```python
import pandas as pd

# Load the CSV file into a pandas DataFrame
df = pd.read_csv('HousingPrices_New.csv')

# Display the first few rows of the DataFrame
df.head()
```

```
{"type":"dataframe","variable_name":"df"}
```

```python
import numpy as np

# Select the five relevant numerical columns
numerical_cols = ['Sale Price', 'Flat Area (in Sqft)', 'Lot Area (in
Sqft)', 'No of Bedrooms', 'No of Bathrooms']

# Calculate descriptive statistics for the selected columns
for col in numerical_cols:
    mean = np.mean(df[col])
    median = np.median(df[col])
    mode = df[col].mode()[0]  # Mode can have multiple values,
selecting the first
    std = np.std(df[col])
    min_val = np.min(df[col])
    max_val = np.max(df[col])

    print(f"Statistics for {col}:")
    print(f"  Mean: {mean}")
    print(f"  Median: {median}")
    print(f"  Mode: {mode}")
    print(f"  Standard Deviation: {std}")
    print(f"  Minimum: {min_val}")
    print(f"  Maximum: {max_val}")
    print("\n")
```

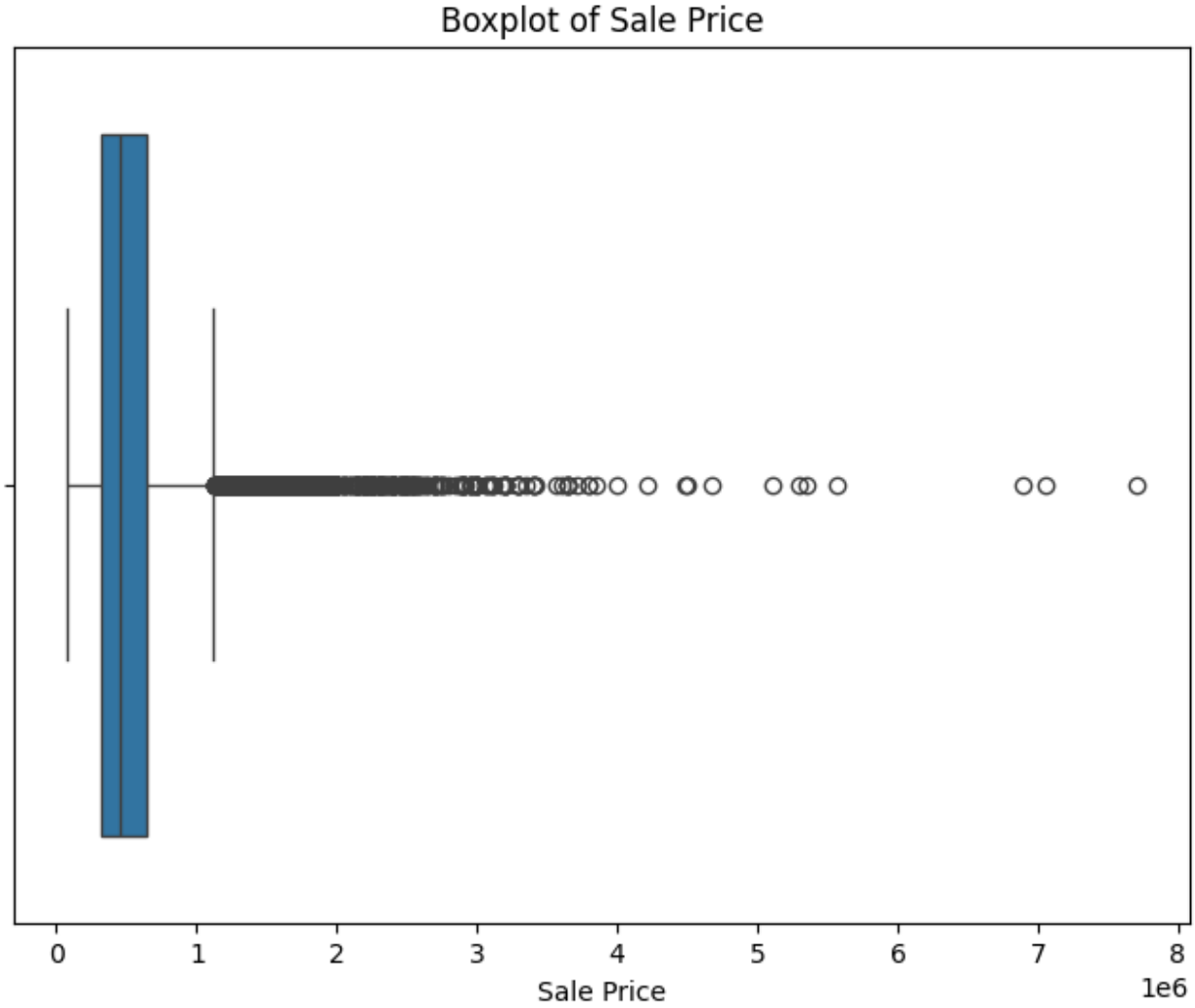
```
Statistics for Sale Price:
  Mean: 540198.4357443658
  Median: nan
  Mode: 350000.0
  Standard Deviation: 367380.4935294451
  Minimum: 75000.0
  Maximum: 7700000.0


Statistics for Flat Area (in Sqft):
  Mean: 2079.931771894094
  Median: nan
  Mode: 1300.0
```

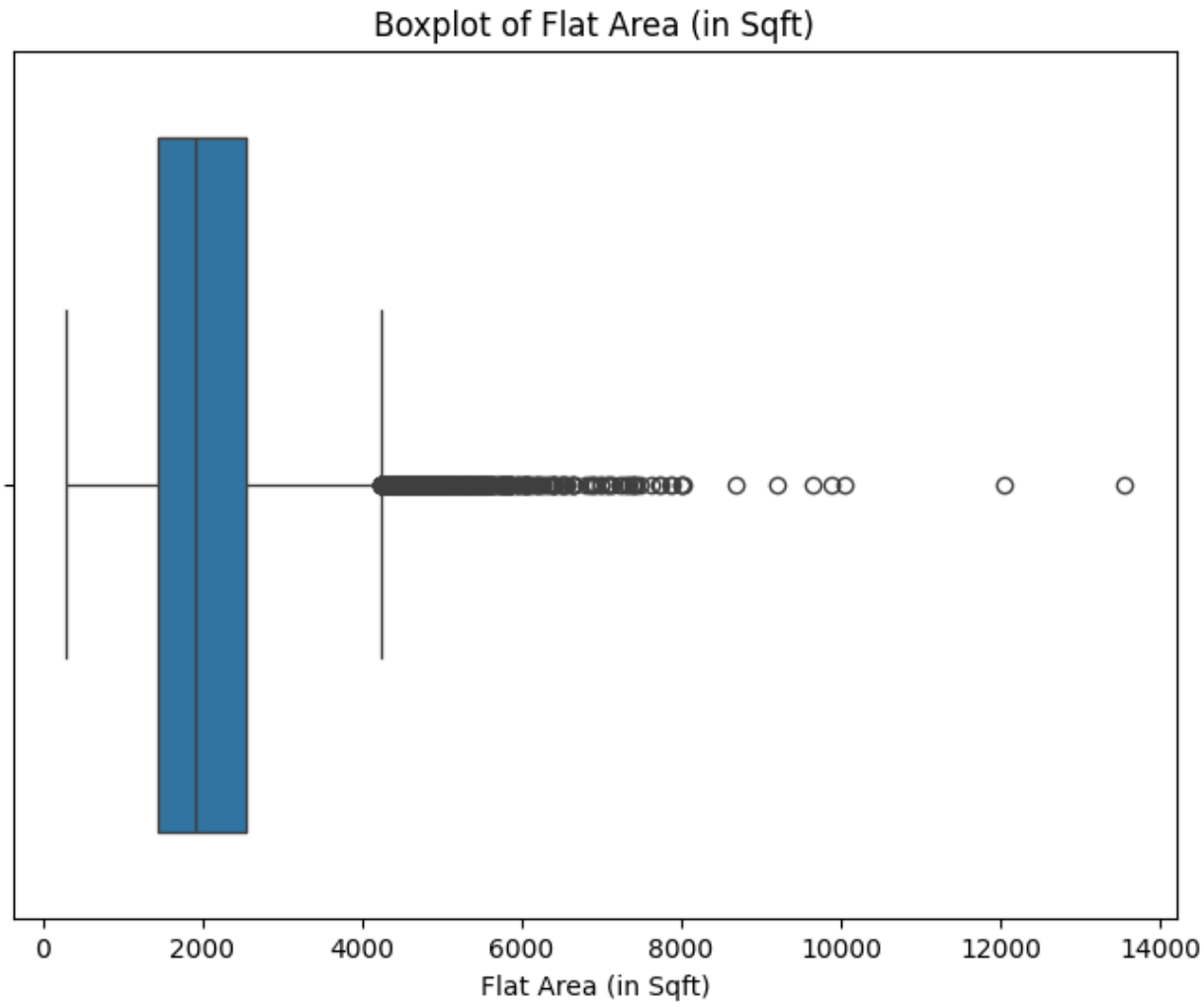
```
  Standard Deviation: 918.4663392228831
  Minimum: 290.0
  Maximum: 13540.0


Statistics for Lot Area (in Sqft):
  Mean: 15107.75689687095
  Median: nan
  Mode: 5000.0
  Standard Deviation: 41427.30736142591
  Minimum: 520.0
  Maximum: 1651359.0


Statistics for No of Bedrooms:
  Mean: 3.37084162309721
  Median: 3.0
  Mode: 3
  Standard Deviation: 0.9300403146391227
  Minimum: 0
  Maximum: 33


Statistics for No of Bathrooms:
  Mean: 2.1147322874728123
  Median: nan
  Mode: 2.5
  Standard Deviation: 0.7701205324604967
  Minimum: 0.0
  Maximum: 8.0
```

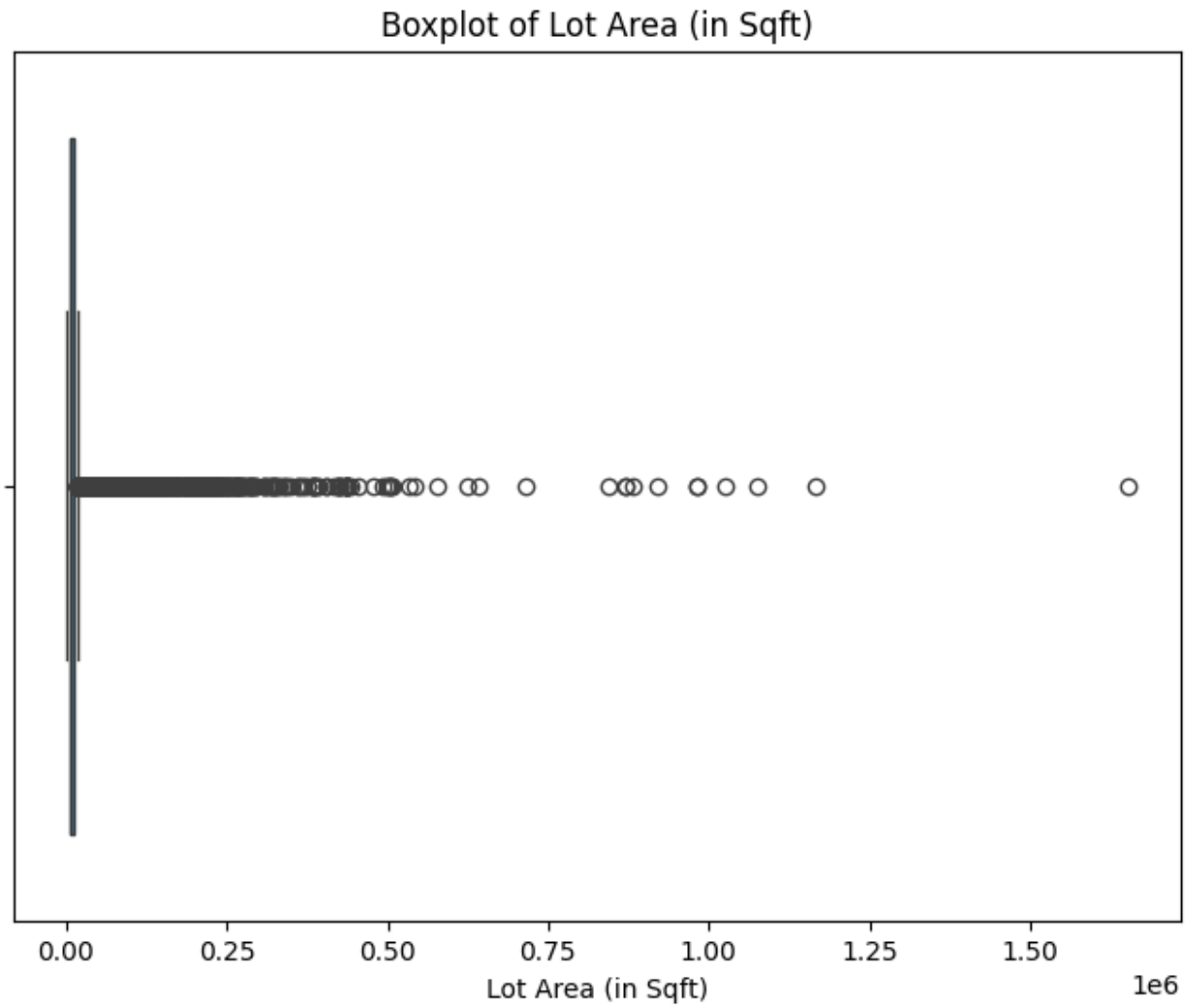
```python
import seaborn as sns
import matplotlib.pyplot as plt

# Check for missing values
print(df.isnull().sum())

# Detect outliers using boxplots
for col in numerical_cols:
    plt.figure(figsize=(8, 6))  # Adjust figure size as needed
    sns.boxplot(x=df[col])
    plt.title(f'Boxplot of {col}')
    plt.show()
```

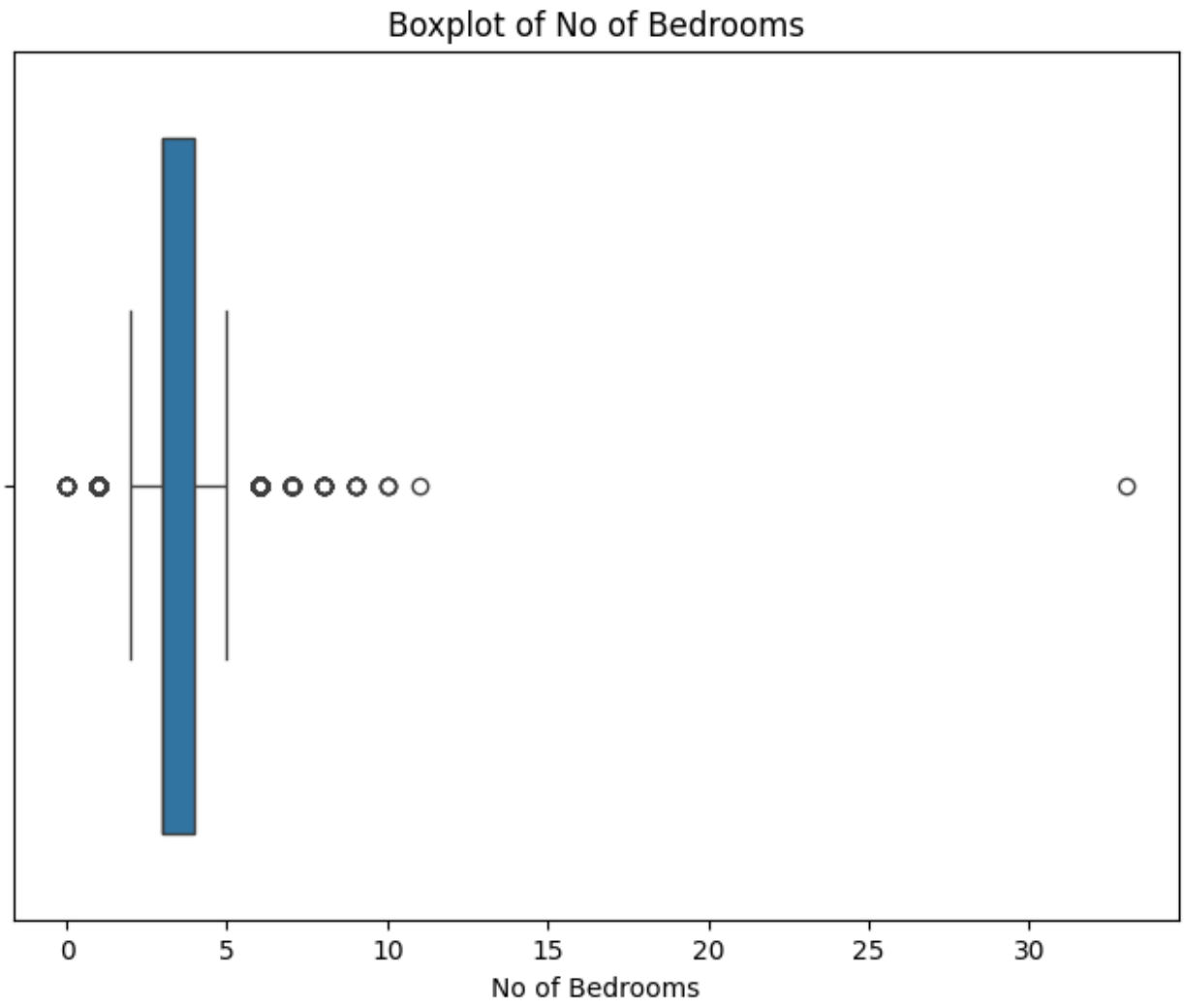
```
ID                                                0
Date House was Sold                               0
Sale Price                                        4
No of Bedrooms                                    0
No of Bathrooms                                   4
```

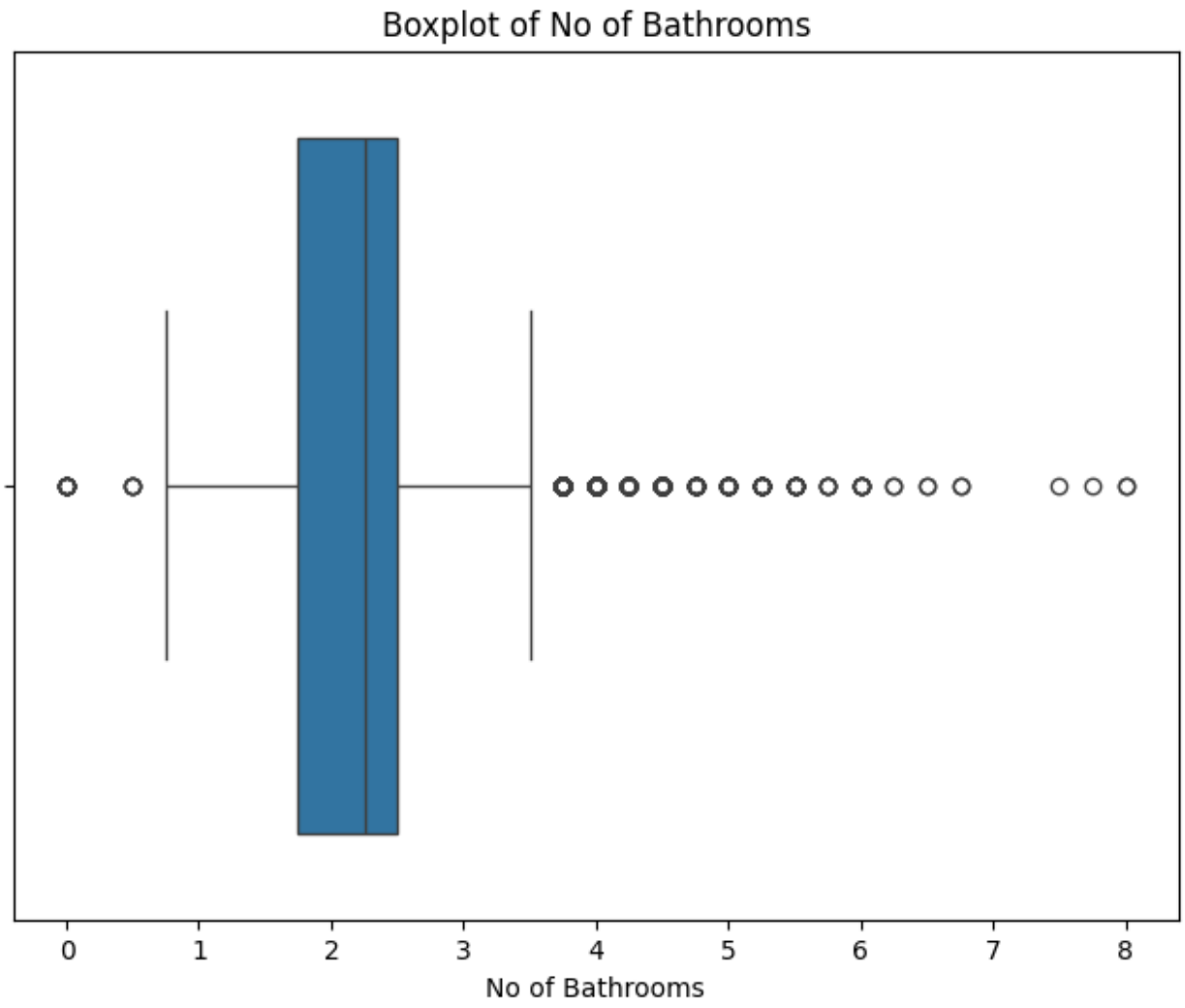
```
Flat Area (in Sqft)                                9
Lot Area (in Sqft)                                 9
No of Floors                                       0
Waterfront View                                    0
No of Times Visited                            19489
Condition of the House                             0
Overall Grade                                      0
Area of the House from Basement (in Sqft)          3
Basement Area (in Sqft)                            0
Age of House (in Years)                            0
Renovated Year                                     0
Zipcode                                            1
Latitude                                           1
Longitude                                          1
Living Area after Renovation (in Sqft)             1
Lot Area after Renovation (in Sqft)                0
dtype: int64
```



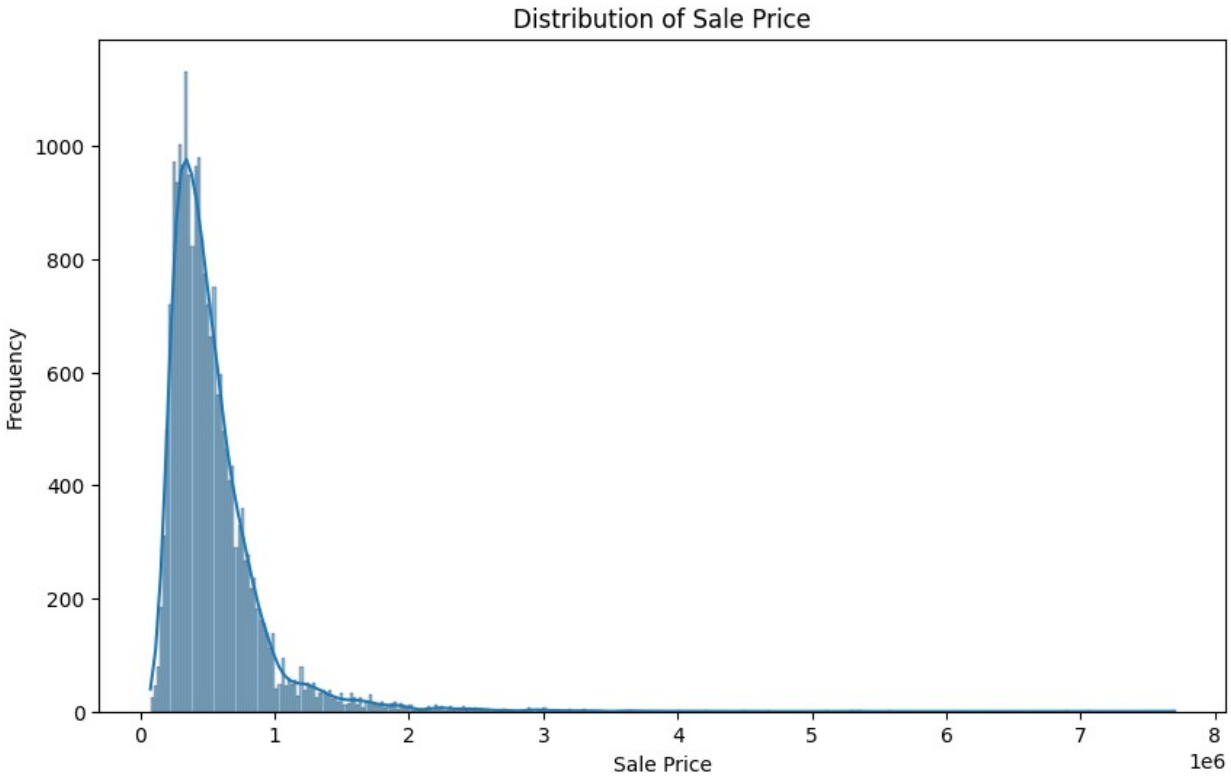

Boxplot of Flat Area (in Sqft)
0
2000
4000
6000
8000
10000
12000
14000
Flat Area (in Sqft)


Boxplot of Lot Area (in Sqft)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
Lot Area (in Sqft)
1e6

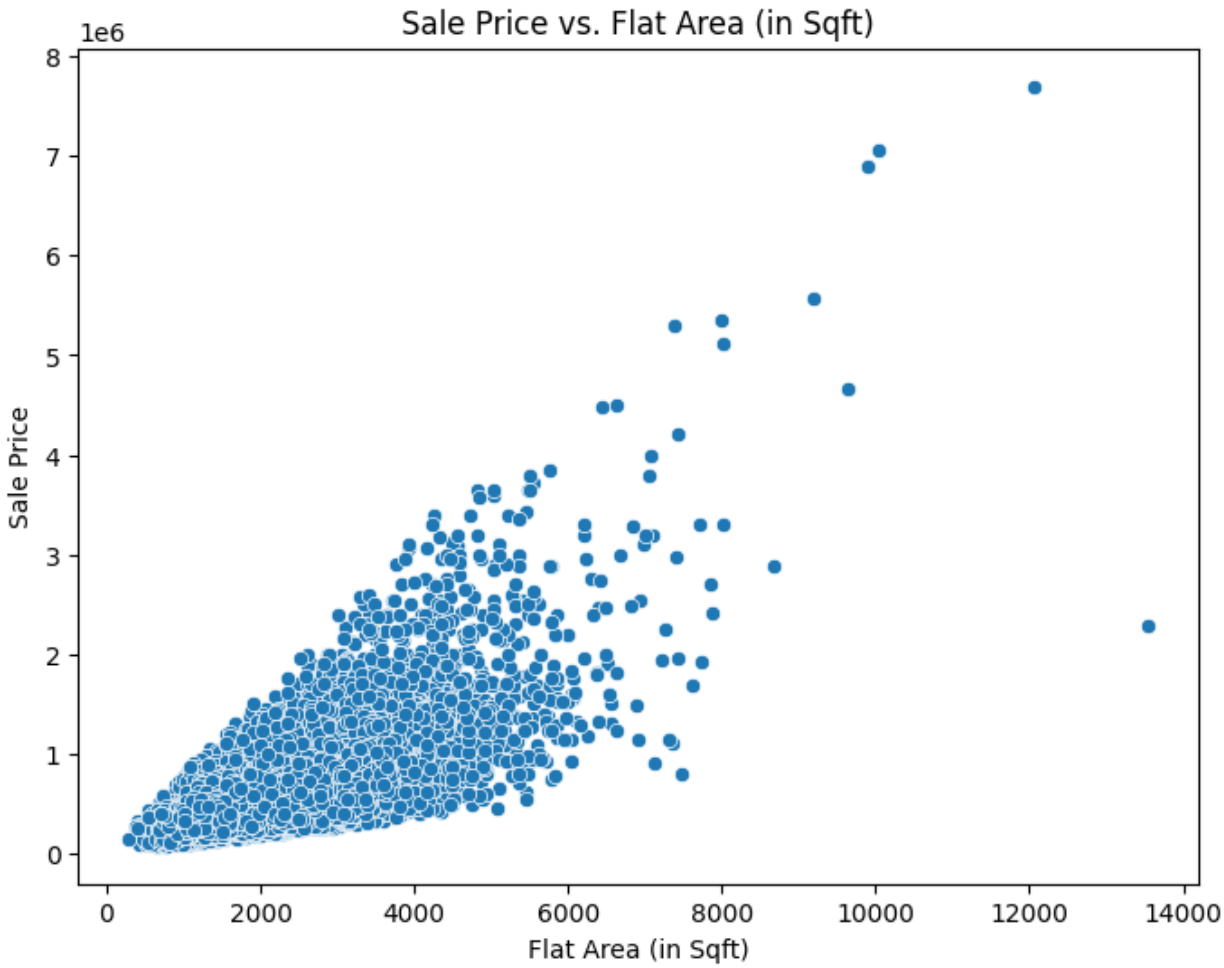

Boxplot of No of Bedrooms
0
5
10
15
20
25
30
No of Bedrooms

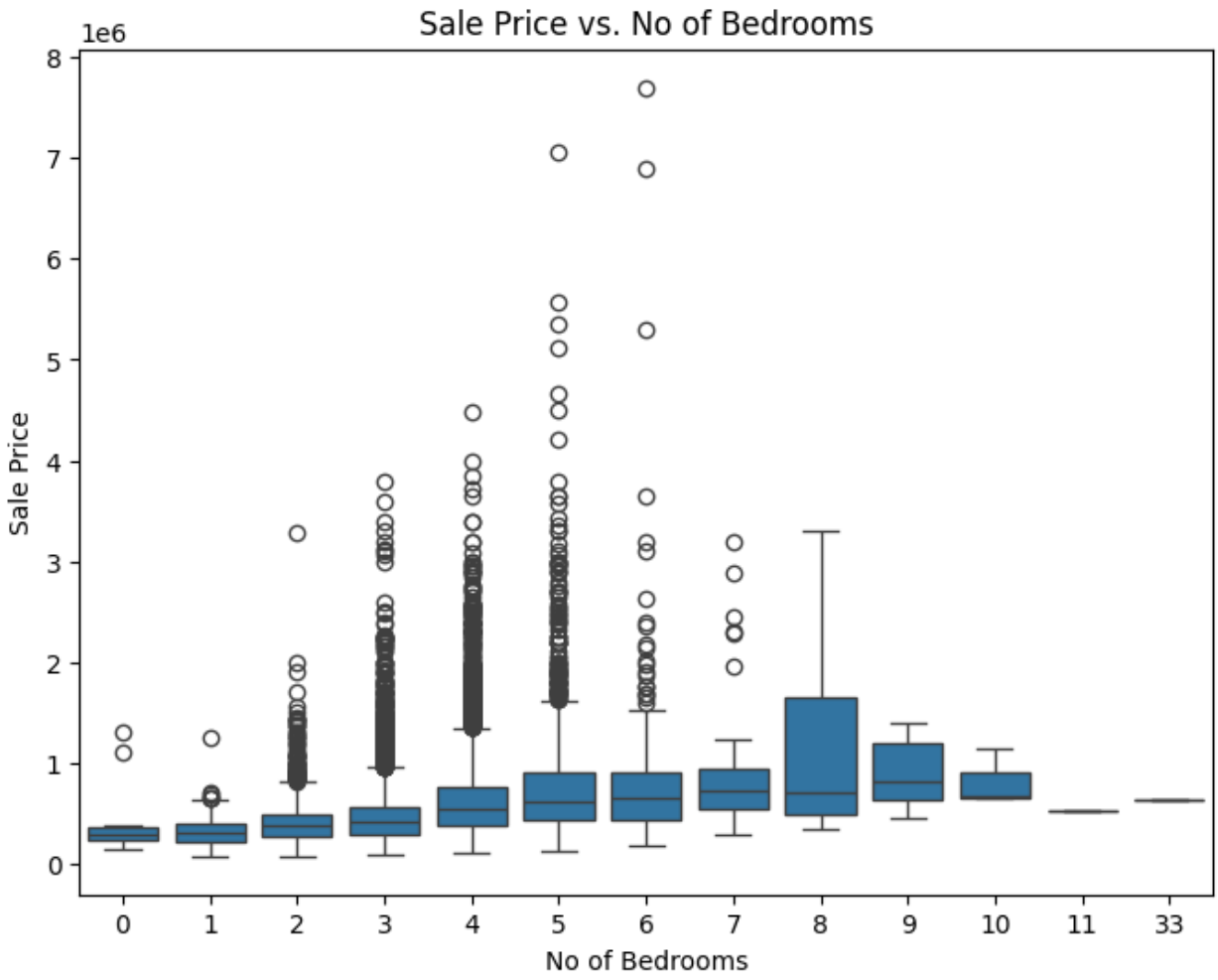



```
import seaborn as sns
import matplotlib.pyplot as plt

# Select the column for univariate analysis (e.g., 'Sale Price')
col = 'Sale Price'

# Plot histogram and KDE
plt.figure(figsize=(10, 6))
sns.histplot(df[col], kde=True)
plt.title(f'Distribution of {col}')
plt.xlabel(col)
plt.ylabel('Frequency')
plt.show()
```




```
import seaborn as sns
import matplotlib.pyplot as plt

# Scatterplot between 'Sale Price' and 'Flat Area (in Sqft)'
plt.figure(figsize=(8, 6))
sns.scatterplot(x='Flat Area (in Sqft)', y='Sale Price', data=df)
plt.title('Sale Price vs. Flat Area (in Sqft)')
plt.show()

# Boxplot between 'Sale Price' and 'No of Bedrooms'
plt.figure(figsize=(8, 6))
sns.boxplot(x='No of Bedrooms', y='Sale Price', data=df)
plt.title('Sale Price vs. No of Bedrooms')
plt.show()
```


Sale Price vs. Flat Area (in Sqft)
1e6
8
7
6
5
4
3
2
1
0
Sale Price
0
2000
4000
6000
8000
10000
12000
14000
Flat Area (in Sqft)


Sale Price vs. No of Bedrooms
1e6
8
7
6
5
4
3
2
1
0
Sale Price
0
1
2
3
4
5
6
7
8
9
10
11
33
No of Bedrooms