In [54]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

In [55]:
```
df = pd.read_csv("D:/USA Housing.csv")
```

In [56]:
```
df
```

Out[56]:

s	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	sqft_above	sqft_baseme
3	1.50	1340	7912	1.5	0	0	3	1340	
5	2.50	3650	9050	2.0	0	4	5	3370	2
3	2.00	1930	11947	1.0	0	0	4	1930	
3	2.25	2000	8030	1.0	0	0	4	1000	10
4	2.50	1940	10500	1.0	0	0	4	1140	8
..	...	...	...	...	...	...	...	...	
3	1.75	1510	6360	1.0	0	0	4	1510	
3	2.50	1460	7573	2.0	0	0	3	1460	
3	2.50	3010	7014	2.0	0	0	3	3010	
4	2.00	2090	6630	1.0	0	0	3	1070	10
3	2.50	1490	8102	2.0	0	0	4	1490	

In [57]:
```
print(df.head())
```

```
                 date      price  bedrooms  bathrooms  sqft_living  sqft_l
ot  \
0  2014-05-02 0:00:00   313000.0         3       1.50         1340      79
12
1  2014-05-02 0:00:00  2384000.0         5       2.50         3650      90
50
2  2014-05-02 0:00:00   342000.0         3       2.00         1930     119
47
3  2014-05-02 0:00:00   420000.0         3       2.25         2000      80
30
4  2014-05-02 0:00:00   550000.0         4       2.50         1940     105
00

   floors  waterfront  view  condition  sqft_above  sqft_basement  yr_buil
t  \
0     1.5           0     0          3        1340              0      195
5
1     2.0           0     4          5        3370            280      192
1
2     1.0           0     0          4        1930              0      196
6
3     1.0           0     0          4        1000           1000      196
3
4     1.0           0     0          4        1140            800      197
6

   yr_renovated                    street       city  statezip country
0          2005      18810 Densmore Ave N  Shoreline  WA 98133     USA
1             0           709 W Blaine St    Seattle  WA 98119     USA
2             0  26206-26214 143rd Ave SE       Kent  WA 98042     USA
3             0           857 170th Pl NE   Bellevue  WA 98008     USA
4          1992         9105 170th Ave NE    Redmond  WA 98052     USA
```

In [58]:
```
print(df.describe())
```

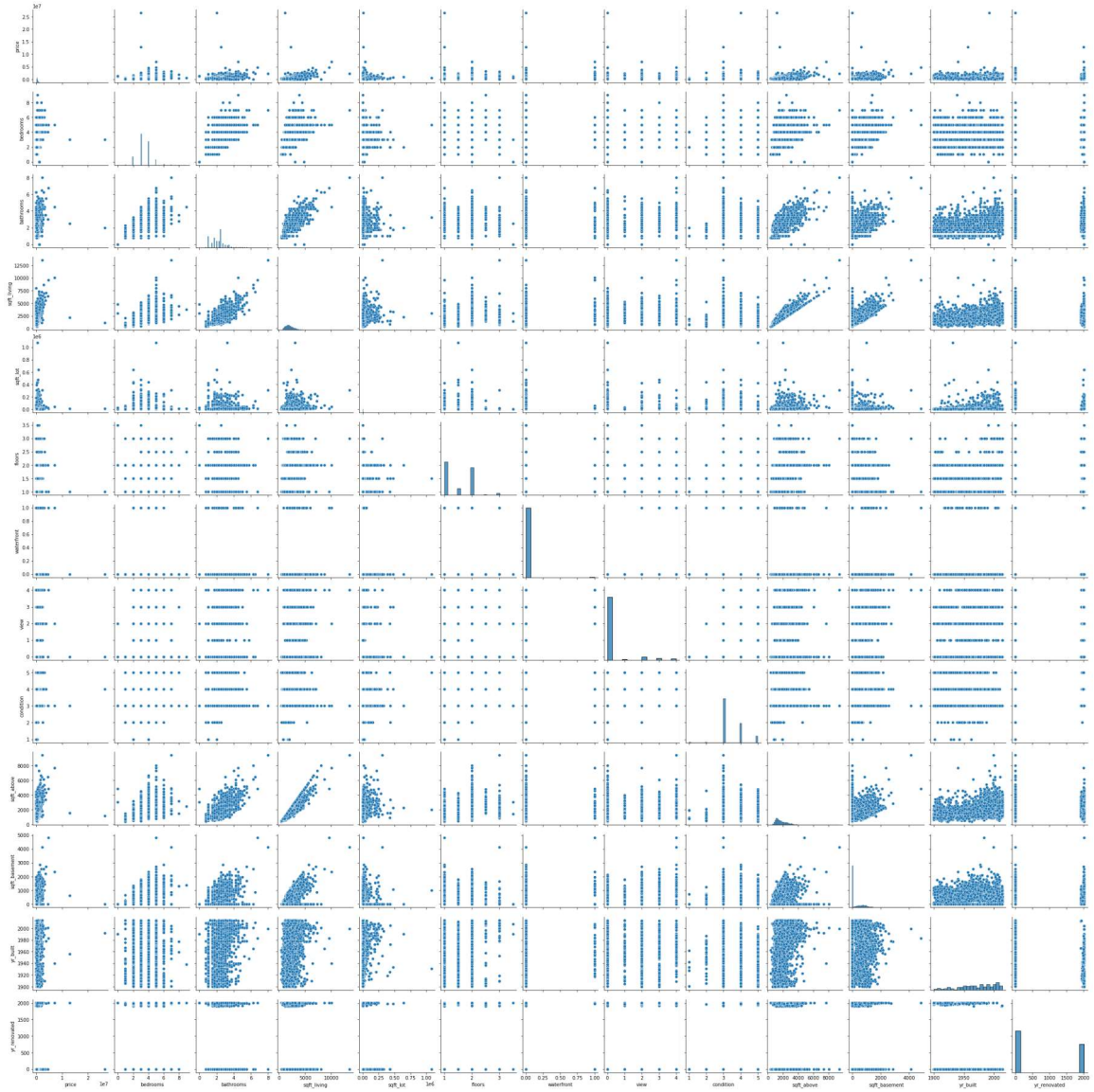
```
              price     bedrooms    bathrooms   sqft_living      sqft_lot
\
count  4.600000e+03  4600.000000  4600.000000   4600.000000  4.600000e+03
mean   5.519630e+05     3.400870     2.160815   2139.346957  1.485252e+04
std    5.638347e+05     0.908848     0.783781    963.206916  3.588444e+04
min    0.000000e+00     0.000000     0.000000    370.000000  6.380000e+02
25%    3.228750e+05     3.000000     1.750000   1460.000000  5.000750e+03
50%    4.609435e+05     3.000000     2.250000   1980.000000  7.683000e+03
75%    6.549625e+05     4.000000     2.500000   2620.000000  1.100125e+04
max    2.659000e+07     9.000000     8.000000  13540.000000  1.074218e+06

             floors   waterfront         view    condition   sqft_above  \
count  4600.000000  4600.000000  4600.000000  4600.000000  4600.000000
mean      1.512065     0.007174     0.240652     3.451739  1827.265435
std       0.538288     0.084404     0.778405     0.677230   862.168977
min       1.000000     0.000000     0.000000     1.000000   370.000000
25%       1.000000     0.000000     0.000000     3.000000  1190.000000
50%       1.500000     0.000000     0.000000     3.000000  1590.000000
75%       2.000000     0.000000     0.000000     4.000000  2300.000000
max       3.500000     1.000000     4.000000     5.000000  9410.000000

       sqft_basement     yr_built  yr_renovated
count    4600.000000  4600.000000   4600.000000
mean      312.081522  1970.786304    808.608261
std       464.137228    29.731848    979.414536
min         0.000000  1900.000000      0.000000
25%         0.000000  1951.000000      0.000000
50%         0.000000  1976.000000      0.000000
75%       610.000000  1997.000000   1999.000000
max      4820.000000  2014.000000   2014.000000
```

In [59]:
```
print(df.isnull().sum())
```

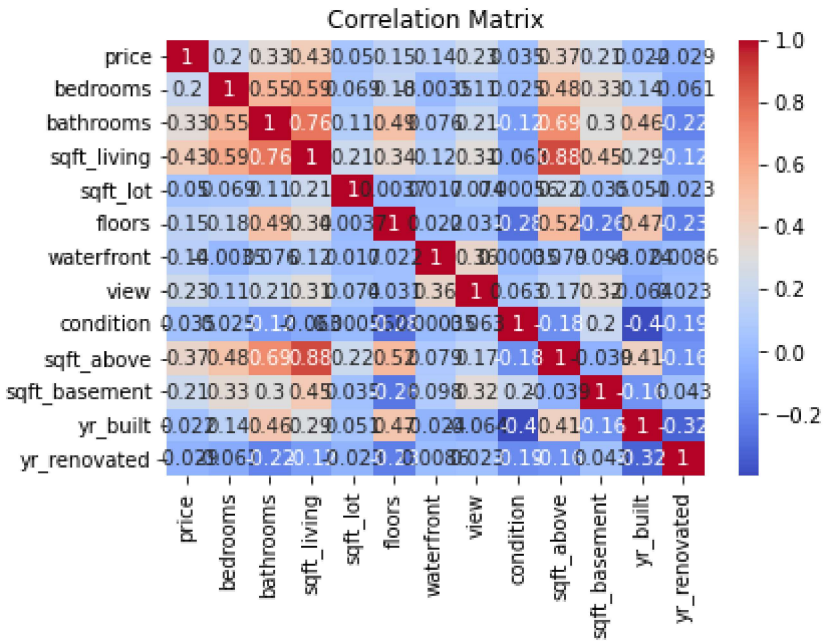
```
date             0
price            0
bedrooms         0
bathrooms        0
sqft_living      0
sqft_lot         0
floors           0
waterfront       0
view             0
condition        0
sqft_above       0
sqft_basement    0
yr_built         0
yr_renovated     0
street           0
city             0
statezip         0
country          0
dtype: int64
```

In [60]:
```
sns.pairplot(df)
```

Out[60]:
```
<seaborn.axisgrid.PairGrid at 0x162e5113ee0>
```



In [61]:
```
correlation_matrix = df.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title("Correlation Matrix")
plt.show()
```



In [62]:
```
X = df[['bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot', 'floors', 'wate
y = df['price']
```

In [63]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, ra
```

In [64]:
```
model = LinearRegression()
```

In [65]:
```
model.fit(X_train, y_train)
```

Out[65]:
```
LinearRegression()
```

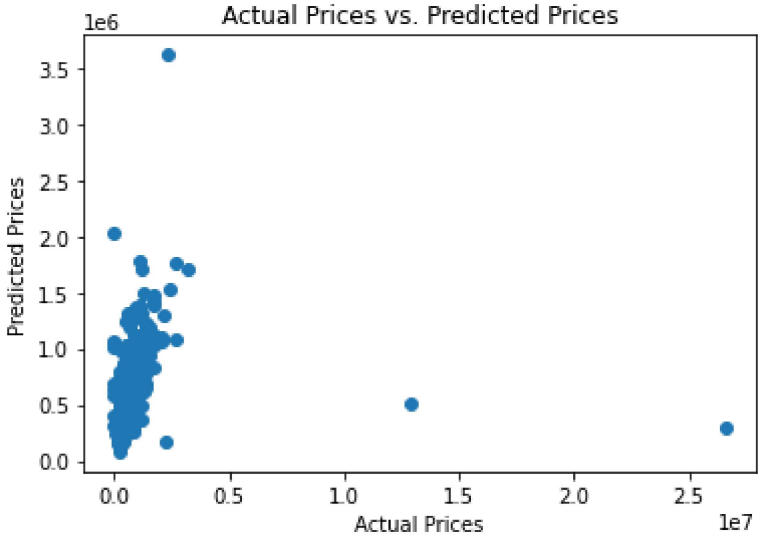
In [66]:
```
y_pred = model.predict(X_test)
```

In [67]:
```
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
```

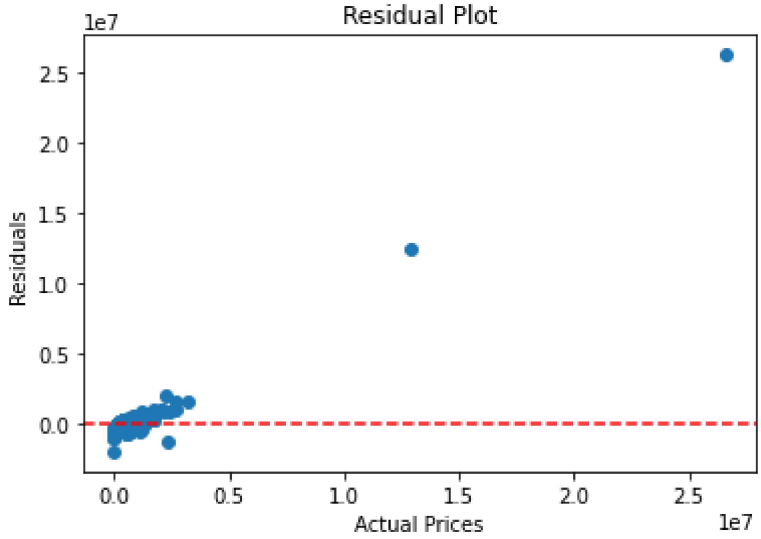
In [68]:
```
print("Mean Squared Error:", mse)
print("R-squared:", r2)
```

```
Mean Squared Error: 986869414953.9674
R-squared: 0.03233518995626383
```

In [69]:
```
plt.scatter(y_test, y_pred)
plt.xlabel("Actual Prices")
plt.ylabel("Predicted Prices")
plt.title("Actual Prices vs. Predicted Prices")
plt.show()
```



In [70]:
```
residuals = y_test - y_pred
plt.scatter(y_test, residuals)
plt.axhline(y=0, color='red', linestyle='--')
plt.xlabel("Actual Prices")
plt.ylabel("Residuals")
plt.title("Residual Plot")
plt.show()
```



In [73]:
```
sns.displot((y_test-predicted_price),bins=50)
```

Out[73]:
```
<seaborn.axisgrid.FacetGrid at 0x162f223d3a0>
```