opnw0xphy

September 26, 2023

```python
[1]:  # Load the dataset from CSV
      import numpy as np
      import matplotlib.pyplot as plt
      import seaborn as sns
      import pandas as pd
      %matplotlib inline
```

```python
[13]: # Load the data
      df = pd.read_csv('house.CSV')
      df.head()
```

```
[13]:                   date       price  bedrooms  bathrooms  sqft_living  sqft_lot  \
      0  2014-05-02 00:00:00    313000.0       3.0       1.50         1340      7912
      1  2014-05-02 00:00:00   2384000.0       5.0       2.50         3650      9050
      2  2014-05-02 00:00:00    342000.0       3.0       2.00         1930     11947
      3  2014-05-02 00:00:00    420000.0       3.0       2.25         2000      8030
      4  2014-05-02 00:00:00    550000.0       4.0       2.50         1940     10500

         floors  waterfront  view  condition  sqft_above  sqft_basement  yr_built  \
      0     1.5           0     0          3        1340              0      1955
      1     2.0           0     4          5        3370            280      1921
      2     1.0           0     0          4        1930              0      1966
      3     1.0           0     0          4        1000           1000      1963
      4     1.0           0     0          4        1140            800      1976

         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
```

```python
[14]: # Check for missing values
      df.isnull().sum()
```

```
[14]: 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
```

[15]: ```python
# Summary statistics of the dataset
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
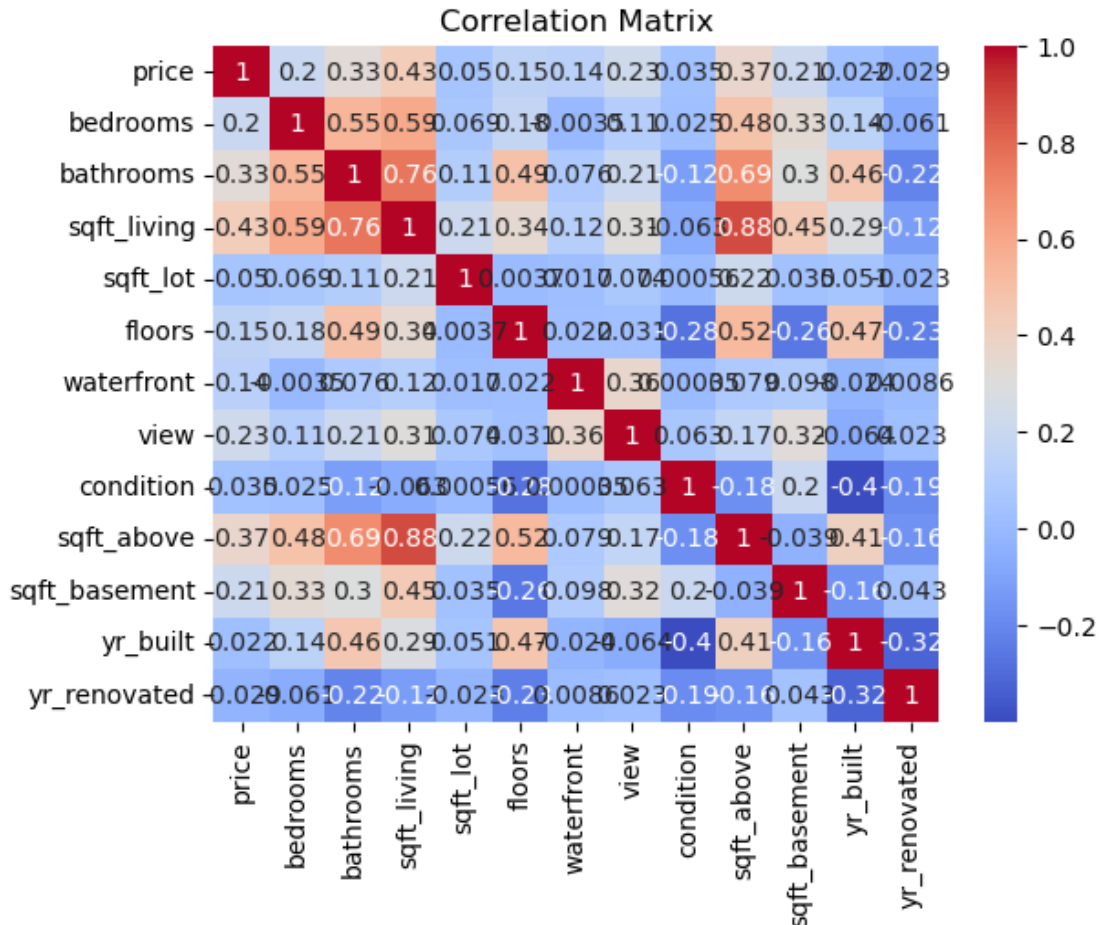25%         0.000000   1951.000000      0.000000
```

```
50%          0.000000   1976.000000      0.000000
75%        610.000000   1997.000000   1999.000000
max       4820.000000   2014.000000   2014.000000
```

[16]:
```python
# Correlation matrix to understand feature relationships
correlation_matrix = df.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title("Correlation Matrix")
plt.show()
```


Correlation Matrix

[17]:
```python
# Preprocessing: Selecting features and target variable
X = df[['bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot', 'floors',
        'waterfront', 'view', 'condition']]
y = df['price']
```

[19]:
```python
# Splitting the dataset into training and testing sets
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=42)
```

[21]:
```
# Building the Linear Regression Model
from sklearn.linear_model import LinearRegression
model = LinearRegression()
```

[22]:
```
model.fit(X_train, y_train)
```

[22]:
```
LinearRegression()
```

[25]:
```
y_pred = model.predict(X_test)
y_pred
```

[25]:
```
array([ 355544.20023003,  412466.5623674 , 1032309.01472662,
        638015.89691969,  386749.03170847,  682878.66774721,
        508131.51104774,  476444.88868517,  528390.38064756,
        536890.28432162,  741611.25431868,  351089.54000972,
        834594.23596635,  507916.36533283,  395783.43614529,
        753604.0107183 ,  591260.93521173,  519685.21315053,
       1097684.01387167,  899350.51526607, 1416428.63269075,
        655027.54695346,  633467.15674151,  508825.54646831,
        241505.58246694,  246422.08147013,  682315.40782021,
        890710.02853515,  343999.53755816,  869713.47838379,
       1706568.09689594,  423119.45957691, 1166554.09905344,
        473485.51556333,  222390.24850546,  389364.98040257,
        837334.18734894,  791445.99802359,  309102.96122697,
        591268.41401444,  412892.26769673,  330863.44338467,
        425901.23000393,  395309.75466822,  321469.19939783,
        278850.88655001,  505322.56349714,  530072.20069295,
        827342.21520756,  301425.32433213, 1238909.153923  ,
        365386.74030001,  441659.23383915,  652940.88315588,
        281396.58930795,  330386.21781283, 1000352.63274041,
        849100.92291191, 1012801.81803242,  332179.52071555,
        571205.44501104,  871454.44218347,  212267.0323663 ,
        891265.86021133,  824687.24892713,  616069.61732021,
        569811.37172089,  559854.10779075,  421212.38840408,
        370559.30276257, 1051942.96914305,  492330.72608398,
        396115.11066693,  945514.71796525,  362793.06521097,
        681450.93054873,  461222.00655153,  656146.05256495,
        495198.94624815,  278322.65936867,  582557.7588189 ,
        346151.92961453,  537667.84476815,  854916.73806702,
        402715.61875398,  343493.82286426,  451658.37309246,
        601056.52758505,  944232.52014678,  425218.75877519,
        850314.40450252,  296788.1261004 ,  312655.01752304,
        507178.30284406,  727933.69304023, 1222226.82418338,
        767332.00725369,  624887.88498518,  521953.60785656,
```

```
  484054.1223268 ,   543448.6715062 ,   347688.46537474,
  145912.29345805,   472337.20453144,   557062.71691043,
  921450.0917354 ,   609542.56854941,   279011.80999432,
  667493.80088901,   639886.13631631,   596544.16047999,
  526461.9239019 ,   330152.09948674,   384656.64352414,
 1024906.31070596,   395969.61837129,   520893.82904682,
  717493.02802647,   449730.56977409,   332435.07130506,
  703422.39258028,   698487.11338821,   791775.34246741,
  463598.24969414,   258051.24042164,   246544.55034443,
  841679.32787877,   536828.58446582,   323342.29079991,
  583924.27449495,   379822.37883361,   483186.73813136,
  696822.41308068,   417956.1392333 ,   318689.59080843,
  392305.11044797,   276749.3765998 ,   258078.71362994,
  647832.54747238,   777728.92243834,   231230.07556114,
  473774.26809724,   447329.64510657,   376761.24642435,
  985274.97521175,   761764.27009758,   241962.55998918,
  344185.75641266,   297095.67972873,   347082.58600061,
  371295.22853319,   765131.9216654 ,   700061.06296691,
  331178.50313188,   726541.89839637,   657008.15970065,
 1261926.59608013,   750282.29743902,   427937.63742424,
  270607.44227879,   338179.40342105,   826575.52153377,
  428441.5420932 ,   626189.08119414,   563843.87938537,
  605927.02591799,   607765.86297932,   290309.9984862 ,
  612246.08721502,   757287.20920595,   274136.45299872,
  484411.58654591,   782683.79721682,   371033.20223174,
  728692.28795929,   817922.96185043,   319719.75135819,
  707828.19488688,   724159.59764569,   419872.27074796,
  430703.89791407,   268953.75433671,   682980.49969034,
  417207.86692684,   419229.36902182,  1763931.66734267,
  617947.89383672,   397313.7315187 ,   589700.82091719,
  422791.56952965,   870577.48480884,   303816.26371375,
  255046.34981618,   679384.68778678,   411349.72904837,
 1533635.57787426,   807248.23482022,  1020644.82190137,
  283838.97245615,   212284.58986112,   385787.77322363,
  597949.42549026,   182799.90241277,   611870.70807075,
  477464.99525726,   576965.29351616,   514825.18796905,
  588965.34957496,   323976.03196657,   880704.64430508,
  594914.26235949,   372398.18052102,   942633.45369853,
  517476.08483946,   573476.19022391,   346789.43489477,
  191218.50301905,   389432.6038724 ,   312899.49918802,
  407181.7730655 ,   769934.15697478,   790553.89228171,
  268327.02007125,   419111.02786738,   730487.01351392,
  585630.23345837,   370252.52547945,   336647.49802711,
  573008.08466814,   731061.11929896,   299022.51014928,
  416000.44297077,   268350.09569218,   308755.78958717,
  693613.0691148 ,   741784.43481469,  1317816.14425297,
  314797.60336941,   300600.06786734,   262259.79250595,
```

```
505526.26168439,   543651.02965507,   601477.61421639,
361221.40233894,   291614.09138192,   567908.50266785,
496026.70531212,   588805.44013718,   392766.95895139,
352177.32446498,   275593.36237382,   261036.79726708,
398128.22131417,   404926.24852354,   388280.02911265,
1315368.95879307,  282316.2733546 ,   815240.48617393,
481483.20681007,   298659.67435872,   201270.94671351,
418864.1186834 ,   411414.07317987,   348144.33298261,
415284.67060244,   471110.1064464 ,   737295.32275552,
359794.91214906,   428724.03484437,   243342.03011795,
1083982.30847548,  753327.40383312,   311040.98638274,
909151.65409543,   415529.07277999,   289247.08357104,
359941.95413945,   610107.1605625 ,   810349.3379766 ,
377412.899774  ,   396994.97607116,   370841.52714802,
534773.7254074 ,   749573.22813575,   520999.55077173,
269786.52567481,   682671.62091789,   569672.03224139,
1108686.41085533,  293862.44828361,   353098.54302923,
490431.82424046,   394855.8874503 ,   756255.98816228,
384182.79153689,   619185.54756485,   797341.4114393 ,
718699.78941244,   291718.40069211,   441738.11965386,
388590.04891943,   761908.85590266,   518808.45581207,
638262.86212386,   605077.59229167,   418063.38796941,
220251.29852052,   346261.72674036,   486711.9941501 ,
532087.87812632,  3615850.10194565,  1715796.04772546,
1234880.41469814, 1048198.70027788,   399330.71171391,
1055117.87549818,  244522.04612783,   490602.82678888,
964321.10911003,   367616.88697963,   577081.80755115,
324405.45526129,   458012.61846184,   796764.23489838,
471132.05484278,   373370.87381181,   311577.11206275,
305913.76050656,   731546.57542177,   445994.64692039,
336809.90434733,   510164.87629557,   316467.51505148,
492794.12333073,   942451.64608705,   288747.1418794 ,
524345.12091019,   648797.04886371,   516998.1984765 ,
1286362.54591364,  386567.13671704,   350590.31730327,
312859.0559236 ,   608848.95094081,   860686.69911459,
364640.89523182,   508091.86801324,   854198.24972641,
373740.20033406,   341331.98590632,   450464.91389575,
223484.7788372 ,  1176445.04666214,   409476.56801421,
296780.99211865,   118889.94385016,   379700.75213958,
679239.70270944,   516752.4751368 ,   631255.97137443,
567806.33470192,   499153.0449227 ,   936900.89800524,
390082.89698434,   591904.98793262,   760306.67792209,
785119.37409601,   596952.42949073,   472069.51432991,
366112.24537331,   477241.2982881 ,   622065.85384176,
417887.41664733,   306538.10089968,   456904.65834702,
424213.01484943,   254390.52748766,   784297.20897053,
421548.89240178,   531215.1316448 ,   462000.93789248,
```

281268.71008598,   303349.97773134,   453936.87177633,
239847.52356175,   503262.12754696,   429287.09578059,
881949.9668716 ,   395586.07765309,   363300.68785709,
299513.53575163,   340945.42693832,   252661.89664184,
539677.74815683,   187245.4396402 ,   842441.70351089,
421959.59314872,   624603.25345657,   509811.55603887,
550200.36814066,   427784.31884093,   300642.72254725,
948129.76205023,   656695.87191022,   772836.27940962,
449532.64841189,   883831.0976331 ,   626616.47279741,
357422.68422825,   588538.79070921,   440850.40149016,
840718.97740887,   391985.2683022 ,   699253.66201054,
591363.11764703,   582557.7588189 ,   347098.1686108 ,
893356.6094896 ,   435264.75734233,   387555.69497214,
545398.78555813,   207109.02964473,   483123.56346793,
511482.37783308,   695714.14676267,   458931.52929117,
758414.57963185,   332865.80975857,   254220.08509034,
307410.89820232,   257453.91804725,   687656.61183882,
583711.67252858,   373873.73294928,   515487.02954269,
464058.0672629 ,   574640.66200548,   328719.60218863,
240569.5576377 ,   617035.59776028,   680997.47833962,
370973.88373306,   313684.79766709,   304912.52443706,
318951.43655938,   268883.18295357,   176854.20429542,
678592.78662907,   1069623.88889817,   1110012.6172728 ,
647120.19285978,   794058.52605254,   630908.58625942,
218084.65431652,   545868.83499834,   649442.50745941,
444911.6968737 ,   738306.99535734,   710996.21462728,
668901.89334988,   307852.64315507,   537947.14241469,
286682.46786765,   674674.61407839,   320170.87186672,
1373166.03842917,   362716.56607221,   1197569.59430267,
483840.26747507,   1239247.94248386,   195031.49539228,
625870.99065701,   394932.50650794,   219694.89949017,
325366.69817468,   684137.29177482,   701358.66592535,
1468844.76437964,   437529.28045449,   487991.32037496,
420782.02608084,   1081975.28256055,   739183.94105351,
583377.88746877,   456394.99665291,   563468.17792639,
684820.46382373,   173987.15588859,   278237.10593135,
428924.70852703,   609742.07687782,   413995.22841242,
653139.68411116,   293469.59128896,   327319.33504794,
281853.99404704,   704384.0463391 ,   634643.91647906,
509613.28066548,   461995.97825828,   687383.77650187,
372226.72860134,   326442.96841943,   311842.21302853,
429768.39564055,   763840.04595892,   366873.70365399,
450494.13199773,   476646.64749911,   269704.56506136,
292894.07102197,   621439.0505139 ,   1117213.24985635,
762410.41310866,   563661.37784445,   502956.19817195,
431173.00832187,   340229.43916991,   578548.72198906,
457016.89105714,   258256.91348024,   575467.71169516,

```
254396.58185948,    362803.92584827,    352546.17816243,
364853.37453385,    653428.16118099,    195011.88520225,
226330.73330689,    319887.69795931,    278669.22250493,
523211.98178045,    387063.98327755,    803806.94537412,
561980.19427619,    513002.74744743,    435107.58188684,
841423.73858939,    549167.21843407,    284208.98984309,
579678.48409748,    401921.72298962,    188762.82984757,
489550.43827829,    213479.25396505,    651505.23624849,
382608.80502299,    671451.01193511,    438127.98388474,
642855.22057895,    336500.01444862,    473125.57983583,
713315.32968182,    486304.52484354,    333577.64208113,
554574.52891553,    331703.8680436 ,    606994.38652235,
339281.71315615,    438180.54975611,    444951.36256418,
368973.92356199,    365468.16247322,    620637.93642256,
278563.18475323,    353527.1845196 ,   1126166.8092668 ,
602078.73327261,   2035254.87380864,    398213.60046012,
680922.13313651,    484676.41118379,    799218.89995472,
491077.04865443,    321429.05829081,    512864.05182913,
715746.51776861,    926155.59161703,    767430.95417627,
169395.63839531,    271668.11736412,    202392.6609656 ,
555291.80494758,   1287983.51674125,    541150.31455517,
392074.82739676,    582526.3597578 ,    505356.54497691,
1488506.56264736,    391239.39104192,    309448.00243435,
1294626.46879701,    329653.2288686 ,    298462.43550468,
600305.70747915,    618907.27881349,    522810.99287346,
205952.09553001,    692798.15317017,    545762.34196593,
409746.23211939,    306548.66390401,    313477.06354308,
412973.28353228,    293758.01958328,    230105.20071592,
650980.76294737,    802865.44269323,    782564.04733742,
615626.72627998,    235691.33155267,    863552.34302864,
307880.77771201,    625867.68886475,    760615.12264199,
607315.29711292,    420334.89368717,    795223.15211585,
637481.6173025 ,   1022458.95757713,    375700.8955897 ,
604691.56054971,    968258.56741482,    536063.97970416,
519336.25003959,    409971.83634847,    366705.03615489,
1005419.90773272,   1065807.87583918,    646609.77705414,
434531.97844663,    263002.44218541,    651138.93012367,
600340.03760906,    649715.35997527,    454638.60316493,
366510.75408792,    209770.24831827,    509458.61440438,
692940.66151759,    808198.37626568,    267102.61900138,
383280.59635017,    846863.15587974,    227239.09656482,
326063.09920265,    560192.32033285,    480208.12097769,
1126806.14687029,    628502.2370342 ,    275830.13960599,
540863.64954081,    374467.61503029,    580199.84709031,
827506.85741688,    985240.87942772,    348436.15089508,
304961.0223209 ,    781309.56099187,    198268.96696977,
579260.55420596,    506248.81765426,    257993.97682377,
```

```
  475000.61319747,   426393.66205629,   740143.10267877,
  494205.18892293,   369619.8239934 ,   581717.53136424,
  461554.416029  ,   580238.91070767,   710849.04999574,
  391511.56213586,   910931.0421943 ,   680837.59711117,
  388174.26885201,   891924.51813522,   980270.86484424,
  230227.31406615,   537133.11879083,   575780.70464338,
  494060.88954058,   354112.87264364,   566444.09193698,
  589063.11118143,   309650.12758115,   386163.4462722 ,
  372774.77249902,   374264.16200898,   796415.08137956,
  563032.73849595,  1114938.29077021,   614603.17681498,
  585771.89660745,   594708.28089805,   347892.75564033,
  537994.687362  ,   316727.04322684,    83935.06384781,
  492055.17725678,   799937.10422055,  1071944.26566475,
  389463.79860882,   775042.96645553,   834379.9720426 ,
  735572.50850346,   527926.20015886,   312109.25798442,
  500053.39014208,   623238.17367653,  1368445.06922171,
  427127.1584122 ,   438398.15888806,   281492.85039307,
  286033.42080893,   466402.77513887,   238154.68323609,
  436689.69398769,   594361.60599159,   431599.56006366,
  452896.09810891,   782501.14120286,   700737.41668815,
 1238354.29083392,   269782.43550612,   437110.85410242,
  225967.23274841,   514368.09571551,   245856.82019779,
  433689.78639231,   771093.09011959,   658791.01336346,
  705899.57642546,   587114.69446515,   632244.45367087,
  490403.37897297,   507934.03945468,  1001120.72912544,
  328910.46337158,   344934.83139814,   568788.27913791,
  466436.08855886,   580457.85085401,   428024.42692561,
  485059.78760375,   421907.01276656,   421439.33934955,
  428081.2195833 ,   463556.87549085,   324640.05657601,
  273930.42236388,   562075.38031117,   534985.64362949,
  493025.08766903,   414641.10425046,   441044.00420601,
  637282.00611102,   409762.5667475 ,   226271.96715043,
  575182.99602643,   553936.17995321,  1120258.33746619,
  618574.96528549,   836222.05390988,   589076.46993373,
  433753.8591103 ,   426474.20204482,   477035.66210602,
  270831.84006156,   758699.85720149,   365462.002223  ,
  230211.97896014,   335878.00612297,   396119.30475417,
  635068.75011243,   522228.00458736,   221541.26818857,
  387261.89808876,   829964.98256211,   521265.30173355,
  305004.04140879,   746497.82378234,   818054.31305073,
  395299.86761749,   445932.81429885,   476389.48468313,
  297940.14127804,   368603.55021381,   690584.42146651,
  284419.2202956 ,   639725.47974504,   461943.44143845,
  533407.95094208,   619847.61040715,   772388.09225938,
  468457.23565825,   635968.63521548,   977948.38235969,
  307318.13106454,   483646.74577104,   517165.73573993,
  347665.21246281,   568973.7179283 ,   532627.70651536,
```

```
        151144.35885191,   839490.89956496,   411430.51681228,
       1091039.33299443,   541578.27490614,   651215.03480576,
        440648.35283077,   294803.71193976,   354562.54262837,
        490808.28553364,   613344.56358602,   623715.45594025,
        502026.95940007,   310260.42337721,   600115.92625715,
        300643.28745929,   633262.44618554,   507710.32908732,
        395826.32630966,   617701.79420375,   705993.6053184 ,
        307510.47870749,   540053.35454227,   443469.72897502,
        457071.15170208,   450255.17420434,   603658.97522641,
        450781.06265057,   418883.79800392,   500422.84795087,
        416474.33901326,   857847.16381781,   350506.55612531,
        244101.2292792 ,   529201.80182031,   965225.39909469,
        747573.08632355,   232310.97847327,   276961.53878619,
        387051.54851533,   200647.16471071,   537981.27139121,
       1387782.35751011,   396911.92552127,   329450.27247103,
        384880.79765606,   767756.28412608,   293263.50503537,
        629502.65118339,   567271.56419537,  1019559.79018072,
        674452.07806434,   262107.38536425,   464988.23538791,
        766517.85249257,   809657.589953   ,   340330.16698363,
        457412.43506232,   644273.60552985,   315733.45117291,
        795405.18040929,   746433.9887217 ,   492101.97603748,
        817286.79058413,   566318.84319527,   534556.67023094,
        513228.14500387,   652421.16407805,   499559.98031741,
        994612.36589381,  1182485.98402713,   898409.19027288,
        479213.10914815,   577700.11950749,  1778427.82571384,
        560179.94457321,  1386083.58669447,  1030446.6107682 ,
        800794.62662907,   697079.40872924,   250990.803651   ,
        418392.21199889,   450938.63048803,   280392.16558895,
        337581.77838535,   294842.64824859,   854229.03728654,
        336776.62343315,   606806.06455202,   630282.70686862,
        448812.97354182,   508286.86719294,   484826.9247334 ,
        691097.52386   ,   492357.36163643,   232503.39003618,
        677448.85449031,   325449.85153786,   371542.01752151,
        292423.98157695,   357692.52139224,   620720.76076161,
        286529.97849307,   365802.26424488,   525115.86895286,
        426979.36307833,   439420.02657167,   672953.85431515,
        242312.95268399,   532285.76167167,   452211.13310435,
        387789.68546096,   358379.37549321,   294380.3033871 ,
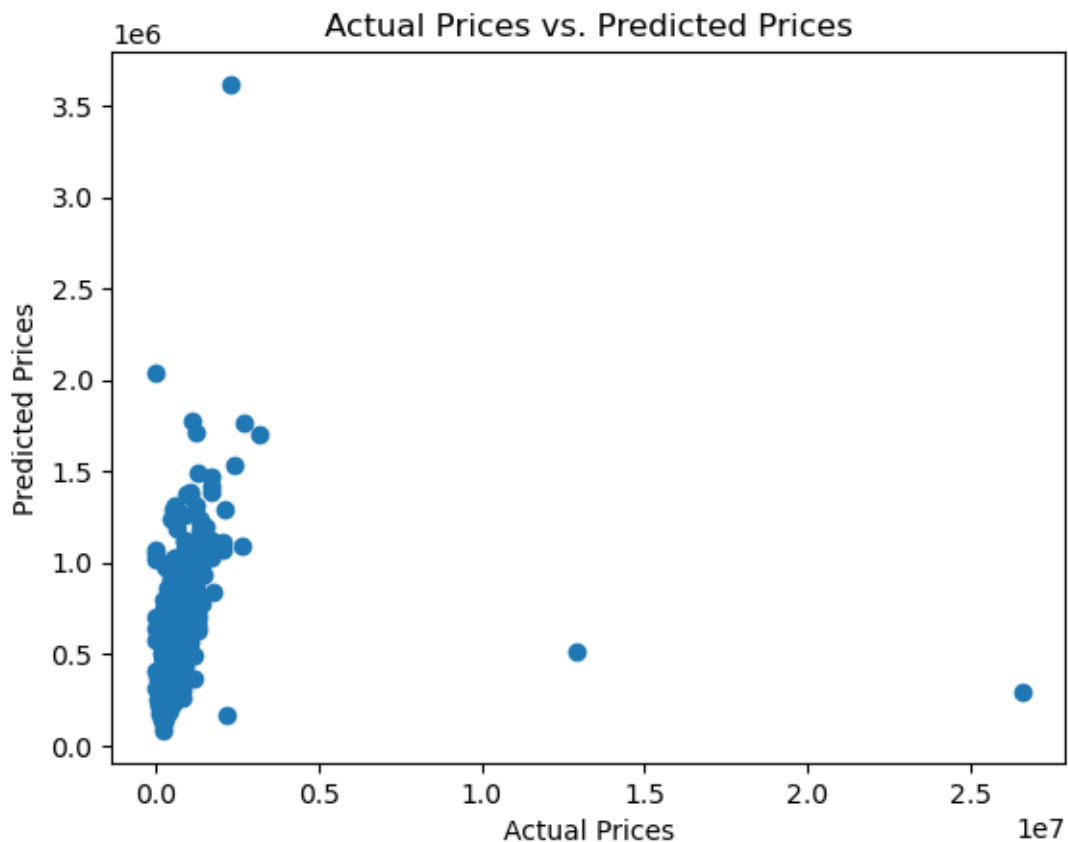        667102.28531202,   582761.60751404])
```

[30]:
```python
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
r2 = r2_score(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
print("Mean Squared Error:", mse)
print("R-squared:", r2)
```

```
Mean Squared Error: 986869414953.9803
R-squared: 0.032335189956324784
```

[31]:
```python
plt.scatter(y_test, y_pred)
plt.xlabel("Actual Prices")
plt.ylabel("Predicted Prices")
plt.title("Actual Prices vs. Predicted Prices")
plt.show()
```


[35]:
```python
new_data = [[3, 2, 1500, 4000, 1, 0, 0, 3]]
predicted_price = model.predict(new_data)
```

```
C:\Users\Logeshwaran\anaconda3\lib\site-packages\sklearn\base.py:450:
UserWarning: X does not have valid feature names, but LinearRegression was
fitted with feature names
  warnings.warn(
```

[36]:
```python
print("Predicted Price:", predicted_price[0])
```

```
Predicted Price: 331038.96876928967
```

[ ]: