Name:- ANMOL VIPIN DEVANSH Registration No.:- 21BCE10314
Assignment 2 Date:- 02-09-23

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

# Task 1: Load the dataset
df = pd.read_csv('HousePriceIndia.csv')
print("Task 1: Load the dataset")
print(df)

columns_to_skip = ['id', 'Date', 'Postal Code', 'Lattitude',
'Longitude']

print("\n\n\n\nTask 2: Univariate Analysis")
for column in df.columns:
    if column in columns_to_skip:
        continue

    # Task 2: Univariate Analysis
    plt.figure(figsize=(10, 4))
    sns.histplot(df[column], bins=30, kde=True)
    plt.title(f'Histogram of {column}')
    plt.xlabel(column)
    plt.ylabel('Frequency')
    plt.show()

print("\n\n\n\nTask 3: Bivariate Analysis")
for column in df.columns:
    if column in columns_to_skip:
        continue

    # Task 3: Bivariate Analysis (Price vs. each variable)
    plt.figure(figsize=(10, 4))
    sns.barplot(data=df, x=column, y='Price')
    plt.title(f'Bar Plot: Price vs. {column}')
    plt.xlabel(column)
    plt.ylabel('Price')
    plt.show()

# Task 4: Multivariate Analysis
print("\n\n\n\nTask 4: Multivariate Analysis")
correlation_matrix = df.corr()
plt.figure(figsize=(12, 10))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm',
fmt=".2f")
plt.title('Correlation Heatmap')
```

```
plt.show()

# Task 5: Descriptive Statistics
descriptive_stats = df.describe()
print("\n\n\n\nTask 5: Descriptive statistics of 'df':")
print(descriptive_stats)

# Task 6: Handle Missing Values
print("\n\n\n\nTask 6: Handle Missing Values")
missing_values = df.isnull().sum()
# Fill missing values with the mean for numeric columns
numeric_columns = df.select_dtypes(include=[np.number])
df[numeric_columns.columns] =
df[numeric_columns.columns].fillna(numeric_columns.mean())
missing_values_after_fill = df.isnull().sum()
print("Missing values after filling:")
print(missing_values_after_fill)
```

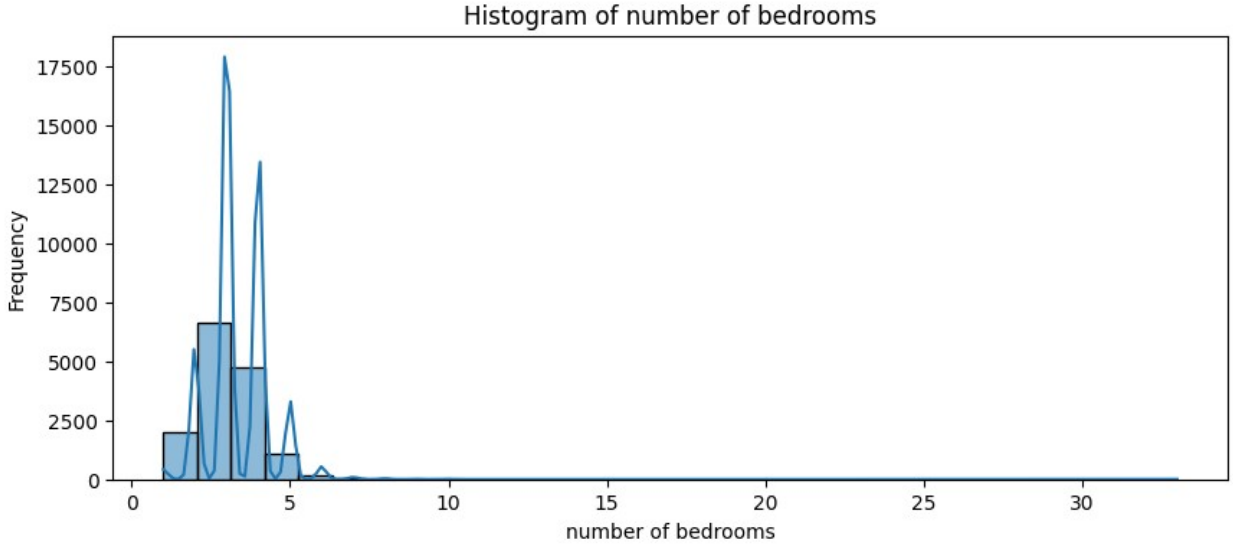
```
Task 1: Load the dataset
                id   Date  number of bedrooms  number of bathrooms
0      6762810145  42491                   5                 2.50  \
1      6762810635  42491                   4                 2.50
2      6762810998  42491                   5                 2.75
3      6762812605  42491                   4                 2.50
4      6762812919  42491                   3                 2.00
...           ...    ...                 ...                  ...
14615  6762830250  42734                   2                 1.50
14616  6762830339  42734                   3                 2.00
14617  6762830618  42734                   2                 1.00
14618  6762830709  42734                   4                 1.00
14619  6762831463  42734                   3                 1.00

       living area  lot area  number of floors  waterfront present
0             3650      9050               2.0                   0  \
1             2920      4000               1.5                   0
2             2910      9480               1.5                   0
3             3310     42998               2.0                   0
4             2710      4500               1.5                   0
...            ...       ...               ...                 ...
14615         1556     20000               1.0                   0
14616         1680      7000               1.5                   0
14617         1070      6120               1.0                   0
14618         1030      6621               1.0                   0
14619          900      4770               1.0                   0

       number of views  condition of the house  ...  Built Year
0                    4                       5  ...        1921  \
1                    0                       5  ...        1909
2                    0                       3  ...        1939
3                    0                       3  ...        2001
```

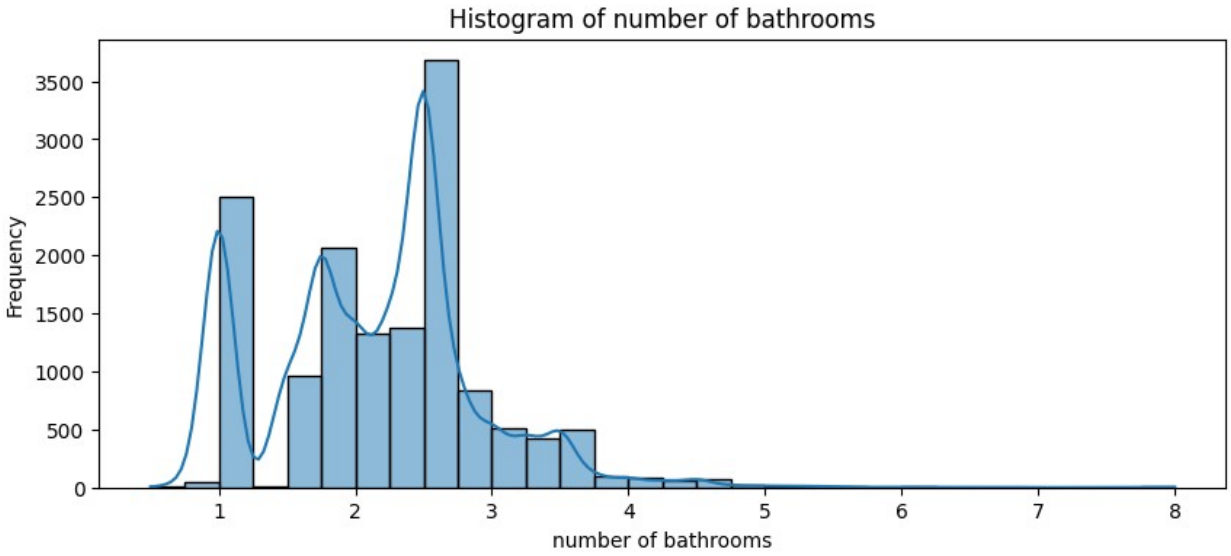
```
4                     0                         4  ...        1929
...                 ...                       ...  ...         ...
14615                 0                         4  ...        1957
14616                 0                         4  ...        1968
14617                 0                         3  ...        1962
14618                 0                         4  ...        1955
14619                 0                         3  ...        1969

       Renovation Year  Postal Code  Lattitude  Longitude
living_area_renov
0                     0       122003    52.8645   -114.557
2880  \
1                     0       122004    52.8878   -114.470
2470
2                     0       122004    52.8852   -114.468
2940
3                     0       122005    52.9532   -114.321
3350
4                     0       122006    52.9047   -114.485
2060
...                 ...          ...        ...        ...
...
14615                 0       122066    52.6191   -114.472
2250
14616                 0       122072    52.5075   -114.393
1540
14617                 0       122056    52.7289   -114.507
1130
14618                 0       122042    52.7157   -114.411
1420
14619              2009       122018    52.5338   -114.552
900

       lot_area_renov  Number of schools nearby  Distance from the
airport
0                5400                         2
58  \
1                4000                         2
51
2                6600                         1
53
3               42847                         3
76
4                4500                         1
51
...               ...                       ...
...
14615           17286                         3
76
```

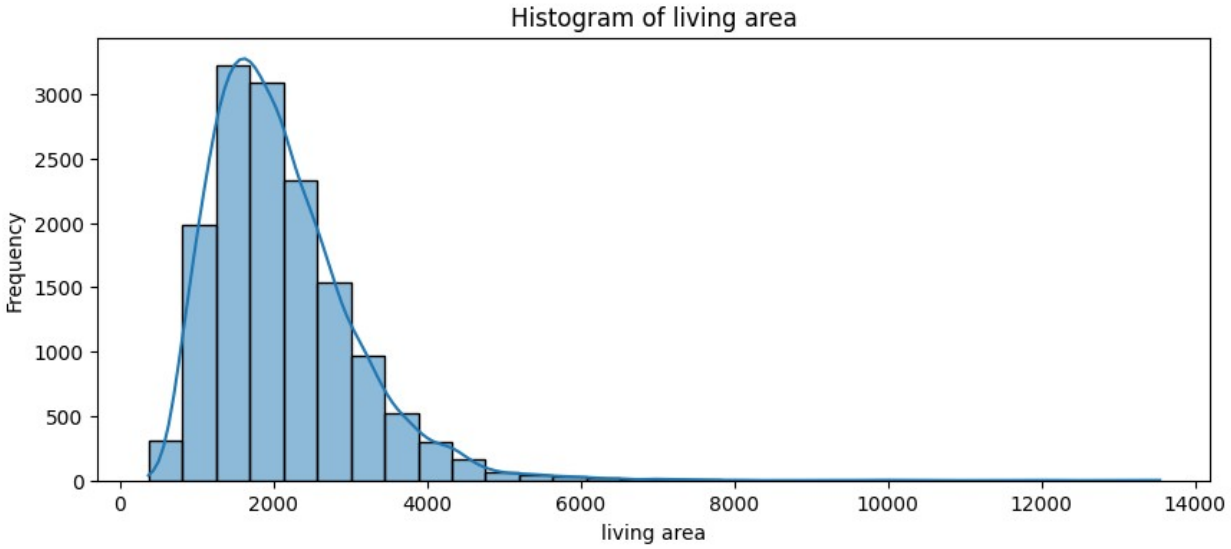
```
14616             7480                           3
59
14617             6120                           2
64
14618             6631                           3
54
14619             3480                           2
55
```

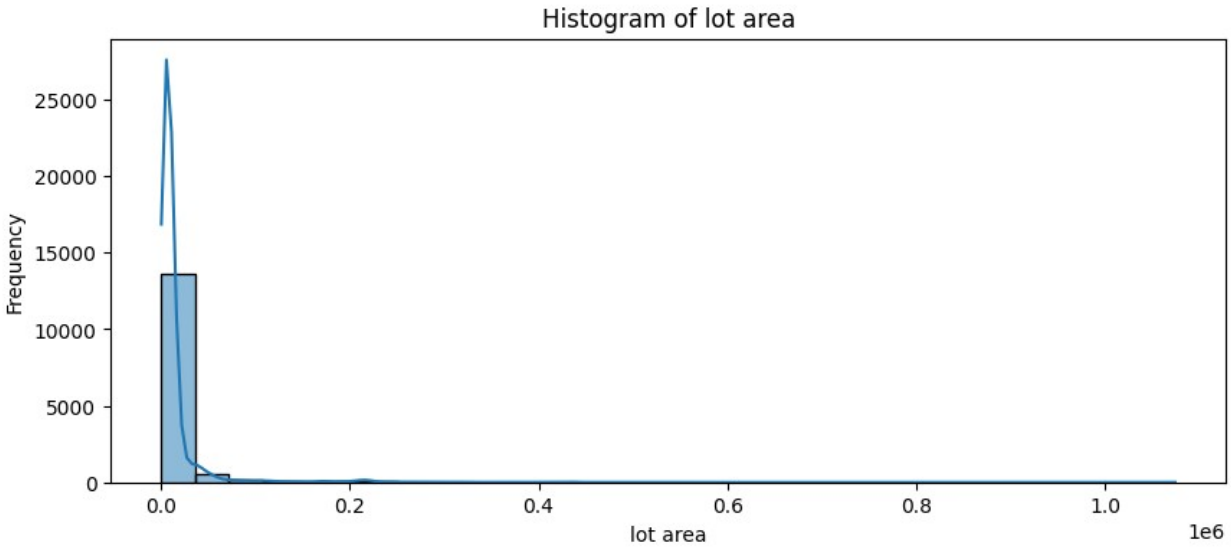
```
         Price
0      2380000
1      1400000
2      1200000
3       838000
4       805000
...        ...
14615   221700
14616   219200
14617   209000
14618   205000
14619   146000

[14620 rows x 23 columns]
```

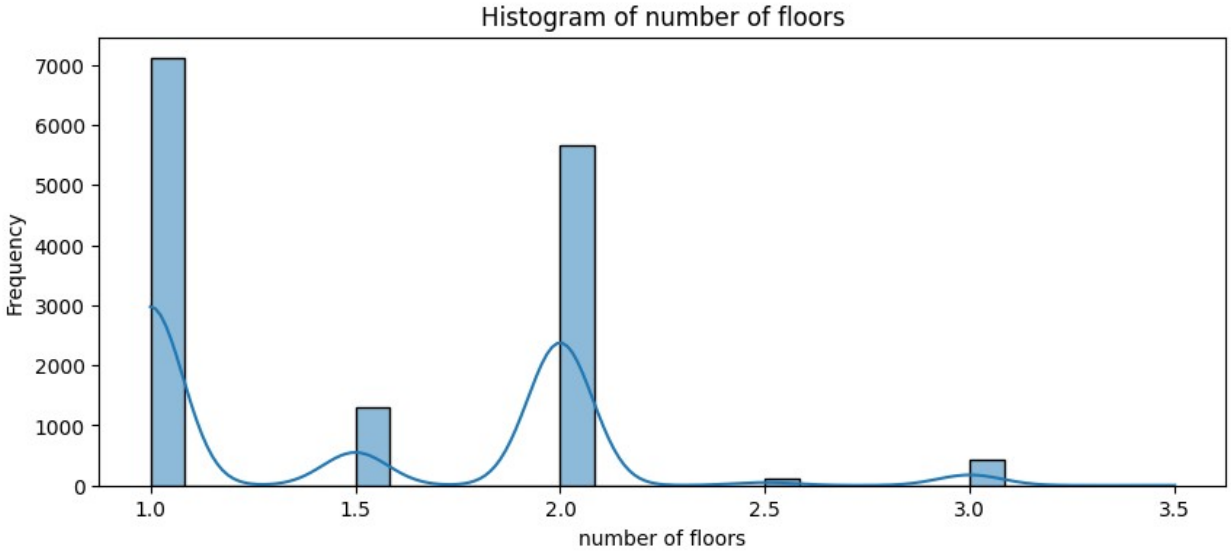
```
Task 2: Univariate Analysis
```

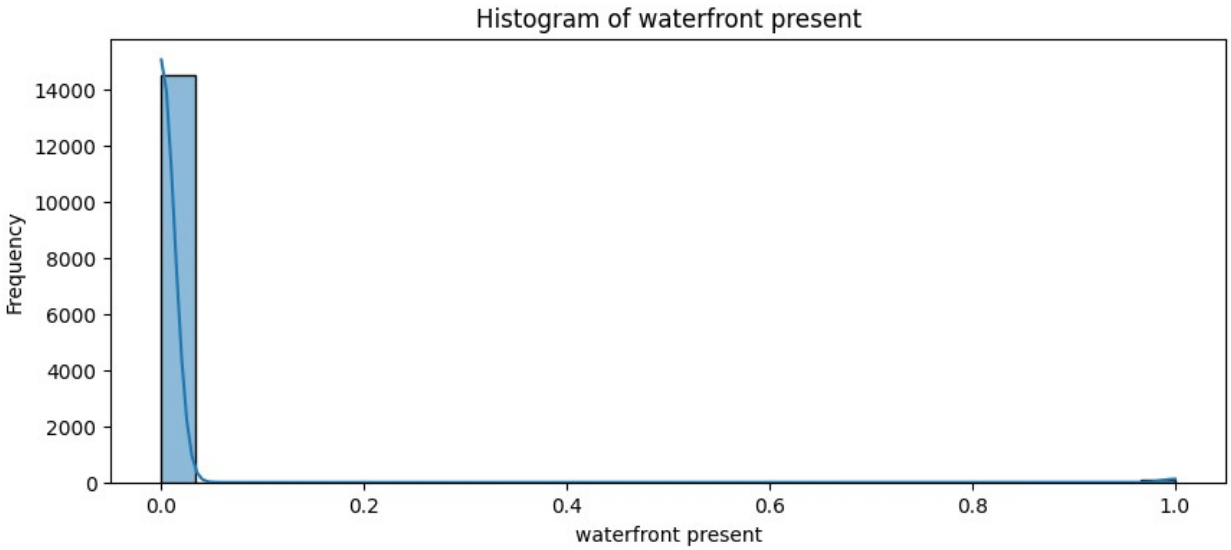


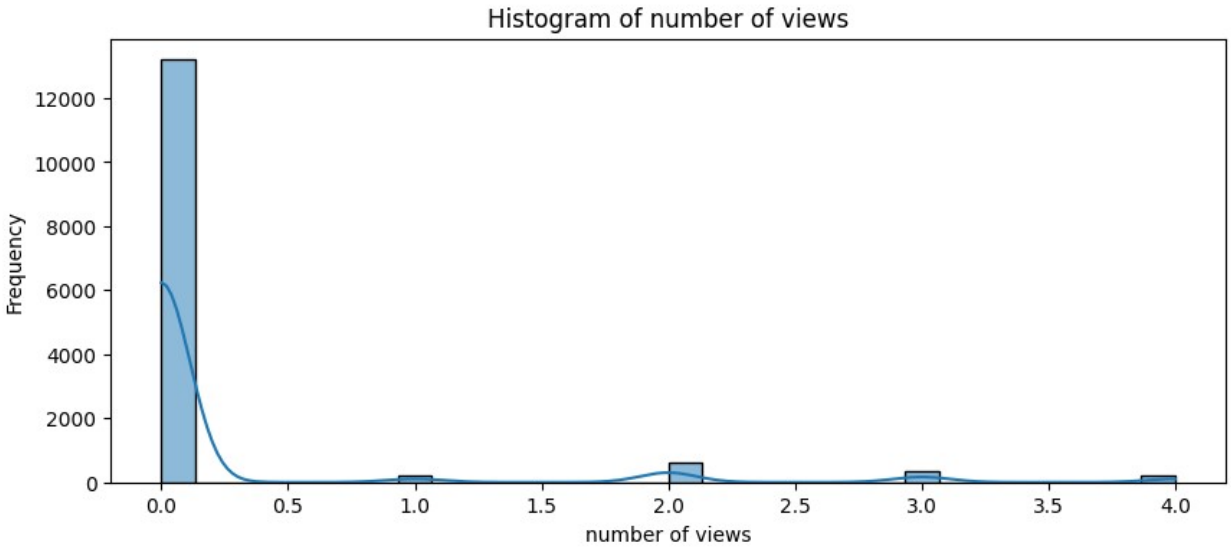

Histogram of number of bathrooms
Frequency
number of bathrooms
0
500
1000
1500
2000
2500
3000
3500
1
2
3
4
5
6
7
8

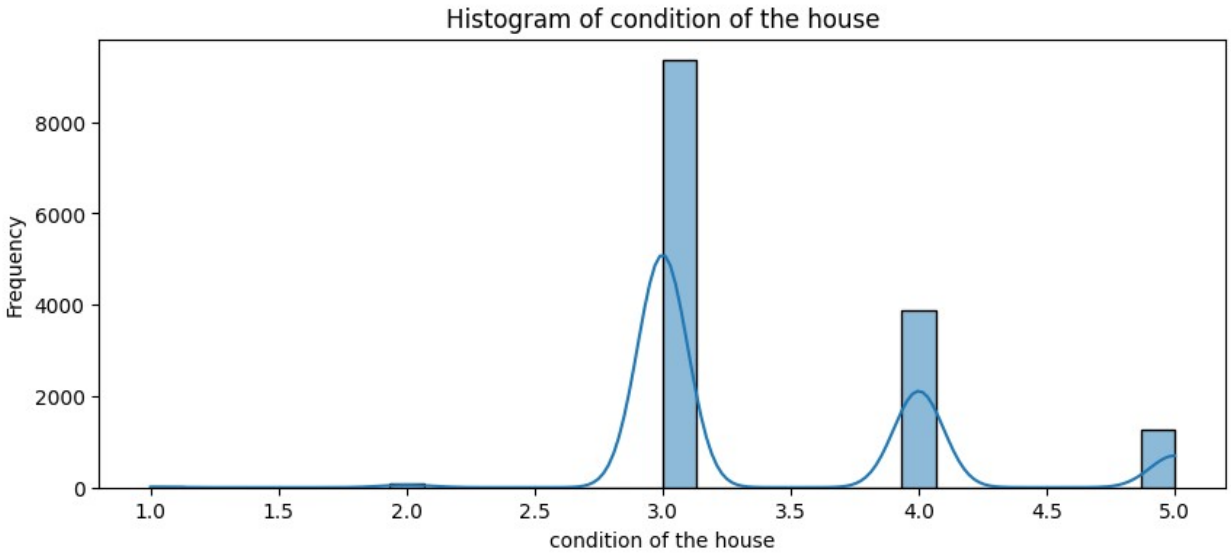

Histogram of living area
Frequency
living area
0
500
1000
1500
2000
2500
3000
0
2000
4000
6000
8000
10000
12000
14000

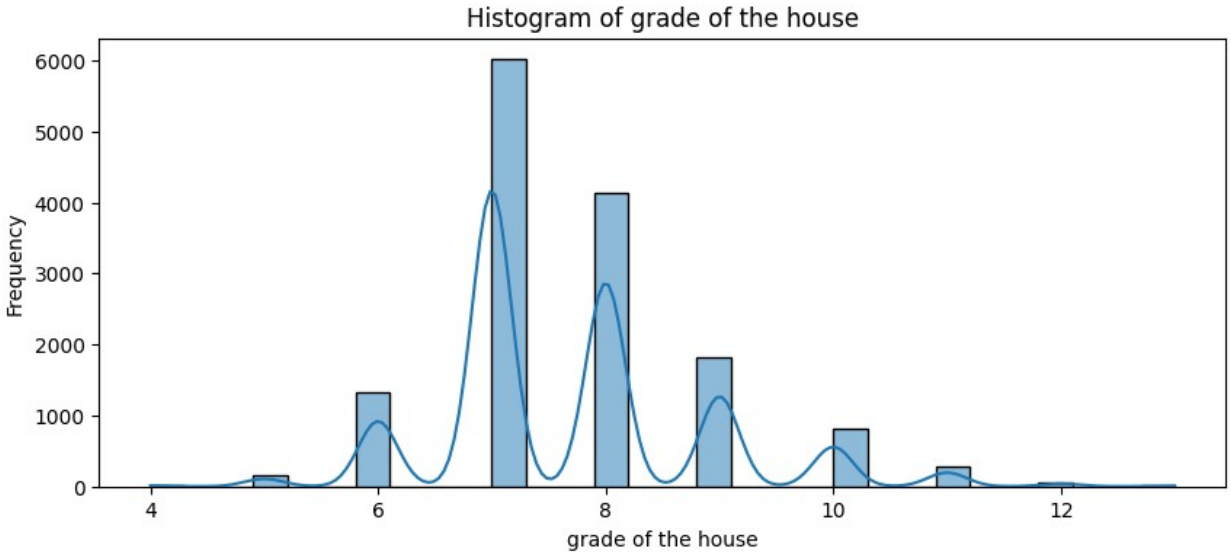

Histogram of lot area
Frequency
25000
20000
15000
10000
5000
0
0.0
0.2
0.4
0.6
0.8
1.0
lot area
1e6

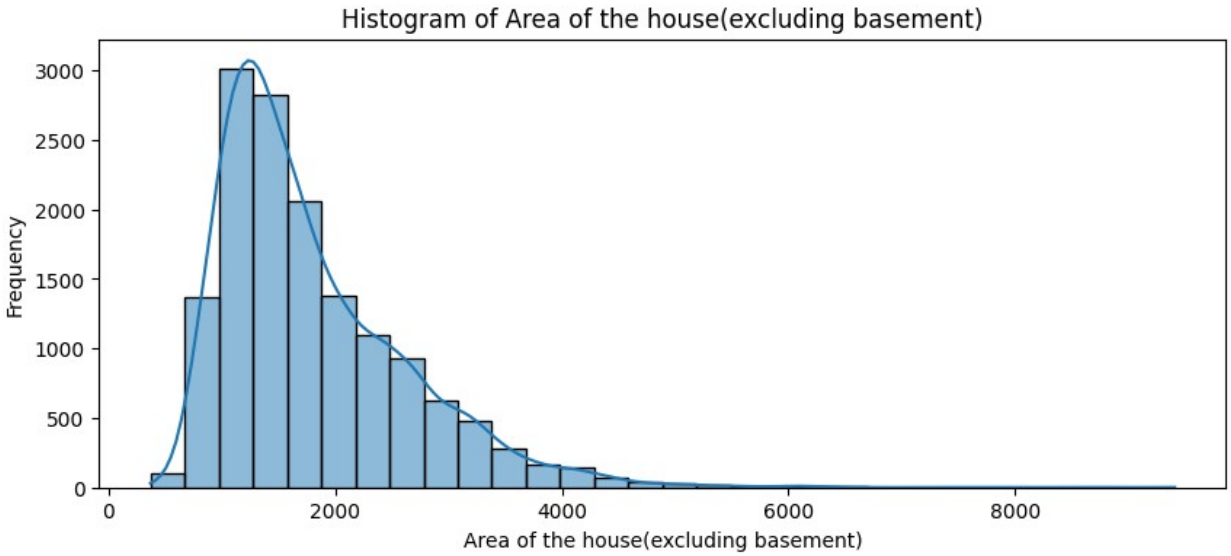

Histogram of number of floors
Frequency
7000
6000
5000
4000
3000
2000
1000
0
1.0
1.5
2.0
2.5
3.0
3.5
number of floors

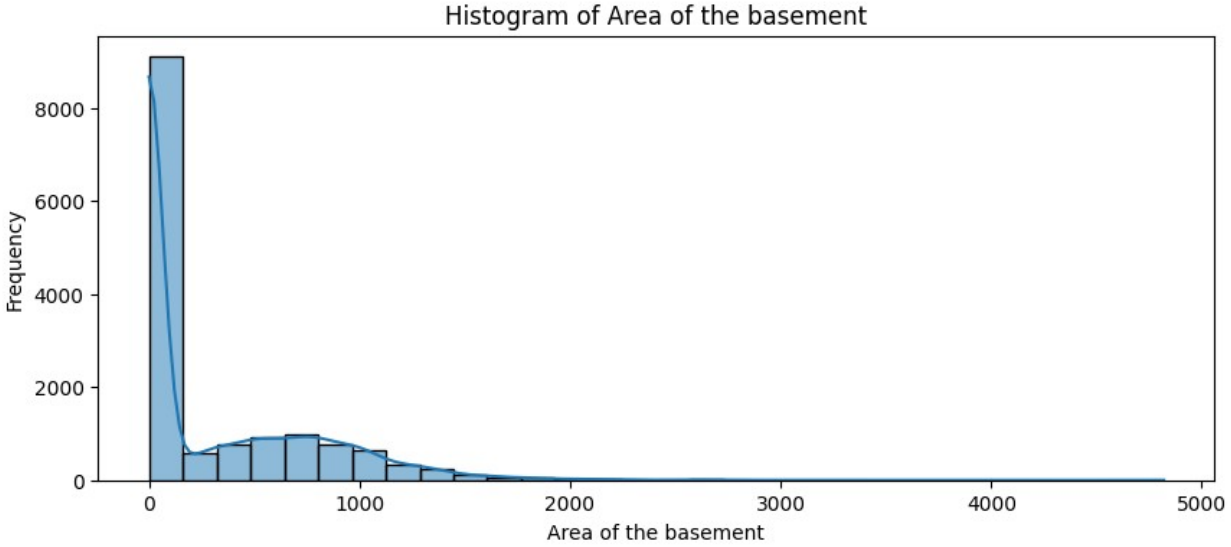

Histogram of waterfront present
Frequency
waterfront present
0
2000
4000
6000
8000
10000
12000
14000
0.0
0.2
0.4
0.6
0.8
1.0

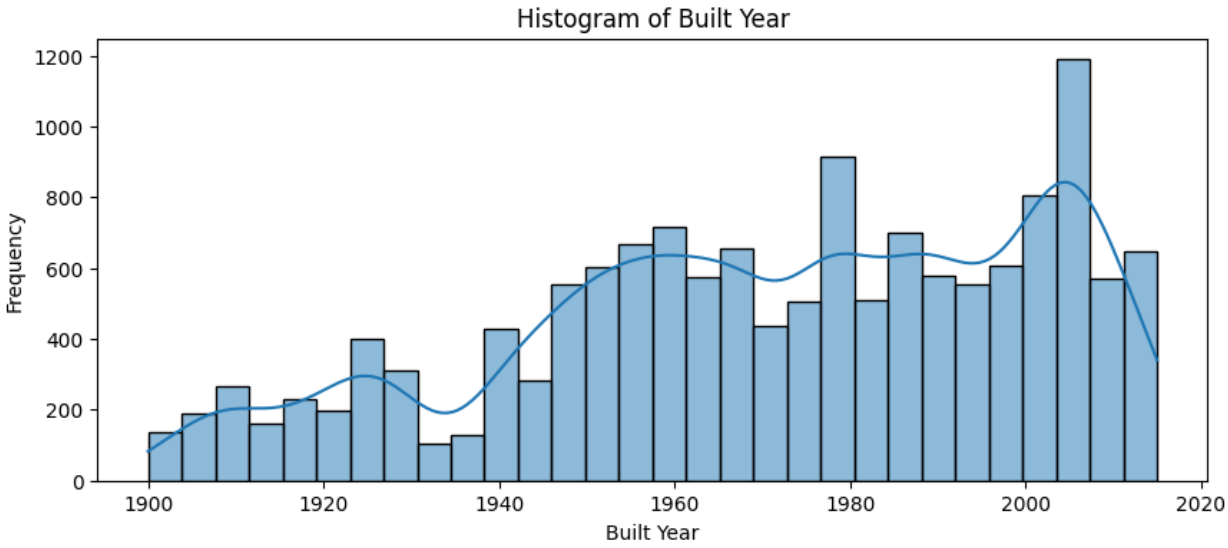

Histogram of number of views
Frequency
number of views
0
2000
4000
6000
8000
10000
12000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

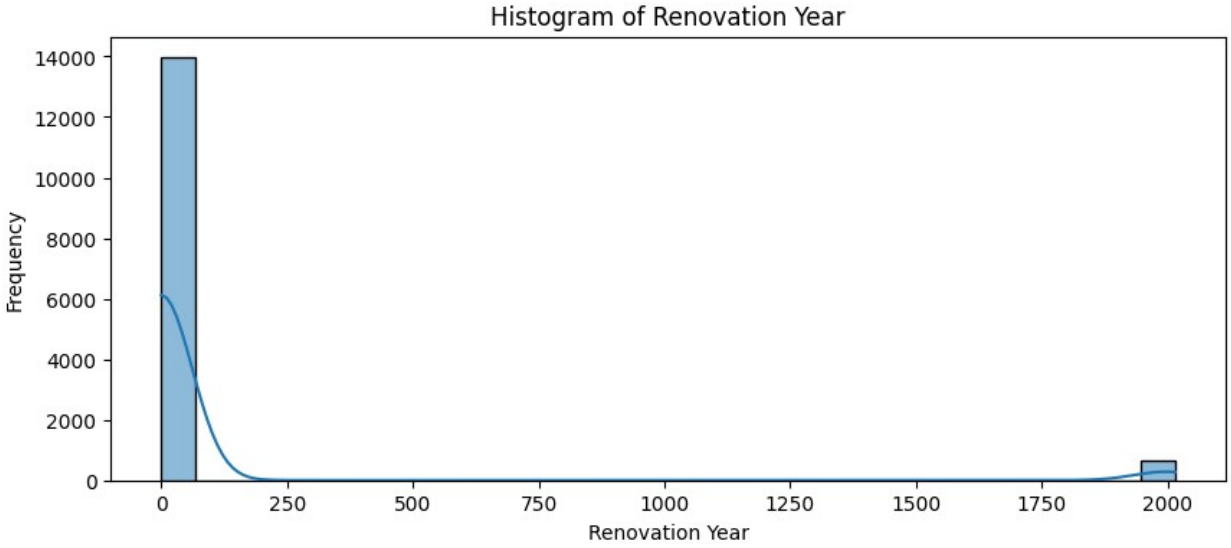

Histogram of condition of the house
Frequency
0
2000
4000
6000
8000
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
condition of the house

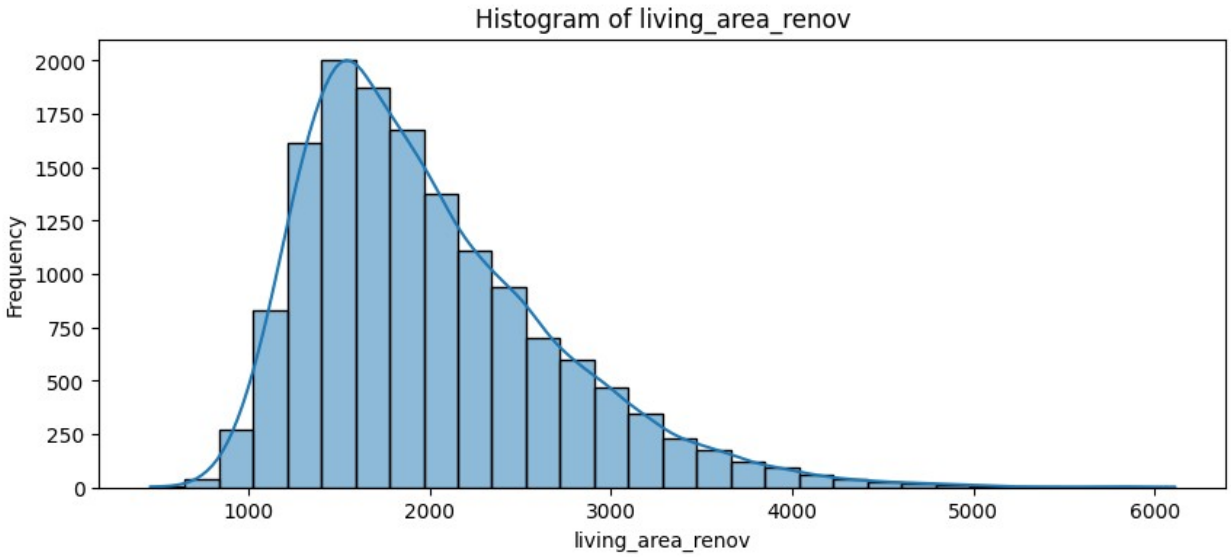

Histogram of grade of the house
Frequency
0
1000
2000
3000
4000
5000
6000
4
6
8
10
12
grade of the house

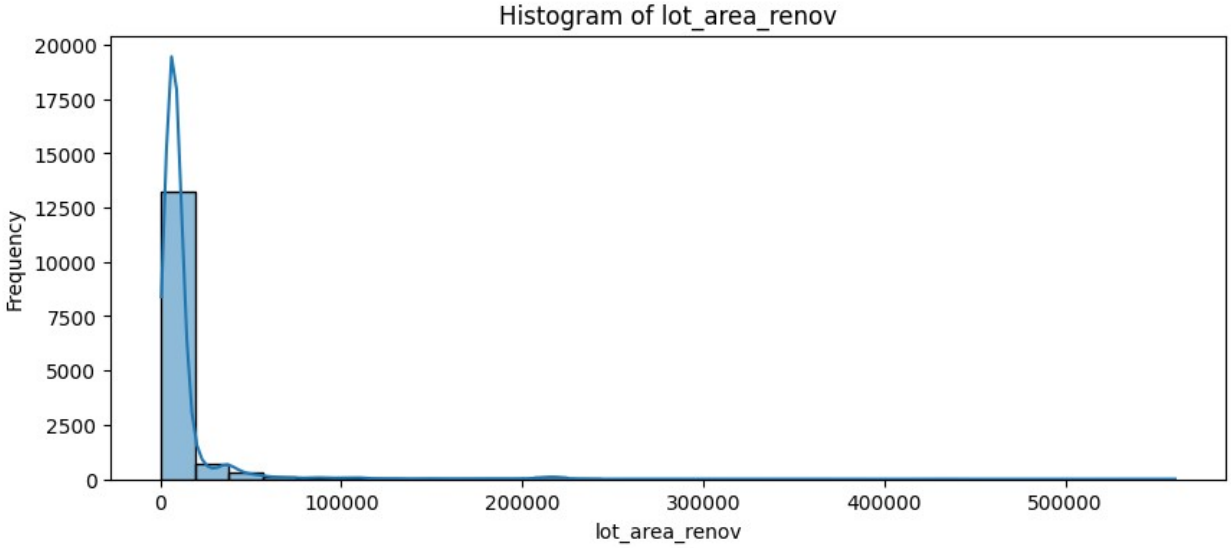

Histogram of Area of the house(excluding basement)
Frequency
Area of the house(excluding basement)
0
500
1000
1500
2000
2500
3000
0
2000
4000
6000
8000

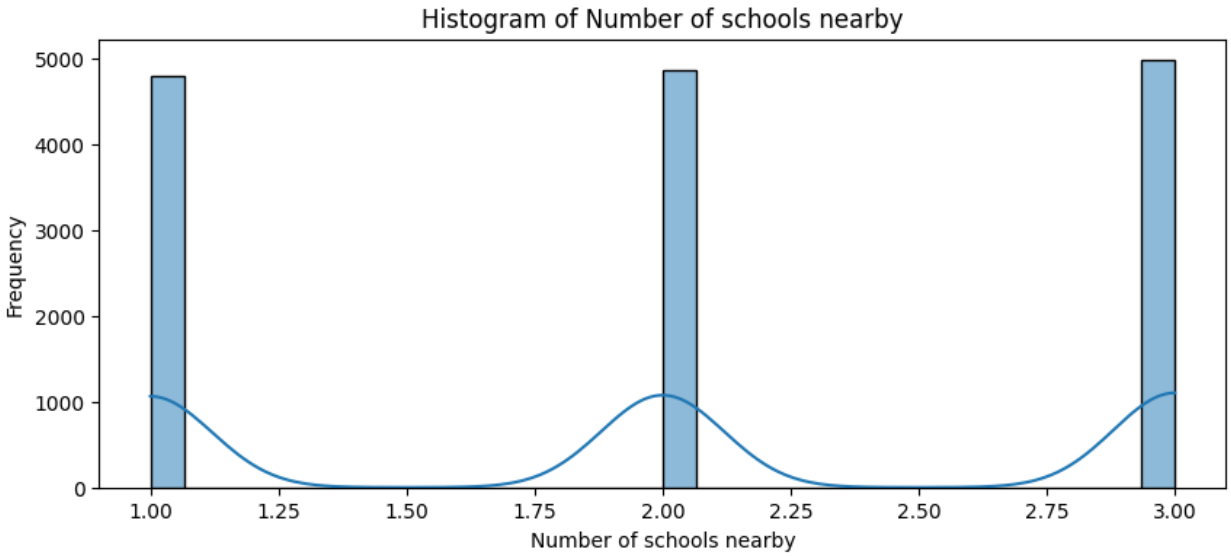

Histogram of Area of the basement
Frequency
Area of the basement
0
2000
4000
6000
8000
0
1000
2000
3000
4000
5000

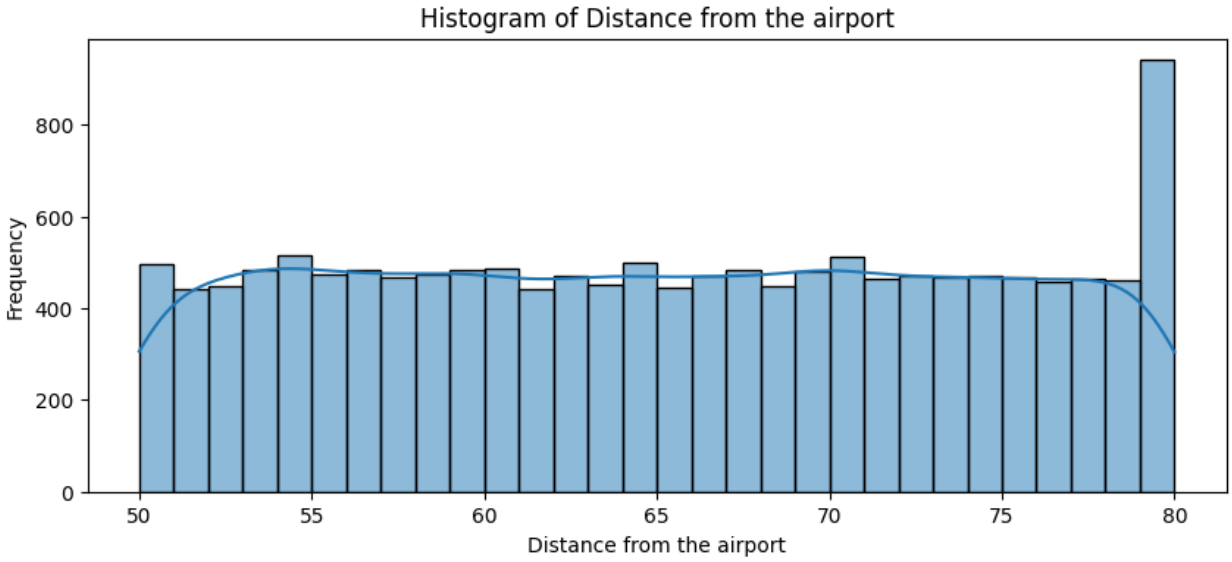

Histogram of Built Year
Frequency
Built Year
0
200
400
600
800
1000
1200
1900
1920
1940
1960
1980
2000
2020

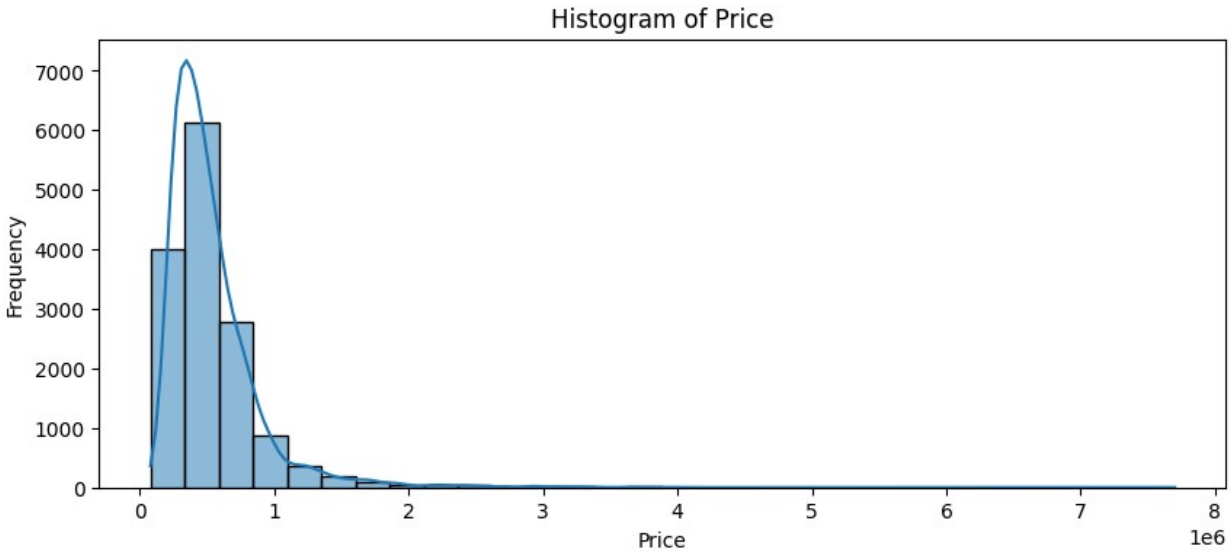

Histogram of Renovation Year
Frequency
Renovation Year
0
2000
4000
6000
8000
10000
12000
14000
0
250
500
750
1000
1250
1500
1750
2000


Histogram of living_area_renov
Frequency
living_area_renov
0
250
500
750
1000
1250
1500
1750
2000
1000
2000
3000
4000
5000
6000

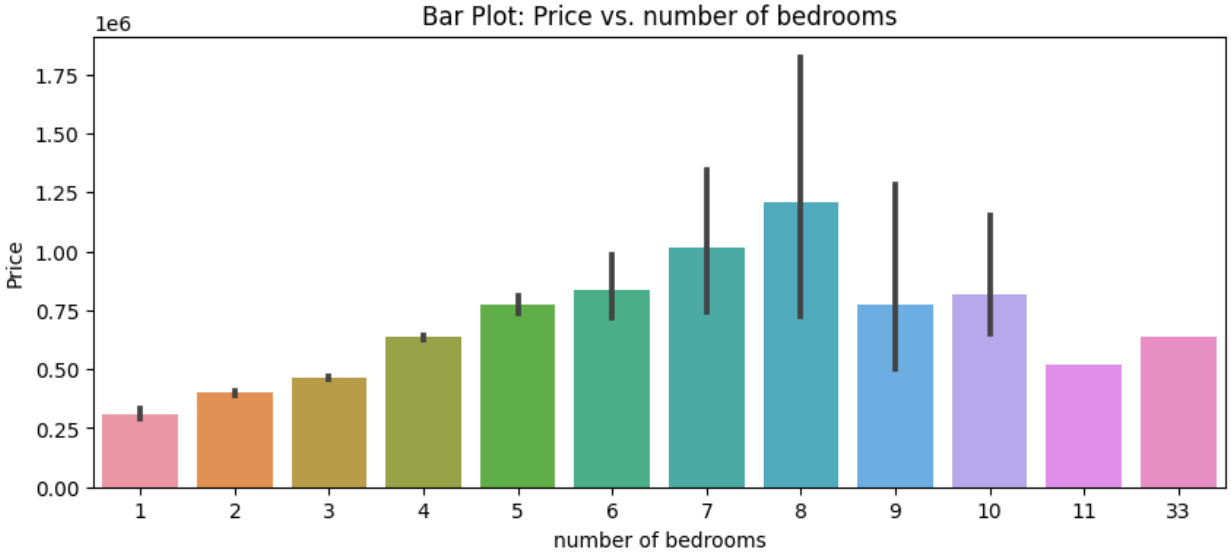

Histogram of lot_area_renov
Frequency
lot_area_renov
0
2500
5000
7500
10000
12500
15000
17500
20000
0
100000
200000
300000
400000
500000

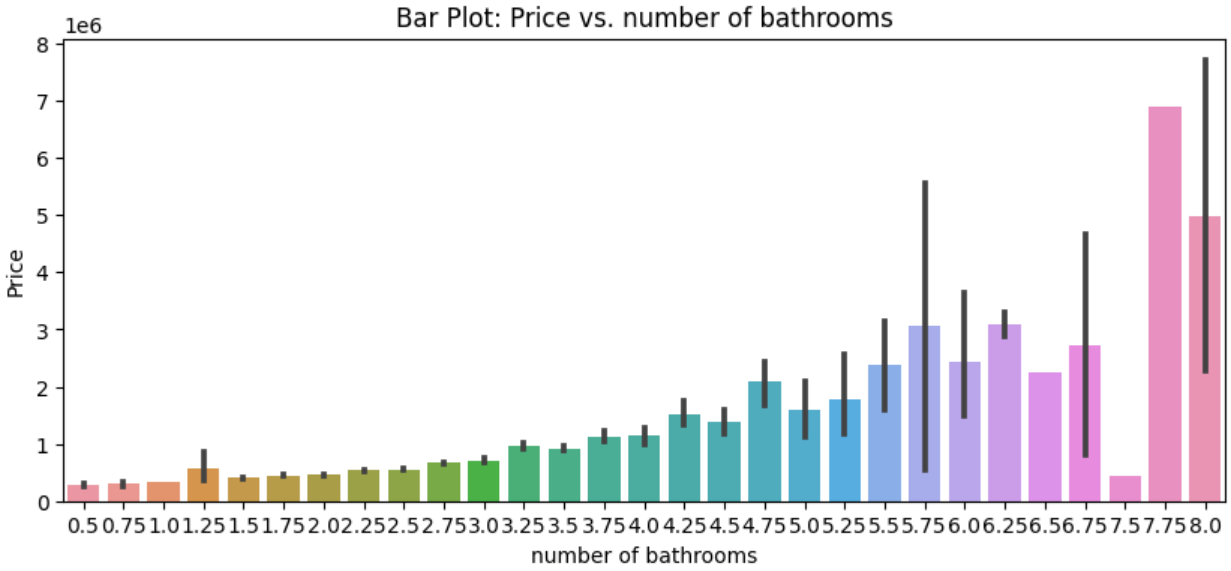

Histogram of Number of schools nearby
Frequency
0
1000
2000
3000
4000
5000
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
Number of schools nearby

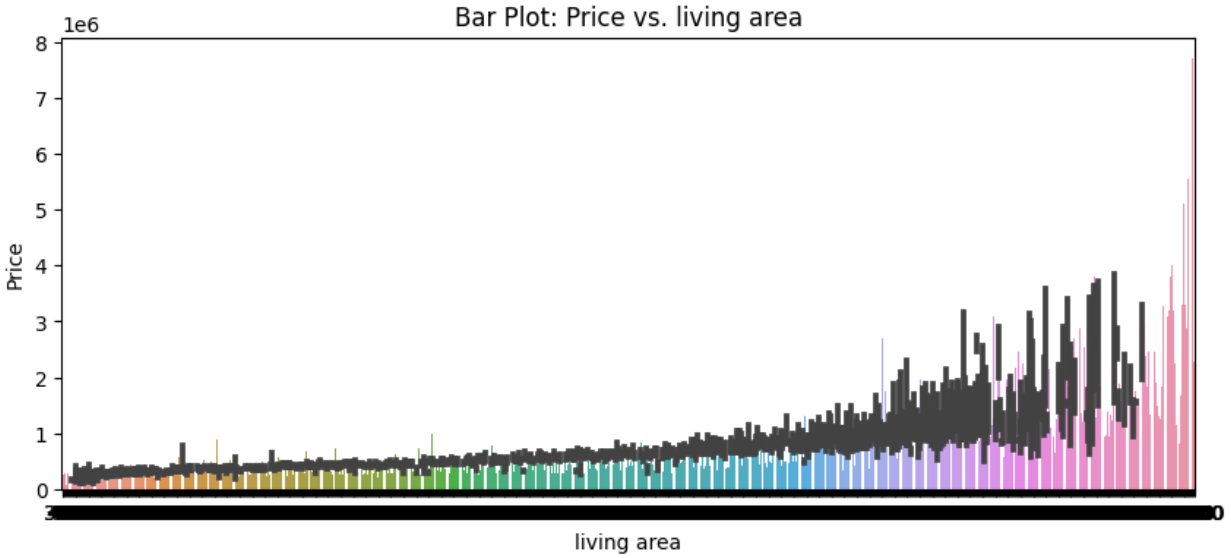

Histogram of Distance from the airport
Frequency
0
200
400
600
800
50
55
60
65
70
75
80
Distance from the airport

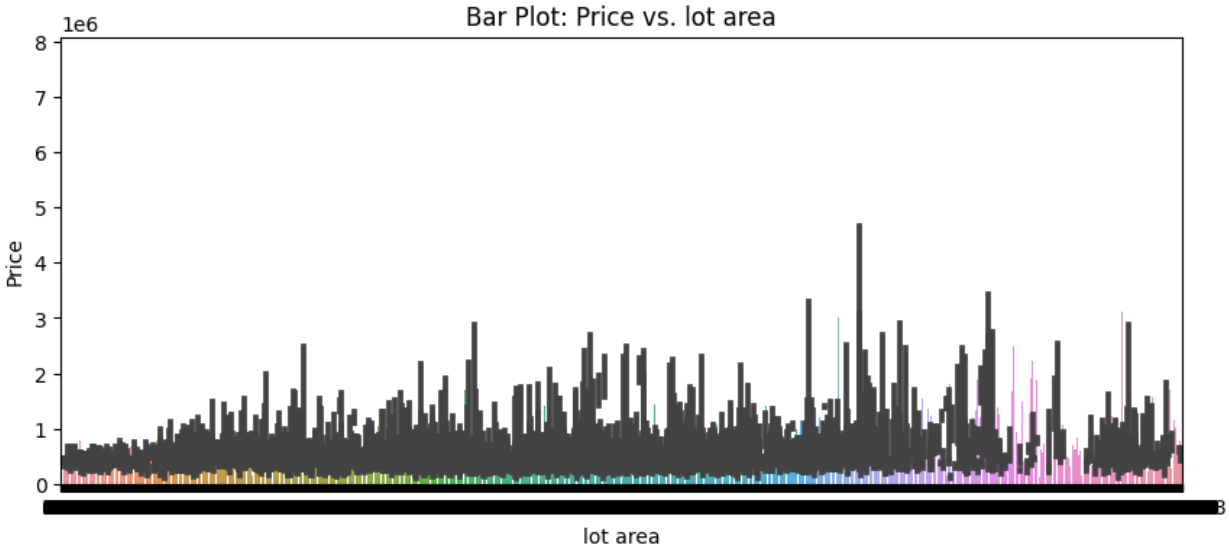


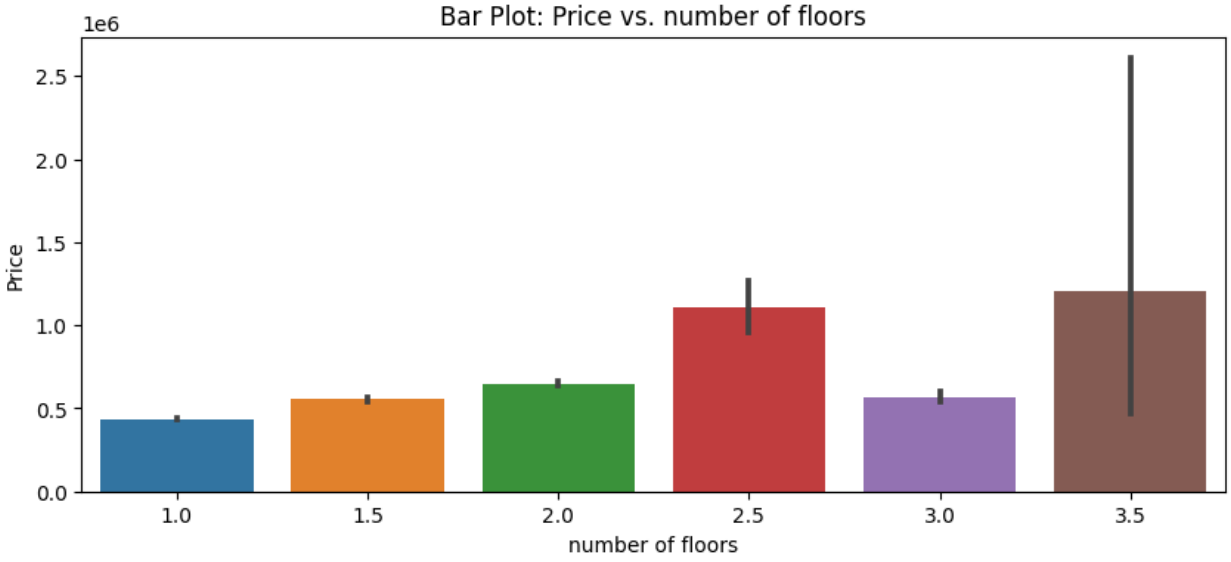

```
Task 3: Bivariate Analysis
```

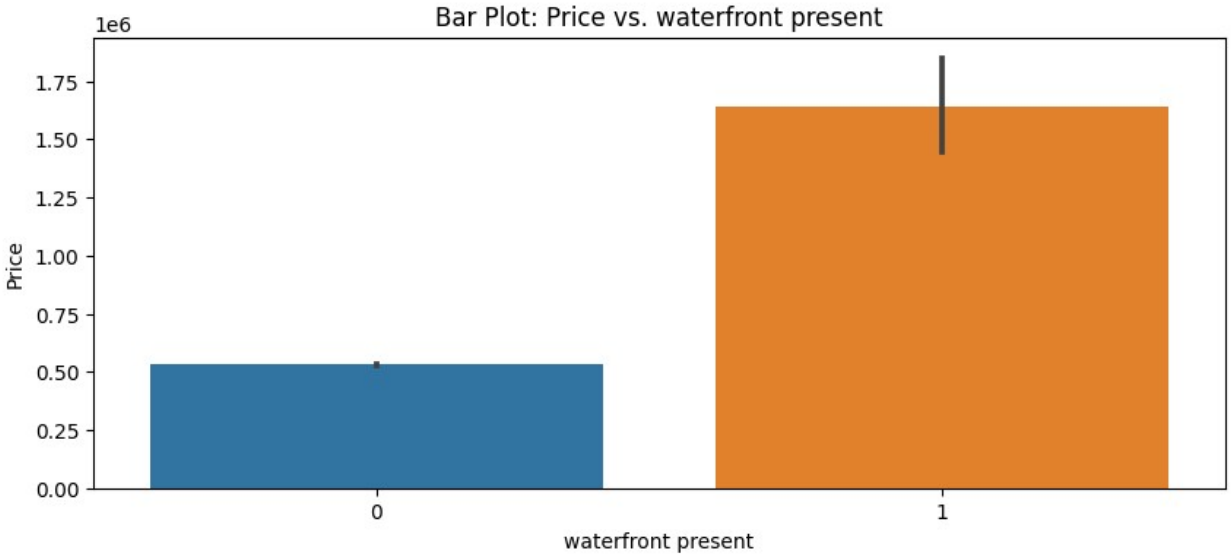


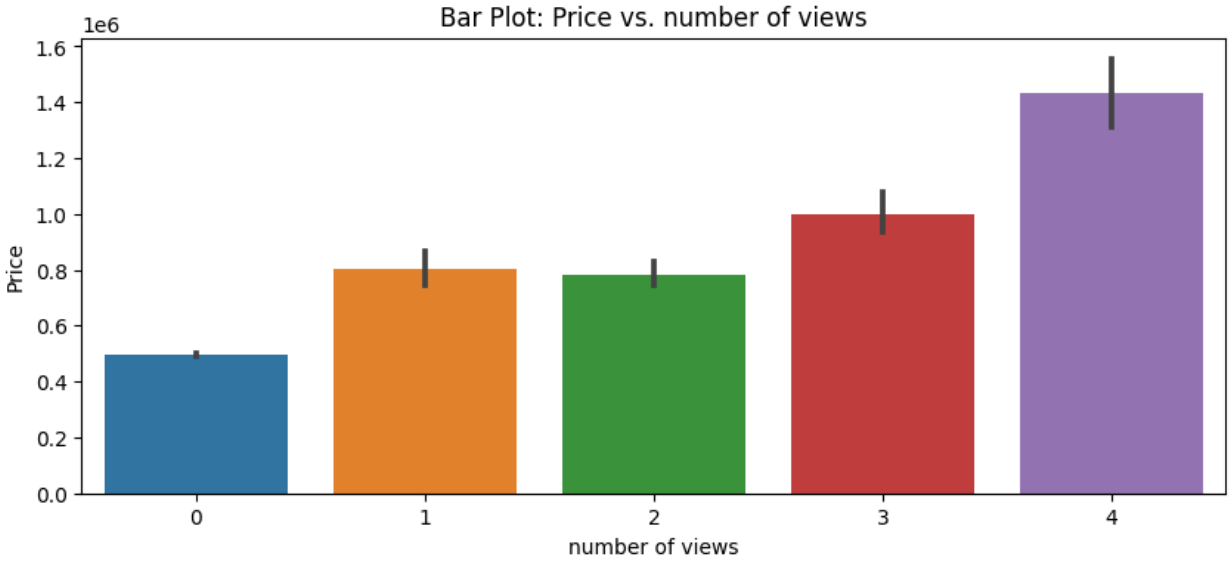

Bar Plot: Price vs. number of bathrooms
1e6
Price
number of bathrooms
0.5 0.75 1.0 1.25 1.5 1.75 2.0 2.25 2.5 2.75 3.0 3.25 3.5 3.75 4.0 4.25 4.5 4.75 5.0 5.25 5.5 5.75 6.0 6.25 6.5 6.75 7.5 7.75 8.0

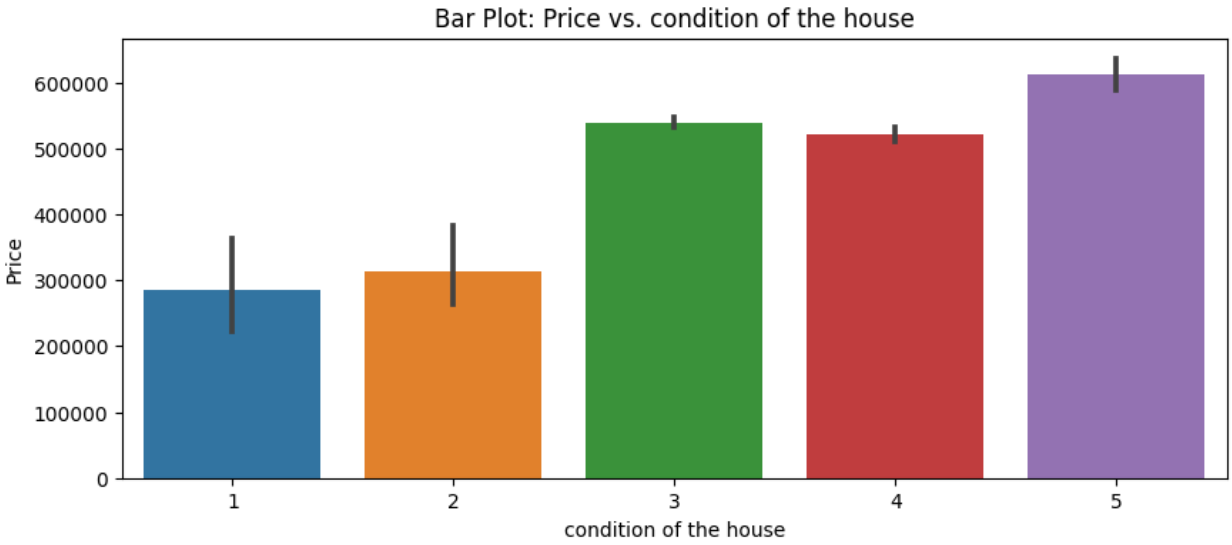

Bar Plot: Price vs. living area
1e6
Price
living area

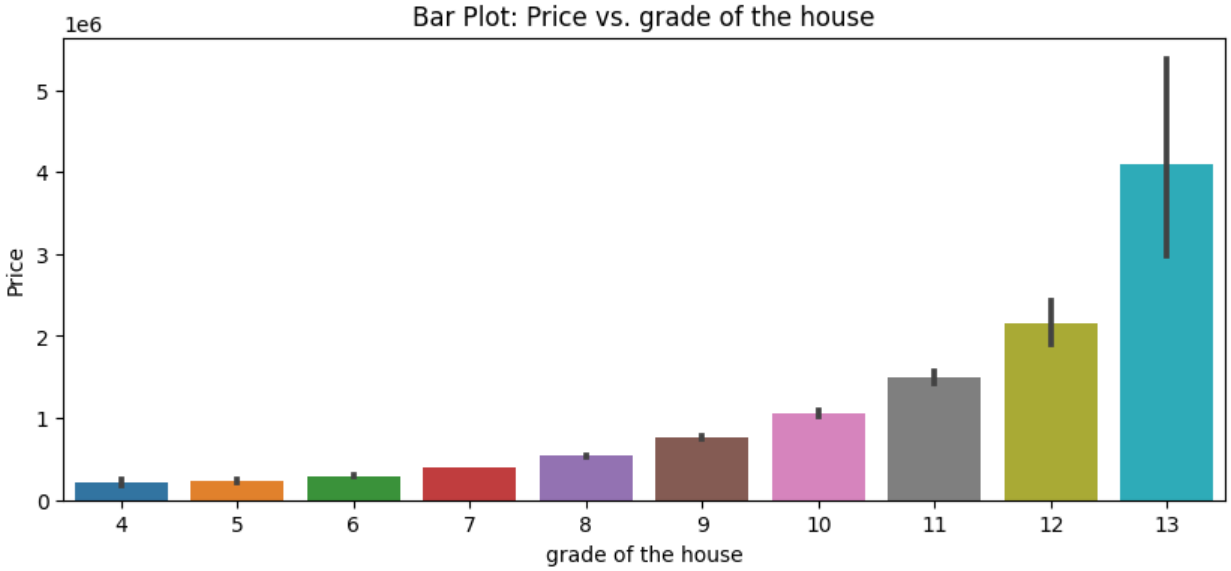

Bar Plot: Price vs. lot area
1e6
8
7
6
5
4
3
2
1
0
Price
lot area

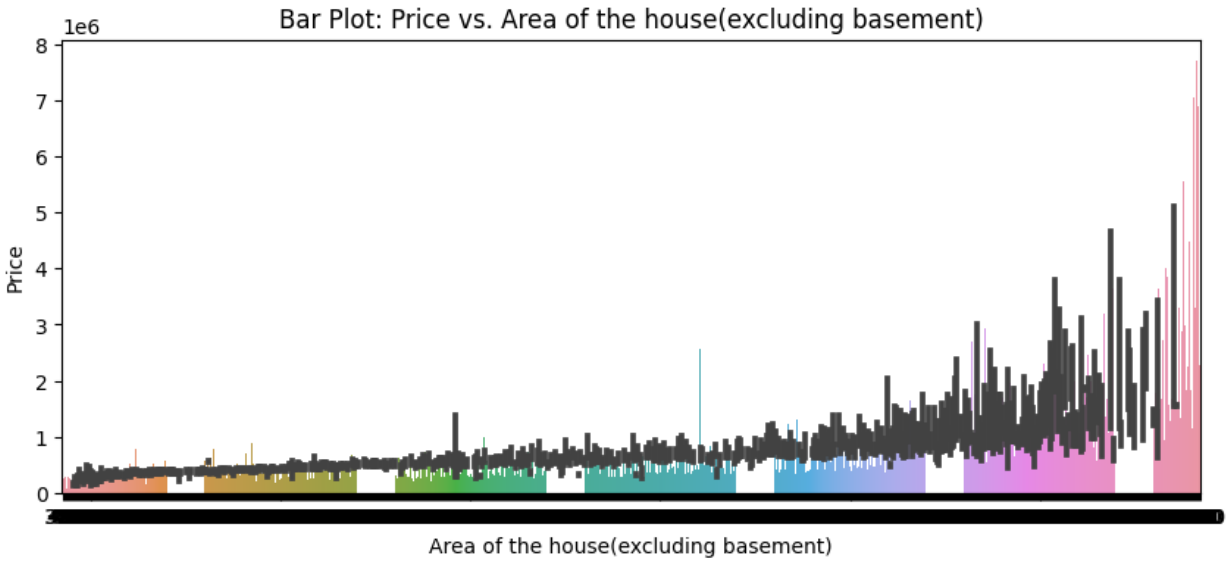

Bar Plot: Price vs. number of floors
1e6
2.5
2.0
1.5
1.0
0.5
0.0
Price
1.0
1.5
2.0
2.5
3.0
3.5
number of floors

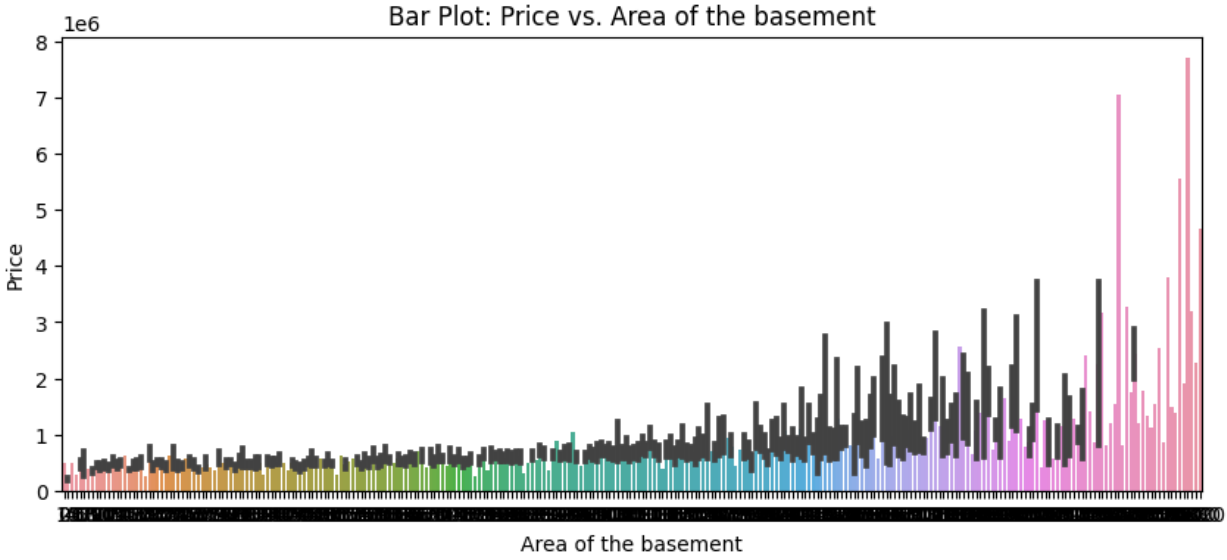

Bar Plot: Price vs. waterfront present
1e6
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Price
0
1
waterfront present

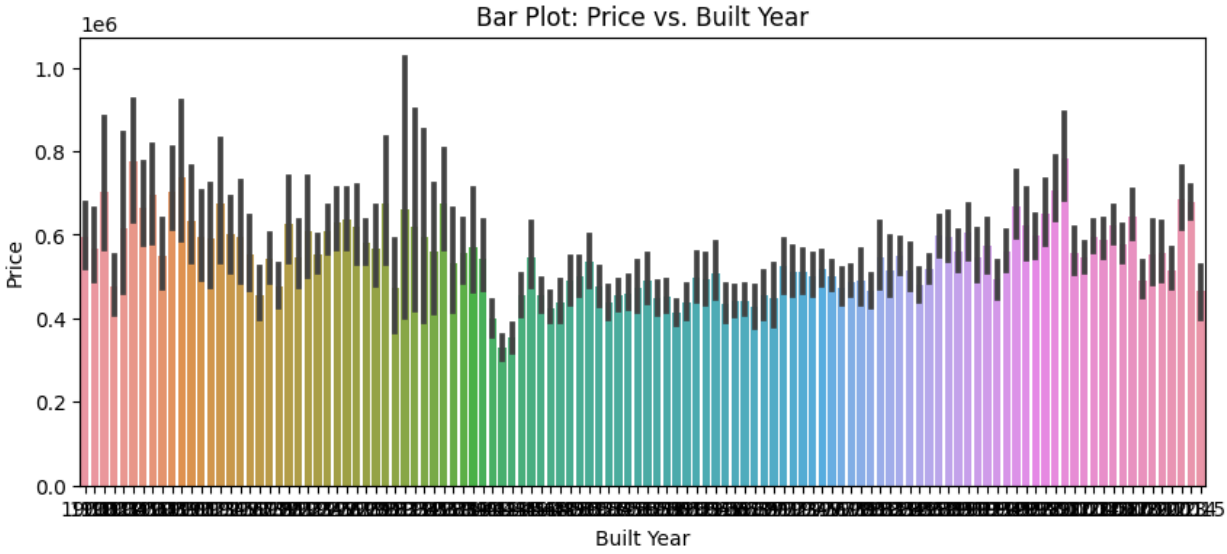

Bar Plot: Price vs. number of views
1e6
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Price
0
1
2
3
4
number of views

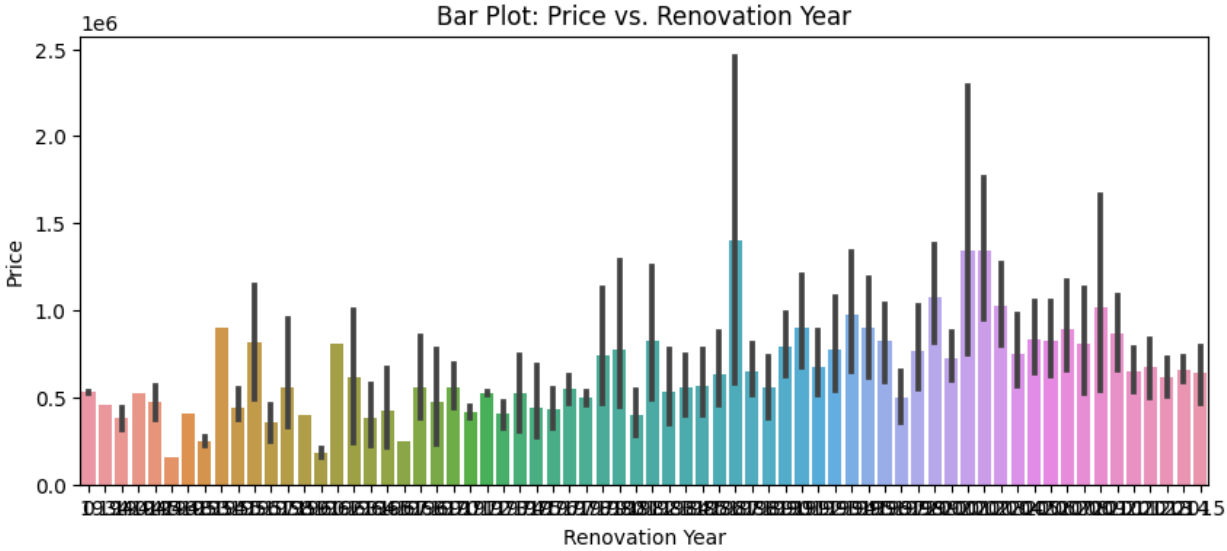

Bar Plot: Price vs. condition of the house
Price
600000
500000
400000
300000
200000
100000
0
1
2
3
4
5
condition of the house

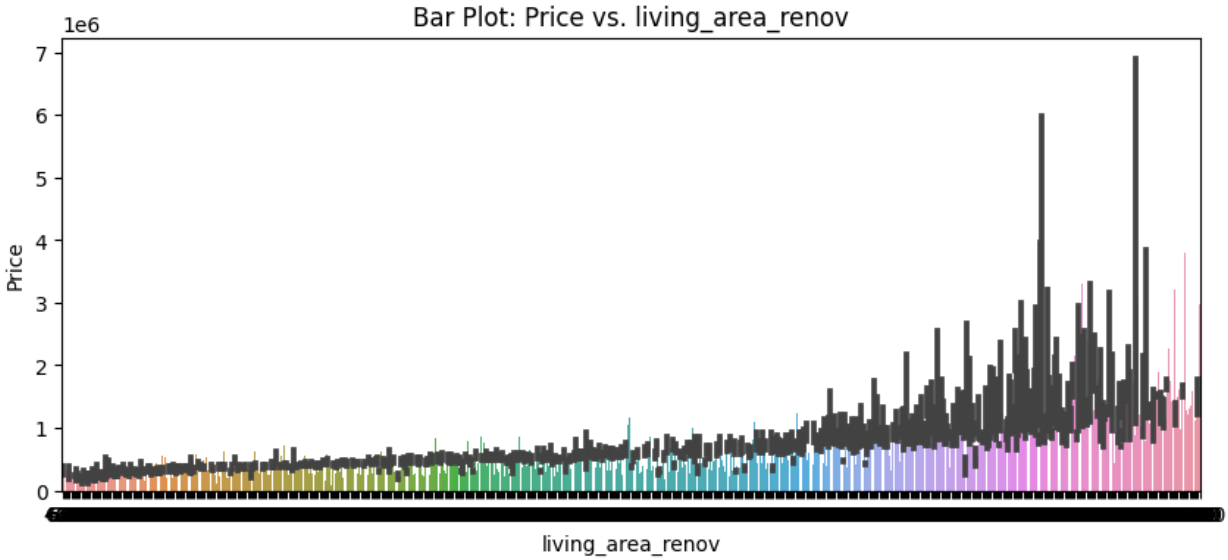

Bar Plot: Price vs. grade of the house
1e6
Price
5
4
3
2
1
0
4
5
6
7
8
9
10
11
12
13
grade of the house

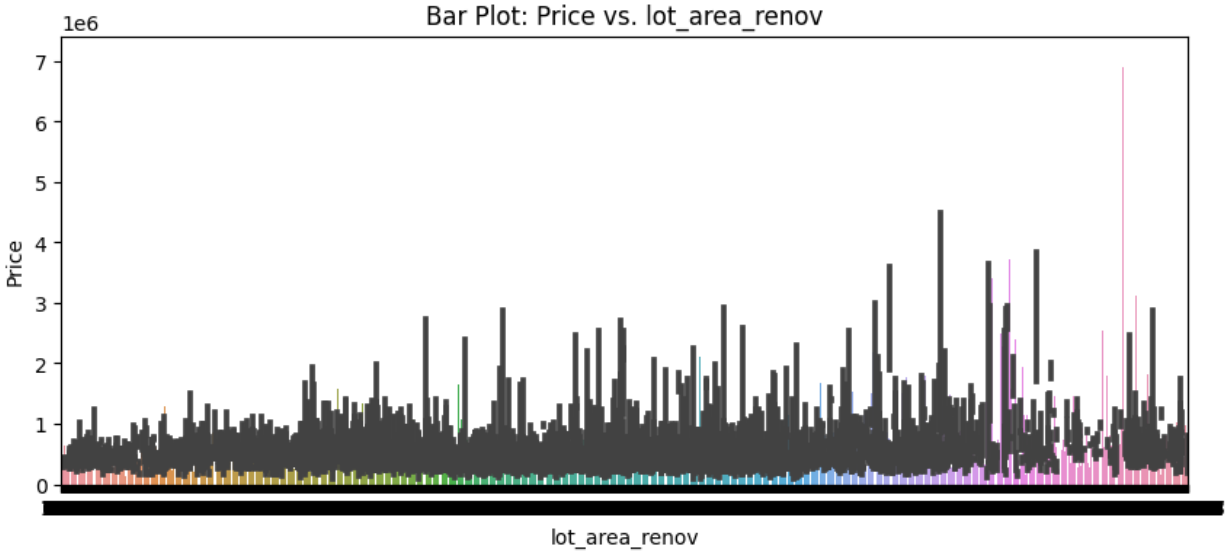

Bar Plot: Price vs. Area of the house(excluding basement)
1e6
8
7
6
5
4
3
2
1
0
Price
Area of the house(excluding basement)

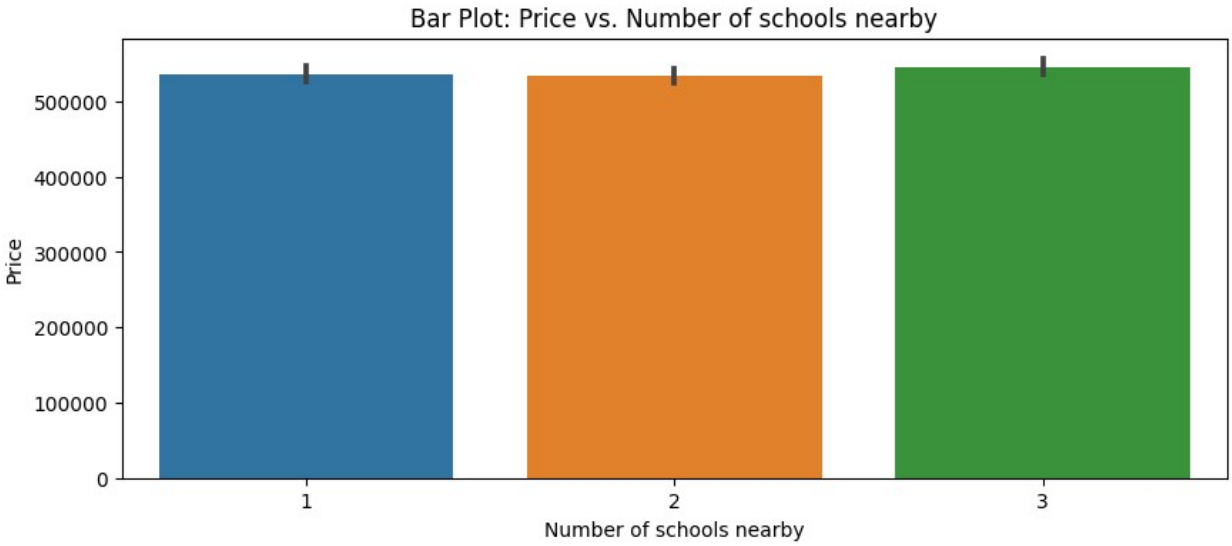

Bar Plot: Price vs. Area of the basement
1e6
8
7
6
5
4
3
2
1
0
Price
Area of the basement

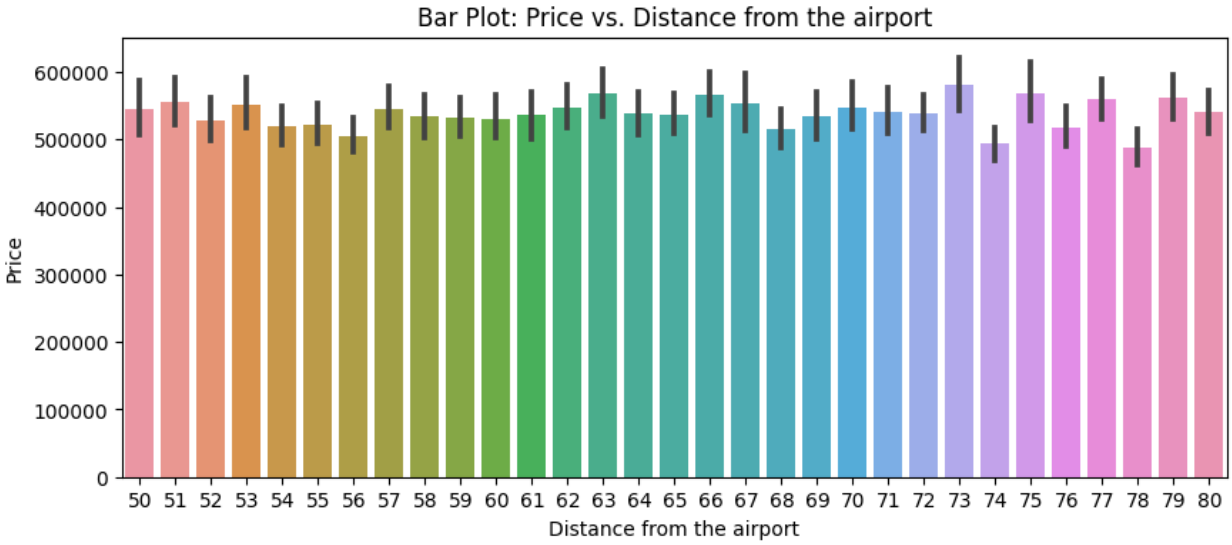

Bar Plot: Price vs. Built Year
1e6
1.0
0.8
0.6
0.4
0.2
0.0
Price
Built Year

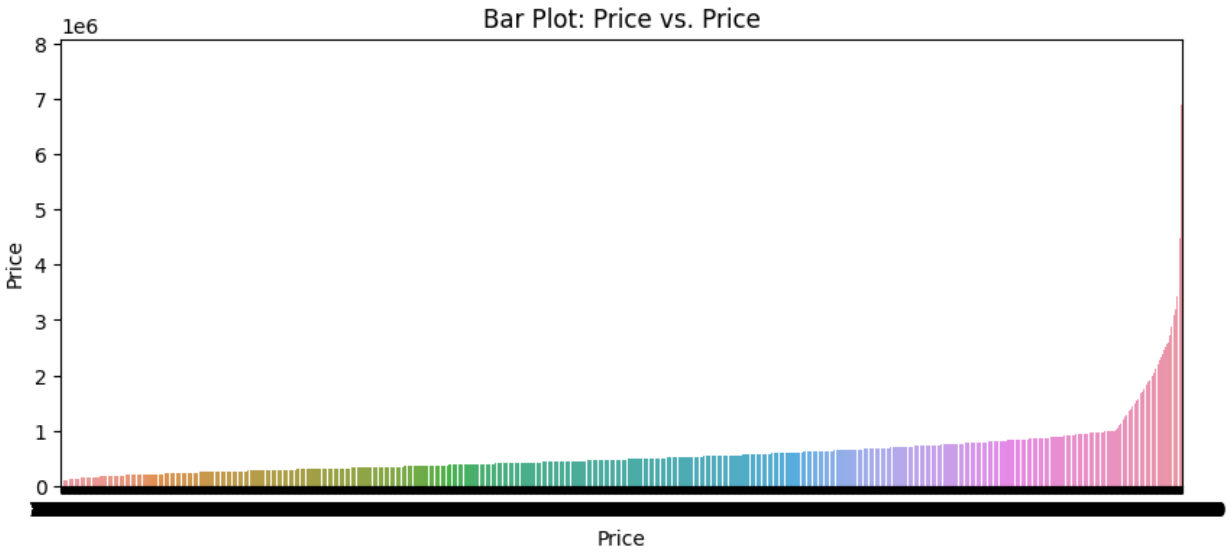

Bar Plot: Price vs. Renovation Year
1e6
2.5
2.0
1.5
1.0
0.5
0.0
Price
Renovation Year


Bar Plot: Price vs. living_area_renov
1e6
7
6
5
4
3
2
1
0
Price
living_area_renov

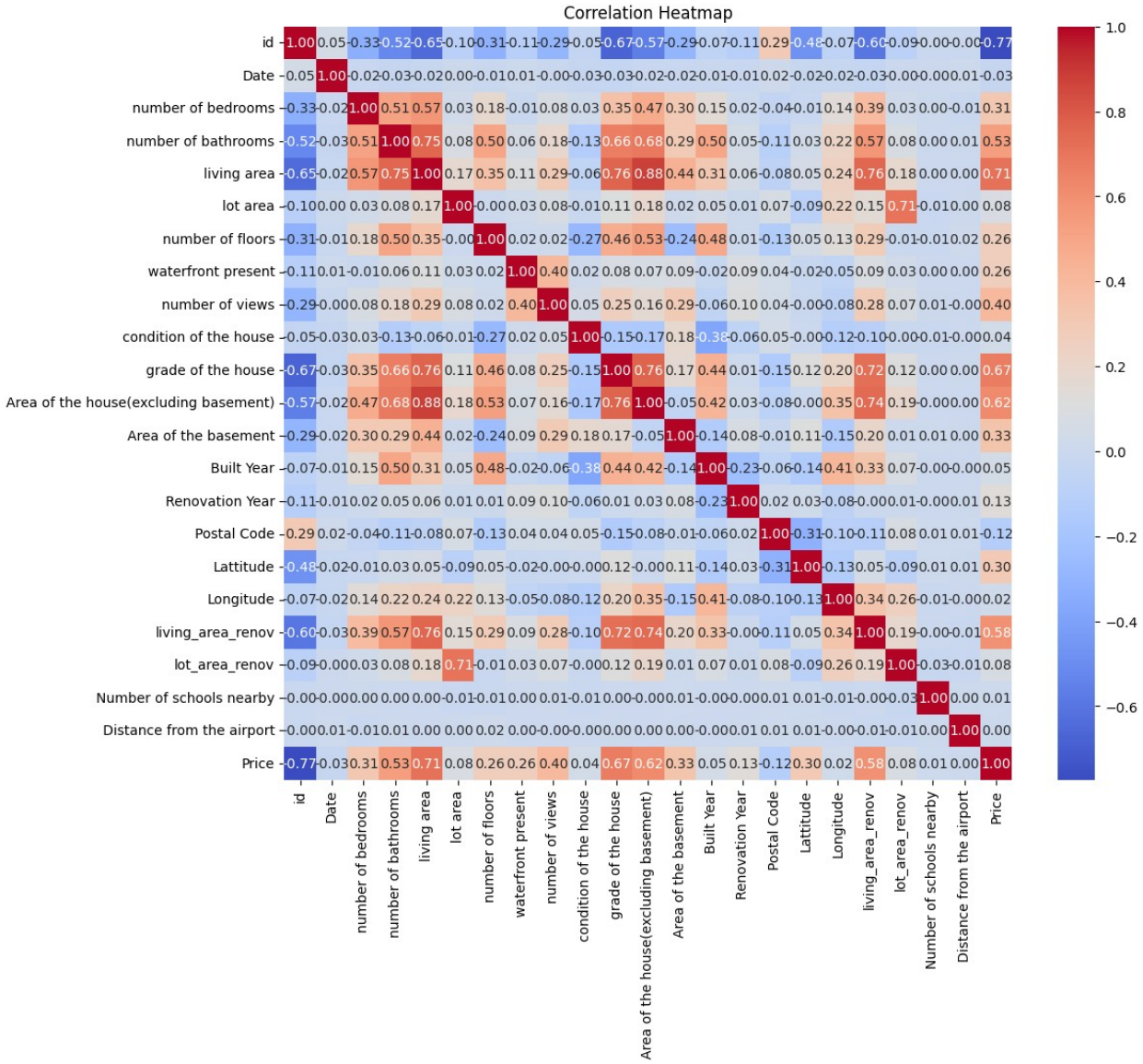

Bar Plot: Price vs. lot_area_renov
1e6
7
6
5
4
3
2
1
0
Price
lot_area_renov


Bar Plot: Price vs. Number of schools nearby
Price
0
100000
200000
300000
400000
500000
1
2
3
Number of schools nearby


Bar Plot: Price vs. Distance from the airport
Price
0
100000
200000
300000
400000
500000
600000
50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
Distance from the airport




```
Task 4: Multivariate Analysis
```




```
Task 5: Descriptive statistics of 'df':
                 id          Date  number of bedrooms  number of
bathrooms
count  1.462000e+04  14620.000000        14620.000000
14620.000000  \
mean   6.762821e+09  42604.538646            3.379343
2.129583
std    6.237575e+03     67.347991            0.938719
0.769934
min    6.762810e+09  42491.000000            1.000000
0.500000
```

```
25%    6.762815e+09  42546.000000            3.000000
1.750000
50%    6.762821e+09  42600.000000            3.000000
2.250000
75%    6.762826e+09  42662.000000            4.000000
2.500000
max    6.762832e+09  42734.000000           33.000000
8.000000

        living area      lot area  number of floors  waterfront
present
count  14620.000000  1.462000e+04      14620.000000
14620.000000  \
mean    2098.262996  1.509328e+04          1.502360
0.007661
std      928.275721  3.791962e+04          0.540239
0.087193
min      370.000000  5.200000e+02          1.000000
0.000000
25%     1440.000000  5.010750e+03          1.000000
0.000000
50%     1930.000000  7.620000e+03          1.500000
0.000000
75%     2570.000000  1.080000e+04          2.000000
0.000000
max    13540.000000  1.074218e+06          3.500000
1.000000

       number of views  condition of the house  ...    Built Year
count     14620.000000            14620.000000  ...  14620.000000  \
mean          0.233105                3.430506  ...   1970.926402
std           0.766259                0.664151  ...     29.493625
min           0.000000                1.000000  ...   1900.000000
25%           0.000000                3.000000  ...   1951.000000
50%           0.000000                3.000000  ...   1975.000000
75%           0.000000                4.000000  ...   1997.000000
max           4.000000                5.000000  ...   2015.000000

       Renovation Year    Postal Code     Lattitude     Longitude
count     14620.000000   14620.000000  14620.000000  14620.000000  \
mean         90.924008  122033.062244     52.792848   -114.404007
std         416.216661      19.082418      0.137522      0.141326
min           0.000000  122003.000000     52.385900   -114.709000
25%           0.000000  122017.000000     52.707600   -114.519000
50%           0.000000  122032.000000     52.806400   -114.421000
75%           0.000000  122048.000000     52.908900   -114.315000
max        2015.000000  122072.000000     53.007600   -113.505000

       living_area_renov  lot_area_renov  Number of schools nearby
count       14620.000000    14620.000000              14620.000000  \
```

```
mean          1996.702257    12753.500068                    2.012244
std            691.093366    26058.414467                    0.817284
min            460.000000      651.000000                    1.000000
25%           1490.000000     5097.750000                    1.000000
50%           1850.000000     7620.000000                    2.000000
75%           2380.000000    10125.000000                    3.000000
max           6110.000000   560617.000000                    3.000000

       Distance from the airport         Price
count                14620.000000  1.462000e+04
mean                    64.950958  5.389322e+05
std                      8.936008  3.675324e+05
min                     50.000000  7.800000e+04
25%                     57.000000  3.200000e+05
50%                     65.000000  4.500000e+05
75%                     73.000000  6.450000e+05
max                     80.000000  7.700000e+06

[8 rows x 23 columns]



Task 6: Handle Missing Values
Missing values after filling:
id                                      0
Date                                    0
number of bedrooms                      0
number of bathrooms                     0
living area                             0
lot area                                0
number of floors                        0
waterfront present                      0
number of views                         0
condition of the house                  0
grade of the house                      0
Area of the house(excluding basement)   0
Area of the basement                    0
Built Year                              0
Renovation Year                         0
Postal Code                             0
Lattitude                               0
Longitude                               0
living_area_renov                       0
lot_area_renov                          0
Number of schools nearby                0
Distance from the airport               0
Price                                   0
dtype: int64
```