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]:

١.	5	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())

- 4-	,	date	price	bedroc	oms	bathroom	s sqft_living	sqft_l
ot 0	2014-05-02 0:00:00		313000.0		3	1.5	0 1340	79
12 1	2014-05-02 0	:00:00	2384000.0	5		2.5	9 3650	90
50 2	2014-05-02 0	:00:00	342000.0		3	2.0	9 1930	119
47 3	2014-05-02 0:00:00		420000.0	3 2.25		2.2	5 2000	80
30 4 00	2014-05-02 0	:00:00	550000.0		4	2.5	9 1940	105
t	floors wate	rfront	view cond	dition	sqf	t_above	sqft_basement	yr_buil
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 6	1.0	0	0	4		1140	800	197
	yr renovated			stree	<b>5</b> +	city	statezip cou	ntrv
0	2005		810 Densmo			Shoreline	WA 98133	USA
1	0			3laine S		Seattle	WA 98119	USA
2	0		262 <b>1</b> 4 143r			Kent	WA 98042	USA
3	0			Oth Pl N		Bellevue	WA 98008	USA
4	1992		9105 1701			Redmond	WA 98052	USA

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

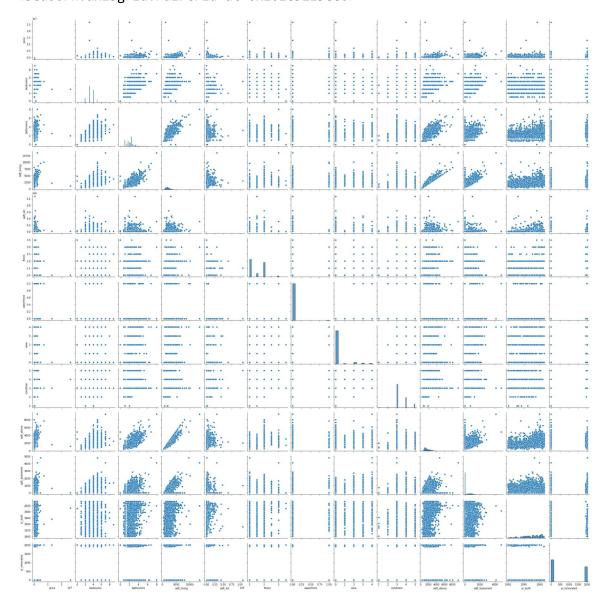
```
sqft_living
               price
                         bedrooms
                                                                     sqft_lot
                                      bathrooms
count
       4.600000e+03
                      4600.000000
                                    4600.000000
                                                   4600.000000
                                                                 4.600000e+03
mean
       5.519630e+05
                         3.400870
                                       2.160815
                                                   2139.346957
                                                                 1.485252e+04
       5.638347e+05
                                       0.783781
std
                         0.908848
                                                    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%
                                       2.250000
                                                                 7.683000e+03
       4.609435e+05
                         3.000000
                                                   1980.000000
75%
       6.549625e+05
                         4.000000
                                       2.500000
                                                   2620.000000
                                                                 1.100125e+04
       2.659000e+07
                         9.000000
                                       8.000000
                                                  13540.000000
                                                                 1.074218e+06
max
                                                                sqft above
            floors
                      waterfront
                                                   condition
                                           view
                                                 4600.000000
       4600.000000
                     4600.000000
                                   4600.000000
                                                               4600.000000
count
          1.512065
                        0.007174
                                      0.240652
                                                               1827.265435
mean
                                                    3.451739
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
          3.500000
                        1.000000
                                      4.000000
                                                    5.000000
                                                               9410.000000
max
       sqft_basement
                          yr_built
                                     yr_renovated
         4600.000000
                       4600.000000
                                      4600.000000
count
          312.081522
                       1970.786304
                                       808.608261
mean
                                       979.414536
std
          464.137228
                         29.731848
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
         4820.000000
                       2014.000000
                                      2014.000000
max
```

## 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	

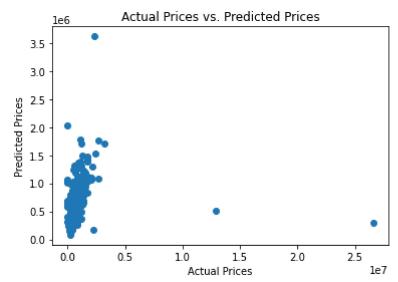
In [60]: sns.pairplot(df)

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

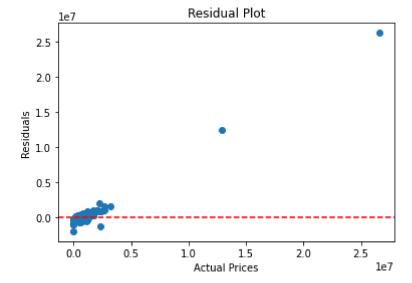


```
correlation matrix = df.corr()
In [61]:
            sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
            plt.title("Correlation Matrix")
            nlt show()
                                         Correlation Matrix
                                                                               - 1.0
                      price - 1 0.2 0.330.430.050.150.140.230.0350.370.210.0240.029
                  bedrooms - 0.2 1 0.550.590.0690.18.0030510.0250.480.330.140.061
                                                                               - 0.8
                 bathrooms -0.330.55 1 0.760.110.490.0760.210.120.69 0.3 0.460.22
                 sqft living -0.430.590.76 1 0.210.340.120.340.060.880.450.290.12
                                                                               - 0.6
                    sqft lot -0.090.0690.110.21 10.00307010.007.4000505220.036.0540.023
                     floors -0.150.18<mark>0.49</mark>0.34.003<mark>71 0.020.031</mark>0.28<mark>0.52</mark>0.260.47<mark>0.2</mark>3
                                                                              - 0.4
                 waterfront -0.14.00B5076.120.010.022 1 0.36000B5079.098.0240086
                      view -0.230.110.210.310.074.0310.36 1 0.0630.170.320.064023
                                                                              - 0.2
                  condition 0.035.0250.1-0.058005560.0008563 1 -0.18 0.2 -0.40.19
                 sqft above -0.370.480.690.880.220.520.0790.170.18 1 0.039.410.1
                                                                               - 0.0
             sqft basement -0.210.33 0.3 0.450.03 0.20.0980.32 0.20.03 1 -0.10.043
                    yr built 9.0220.140.460.290.0510.470.0291.06-0.40.410.16 1 -0.3
               yr renovated -0.020.060.220.1-0.025.20.008060250.190.1-0.0450.32
                                             floors
                                                                aff_basement
                                                 waterfront
                                                         condition
                                                                       yr renovated
In [62]:
            X = df[['bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot', 'floors', 'wate
            v = df['nrice']
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, v train)
Out[65]: LinearRegression()
In [66]: v pred = model.predict(X test)
In [67]: | mse = mean_squared_error(y_test, y_pred)
            r2 = r2 score(v test. v nred)
In [68]: print("Mean Squared Error:", mse)
            nrint("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((v test-predicted price).bins=50)

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

