

▾ Assignment - 2

Pranav Balaji R S - 21BAI1893

VIT -CHENNAI

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import rcParams
import seaborn as sn

data = pd.read_csv("/content/House Price India.csv")

df = pd.DataFrame(data)

df
```

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	..
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	..
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	..
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	..
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	..
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	..
...	...	...	...	...	...	...	...	...	...	...	..
14615	6762830250	42734	2	1.50	1556	20000	1.0	0	0	4	..
14616	6762830339	42734	3	2.00	1680	7000	1.5	0	0	4	..
14617	6762830618	42734	2	1.00	1070	6120	1.0	0	0	3	..
14618	6762830709	42734	4	1.00	1030	6621	1.0	0	0	4	..
14619	6762831463	42734	3	1.00	900	4770	1.0	0	0	3	..

14620 rows × 23 columns

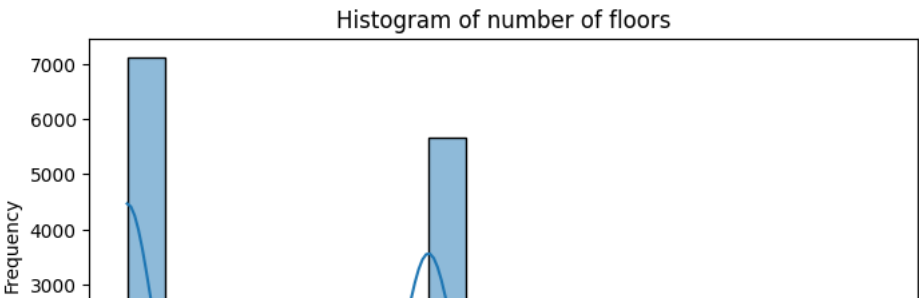
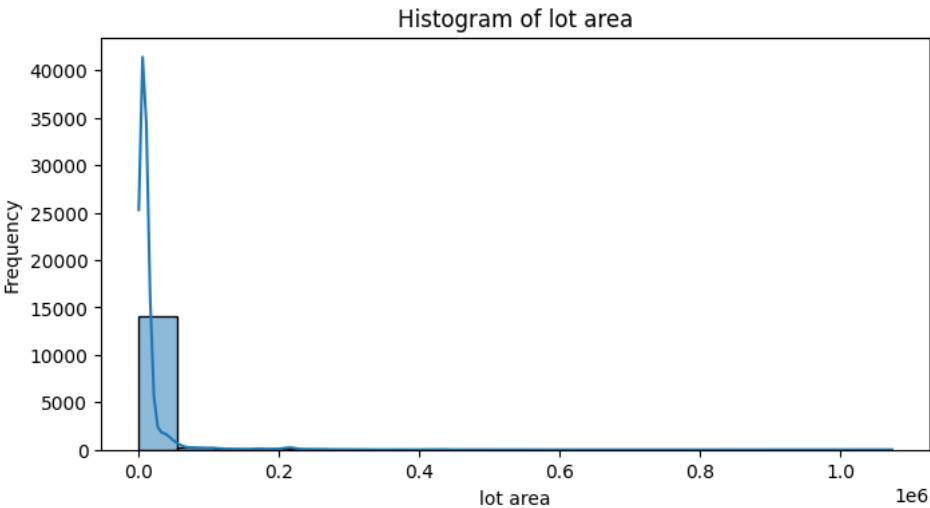
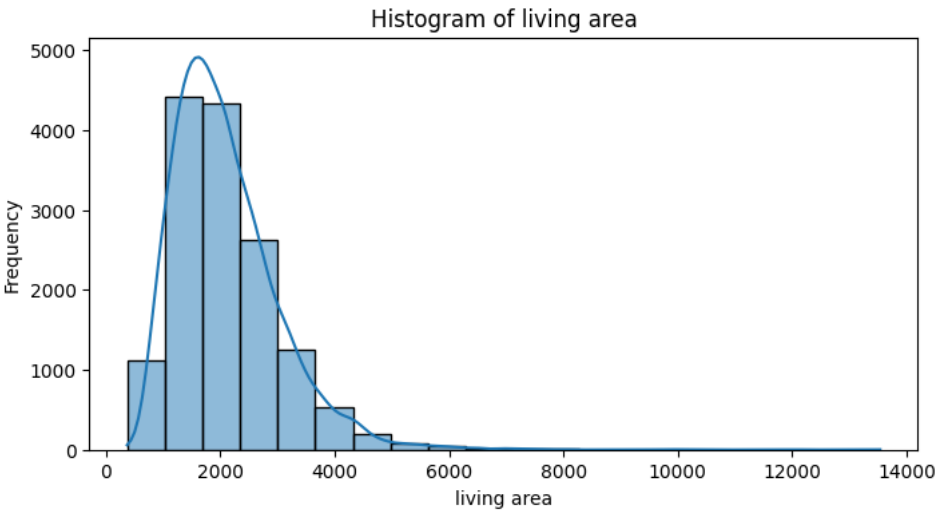
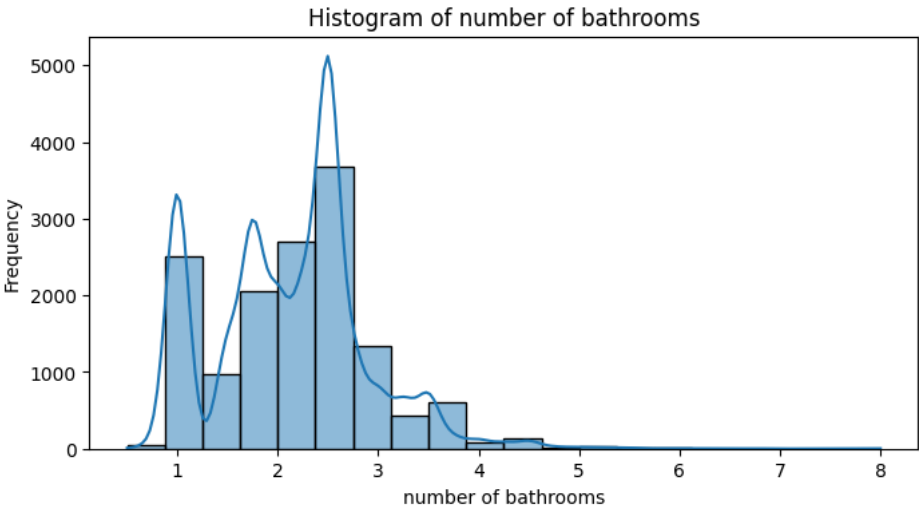
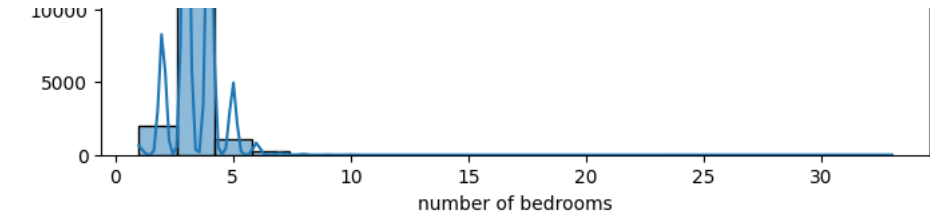
```
df.info()
```

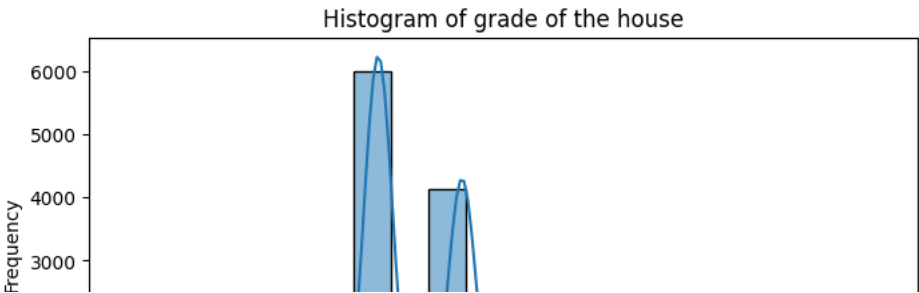
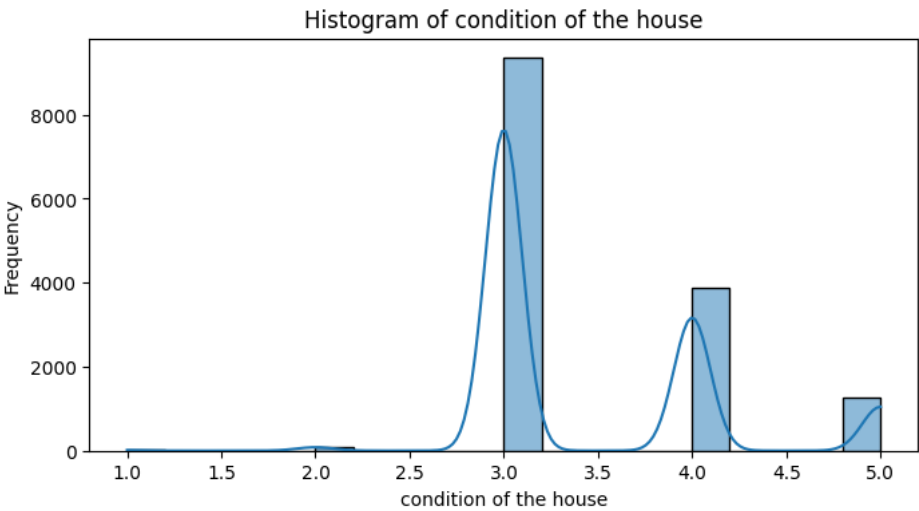
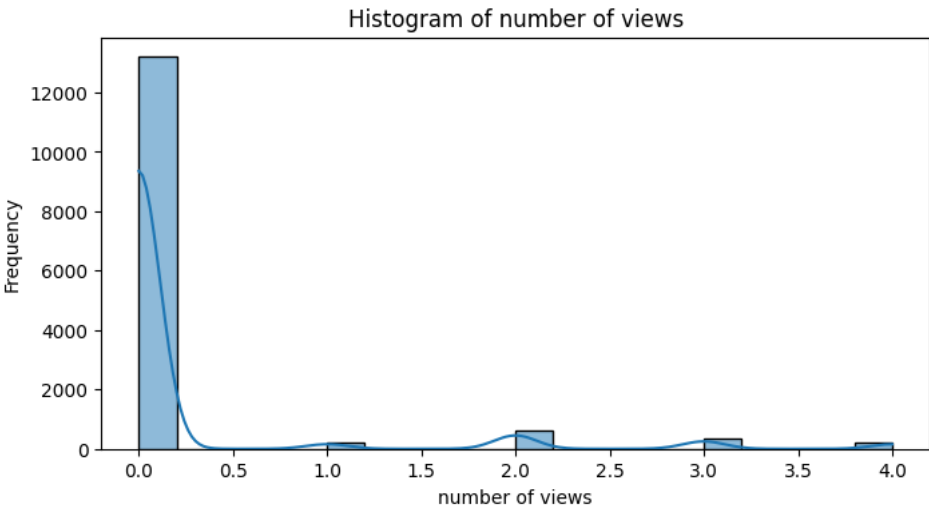
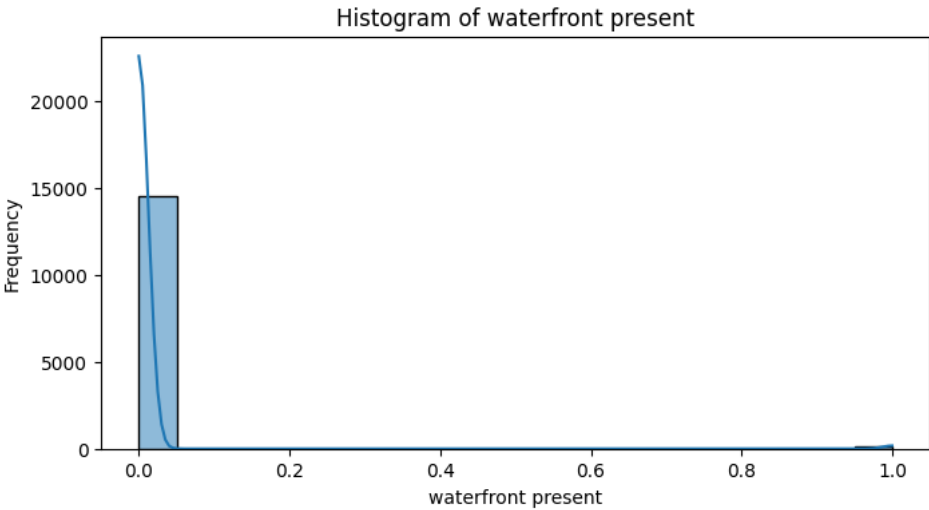
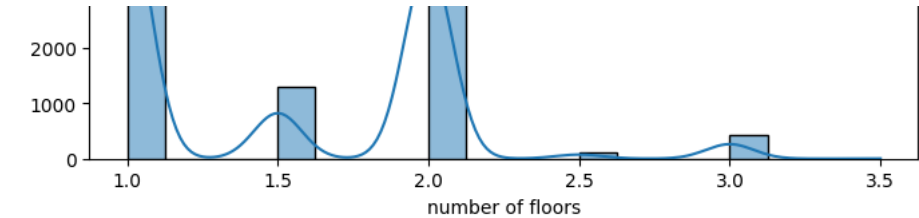
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14620 entries, 0 to 14619
Data columns (total 23 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   id                                         14620 non-null  int64
1   Date                                       14620 non-null  int64
2   number of bedrooms                       14620 non-null  int64
3   number of bathrooms                      14620 non-null  float64
4   living area                              14620 non-null  int64
5   lot area                                 14620 non-null  int64
6   number of floors                         14620 non-null  float64
7   waterfront present                       14620 non-null  int64
8   number of views                          14620 non-null  int64
9   condition of the house                   14620 non-null  int64
10  grade of the house                       14620 non-null  int64
11  Area of the house(excluding basement)    14620 non-null  int64
12  Area of the basement                     14620 non-null  int64
13  Built Year                               14620 non-null  int64
14  Renovation Year                           14620 non-null  int64
15  Postal Code                              14620 non-null  int64
16  Lattitude                                14620 non-null  float64
17  Longitude                                14620 non-null  float64
18  living_area_renov                         14620 non-null  int64
19  lot_area_renov                           14620 non-null  int64
20  Number of schools nearby                  14620 non-null  int64
21  Distance from the airport                 14620 non-null  int64
22  Price                                     14620 non-null  int64
dtypes: float64(4), int64(19)
memory usage: 2.6 MB
```

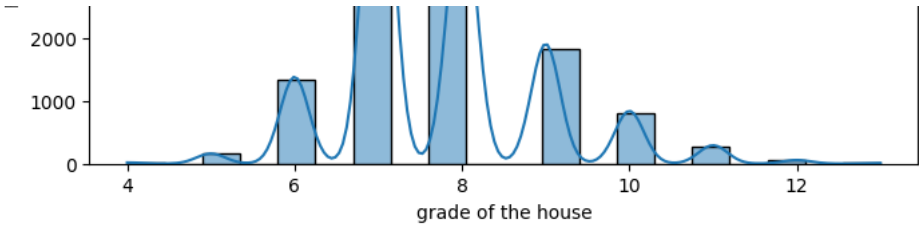
```
columns = df.columns
```

## ▼ Univariate Analysis

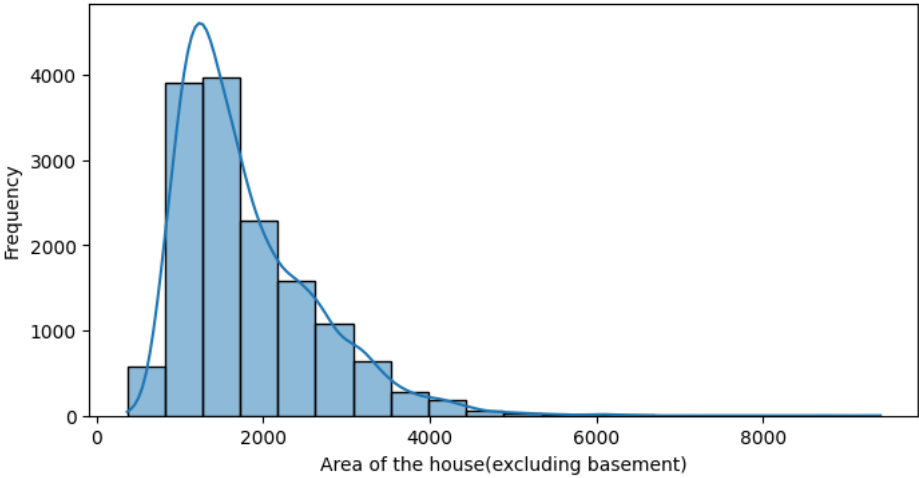
```
for column in columns:
    plt.figure(figsize=(8, 4))
    sn.histplot(df[column], bins=20, kde=True)
    plt.title(f'Histogram of {column}')
    plt.xlabel(column)
    plt.ylabel('Frequency')
    plt.show()
```



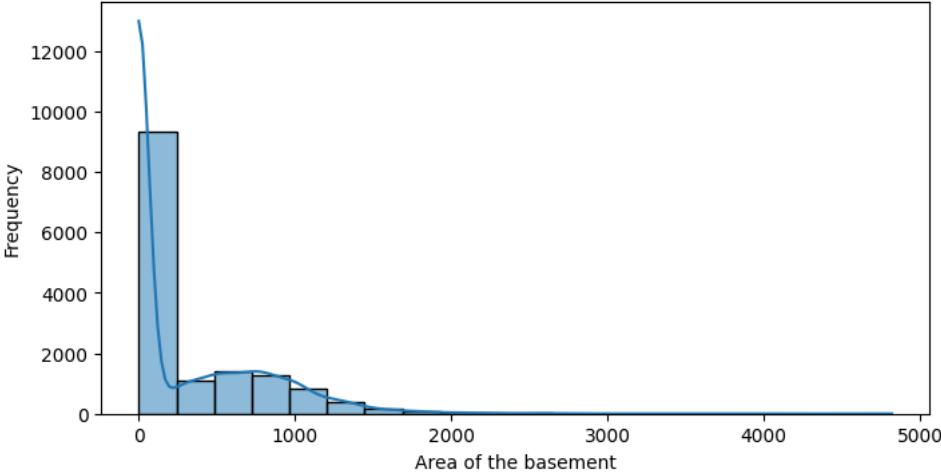




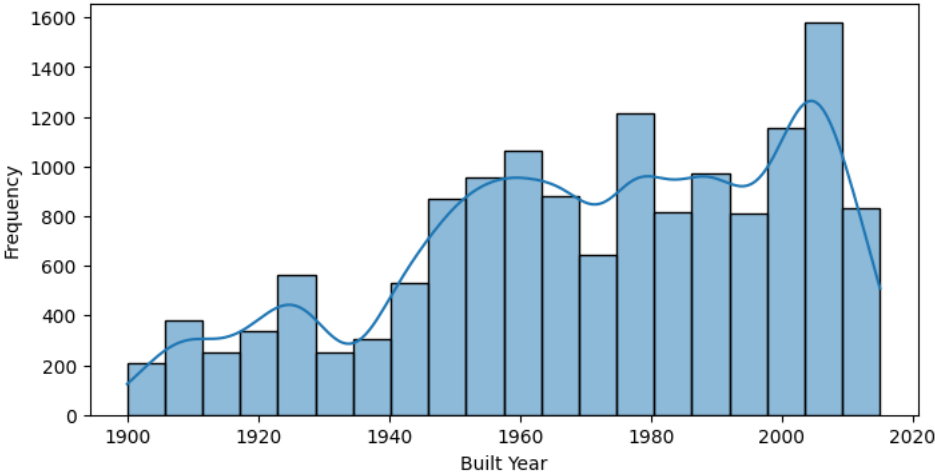
Histogram of Area of the house(excluding basement)



Histogram of Area of the basement

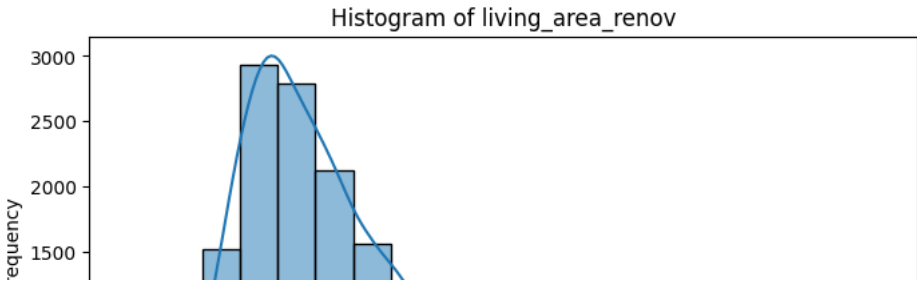
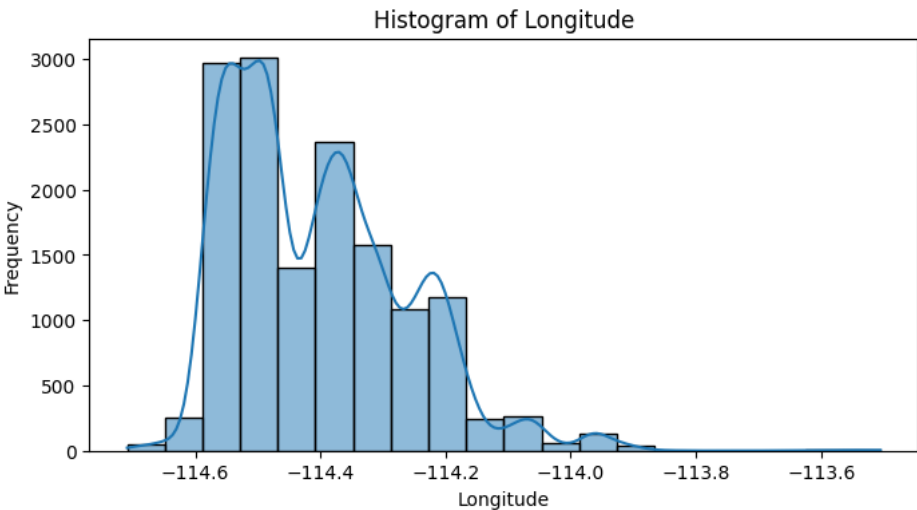
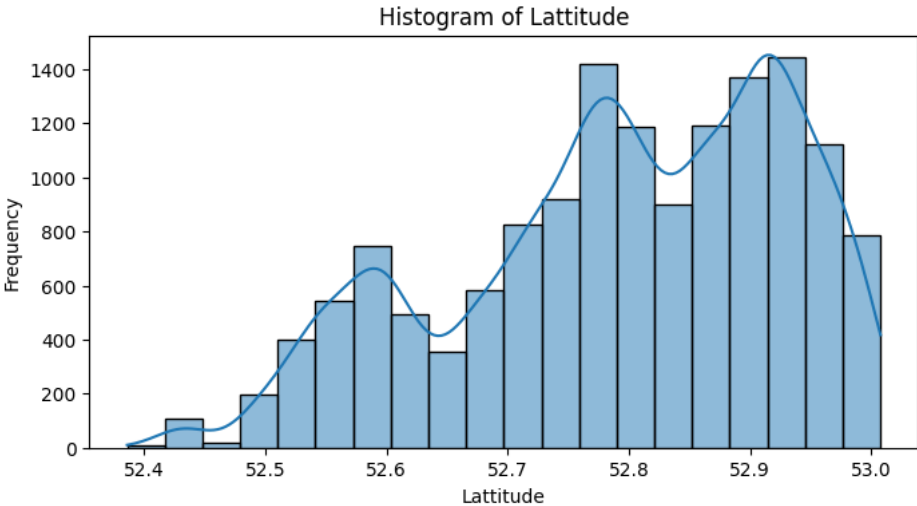
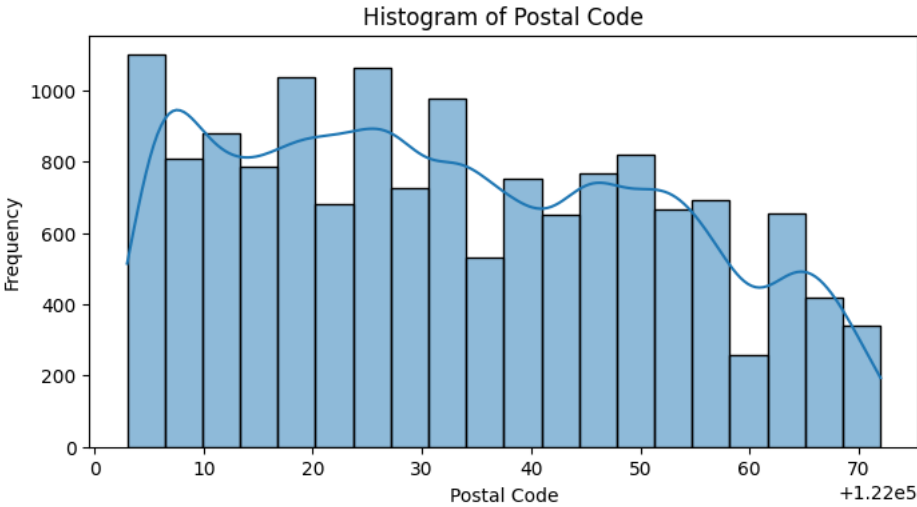
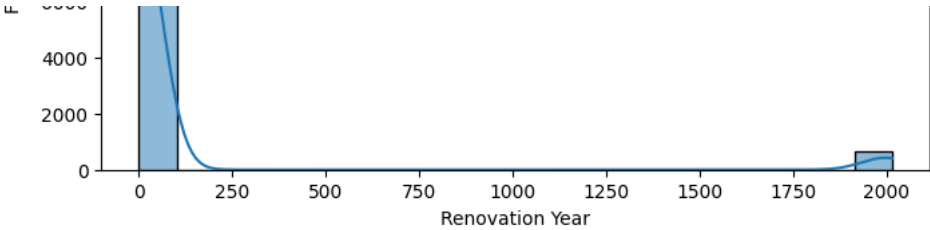


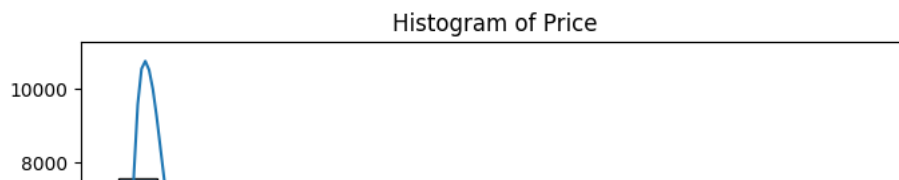
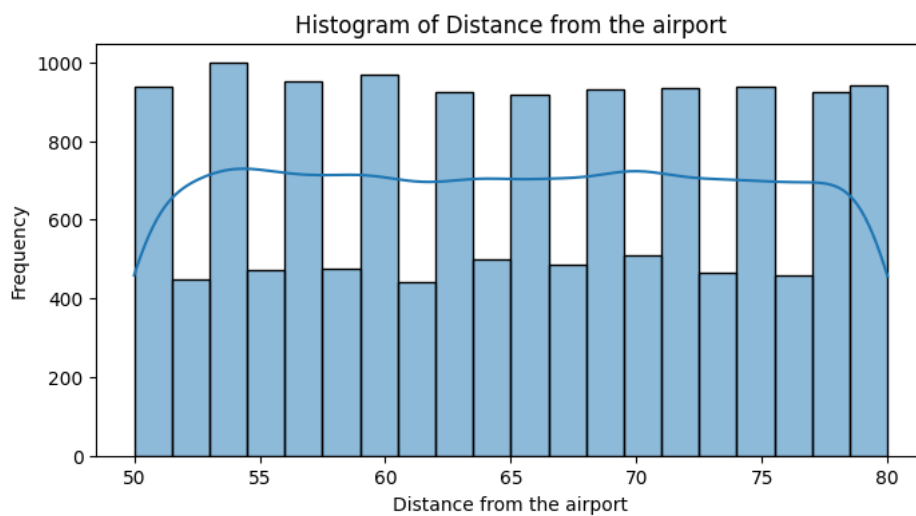
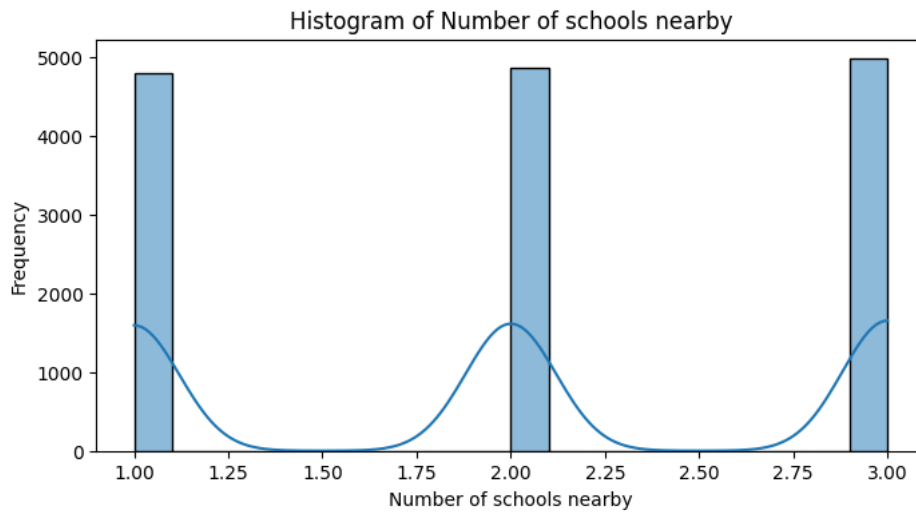
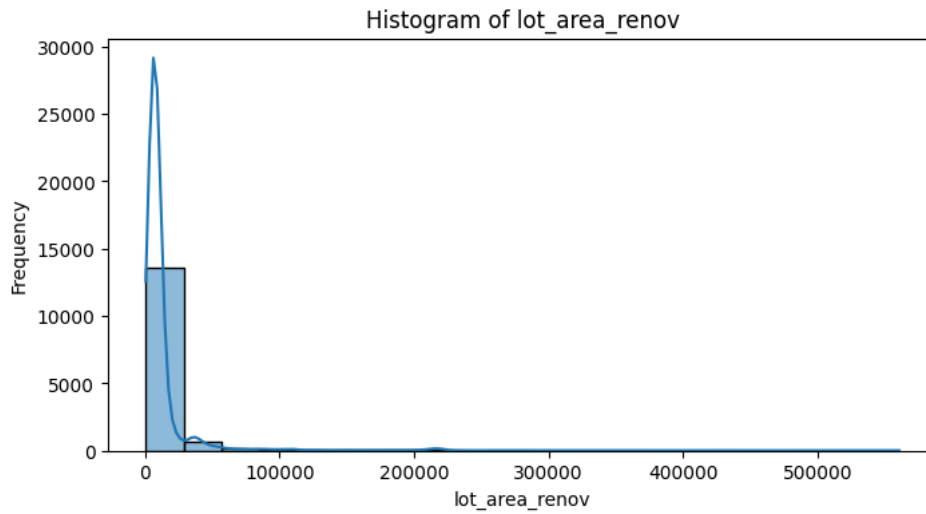
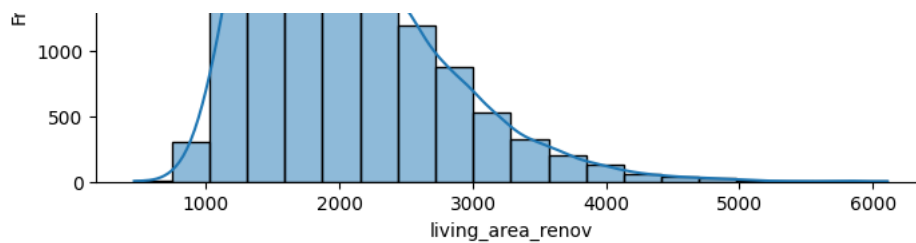
Histogram of Built Year



Histogram of Renovation Year



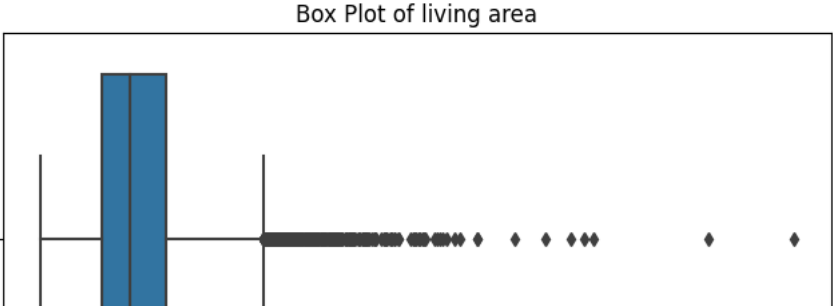
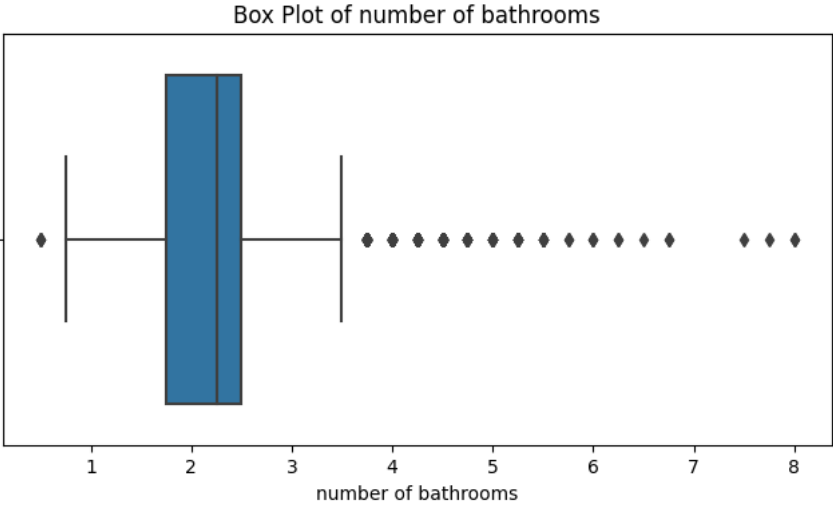
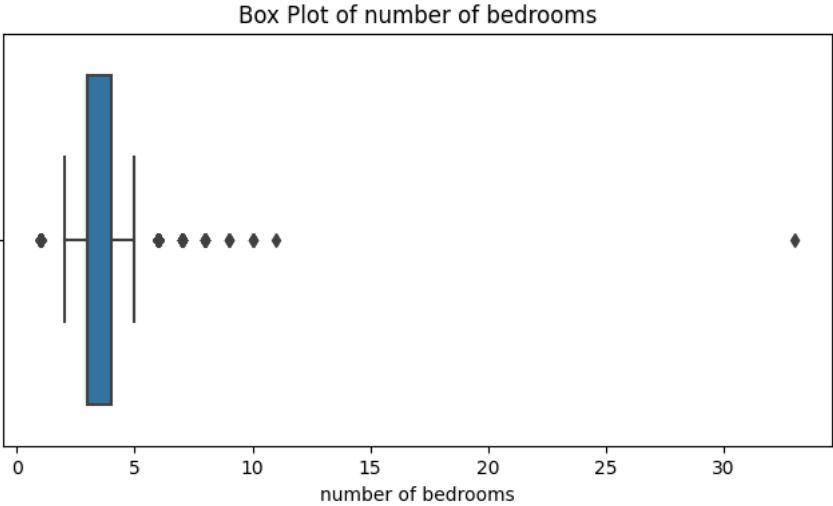
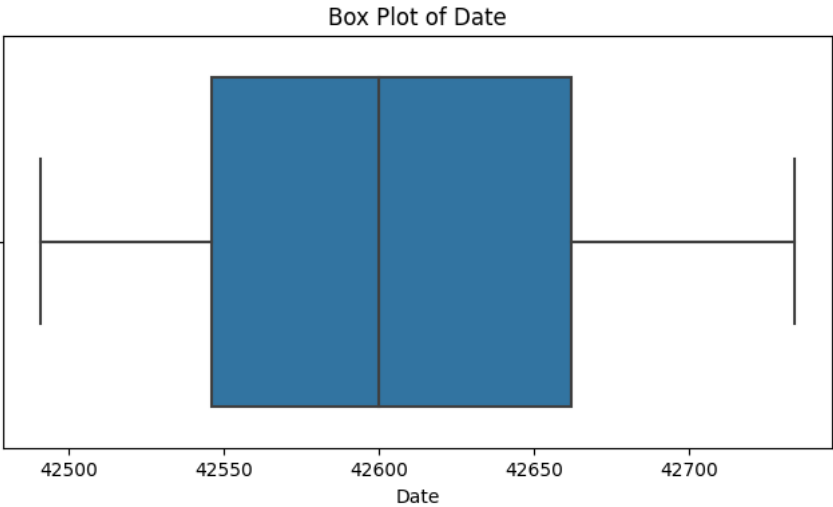
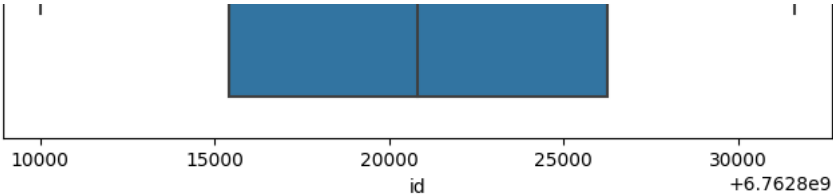


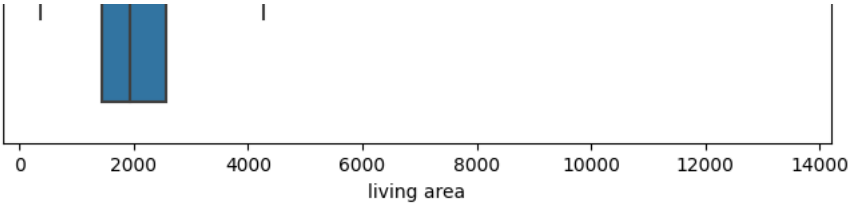


```
for column in columns:
    plt.figure(figsize=(8, 4))
    sn.boxplot(x=df[column])
```

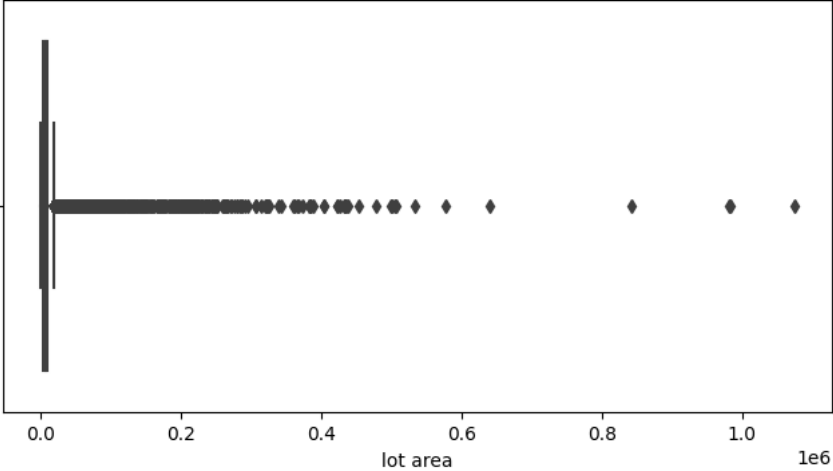
```
plt.title(f'Box Plot of {column}')  
plt.xlabel(column)  
plt.show()
```



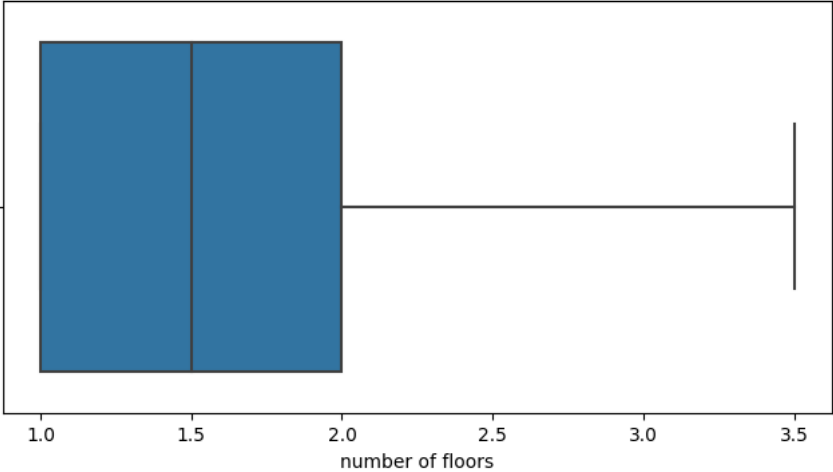




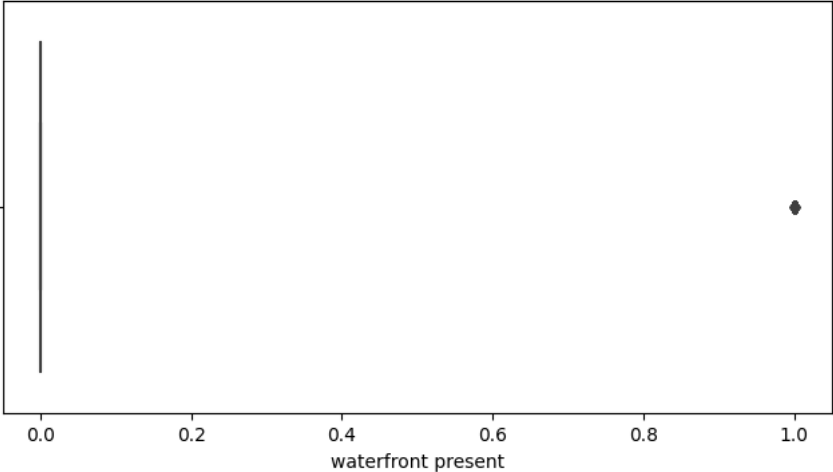
Box Plot of lot area



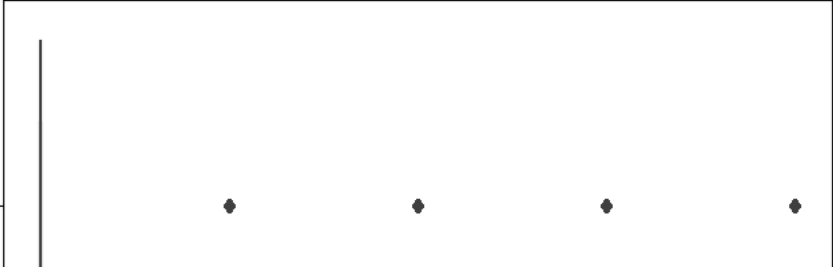
Box Plot of number of floors

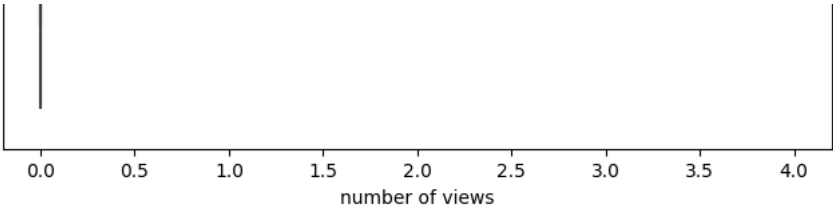


Box Plot of waterfront present

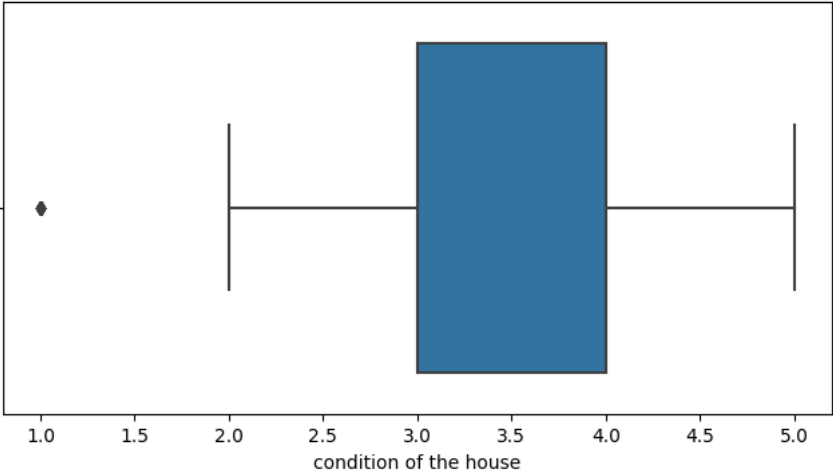


Box Plot of number of views

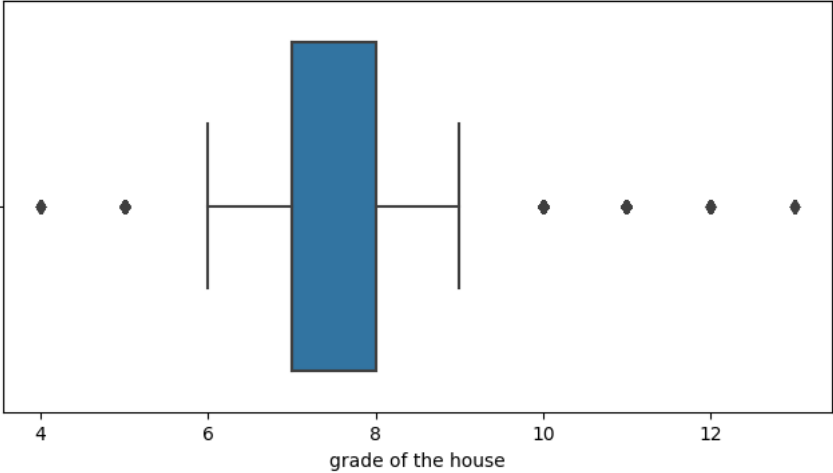




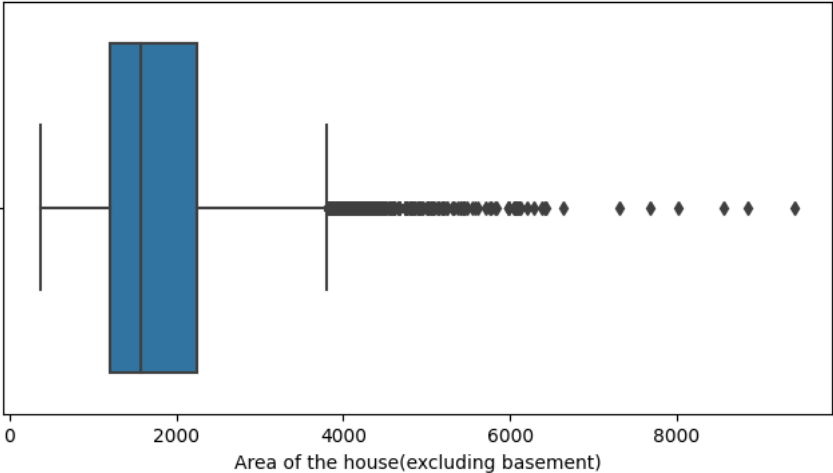
Box Plot of condition of the house



Box Plot of grade of the house

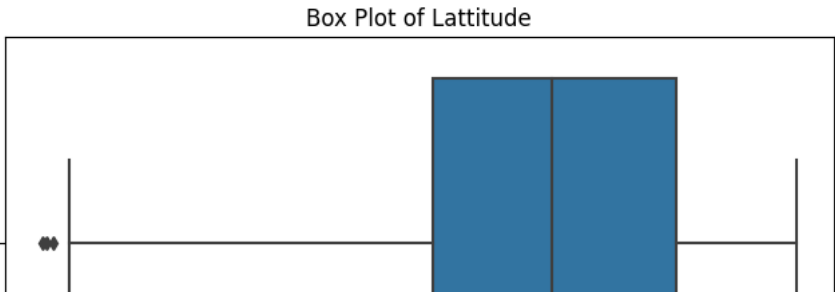
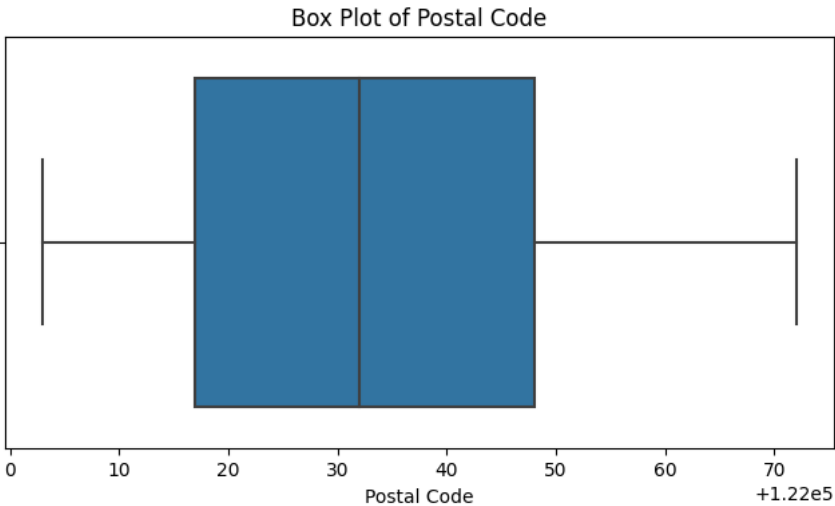
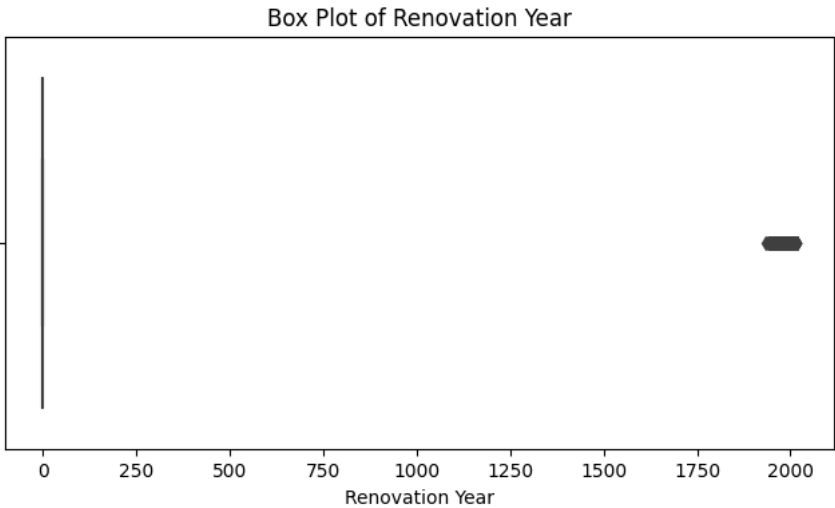
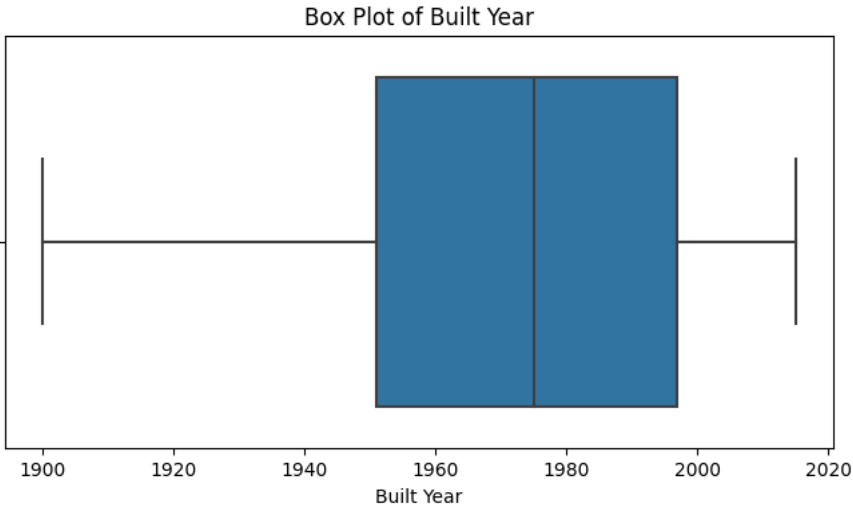
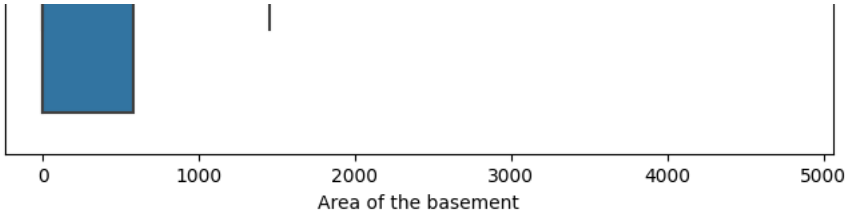


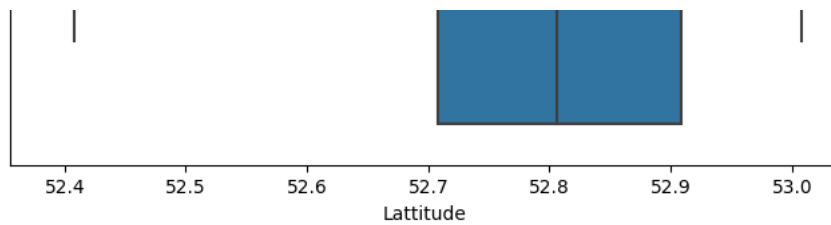
Box Plot of Area of the house(excluding basement)



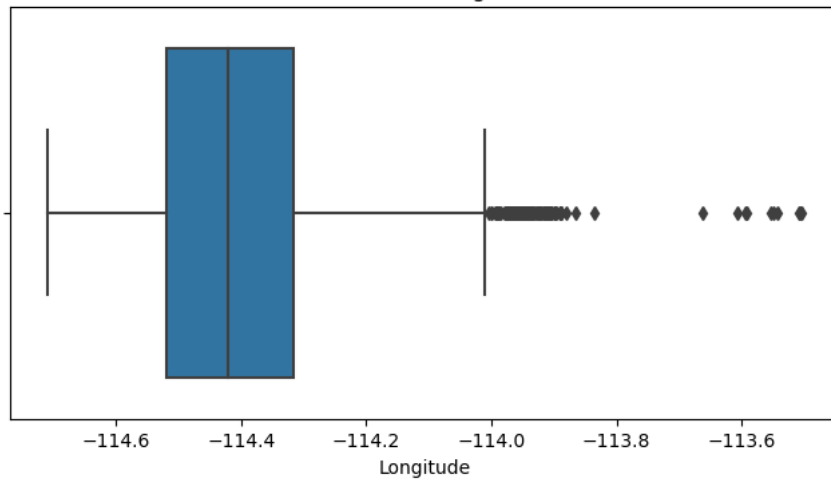
Box Plot of Area of the basement



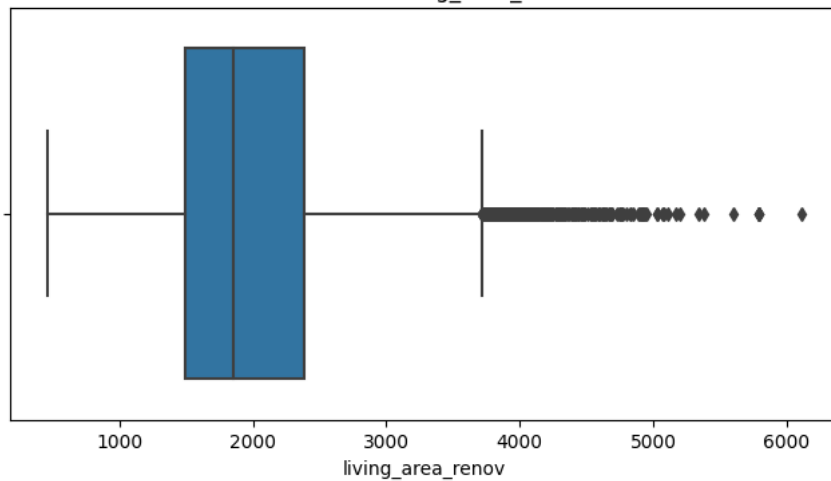




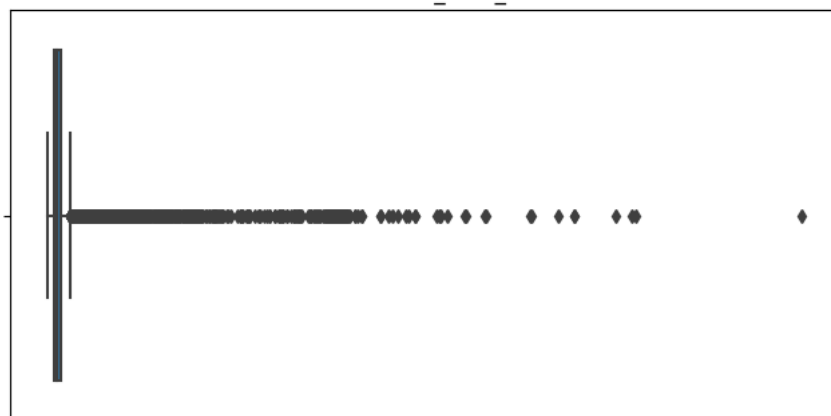
Box Plot of Longitude



Box Plot of living\_area\_renov



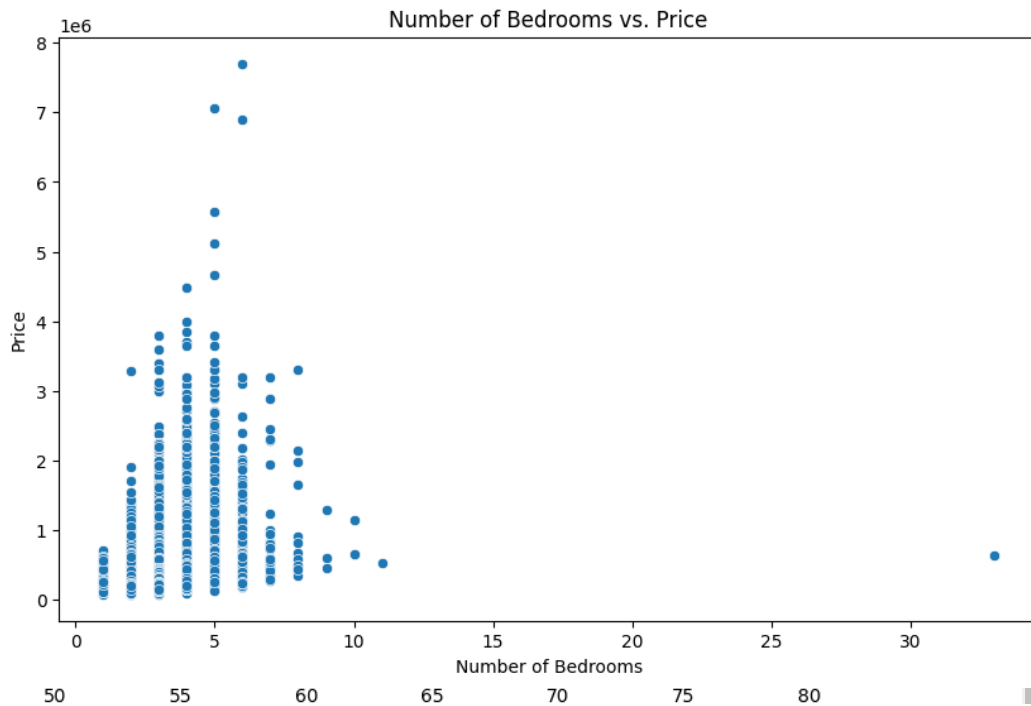
Box Plot of lot\_area\_renov



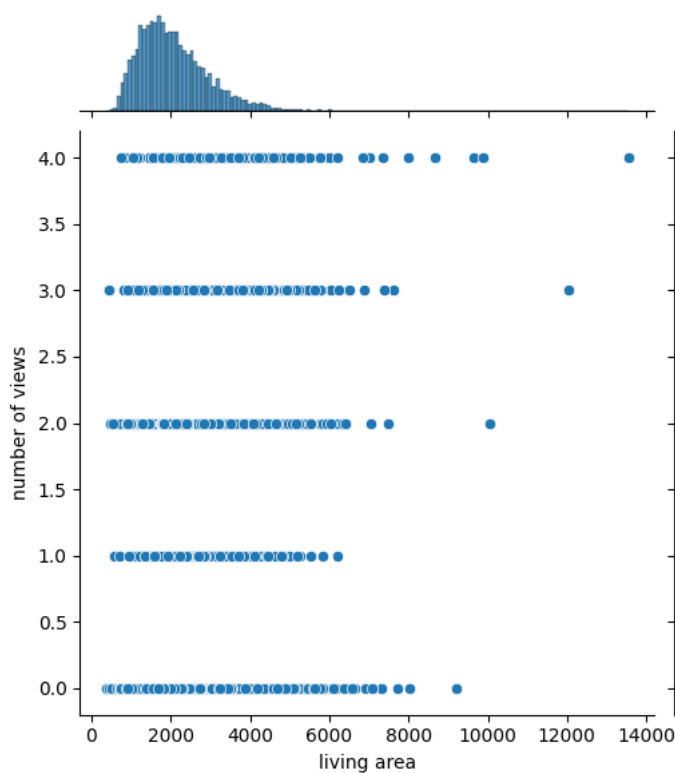
## ▼ Bivariate Analysis

BOX PLOT OF NUMBER OF SCHOOLS NEARBY

```
plt.figure(figsize=(10, 6))
sn.scatterplot(data=df, x='number of bedrooms', y='Price')
plt.title('Number of Bedrooms vs. Price')
plt.xlabel('Number of Bedrooms')
plt.ylabel('Price')
plt.show()
```



```
sn.jointplot(x='living area', y='number of views', data=df, kind='scatter')
plt.show()
```



## ▼ Multivariate Analysis

```
correlation_matrix = df.corr()

plt.figure(figsize=(12, 10))
sn.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f', linewidths=0.5)
plt.title('Correlation Matrix for All Columns')
plt.show()
```

