house-price-prediction

April 9, 2024

ABOUT THE DATASET

The aim of this project is to develop a predictive model for housing prices in Washington State using a data-driven approach. By harnessing the power of machine learning and data analysis, we will explore historical housing data, including factors such as location, square footage, number of bedrooms, and more, to create an accurate prediction model.

1. Date: This column contains the date when a particular property transaction occurred.
2. Price: This column contains the selling price of the house.
3. Bedrooms: This column indicates the number of bedrooms in the house.
4. Bathrooms: This column shows the number of bathrooms in the house.
5. Sqft_living: This column represents the total square footage of the living space (interior) of the house.
6. Sqft_lot: This column is likely the total square footage of the land or plot on which the house is built.
7. Floors: It indicates the number of floors in the house.
8. Waterfront: It's a binary column that could indicate whether the property has a waterfront view or not.
9. View: This column might describe the level of view the property has, typically on a scale from 0 to 4, with 0 being no view and 4 being an excellent view.
10. Condition: This column could represent the overall condition of the property, often rated on a scale from 1 to 5, with 1 being poor and 5 being excellent.
11. Sqft_above: This column likely shows the square footage of the interior living space above ground level.
12. Sqft_basement: This column should contain the square footage of any basement space in the house.
13. Yr_built: This is the year the house was originally built.
14. Yr_renovated: If the house has been renovated, this column may contain the year when the renovation took place.
15. Street: This column might provide information about the street or address of the property.
16. City: It represents the city where the property is located.

17. Statezip: This column could contain information about the state and ZIP code of the property.
18. Country: In this context, it's likely that all entries are from the same country, so this column may not provide much variation.

IMPORTING REQUIRED LIBRARIES

[258]:

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

IMPORTING DATASET

[259]:

```
df=pd.read_csv('/content/data1.csv')
df
```

[259]:

```
                     date         price  bedrooms  bathrooms  sqft_living  \
0     2014-05-02 00:00:00  3.130000e+05       3.0       1.50         1340
1     2014-05-02 00:00:00  2.384000e+06       5.0       2.50         3650
2     2014-05-02 00:00:00  3.420000e+05       3.0       2.00         1930
3     2014-05-02 00:00:00  4.200000e+05       3.0       2.25         2000
4     2014-05-02 00:00:00  5.500000e+05       4.0       2.50         1940
…                   …           …       …        …          …
4595  2014-07-09 00:00:00  3.081667e+05       3.0       1.75         1510
4596  2014-07-09 00:00:00  5.343333e+05       3.0       2.50         1460
4597  2014-07-09 00:00:00  4.169042e+05       3.0       2.50         3010
4598  2014-07-10 00:00:00  2.034000e+05       4.0       2.00         2090
4599  2014-07-10 00:00:00  2.206000e+05       3.0       2.50         1490

      sqft_lot  floors  waterfront  view  condition  sqft_above  \
0         7912     1.5           0     0          3        1340
1         9050     2.0           0     4          5        3370
2        11947     1.0           0     0          4        1930
3         8030     1.0           0     0          4        1000
4        10500     1.0           0     0          4        1140
…        …     …         …   …        …         …
4595      6360     1.0           0     0          4        1510
4596      7573     2.0           0     0          3        1460
4597      7014     2.0           0     0          3        3010
4598      6630     1.0           0     0          3        1070
4599      8102     2.0           0     0          4        1490

      sqft_basement  yr_built  yr_renovated                    street  \
0                 0      1955          2005      18810 Densmore Ave N
1               280      1921             0           709 W Blaine St
2                 0      1966             0  26206-26214 143rd Ave SE
3              1000      1963             0           857 170th Pl NE
4               800      1976          1992         9105 170th Ave NE
```

```
…             …       …           …                       …
4595              0      1954          1979            501 N 143rd St
4596              0      1983          2009          14855 SE 10th Pl
4597              0      2009             0          759 Ilwaco Pl NE
4598           1020      1974             0         5148 S Creston St
4599              0      1990             0         18717 SE 258th St

           city  statezip country
0     Shoreline  WA 98133     USA
1       Seattle  WA 98119     USA
2          Kent  WA 98042     USA
3      Bellevue  WA 98008     USA
4       Redmond  WA 98052     USA
…         …       …     …
4595    Seattle  WA 98133     USA
4596   Bellevue  WA 98007     USA
4597     Renton  WA 98059     USA
4598    Seattle  WA 98178     USA
4599  Covington  WA 98042     USA

[4600 rows x 18 columns]
```

[260]:

```
#printing first 5 rows
df.head()
```

[260]:

```
                  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
```

[261]:
```
#printing last 5 rows
df.tail()
```

[261]:
```
                     date          price  bedrooms  bathrooms  sqft_living  \
4595  2014-07-09 00:00:00  308166.666667       3.0       1.75         1510
4596  2014-07-09 00:00:00  534333.333333       3.0       2.50         1460
4597  2014-07-09 00:00:00  416904.166667       3.0       2.50         3010
4598  2014-07-10 00:00:00  203400.000000       4.0       2.00         2090
4599  2014-07-10 00:00:00  220600.000000       3.0       2.50         1490

      sqft_lot  floors  waterfront  view  condition  sqft_above  \
4595      6360     1.0           0     0          4        1510
4596      7573     2.0           0     0          3        1460
4597      7014     2.0           0     0          3        3010
4598      6630     1.0           0     0          3        1070
4599      8102     2.0           0     0          4        1490

      sqft_basement  yr_built  yr_renovated             street       city  \
4595              0      1954          1979     501 N 143rd St    Seattle
4596              0      1983          2009   14855 SE 10th Pl   Bellevue
4597              0      2009             0   759 Ilwaco Pl NE     Renton
4598           1020      1974             0  5148 S Creston St    Seattle
4599              0      1990             0  18717 SE 258th St  Covington

      statezip country
4595  WA 98133     USA
4596  WA 98007     USA
4597  WA 98059     USA
4598  WA 98178     USA
4599  WA 98042     USA
```

[262]:
```
#printing columns
df.columns
```

[262]:
```
Index(['date', 'price', 'bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot',
       'floors', 'waterfront', 'view', 'condition', 'sqft_above',
       'sqft_basement', 'yr_built', 'yr_renovated', 'street', 'city',
       'statezip', 'country'],
      dtype='object')
```

[263]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4600 entries, 0 to 4599
Data columns (total 18 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
```

```
 0   date           4600 non-null   object
 1   price          4600 non-null   float64
 2   bedrooms       4600 non-null   float64
 3   bathrooms      4600 non-null   float64
 4   sqft_living    4600 non-null   int64
 5   sqft_lot       4600 non-null   int64
 6   floors         4600 non-null   float64
 7   waterfront     4600 non-null   int64
 8   view           4600 non-null   int64
 9   condition      4600 non-null   int64
 10  sqft_above     4600 non-null   int64
 11  sqft_basement  4600 non-null   int64
 12  yr_built       4600 non-null   int64
 13  yr_renovated   4600 non-null   int64
 14  street         4600 non-null   object
 15  city           4600 non-null   object
 16  statezip       4600 non-null   object
 17  country        4600 non-null   object
dtypes: float64(4), int64(9), object(5)
memory usage: 647.0+ KB
```

[264]:

```
df.describe()
```

[264]:

```
              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
```

[265]:

```
#printing datatypes
df.dtypes
```

[265]:

```
date              object
price            float64
bedrooms         float64
bathrooms        float64
sqft_living        int64
sqft_lot           int64
floors           float64
waterfront         int64
view               int64
condition          int64
sqft_above         int64
sqft_basement      int64
yr_built           int64
yr_renovated       int64
street            object
city              object
statezip          object
country           object
dtype: object
```

[266]:

```
#findingout missing values
df.isna().sum()
```

[266]:

```
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
```

```
[267]: df.nunique()
```

```
[267]: date                70
       price             1741
       bedrooms            10
       bathrooms           26
       sqft_living        566
       sqft_lot          3113
       floors               6
       waterfront           2
       view                 5
       condition            5
       sqft_above         511
       sqft_basement      207
       yr_built           115
       yr_renovated        60
       street            4525
       city                44
       statezip            77
       country              1
       dtype: int64
```

DATA VISUALIZATION

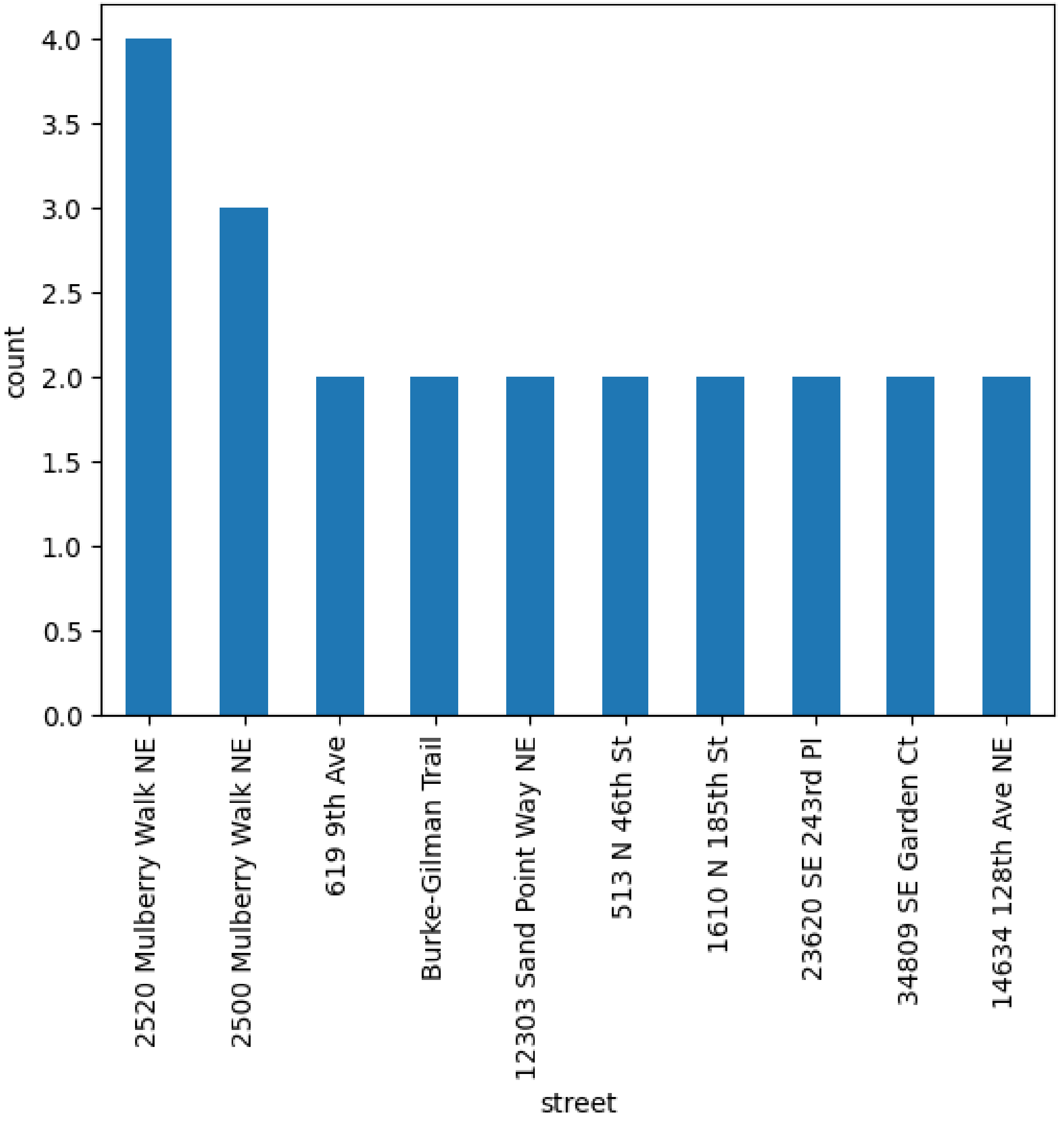
```
[268]: df['street'].value_counts()
```

```
[268]: street
       2520 Mulberry Walk NE     4
       2500 Mulberry Walk NE     3
       9413 34th Ave SW          2
       6008 8th Ave NE           2
       11034 NE 26th Pl          2
                                ..
       1404 Broadmoor Dr E       1
       3249 E Ames Lake Dr NE    1
       6032 35th Ave NE          1
       1006 NE Ravenna Blvd      1
       18717 SE 258th St         1
       Name: count, Length: 4525, dtype: int64
```

```
[269]: df['street'].value_counts().sort_values(ascending=False).head(10).
        ↪plot(kind='bar')
```

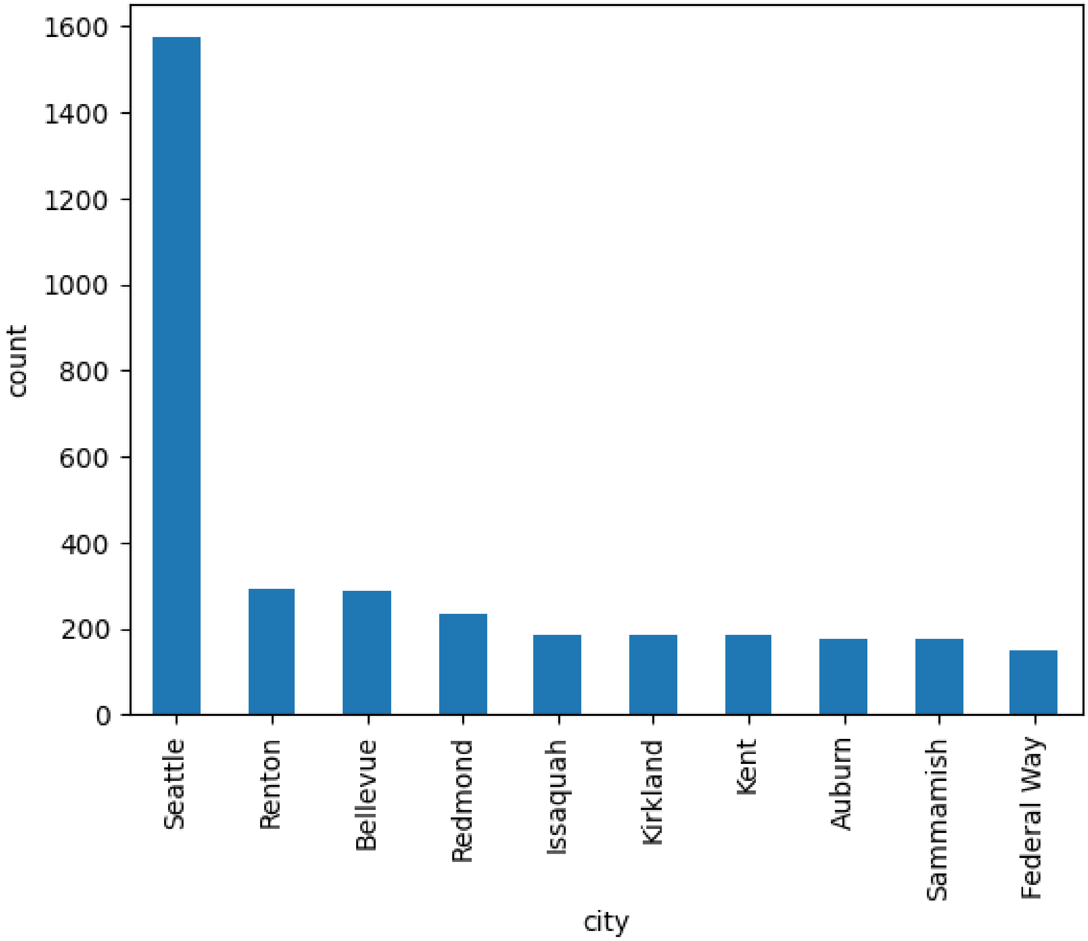
```
plt.xlabel('street')
plt.ylabel('count')
```

```
[269]: Text(0, 0.5, 'count')
```



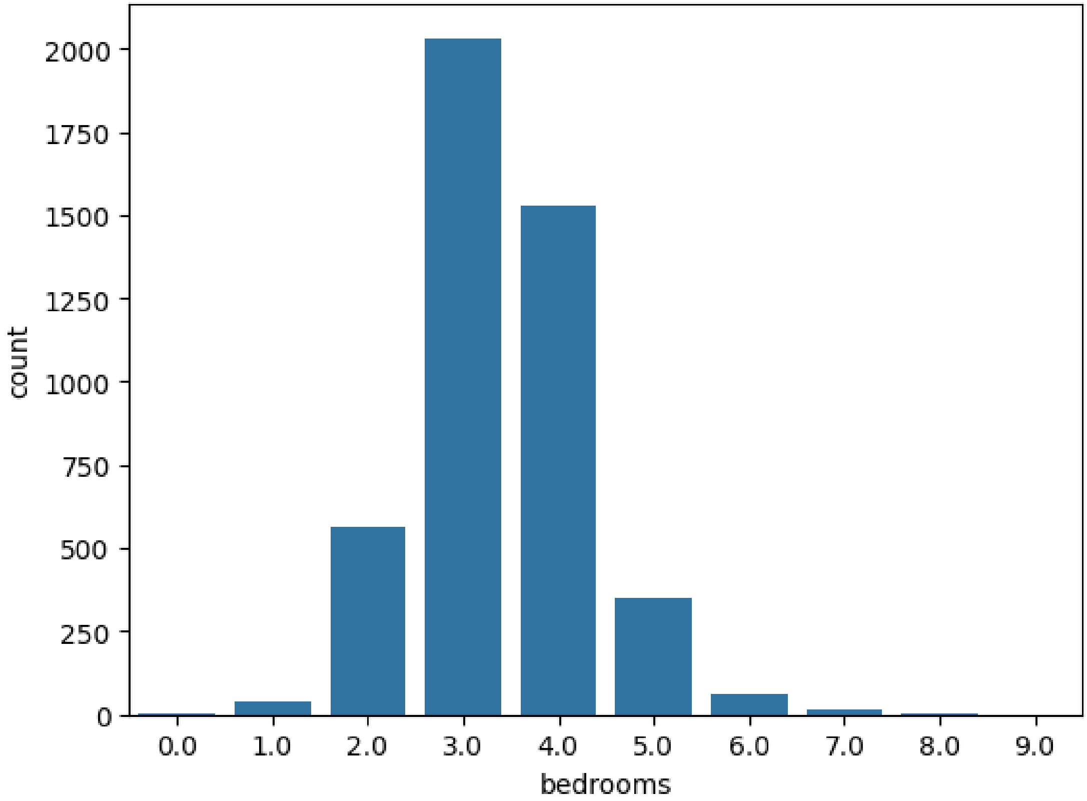
```
[270]: df['city'].value_counts().sort_values(ascending=False).head(10).plot(kind='bar')
plt.xlabel('city')
plt.ylabel('count')
```

```
[270]: Text(0, 0.5, 'count')
```



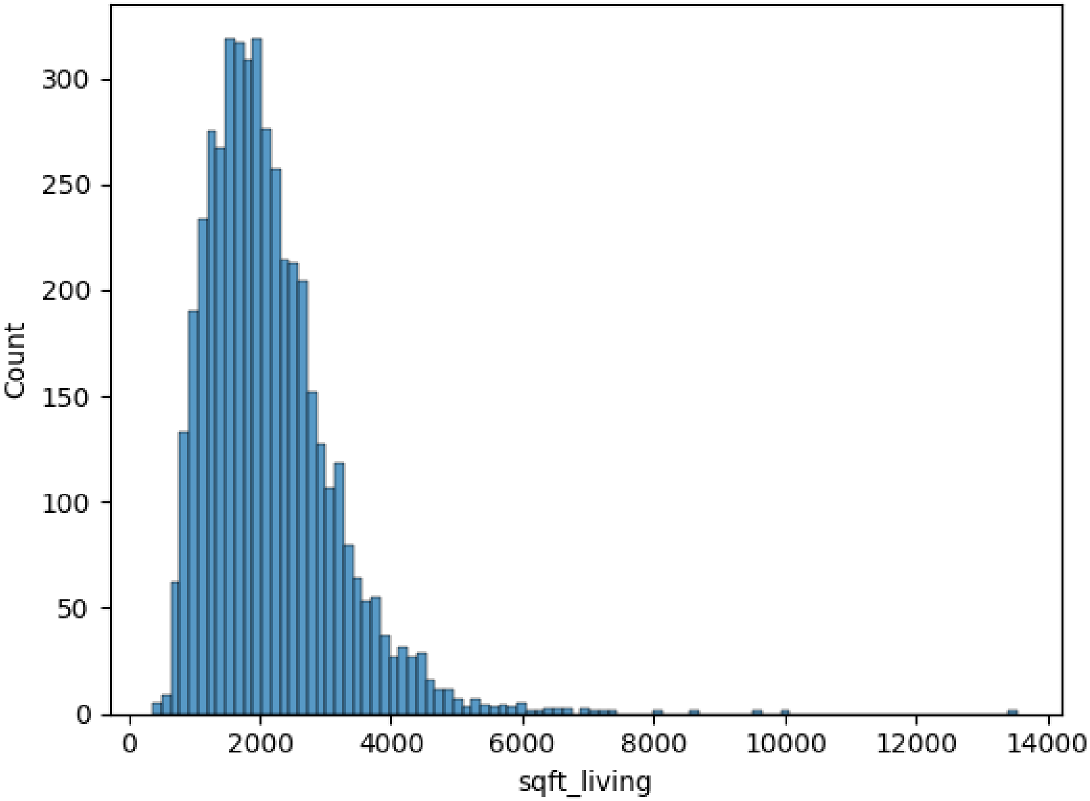
```
[271]: sns.countplot(x='bedrooms',data=df)
```

```
[271]: <Axes: xlabel='bedrooms', ylabel='count'>
```



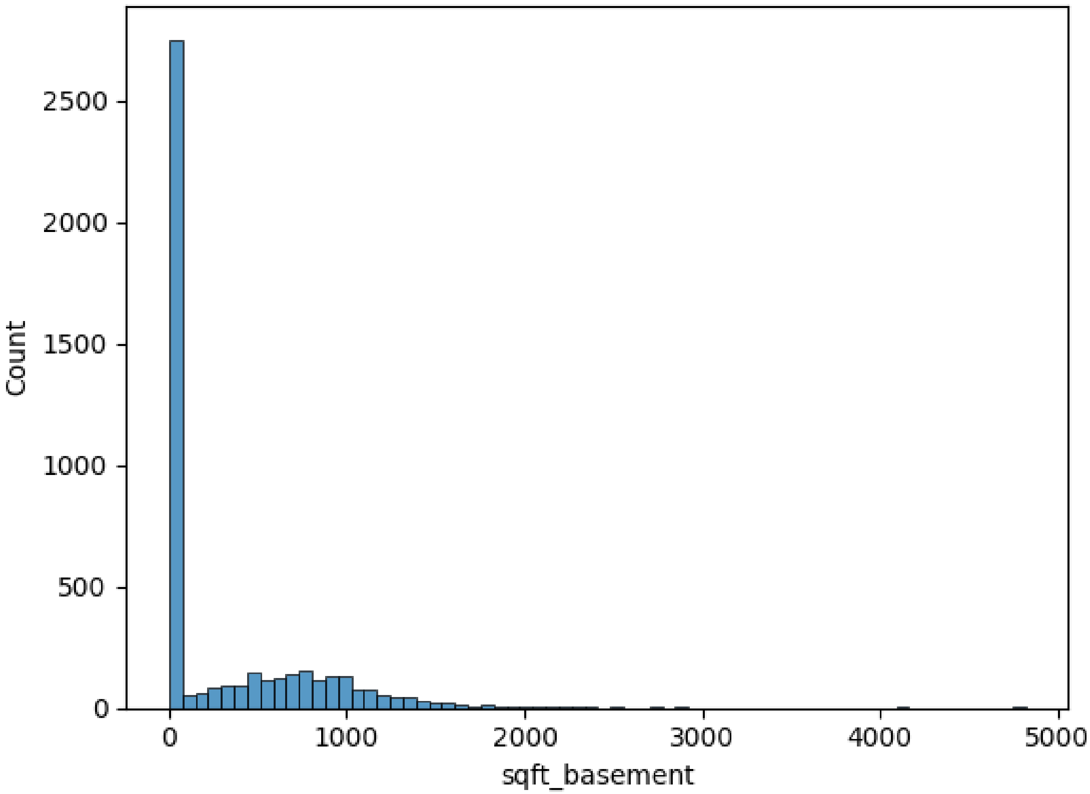
```
[272]: sns.histplot(x='sqft_living',data=df)
```

```
[272]: <Axes: xlabel='sqft_living', ylabel='Count'>
```



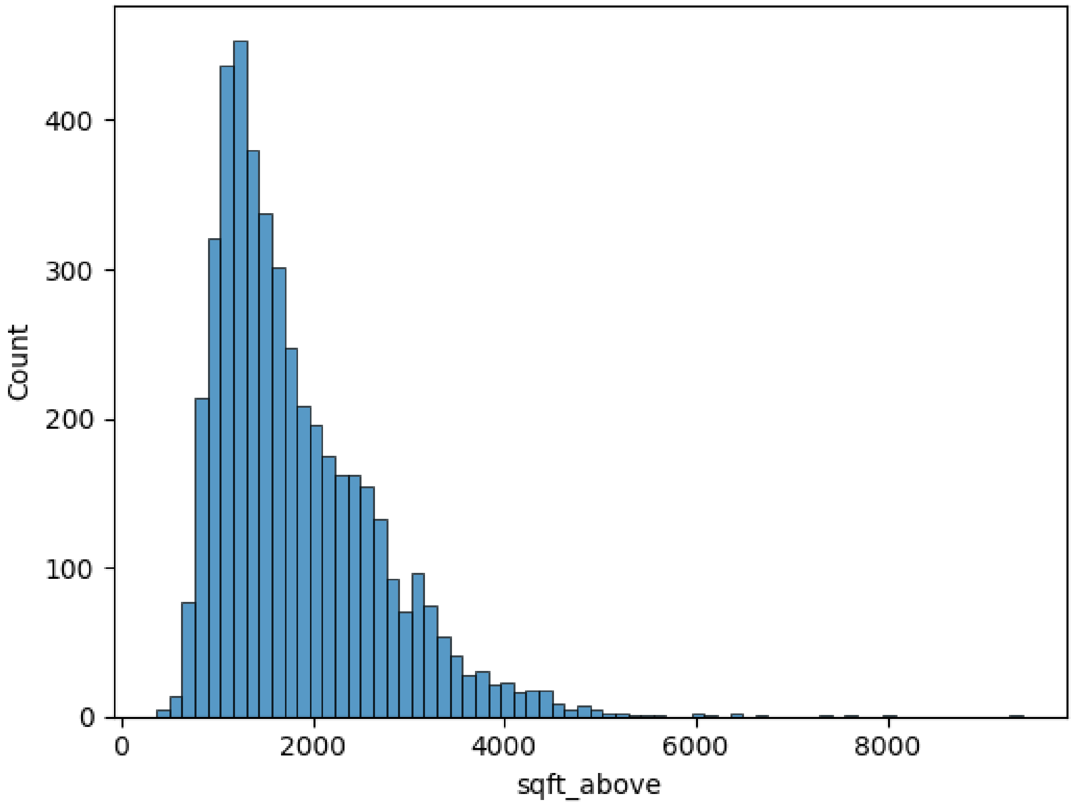
[273]:
```
sns.histplot(x='sqft_basement',data=df)
```

[273]:
```
<Axes: xlabel='sqft_basement', ylabel='Count'>
```



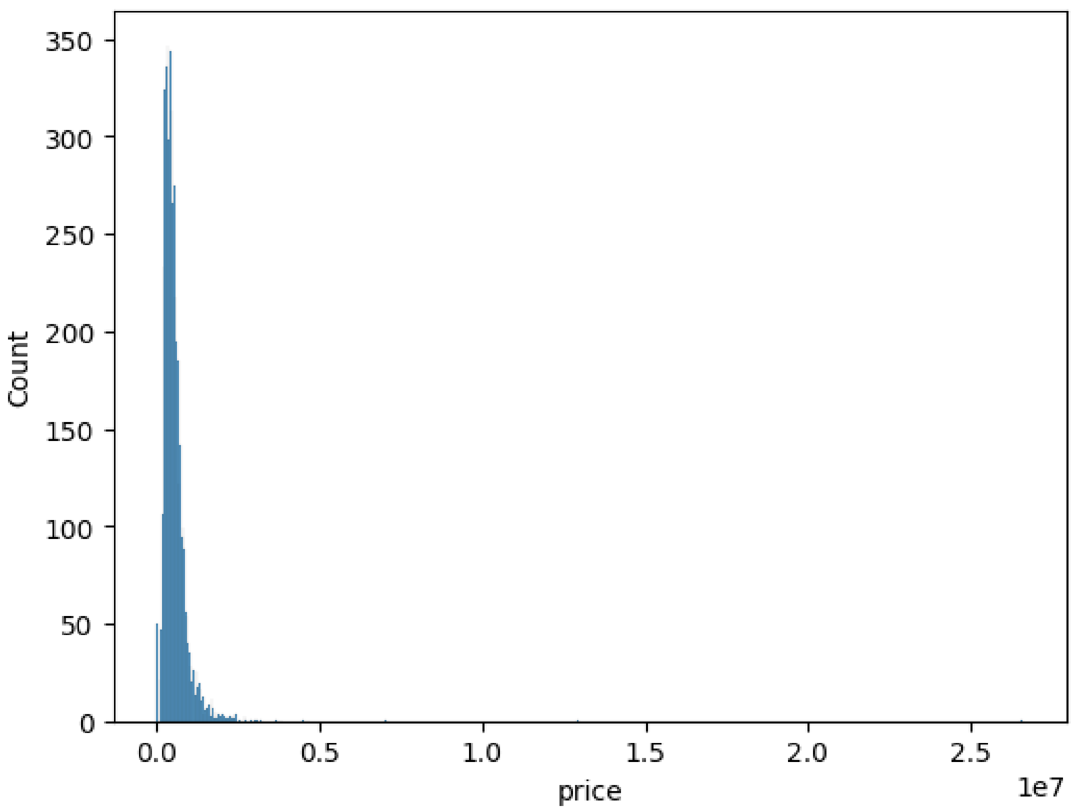
[274]:
```
sns.histplot(x='sqft_above',data=df)
```

[274]:
```
<Axes: xlabel='sqft_above', ylabel='Count'>
```



[275]:
```
sns.histplot(x='price',data=df)
```

[275]:
```
<Axes: xlabel='price', ylabel='Count'>
```



```
[276]: #encoding
       from sklearn.preprocessing import LabelEncoder
       end=LabelEncoder()
       df['city']=end.fit_transform(df['city'])
       df
```

```
[276]:                      date         price  bedrooms  bathrooms  sqft_living  \
       0     2014-05-02 00:00:00  3.130000e+05       3.0       1.50         1340
       1     2014-05-02 00:00:00  2.384000e+06       5.0       2.50         3650
       2     2014-05-02 00:00:00  3.420000e+05       3.0       2.00         1930
       3     2014-05-02 00:00:00  4.200000e+05       3.0       2.25         2000
       4     2014-05-02 00:00:00  5.500000e+05       4.0       2.50         1940
       …                      …             …        …          …            …
       4595  2014-07-09 00:00:00  3.081667e+05       3.0       1.75         1510
       4596  2014-07-09 00:00:00  5.343333e+05       3.0       2.50         1460
       4597  2014-07-09 00:00:00  4.169042e+05       3.0       2.50         3010
       4598  2014-07-10 00:00:00  2.034000e+05       4.0       2.00         2090
       4599  2014-07-10 00:00:00  2.206000e+05       3.0       2.50         1490

             sqft_lot  floors  waterfront  view  condition  sqft_above  \
       0         7912     1.5           0     0          3        1340
```

```
1         9050     2.0           0     4          5        3370
2        11947     1.0           0     0          4        1930
3         8030     1.0           0     0          4        1000
4        10500     1.0           0     0          4        1140
…        …     …         …   …        …         …
4595      6360     1.0           0     0          4        1510
4596      7573     2.0           0     0          3        1460
4597      7014     2.0           0     0          3        3010
4598      6630     1.0           0     0          3        1070
4599      8102     2.0           0     0          4        1490

      sqft_basement  yr_built  yr_renovated                    street  city  \
0                 0      1955          2005      18810 Densmore Ave N    36
1               280      1921             0           709 W Blaine St    35
2                 0      1966             0  26206-26214 143rd Ave SE    18
3              1000      1963             0           857 170th Pl NE     3
4               800      1976          1992         9105 170th Ave NE    31
…             …       …           …                       …   …
4595              0      1954          1979            501 N 143rd St    35
4596              0      1983          2009          14855 SE 10th Pl     3
4597              0      2009             0          759 Ilwaco Pl NE    32
4598           1020      1974             0         5148 S Creston St    35
4599              0      1990             0         18717 SE 258th St     9

      statezip country
0     WA 98133     USA
1     WA 98119     USA
2     WA 98042     USA
3     WA 98008     USA
4     WA 98052     USA
…        …     …
4595  WA 98133     USA
4596  WA 98007     USA
4597  WA 98059     USA
4598  WA 98178     USA
4599  WA 98042     USA

[4600 rows x 18 columns]
```

[277]:
```
df.drop(np.where(df["price"]==0)[0],axis=0,inplace=True)
df['statezip']=df['statezip'].str.replace('WA','')
df
```

[277]:
```
                     date         price  bedrooms  bathrooms  sqft_living  \
0     2014-05-02 00:00:00  3.130000e+05       3.0       1.50         1340
1     2014-05-02 00:00:00  2.384000e+06       5.0       2.50         3650
2     2014-05-02 00:00:00  3.420000e+05       3.0       2.00         1930
```

```
3     2014-05-02 00:00:00  4.200000e+05       3.0       2.25         2000
4     2014-05-02 00:00:00  5.500000e+05       4.0       2.50         1940
…                    …           …       …        …          …
4595  2014-07-09 00:00:00  3.081667e+05       3.0       1.75         1510
4596  2014-07-09 00:00:00  5.343333e+05       3.0       2.50         1460
4597  2014-07-09 00:00:00  4.169042e+05       3.0       2.50         3010
4598  2014-07-10 00:00:00  2.034000e+05       4.0       2.00         2090
4599  2014-07-10 00:00:00  2.206000e+05       3.0       2.50         1490

      sqft_lot  floors  waterfront  view  condition  sqft_above  \
0         7912     1.5           0     0          3        1340
1         9050     2.0           0     4          5        3370
2        11947     1.0           0     0          4        1930
3         8030     1.0           0     0          4        1000
4        10500     1.0           0     0          4        1140
…        …     …          …   …        …         …
4595      6360     1.0           0     0          4        1510
4596      7573     2.0           0     0          3        1460
4597      7014     2.0           0     0          3        3010
4598      6630     1.0           0     0          3        1070
4599      8102     2.0           0     0          4        1490

      sqft_basement  yr_built  yr_renovated                    street  city  \
0                 0      1955          2005      18810 Densmore Ave N    36
1               280      1921             0          709 W Blaine St    35
2                 0      1966             0  26206-26214 143rd Ave SE    18
3              1000      1963             0          857 170th Pl NE     3
4               800      1976          1992         9105 170th Ave NE    31
…             …       …            …                        …   …
4595              0      1954          1979          501 N 143rd St    35
4596              0      1983          2009         14855 SE 10th Pl     3
4597              0      2009             0        759 Ilwaco Pl NE    32
4598           1020      1974             0        5148 S Creston St    35
4599              0      1990             0        18717 SE 258th St     9

     statezip country
0       98133     USA
1       98119     USA
2       98042     USA
3       98008     USA
4       98052     USA
…       …     …
4595    98133     USA
4596    98007     USA
4597    98059     USA
4598    98178     USA
4599    98042     USA
```

```
[4551 rows x 18 columns]
```

[278]:

```
df.drop(['country','date','street'],axis=1,inplace=True)
df
```

[278]:

```
             price  bedrooms  bathrooms  sqft_living  sqft_lot  floors  \
0     3.130000e+05       3.0       1.50         1340      7912     1.5
1     2.384000e+06       5.0       2.50         3650      9050     2.0
2     3.420000e+05       3.0       2.00         1930     11947     1.0
3     4.200000e+05       3.0       2.25         2000      8030     1.0
4     5.500000e+05       4.0       2.50         1940     10500     1.0
…            …       …        …          …       …     …
4595  3.081667e+05       3.0       1.75         1510      6360     1.0
4596  5.343333e+05       3.0       2.50         1460      7573     2.0
4597  4.169042e+05       3.0       2.50         3010      7014     2.0
4598  2.034000e+05       4.0       2.00         2090      6630     1.0
4599  2.206000e+05       3.0       2.50         1490      8102     2.0

      waterfront  view  condition  sqft_above  sqft_basement  yr_built  \
0              0     0          3        1340              0      1955
1              0     4          5        3370            280      1921
2              0     0          4        1930              0      1966
3              0     0          4        1000           1000      1963
4              0     0          4        1140            800      1976
…          …   …        …         …            …       …
4595           0     0          4        1510              0      1954
4596           0     0          3        1460              0      1983
4597           0     0          3        3010              0      2009
4598           0     0          3        1070           1020      1974
4599           0     0          4        1490              0      1990

      yr_renovated  city statezip
0             2005    36    98133
1                0    35    98119
2                0    18    98042
3                0     3    98008
4             1992    31    98052
…            …   …      …
4595          1979    35    98133
4596          2009     3    98007
4597             0    32    98059
4598             0    35    98178
4599             0     9    98042

[4551 rows x 15 columns]
```

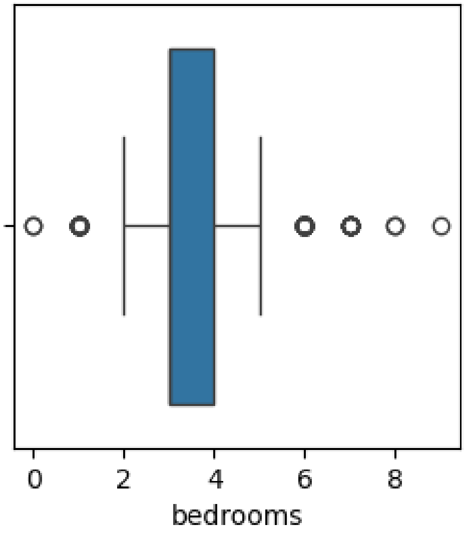
[279]:

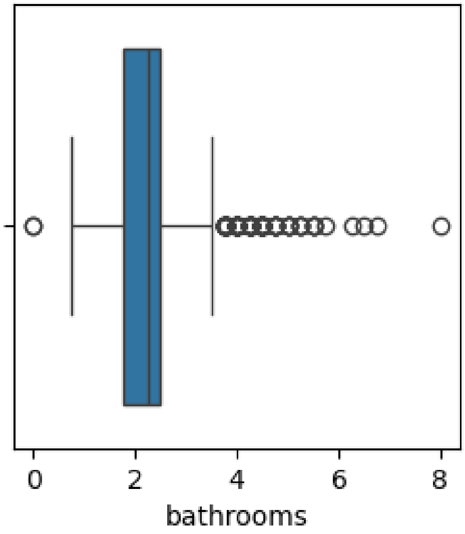
```
print('Before transformation skew : ',df['price'].skew())
df['price'] = np.log(df['price'])
print('After transformation skew : ',df['price'].skew())
```

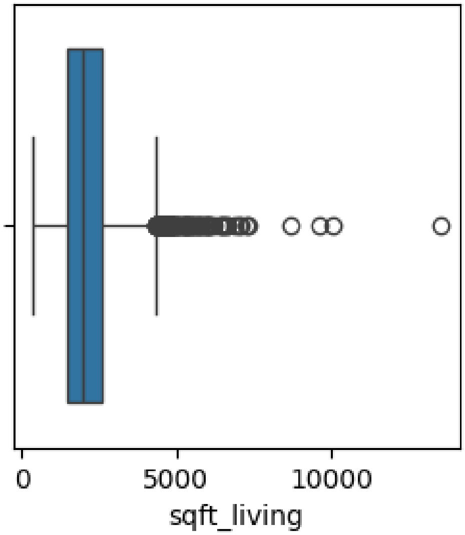
```
Before transformation skew :  25.023817262008482
After transformation skew :  0.3299813838090415
```

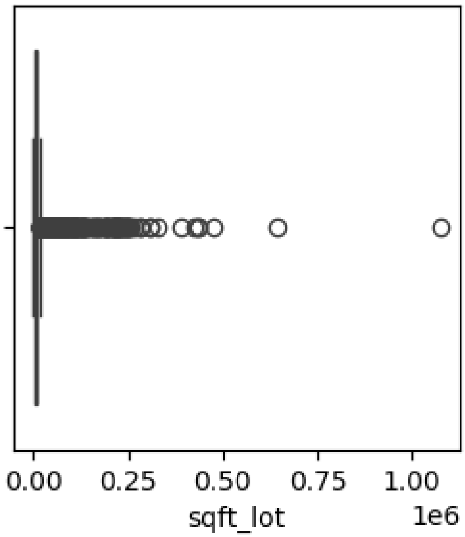
[280]:

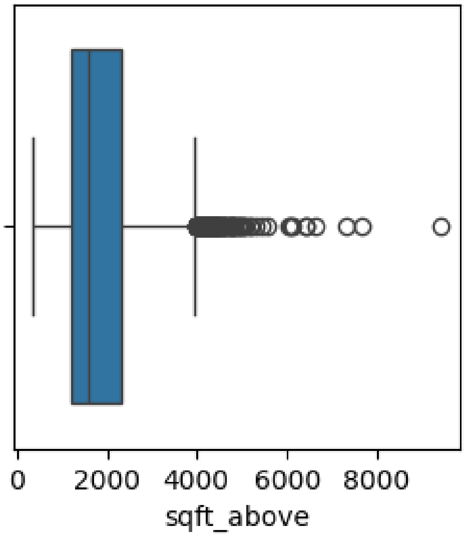
```
for i in␣
 ↪['bedrooms','bathrooms','sqft_living','sqft_lot','sqft_above','sqft_basement']:
 ↪
  plt.figure(figsize=(3,3))
  sns.boxplot(x=i,data=df)
  plt.show()
```

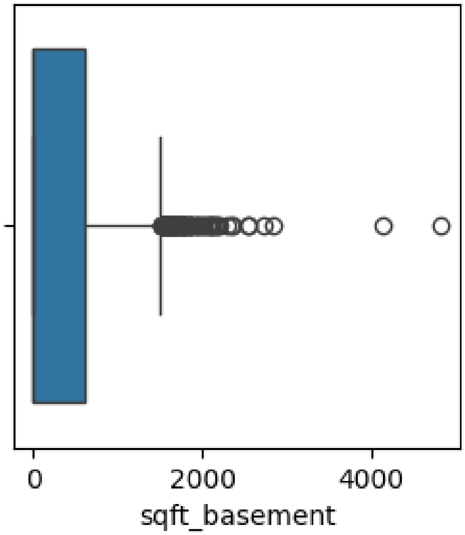



0
2
4
6
8
bathrooms


0
5000
10000
sqft_living


0.00
0.25
0.50
0.75
1.00
1e6
sqft_lot


0
2000
4000
6000
8000
sqft_above



```
[281]: for i in␣
         ↪['bedrooms','bathrooms','sqft_living','sqft_lot','sqft_above','sqft_basement']:
         ↪
           Q1=df[i].quantile(0.25)
           Q3=df[i].quantile(0.75)
           IQR=Q3-Q1
           LOWER=Q1-(1.5*IQR)
           UPPER=Q3+(1.5*IQR)
           dfe = df[(df[i]>=LOWER)&(df[i]<=UPPER)]
       dfe
```

```
[281]:            price  bedrooms  bathrooms  sqft_living  sqft_lot  floors  \
       0      12.653958       3.0       1.50         1340      7912     1.5
       1      14.684290       5.0       2.50         3650      9050     2.0
       2      12.742566       3.0       2.00         1930     11947     1.0
       3      12.948010       3.0       2.25         2000      8030     1.0
       4      13.217674       4.0       2.50         1940     10500     1.0
       …            …         …          …           …         …       …
       4595   12.638396       3.0       1.75         1510      6360     1.0
       4596   13.188775       3.0       2.50         1460      7573     2.0
       4597   12.940612       3.0       2.50         3010      7014     2.0
       4598   12.222930       4.0       2.00         2090      6630     1.0
       4599   12.304106       3.0       2.50         1490      8102     2.0

             waterfront  view  condition  sqft_above  sqft_basement  yr_built  \
       0              0     0          3        1340              0      1955
       1              0     4          5        3370            280      1921
       2              0     0          4        1930              0      1966
```

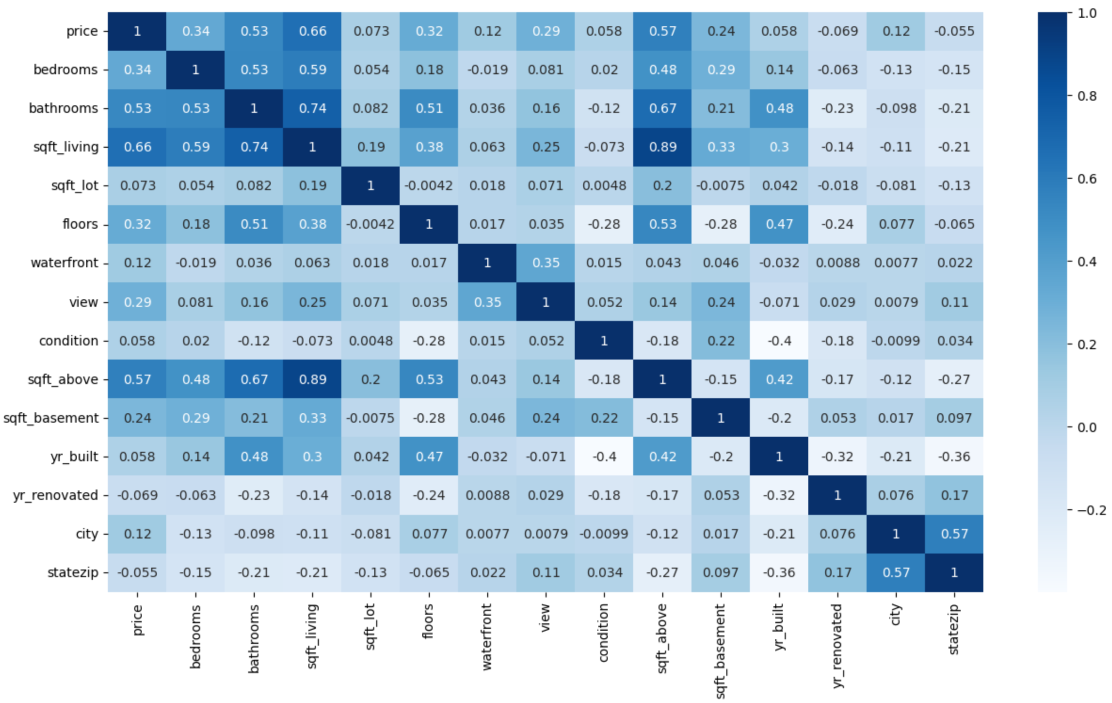
```
3              0     0          4        1000           1000      1963
4              0     0          4        1140            800      1976
…          …   …        …         …            …       …
4595           0     0          4        1510              0      1954
4596           0     0          3        1460              0      1983
4597           0     0          3        3010              0      2009
4598           0     0          3        1070           1020      1974
4599           0     0          4        1490              0      1990

      yr_renovated  city statezip
0             2005    36    98133
1                0    35    98119
2                0    18    98042
3                0     3    98008
4             1992    31    98052
…            …   …      …
4595          1979    35    98133
4596          2009     3    98007
4597             0    32    98059
4598             0    35    98178
4599             0     9    98042

[4468 rows x 15 columns]
```

[282]:

```
plt.figure(figsize=(15,8))
sns.heatmap(dfe.corr(),annot=True,cmap='Blues')
plt.plot()
```

[282]:

```
[]
```



[283]:
```
x=dfe.drop('price',axis=1)
x
```

[283]:
```
      bedrooms  bathrooms  sqft_living  sqft_lot  floors  waterfront  view  \
0          3.0       1.50         1340      7912     1.5           0     0
1          5.0       2.50         3650      9050     2.0           0     4
2          3.0       2.00         1930     11947     1.0           0     0
3          3.0       2.25         2000      8030     1.0           0     0
4          4.0       2.50         1940     10500     1.0           0     0
…            …          …            …         …       …   …
4595       3.0       1.75         1510      6360     1.0           0     0
4596       3.0       2.50         1460      7573     2.0           0     0
4597       3.0       2.50         3010      7014     2.0           0     0
4598       4.0       2.00         2090      6630     1.0           0     0
4599       3.0       2.50         1490      8102     2.0           0     0

      condition  sqft_above  sqft_basement  yr_built  yr_renovated  city  \
0             3        1340              0      1955          2005    36
1             5        3370            280      1921             0    35
2             4        1930              0      1966             0    18
3             4        1000           1000      1963             0     3
4             4        1140            800      1976          1992    31
…           …           …              …         …           …   …
4595          4        1510              0      1954          1979    35
```

```
4596           3        1460             0      1983          2009     3
4597           3        3010             0      2009             0    32
4598           3        1070          1020      1974             0    35
4599           4        1490             0      1990             0     9

     statezip
0       98133
1       98119
2       98042
3       98008
4       98052
…         …
4595    98133
4596    98007
4597    98059
4598    98178
4599    98042

[4468 rows x 14 columns]
```

```
[284]: y=dfe['price']
       y
```

```
[284]: 0       12.653958
       1       14.684290
       2       12.742566
       3       12.948010
       4       13.217674
                 …
       4595    12.638396
       4596    13.188775
       4597    12.940612
       4598    12.222930
       4599    12.304106
       Name: price, Length: 4468, dtype: float64
```

CONVERT INTO TRAINING AND TESTING DATA

```
[285]: from sklearn.model_selection import train_test_split
       x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.
         ↪30,random_state=42)
       x_train
```

```
[285]:       bedrooms  bathrooms  sqft_living  sqft_lot  floors  waterfront  view  \
       737        5.0       1.75         2000      3750     2.0           0     0
       3092       3.0       1.00         1150      8145     1.0           0     0
       991        3.0       2.50         2090      4700     2.0           0     0
```

```
3758       3.0       3.50         2080      5100     2.0           0     0
3904       3.0       2.50         3230      5000     2.0           0     2
…        …        …          …        …     …         …   …
4549       3.0       1.00         1750      7800     1.0           0     0
470        3.0       2.50         1370      1350     2.0           0     0
3153       3.0       2.75         2810     18731     2.0           1     4
3844       3.0       2.25         1550      5511     2.0           0     0
874        2.0       1.00         1120      9912     1.0           0     0

      condition  sqft_above  sqft_basement  yr_built  yr_renovated  city  \
737           4        2000              0      1921             0    35
3092          4         990            160      1932          1958    36
991           3        2090              0      2002             0     9
3758          3        2080              0      2004          2003    21
3904          5        2430            800      1945             0    35
…         …         …             …       …           …   …
4549          4        1150            600      1956             0     6
470           3        1010            360      2007             0    35
3153          4        2810              0      1974             0    31
3844          3        1550              0      1987          2000    19
874           4        1120              0      1980             0    41

     statezip
737     98103
3092    98155
991     98042
3758    98038
3904    98117
…       …
4549    98166
470     98136
3153    98052
3844    98033
874     98070

[3127 rows x 14 columns]
```

[286]:

```
x_test
```

[286]:

```
      bedrooms  bathrooms  sqft_living  sqft_lot  floors  waterfront  view  \
4252       4.0       3.25         4280     47179     2.0           0     0
152        2.0       2.00         1100      3000     1.5           0     0
3073       3.0       2.25         1650      2958     2.0           0     0
611        5.0       3.25         4860     23723     2.0           0     2
3278       3.0       2.00         1410      5760     1.0           0     0
…        …        …          …        …     …         …   …
344        5.0       2.25         2440     20828     1.5           0     0
```

```
2617       3.0       2.25         1350      1186     2.0           0     0
3653       3.0       2.25         2310      7200     2.0           0     0
1107       4.0       2.50         3240      3600     2.0           0     0
484        4.0       2.25         2540    228254     1.0           0     0

      condition  sqft_above  sqft_basement  yr_built  yr_renovated  city  \
4252          3        3050           1230      2002             0     1
152           3        1100              0      1912          2005    35
3073          3        1650              0      1985             0    18
611           4        3820           1040      1989             0    23
3278          3        1410              0      1985             0    18
…         …         …            …       …           …   …
344           4        2440              0      1975             0    18
2617          3        1120            230      2007             0    35
3653          3        2310              0      1990          2009    18
1107          3        2060           1180      2008             0    35
484           3        1450           1090      1990          2009    11

     statezip
4252    98092
152     98117
3073    98031
611     98040
3278    98030
…       …
344     98042
2617    98117
3653    98031
1107    98109
484     98019

[1341 rows x 14 columns]
```

[287]:

```
y_train
```

[287]:

```
737     13.279367
3092    12.577636
991     12.574182
3758    12.786891
3904    13.652992
          …
4549    12.435419
470     12.815838
3153    14.346139
3844    13.083654
874     12.454884
Name: price, Length: 3127, dtype: float64
```

[288]:

```
y_test
```

[288]:

```
4252    13.507558
152     13.017003
3073    12.360937
611     14.334304
3278    12.409013
          …
344     13.102140
2617    13.057291
3653    12.699243
1107    14.014361
484     13.199324
Name: price, Length: 1341, dtype: float64
```

MODEL CREATION

[289]:

```
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import␣
 ↪mean_absolute_error,mean_absolute_percentage_error,mean_squared_error,r2_score
lr=LinearRegression()
rf=RandomForestRegressor(n_estimators=100,random_state=42)
tree=DecisionTreeRegressor()
lst=[lr,rf,tree]
```

[290]:

```
for i in lst:
  print('MODEL IS',i)
  print('-'*80)
  i.fit(x_train,y_train)
  y_pred=i.predict(x_test)
  print('MAPE IS',mean_absolute_percentage_error(y_test,y_pred))
  print('MAE IS',mean_absolute_error(y_test,y_pred))
  print('score IS',r2_score(y_test,y_pred))
  print('MSE IS',mean_squared_error(y_test,y_pred))
  print('-'*80)
  print()
```

```
MODEL IS LinearRegression()
--------------------------------------------------------------------------------
MAPE IS 0.02132980586113901
MAE IS 0.2780404967521377
score IS 0.539982683551661
MSE IS 0.12529611642811853
--------------------------------------------------------------------------------

MODEL IS RandomForestRegressor(random_state=42)
```

```
--------------------------------------------------------------------------------
MAPE IS 0.01482196740834958
MAE IS 0.1938454839419827
score IS 0.7081706309167244
MSE IS 0.0794863264889917
--------------------------------------------------------------------------------

MODEL IS DecisionTreeRegressor()
--------------------------------------------------------------------------------
MAPE IS 0.022223240779722094
MAE IS 0.2895947082125479
score IS 0.26192088118368395
MSE IS 0.20103253485840855
--------------------------------------------------------------------------------
```