

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: df = pd.read_csv("./House_Price_India.csv")
```

```
In [3]: df.head()
```

Out[3]:

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0

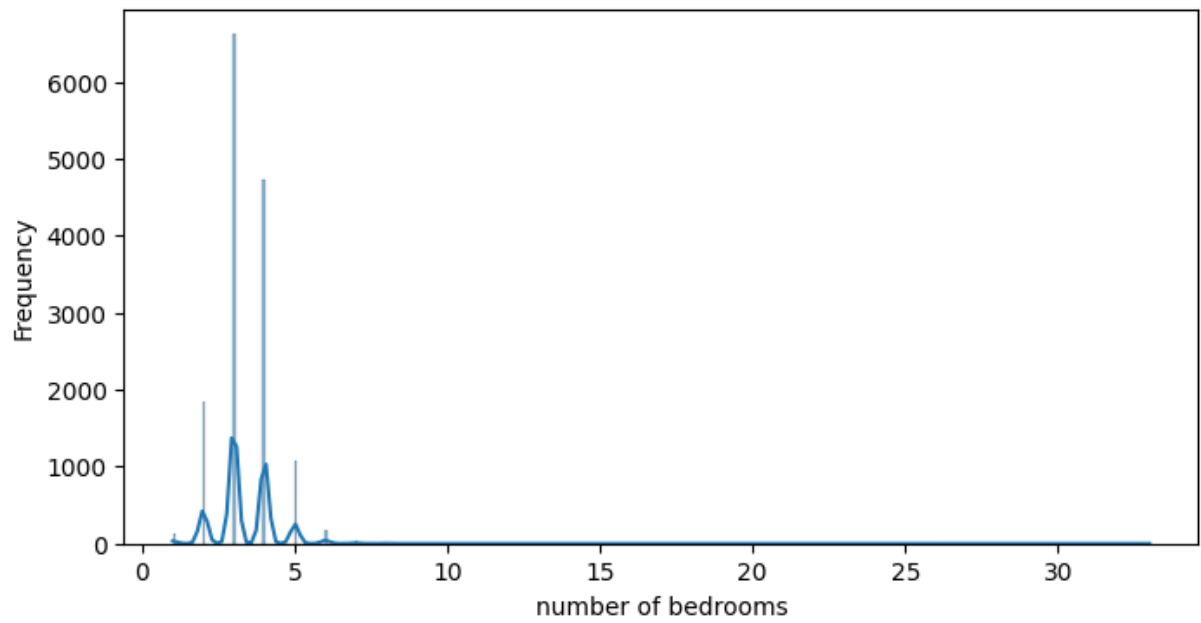
5 rows × 23 columns

```
In [6]: df.columns
```

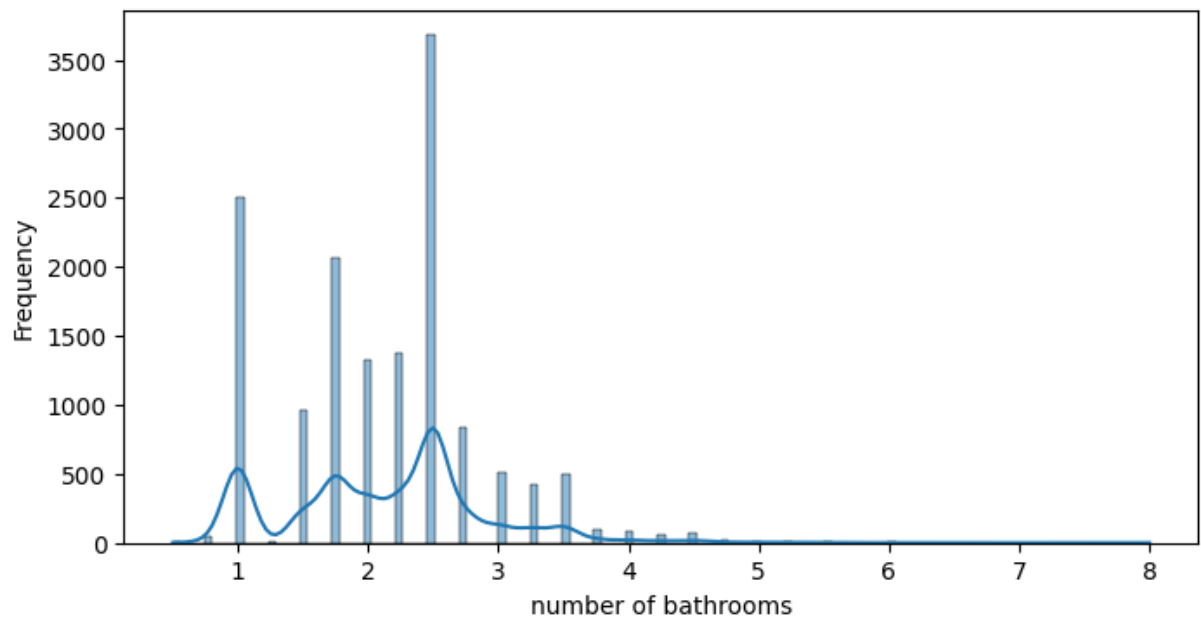
```
Out[6]: Index(['id', 'Date', 'number of bedrooms', 'number of bathrooms',
              'living area', 'lot area', 'number of floors', 'waterfront present',
              'number of views', 'condition of the house', 'grade of the house',
              'Area of the house(excluding basement)', 'Area of the basement',
              'Built Year', 'Renovation Year', 'Postal Code', 'Latitude',
              'Longitude', 'living_area_renov', 'lot_area_renov',
              'Number of schools nearby', 'Distance from the airport', 'Price'],
              dtype='object')
```

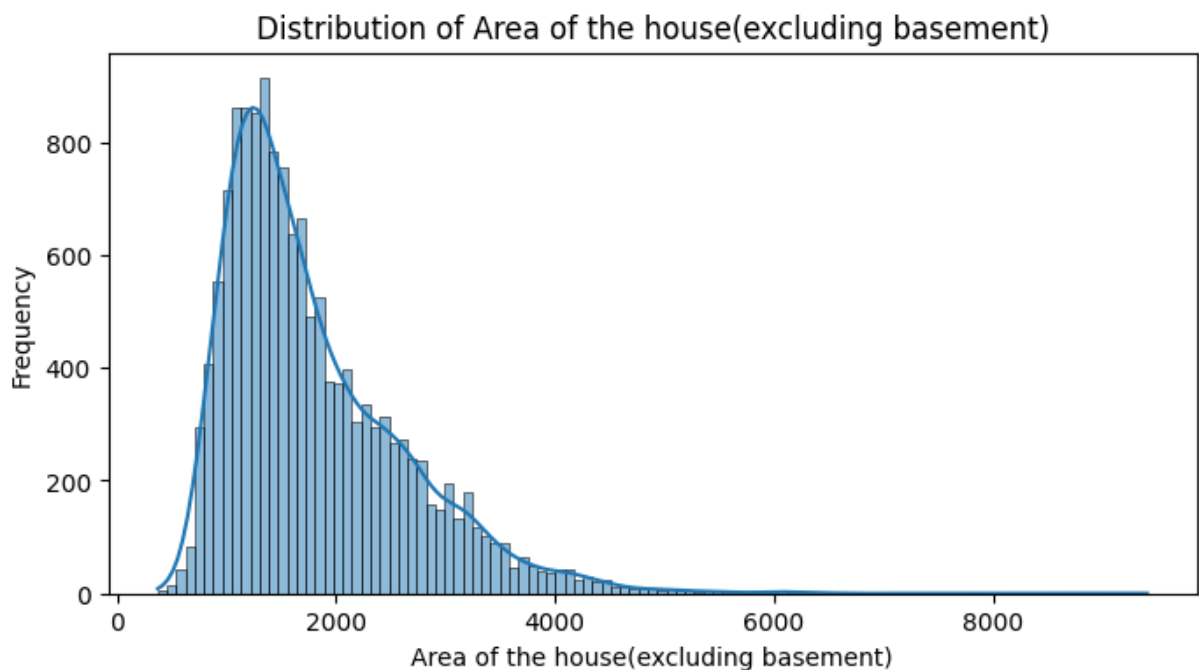
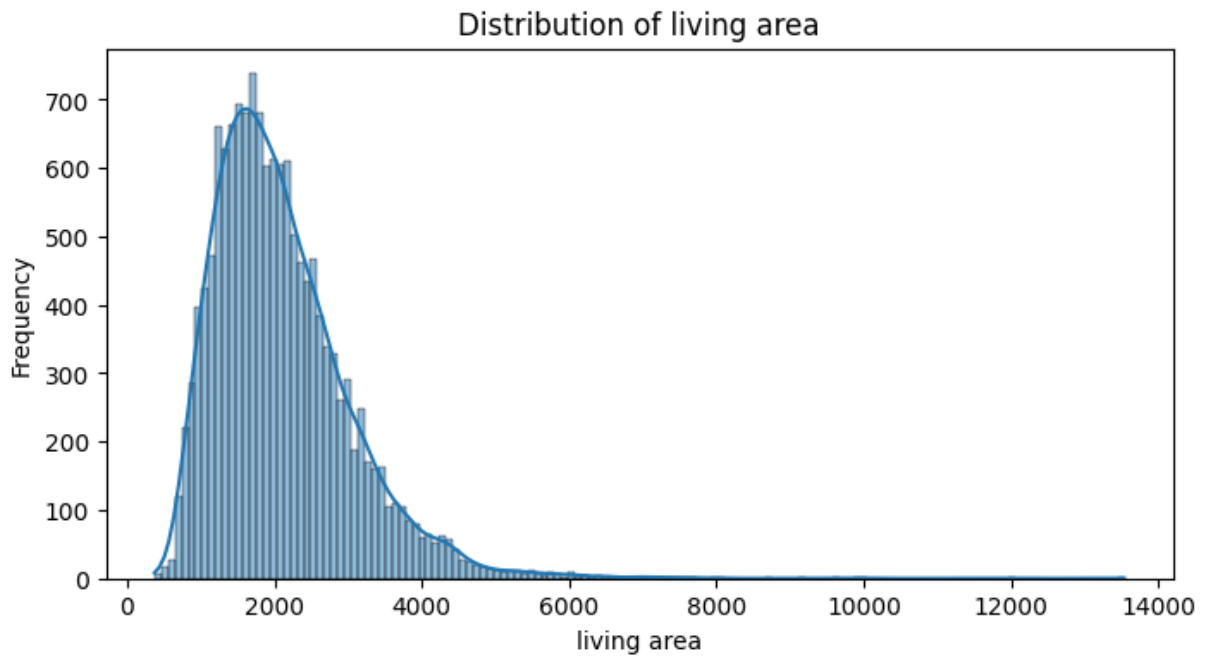
```
In [7]: numerical_features = ['number of bedrooms', 'number of bathrooms', 'living area', '
for feature in numerical_features:
    plt.figure(figsize=(8, 4))
    sns.histplot(df[feature], kde=True)
    plt.title(f'Distribution of {feature}')
    plt.xlabel(feature)
    plt.ylabel('Frequency')
    plt.show()
```

Distribution of number of bedrooms



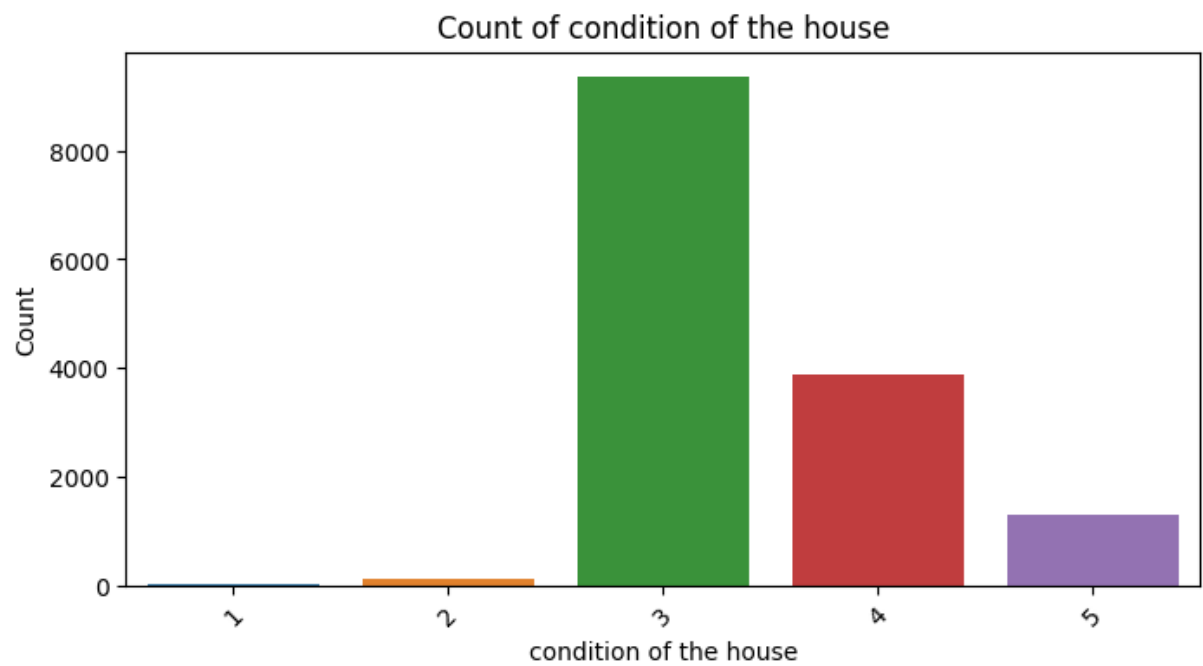
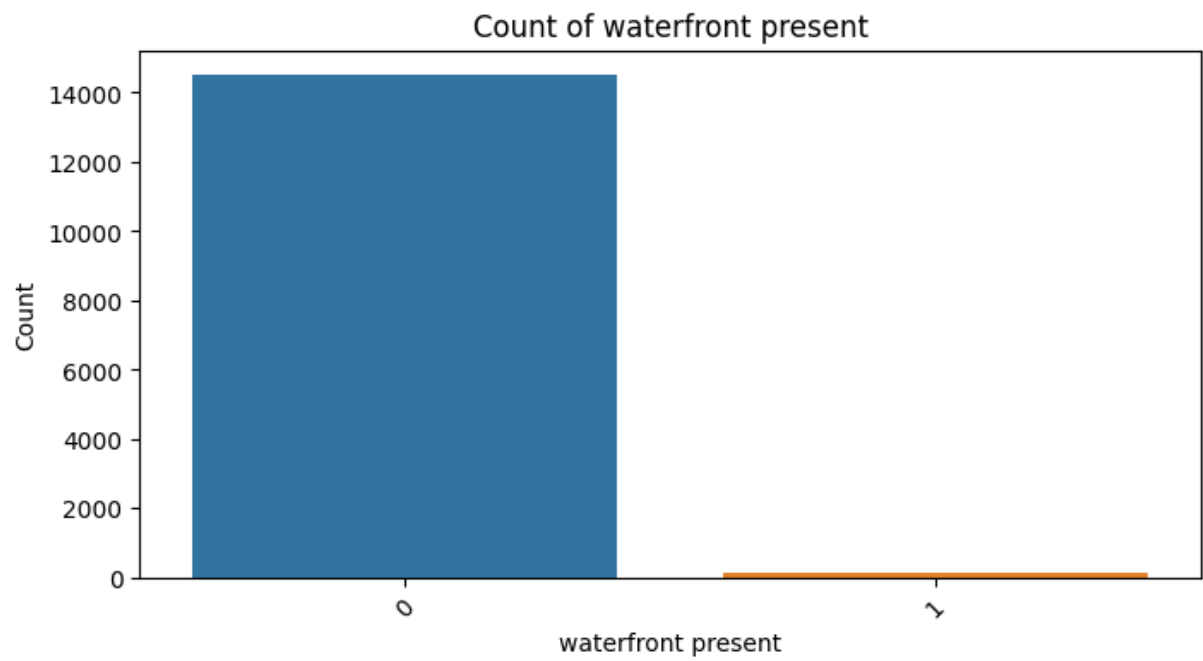
Distribution of number of bathrooms

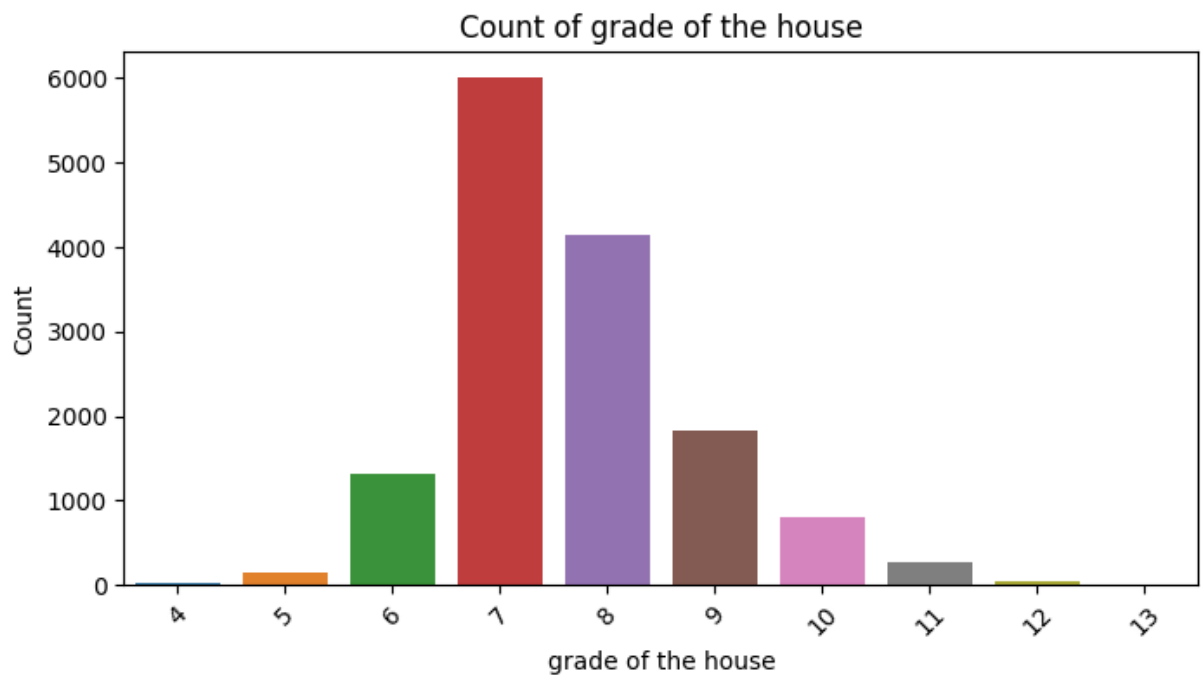




```
In [8]: categorical_features = ['waterfront present', 'condition of the house', 'grade of t

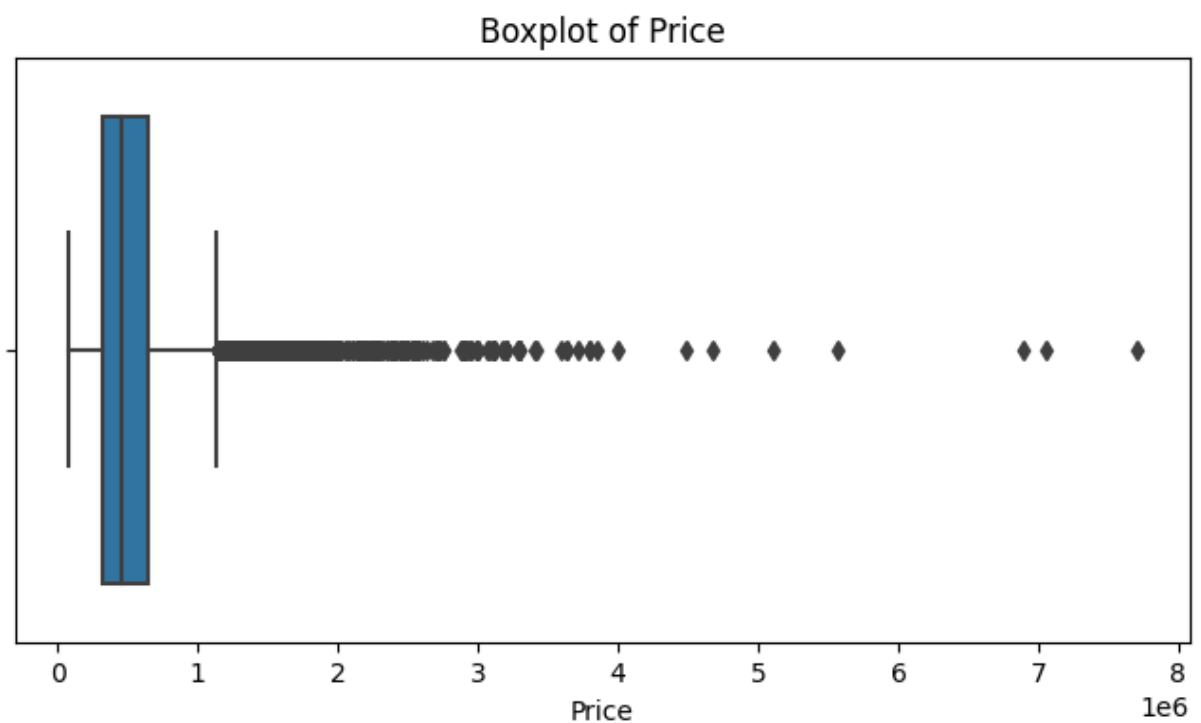
for feature in categorical_features:
    plt.figure(figsize=(8, 4))
    sns.countplot(data=df, x=feature)
    plt.title(f'Count of {feature}')
    plt.xlabel(feature)
    plt.ylabel('Count')
    plt.xticks(rotation=45)
    plt.show()
```

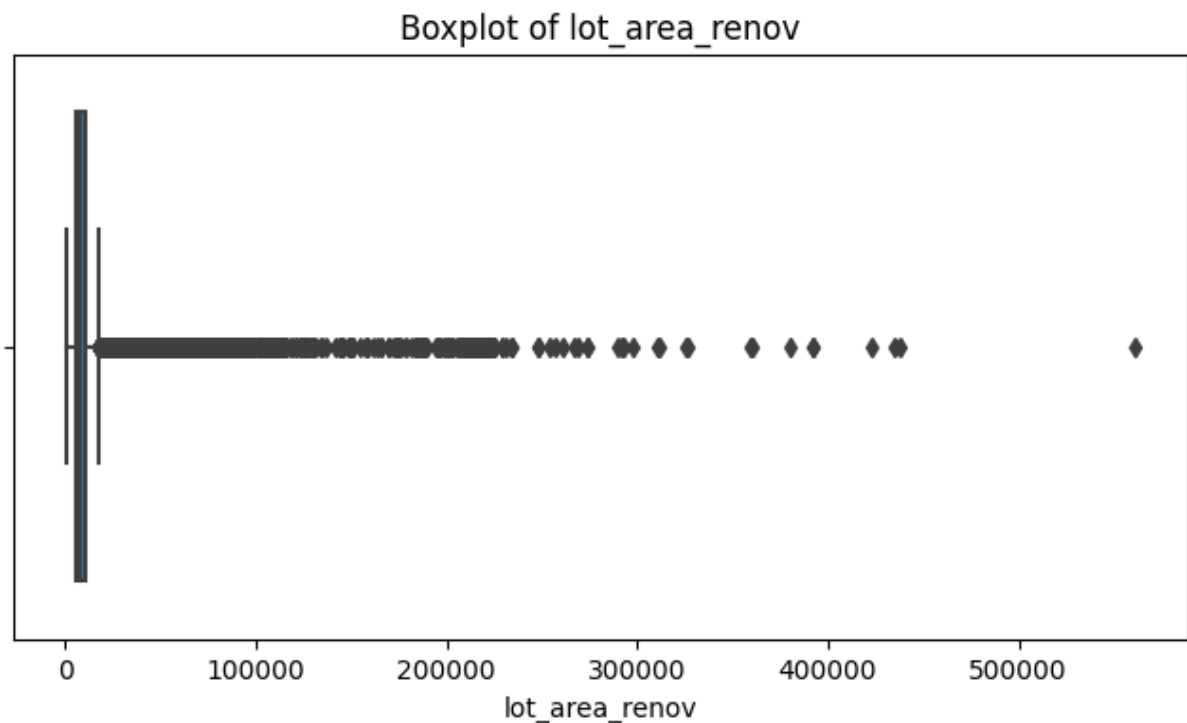
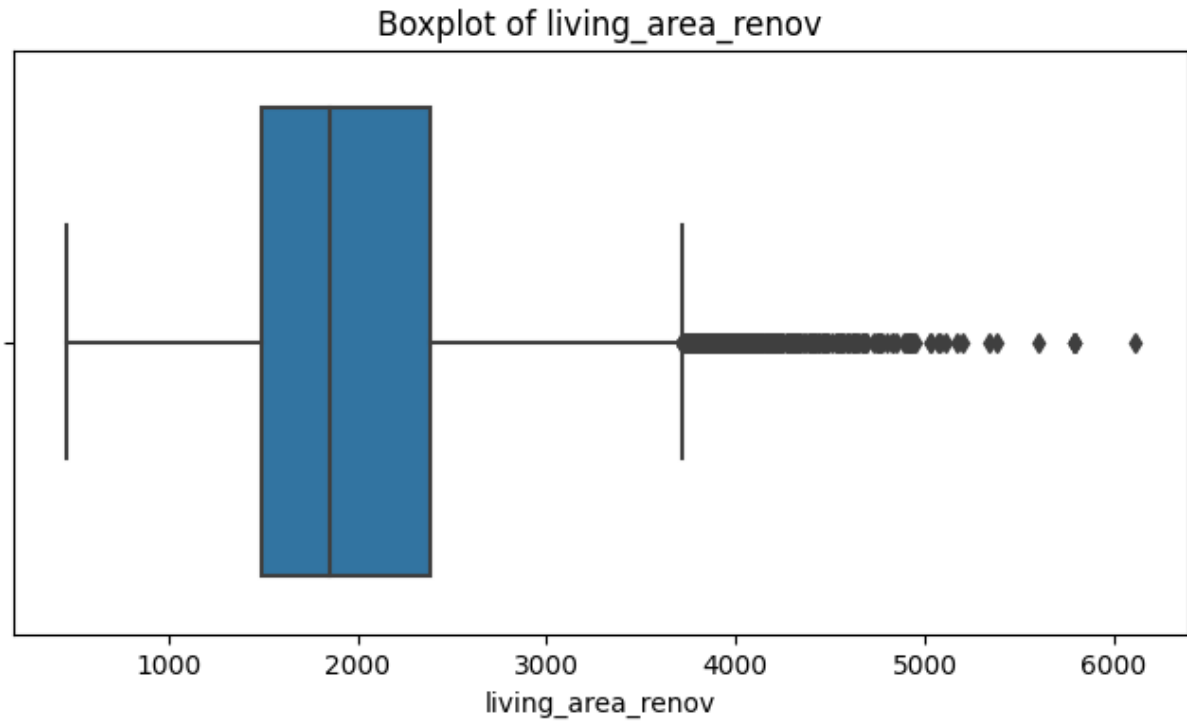




```
In [9]: outlier_features = ['Price', 'living_area_renov', 'lot_area_renov']

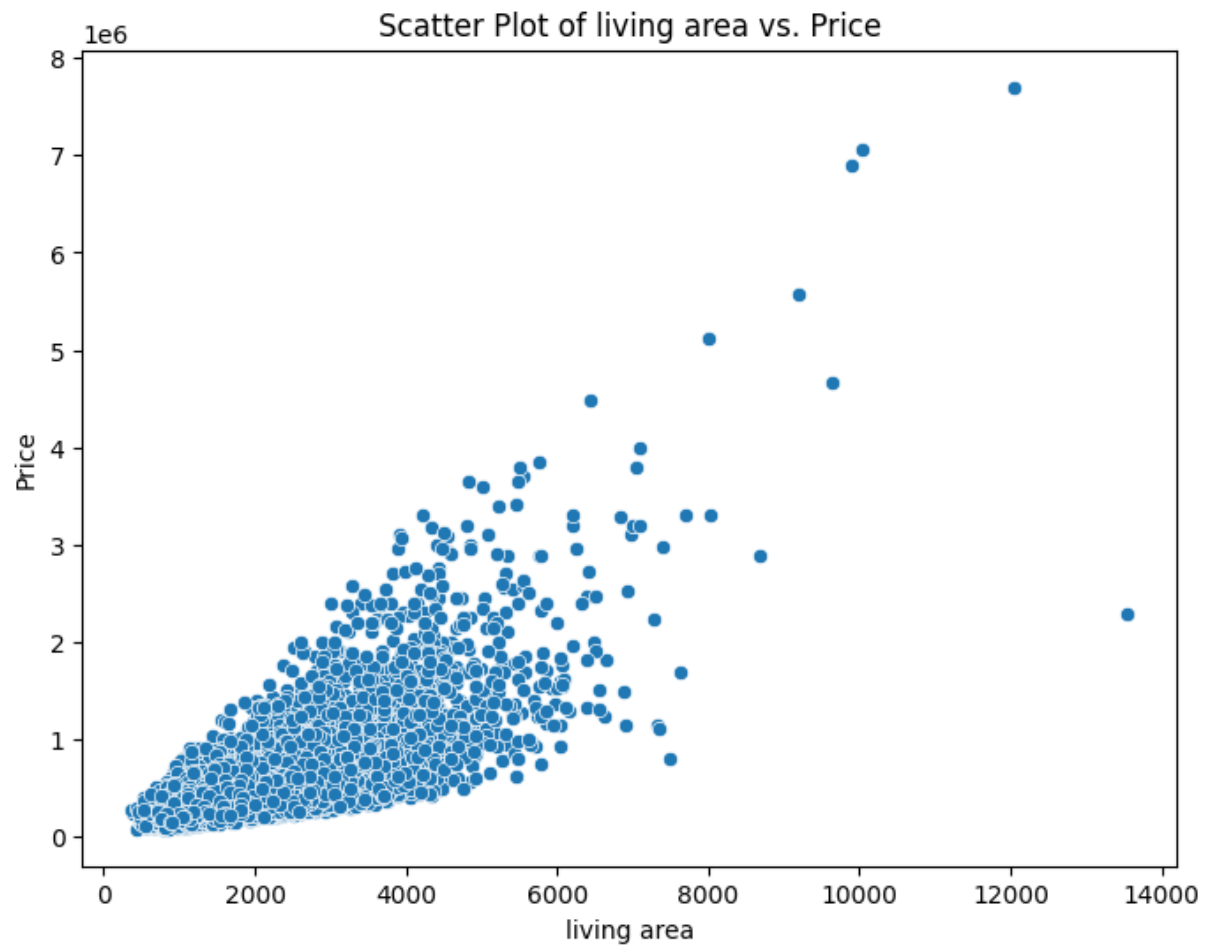
for feature in outlier_features:
    plt.figure(figsize=(8, 4))
    sns.boxplot(x=df[feature])
    plt.title(f'Boxplot of {feature}')
    plt.xlabel(feature)
    plt.show()
```

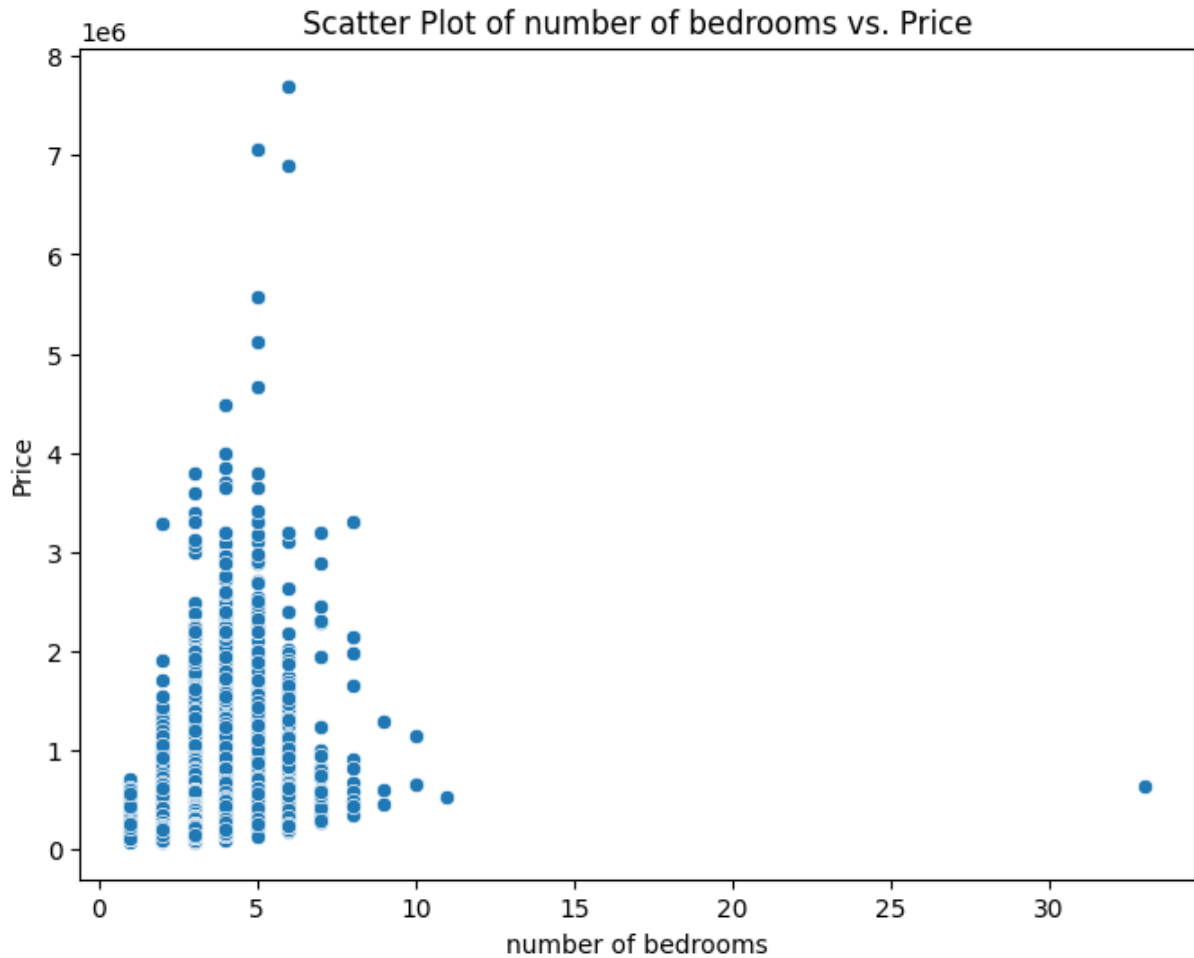




```
In [11]: numerical_features = ['living area', 'number of bedrooms']

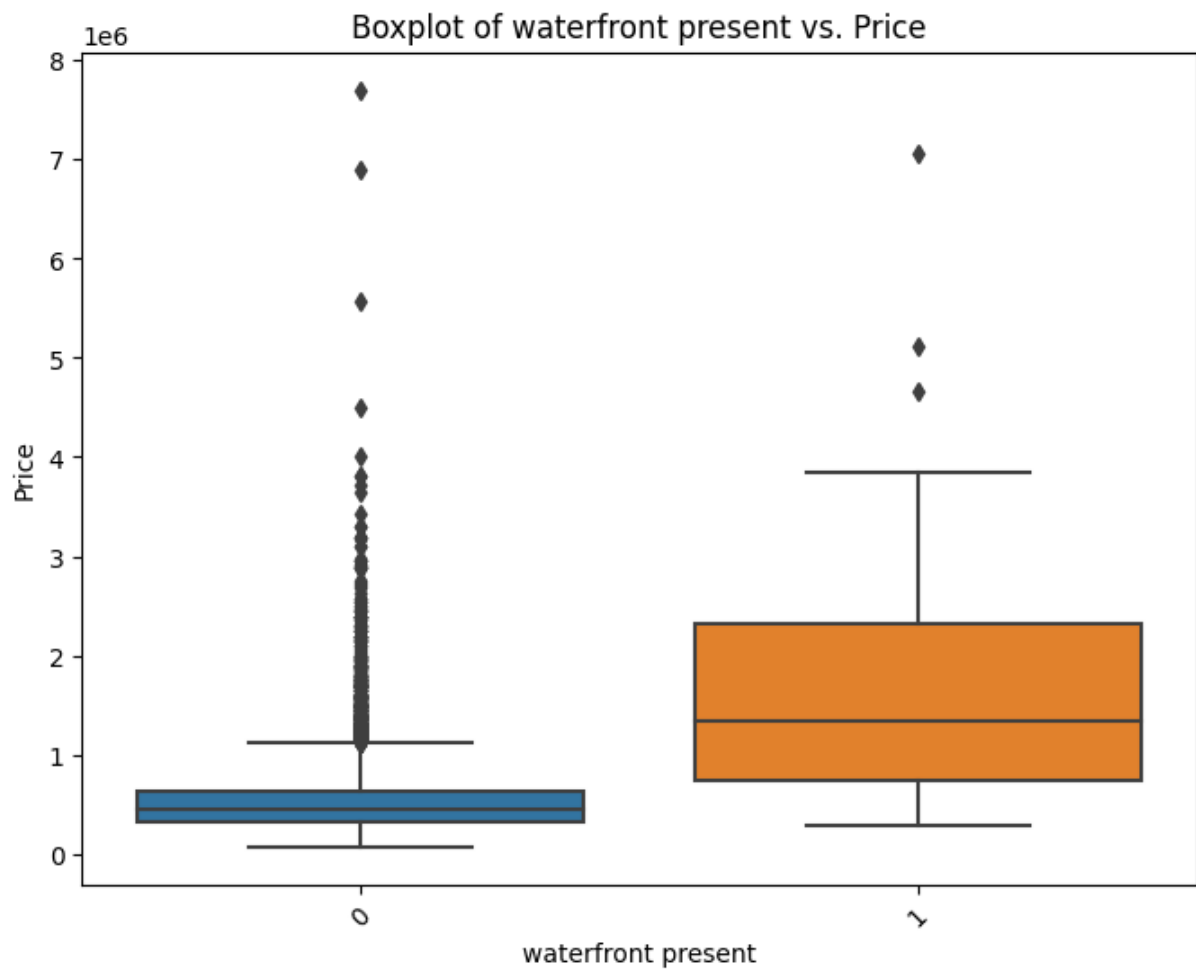
for feature in numerical_features:
    plt.figure(figsize=(8, 6))
    sns.scatterplot(data=df, x=feature, y='Price')
    plt.title(f'Scatter Plot of {feature} vs. Price')
    plt.xlabel(feature)
    plt.ylabel('Price')
    plt.show()
```

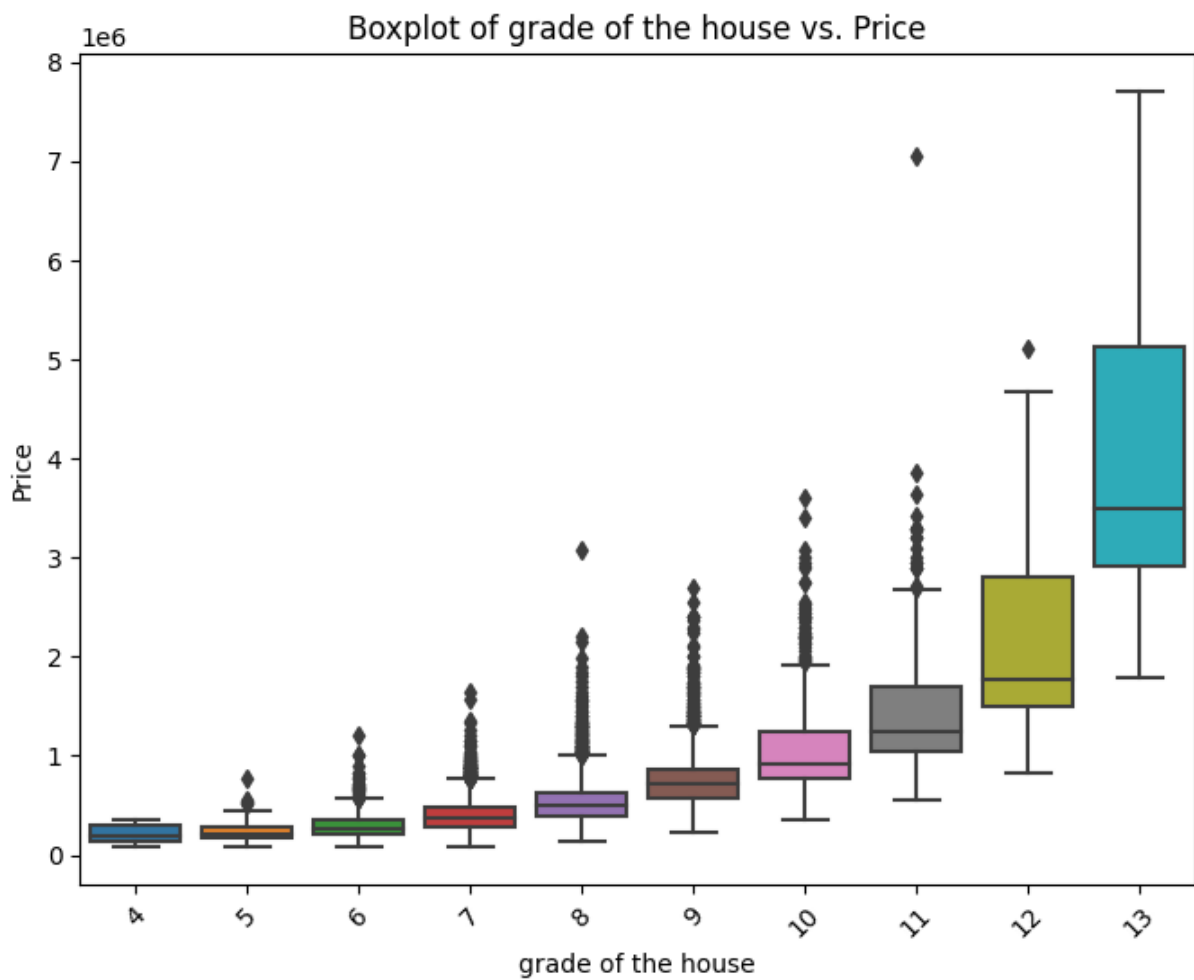




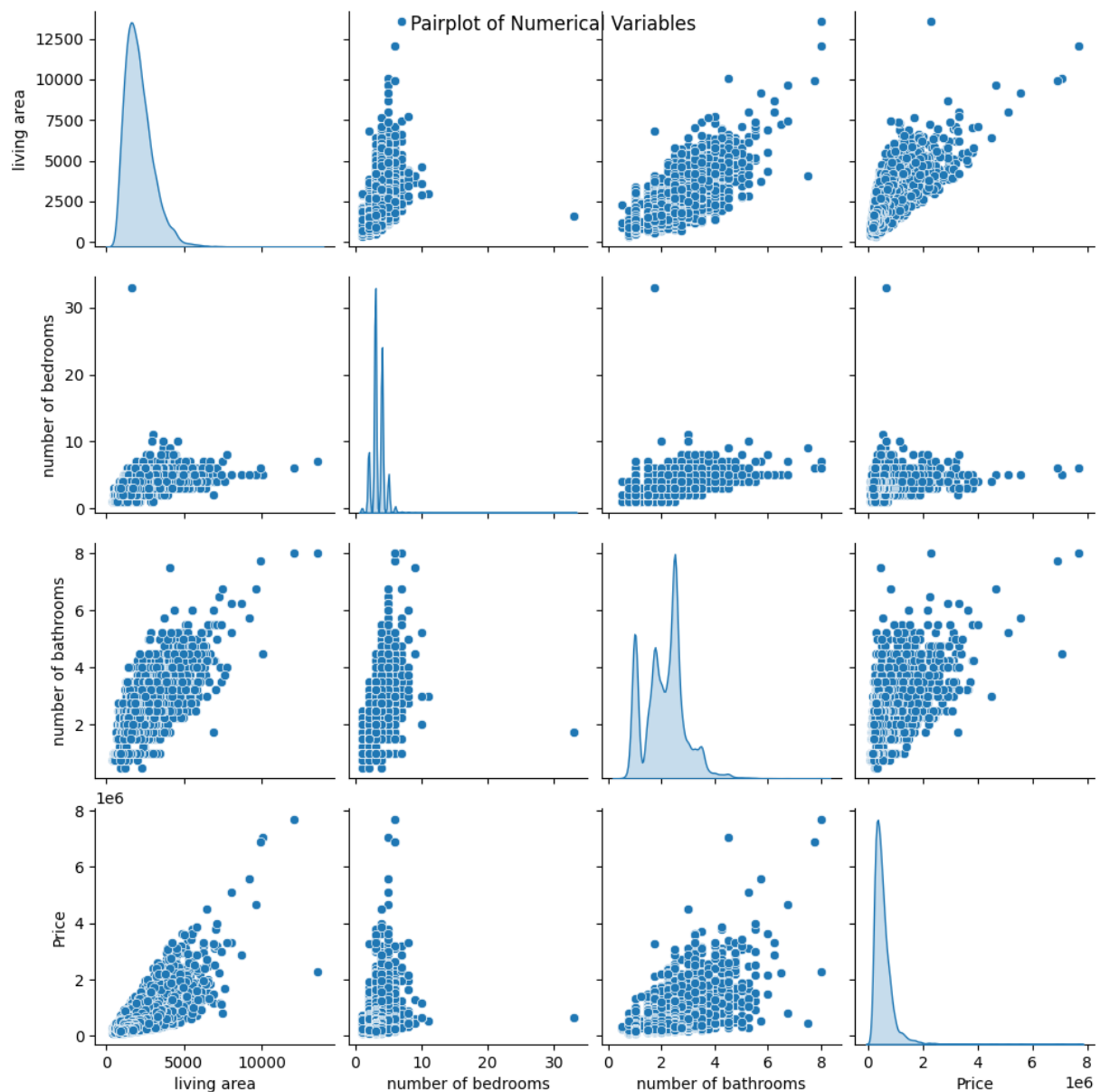
```
In [12]: categorical_features = ['waterfront present', 'grade of the house']

for feature in categorical_features:
    plt.figure(figsize=(8, 6))
    sns.boxplot(data=df, x=feature, y='Price')
    plt.title(f'Boxplot of {feature} vs. Price')
    plt.xlabel(feature)
    plt.ylabel('Price')
    plt.xticks(rotation=45)
    plt.show()
```

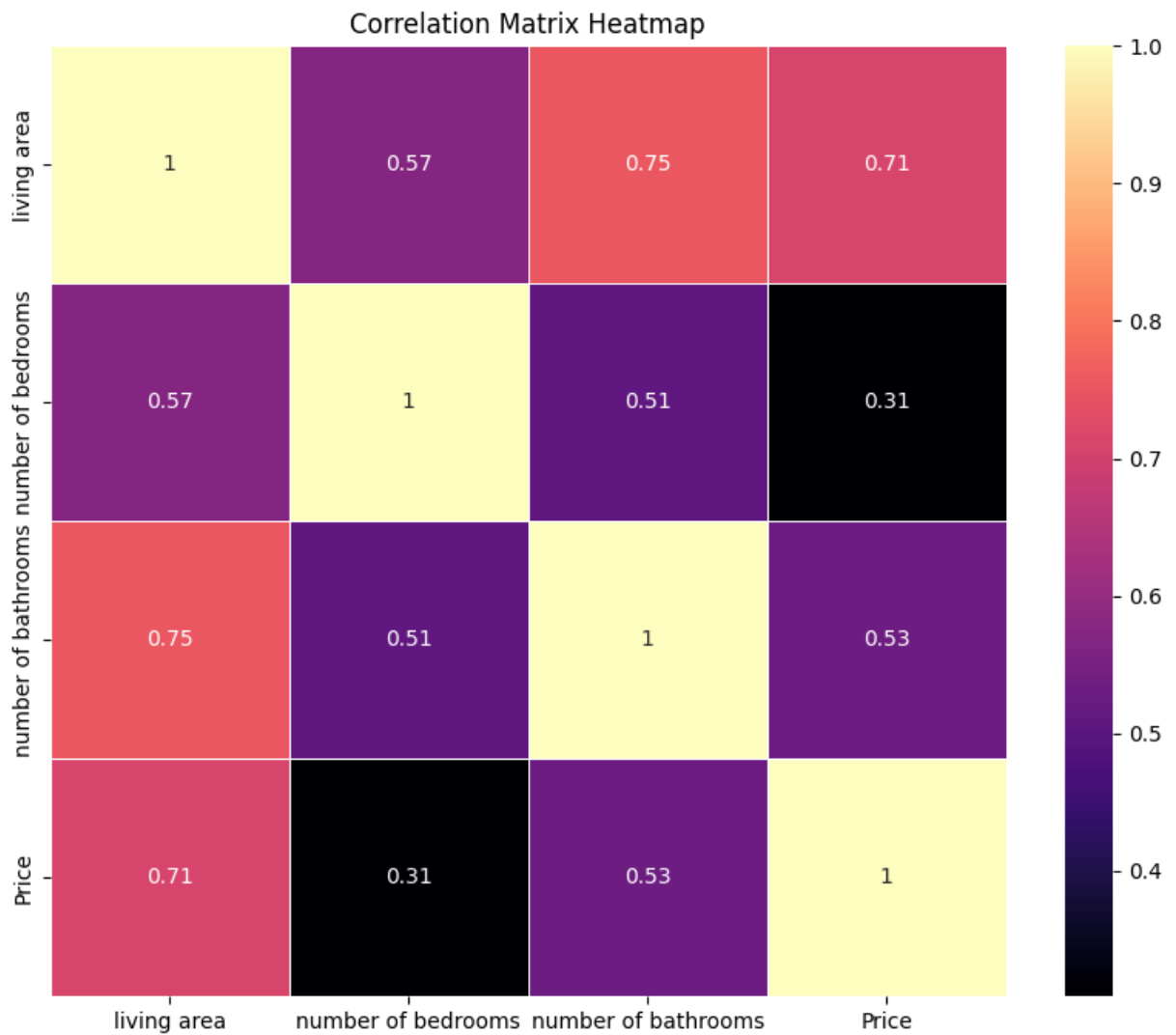


```
In [13]: numerical_features = ['living area', 'number of bedrooms', 'number of bathrooms', '
sns.pairplot(data=df[numerical_features], diag_kind='kde')
plt.suptitle('Pairplot of Numerical Variables')
plt.show()
```

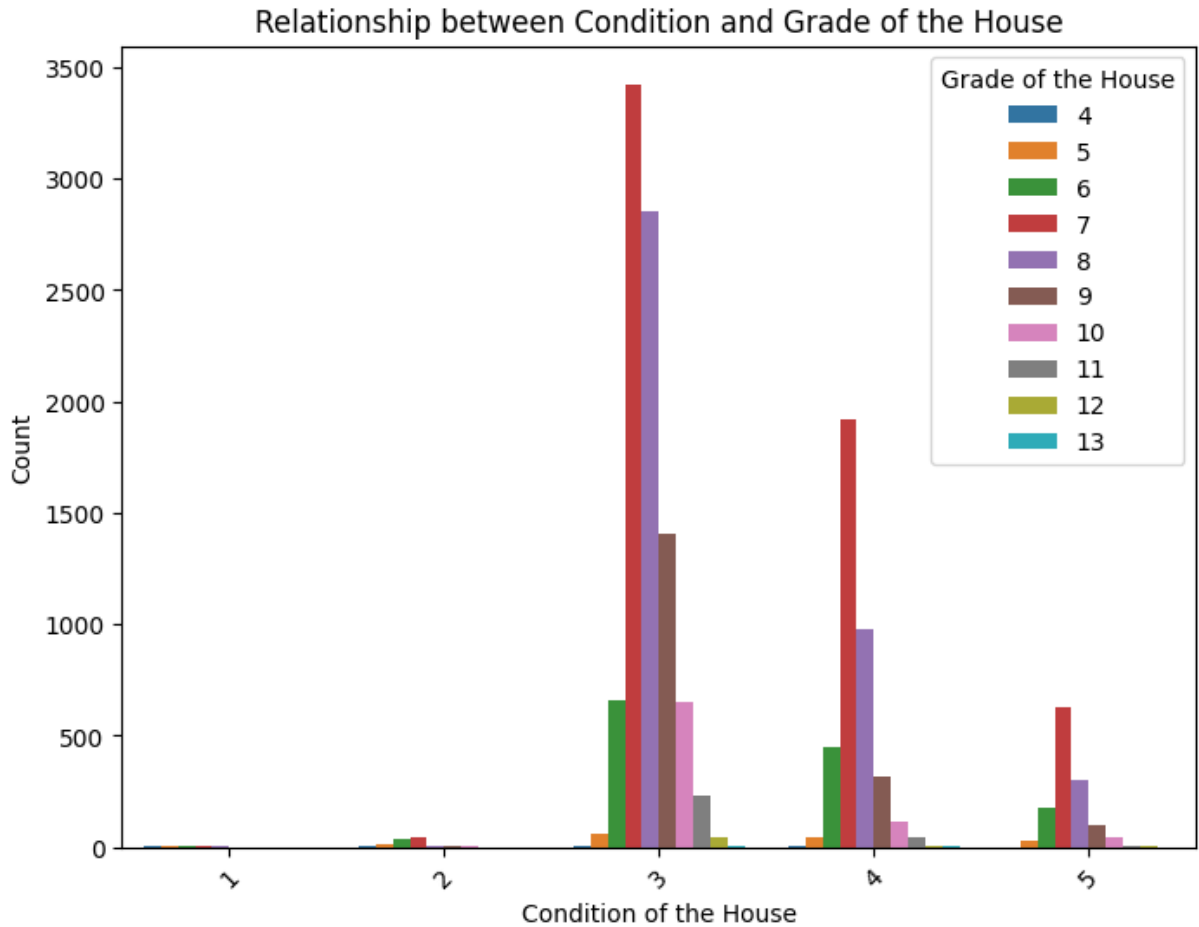


```
In [15]: correlation_matrix = df[numerical_features].corr()

plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='magma', linewidths=0.5)
plt.title('Correlation Matrix Heatmap')
plt.show()
```



```
In [16]: plt.figure(figsize=(8, 6))
sns.countplot(data=df, x='condition of the house', hue='grade of the house')
plt.title('Relationship between Condition and Grade of the House')
plt.xlabel('Condition of the House')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.legend(title='Grade of the House')
plt.show()
```



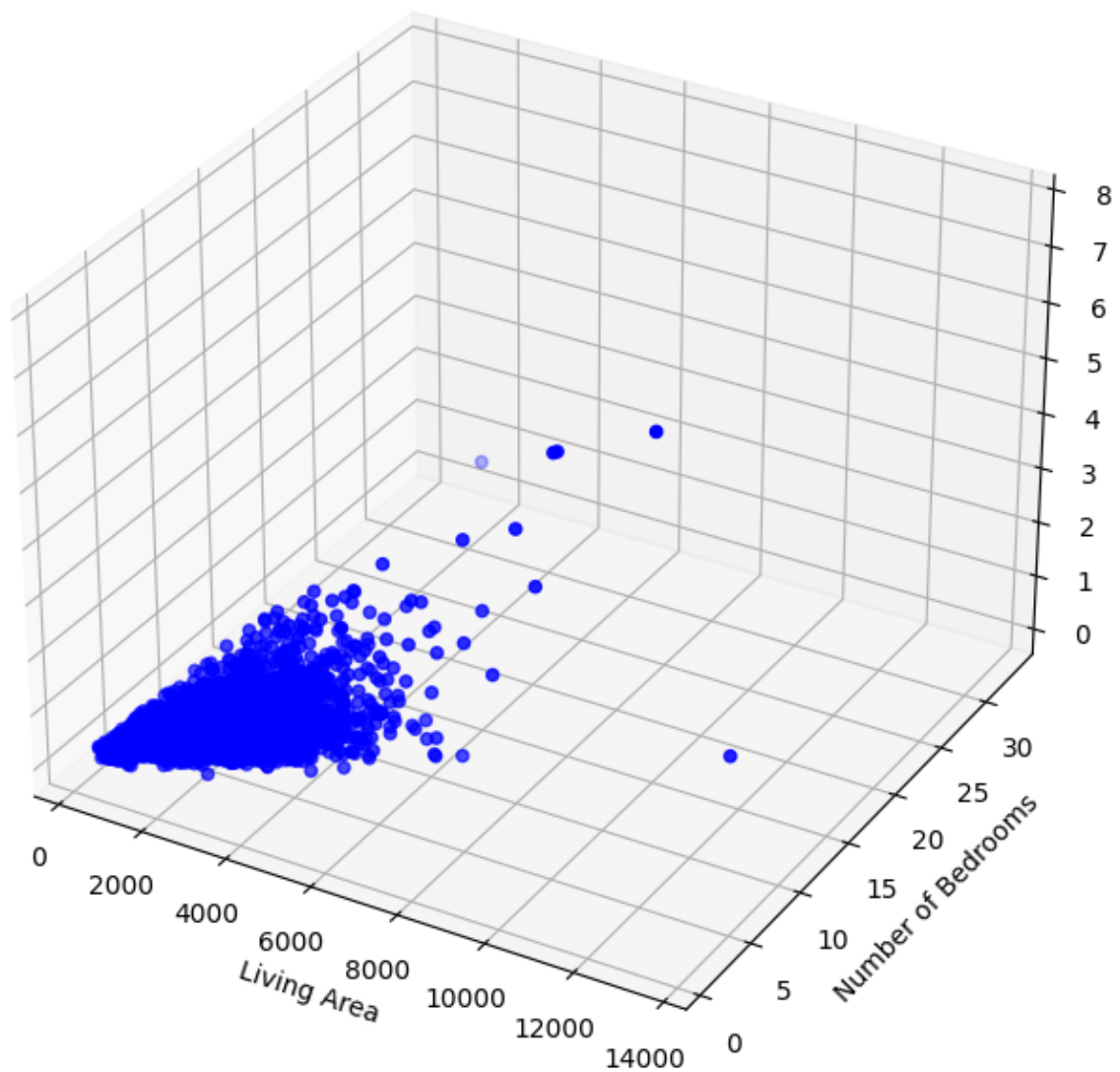
```
In [17]: from mpl_toolkits.mplot3d import Axes3D

fig = plt.figure(figsize=(10, 8))
ax = fig.add_subplot(111, projection='3d')

ax.scatter(df['living area'], df['number of bedrooms'], df['Price'], c='b', marker='b')
ax.set_xlabel('Living Area')
ax.set_ylabel('Number of Bedrooms')
ax.set_zlabel('Price')

plt.title('3D Scatter Plot of Living Area, Bedrooms, and Price')
plt.show()
```

3D Scatter Plot of Living Area, Bedrooms, and Price



```
In [18]: numerical_features = ['living area', 'number of bedrooms', 'number of bathrooms', '  
  
sns.pairplot(data=df[numerical_features], diag_kind='kde')  
plt.suptitle('Pairplot of Numerical Variables for Multivariate Analysis')  
plt.show()
```

