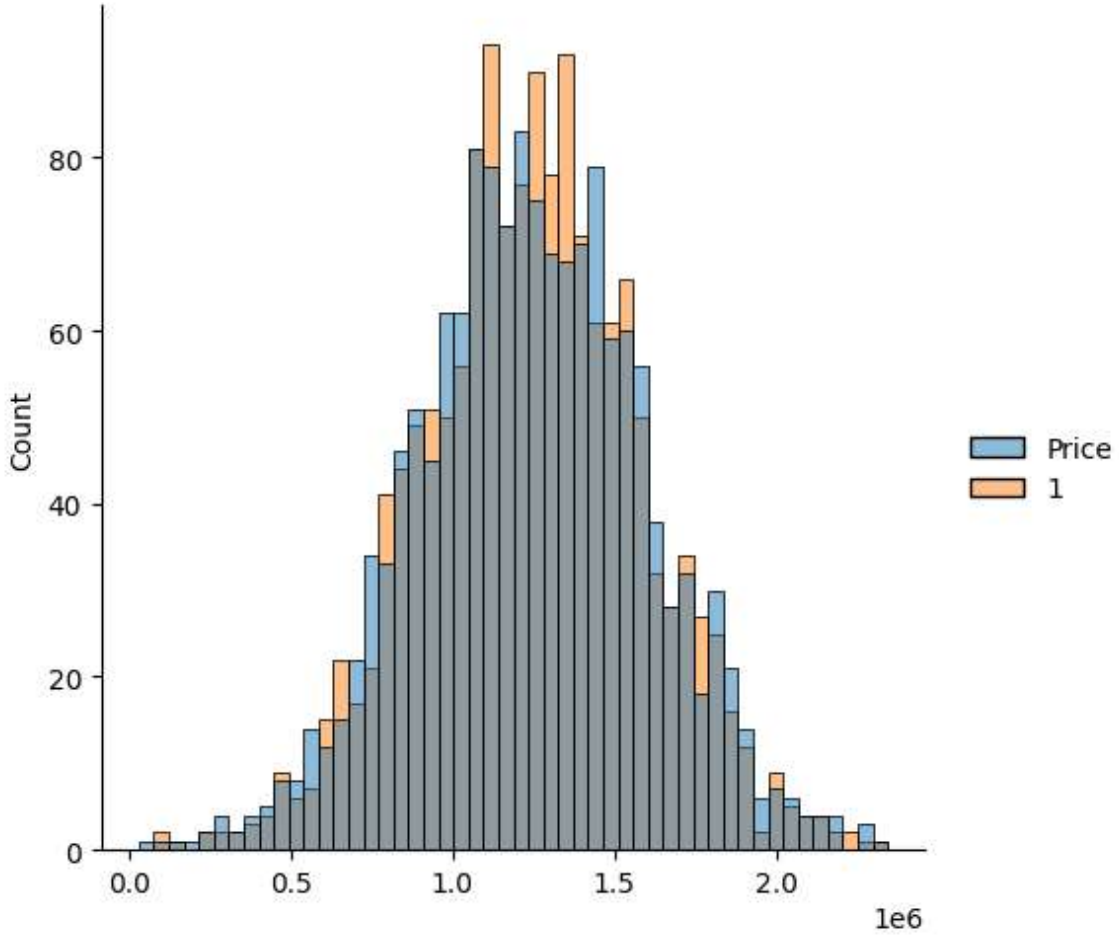
Out[31]:

```
<matplotlib.collections.PathCollection at 0x24df022b0d0>
```



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('MAE:',metrics.mean_squared_error(y_test,predictions))
print('RMSE:',np.sqrt(metrics.mean_squared_error(y_test,predictions)))
```

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

In [ ]:

```
P
```