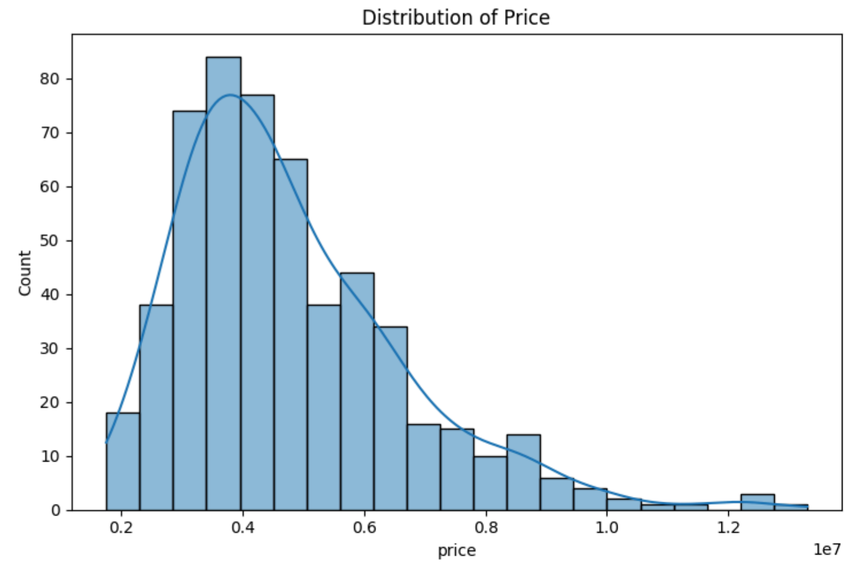
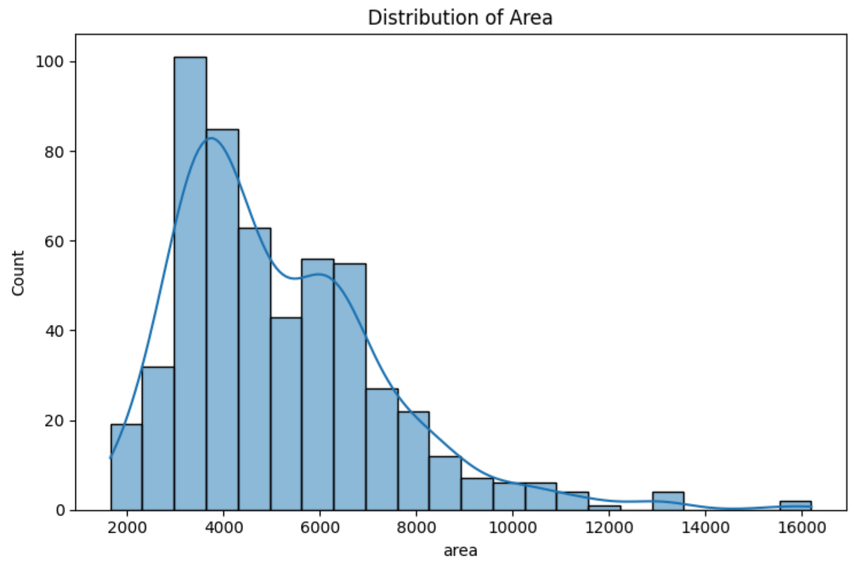
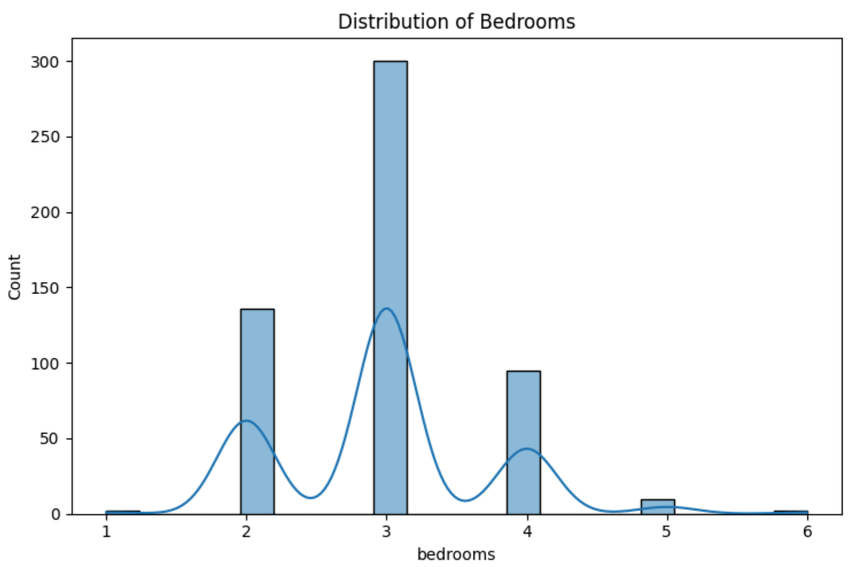
EDA (Exploratory Data Analysis) of Housing data

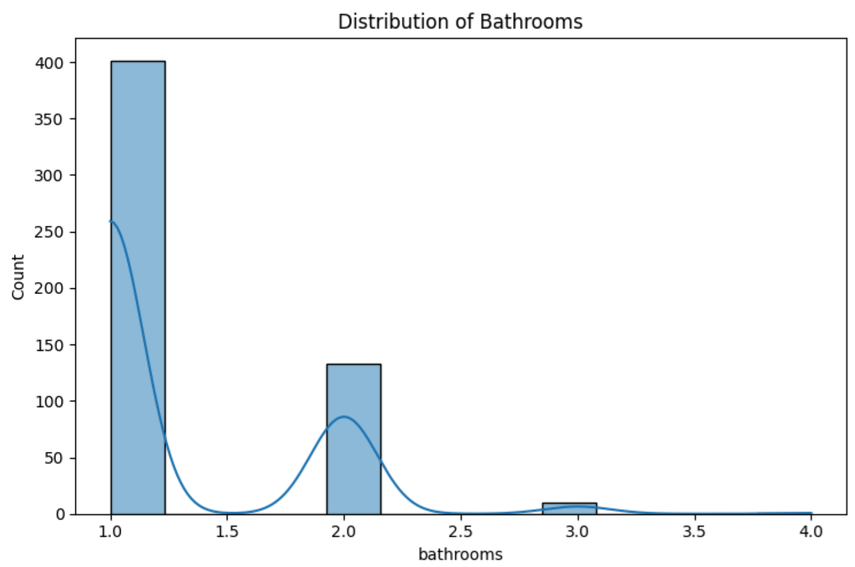
In [16]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
file_path = 'Housing.csv'
housing_data = pd.read_csv(file_path)

# Descriptive statistics for numerical columns
desc_stats = housing_data.describe()
print(desc_stats)

# Plot distribution of key numerical variables
fig, axes = plt.subplots(3, 2, figsize=(15, 15))
sns.histplot(housing_data['price'], kde=True, ax=axes[0, 0])
axes[0, 0].set_title('Distribution of Price')

sns.histplot(housing_data['area'], kde=True, ax=axes[0, 1])
axes[0, 1].set_title('Distribution of Area')

sns.histplot(housing_data['bedrooms'], kde=True, ax=axes[1, 0])
axes[1, 0].set_title('Distribution of Bedrooms')

sns.histplot(housing_data['bathrooms'], kde=True, ax=axes[1, 1])
axes[1, 1].set_title('Distribution of Bathrooms')

sns.histplot(housing_data['stories'], kde=True, ax=axes[2, 0])
axes[2, 0].set_title('Distribution of Stories')

sns.histplot(housing_data['parking'], kde=True, ax=axes[2, 1])
axes[2, 1].set_title('Distribution of Parking')

plt.tight_layout()
plt.show()

# Convert Symbolic to numerical values
housing_data['mainroad'] = housing_data['mainroad'].map({'no': 0, 'yes': 1})
housing_data['guestroom'] = housing_data['guestroom'].map({'no': 0, 'yes': 1})
housing_data['basement'] = housing_data['basement'].map({'no': 0, 'yes': 1})
housing_data['hotwaterheating'] = housing_data['hotwaterheating'].map({'no': 0, 'yes': 1})
housing_data['airconditioning'] = housing_data['airconditioning'].map({'no': 0, 'yes': 1})
housing_data['prefarea'] = housing_data['prefarea'].map({'no': 0, 'yes': 1})
housing_data['furnishingstatus'] = housing_data['furnishingstatus'].map({'unfurnished': 0, 'semi-furnished': 1, 'furnished': 2})


# Compute the correlation matrix
correlation_matrix = housing_data.corr()

# Plot the heatmap for correlation matrix
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()

# Relationship Analysis: Scatter plots with regression lines
fig, axes = plt.subplots(3, 2, figsize=(15, 18))

sns.regplot(x='area', y='price', data=housing_data, ax=axes[0, 0])
axes[0, 0].set_title('Price vs Area')

sns.regplot(x='bedrooms', y='price', data=housing_data, ax=axes[0, 1])
axes[0, 1].set_title('Price vs Bedrooms')

sns.regplot(x='bathrooms', y='price', data=housing_data, ax=axes[1, 0])
axes[1, 0].set_title('Price vs Bathrooms')

sns.regplot(x='stories', y='price', data=housing_data, ax=axes[1, 1])
axes[1, 1].set_title('Price vs Stories')

sns.regplot(x='parking', y='price', data=housing_data, ax=axes[2, 0])
axes[2, 0].set_title('Price vs Parking')

plt.tight_layout()
plt.show()

# Categorical Variables Analysis
fig, axes = plt.subplots(4, 2, figsize=(15, 20))

sns.boxplot(x='mainroad', y='price', data=housing_data, ax=axes[0, 0])
axes[0, 0].set_title('Price vs Mainroad')

sns.boxplot(x='guestroom', y='price', data=housing_data, ax=axes[0, 1])
axes[0, 1].set_title('Price vs Guestroom')

sns.boxplot(x='basement', y='price', data=housing_data, ax=axes[1, 0])
axes[1, 0].set_title('Price vs Basement')

sns.boxplot(x='hotwaterheating', y='price', data=housing_data, ax=axes[1, 1])
axes[1, 1].set_title('Price vs Hotwaterheating')

sns.boxplot(x='airconditioning', y='price', data=housing_data, ax=axes[2, 0])
axes[2, 0].set_title('Price vs Airconditioning')

sns.boxplot(x='prefarea', y='price', data=housing_data, ax=axes[2, 1])
axes[2, 1].set_title('Price vs Prefarea')

sns.boxplot(x='furnishingstatus', y='price', data=housing_data, ax=axes[3, 0])
axes[3, 0].set_title('Price vs Furnishingstatus')

plt.tight_layout()
plt.show()
```

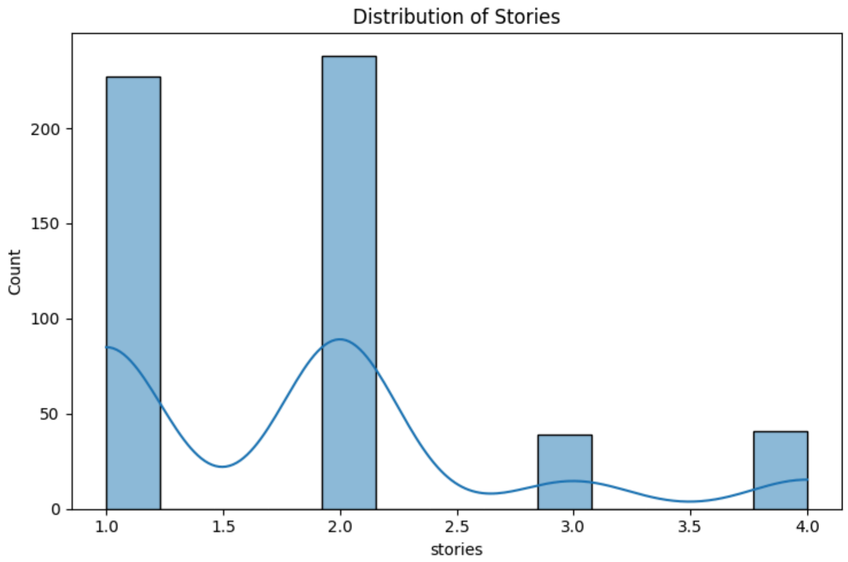
```
              price          area    bedrooms   bathrooms     stories  \
count  5.450000e+02    545.000000  545.000000  545.000000  545.000000
mean   4.766729e+06   5150.541284    2.965138    1.286239    1.805505
std    1.870440e+06   2170.141023    0.738064    0.502470    0.867492
min    1.750000e+06   1650.000000    1.000000    1.000000    1.000000
25%    3.430000e+06   3600.000000    2.000000    1.000000    1.000000
50%    4.340000e+06   4600.000000    3.000000    1.000000    2.000000
75%    5.740000e+06   6360.000000    3.000000    2.000000    2.000000
max    1.330000e+07  16200.000000    6.000000    4.000000    4.000000

          parking
count  545.000000
mean     0.693578
std      0.861586
min      0.000000
25%      0.000000
50%      0.000000
75%      1.000000
max      3.000000
```

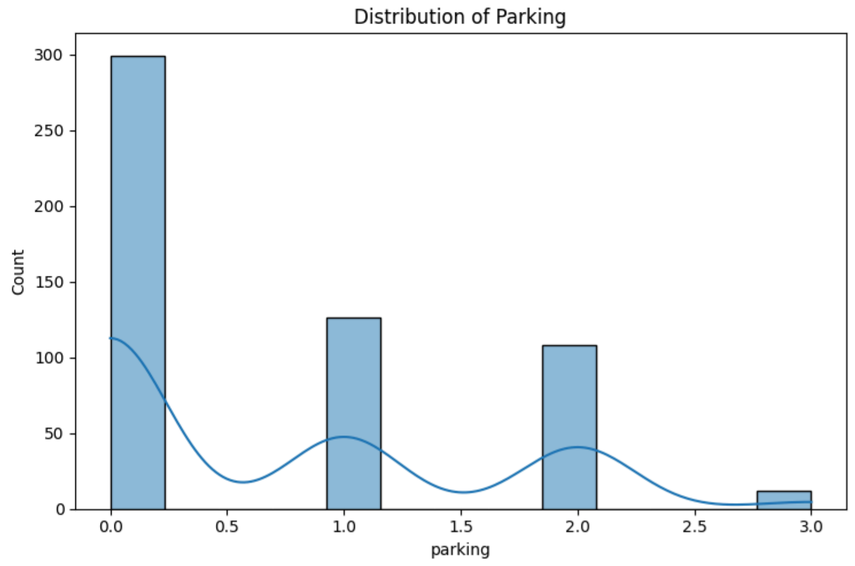

Distribution of Price
Count
price
1e7

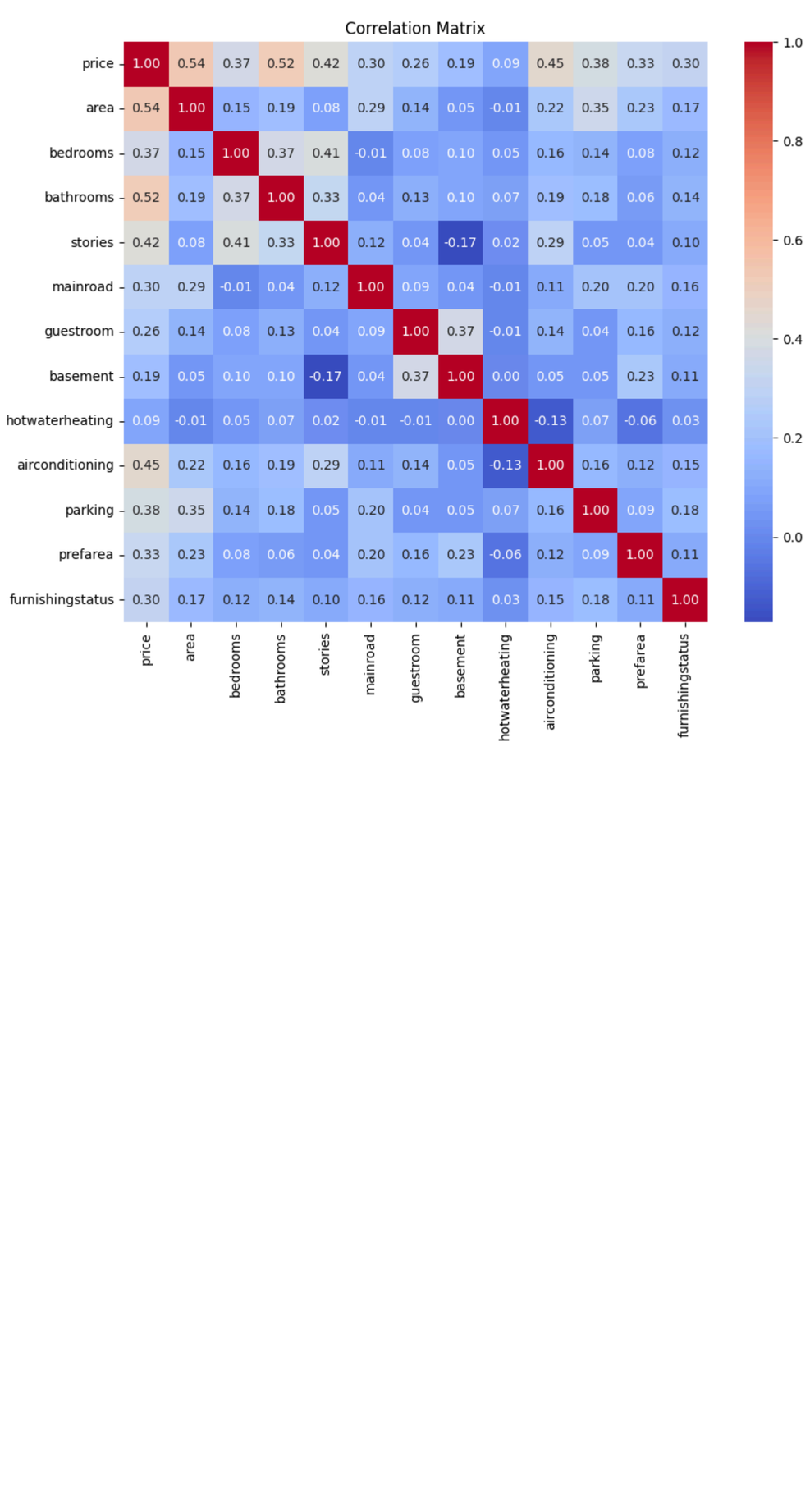

Distribution of Area
Count
area


Distribution of Bedrooms
Count
bedrooms

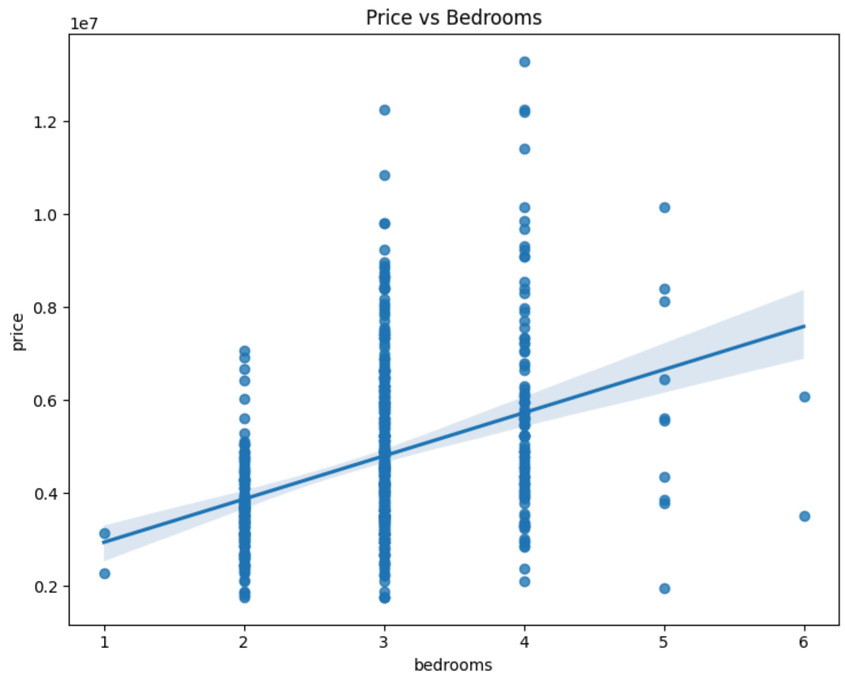

Distribution of Bathrooms
Count
bathrooms

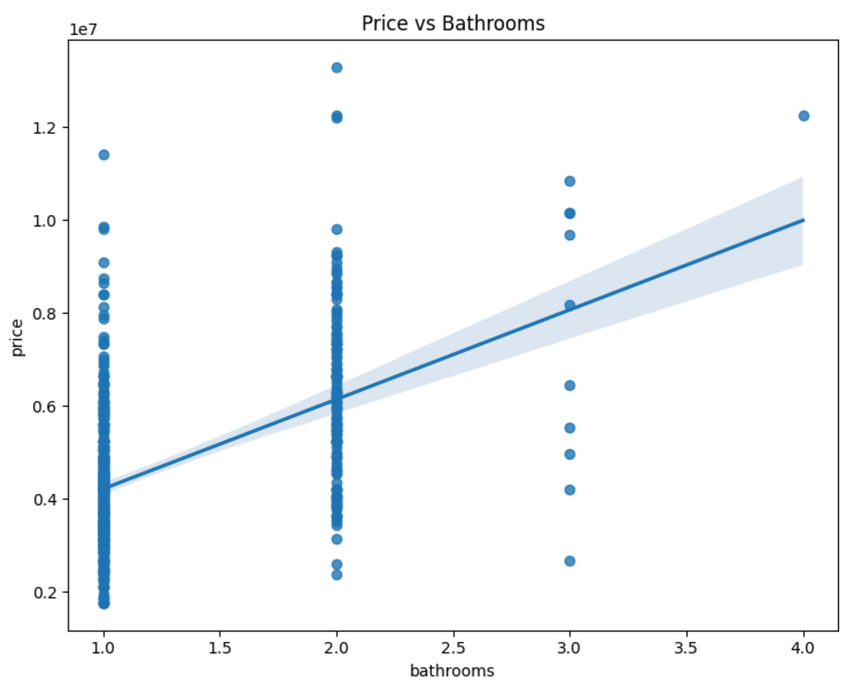

Distribution of Stories
Count
stories

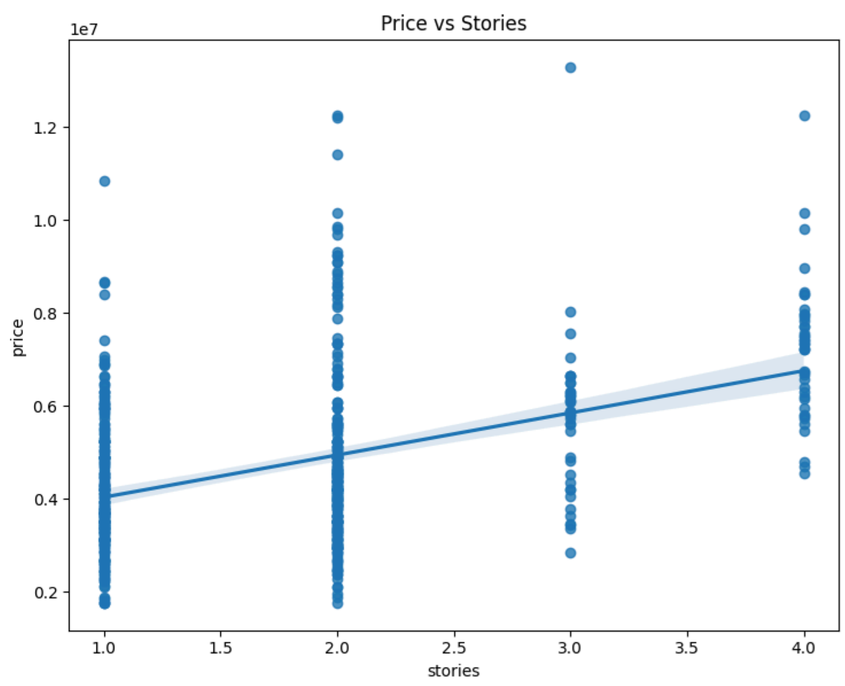

Distribution of Parking
Count
parking

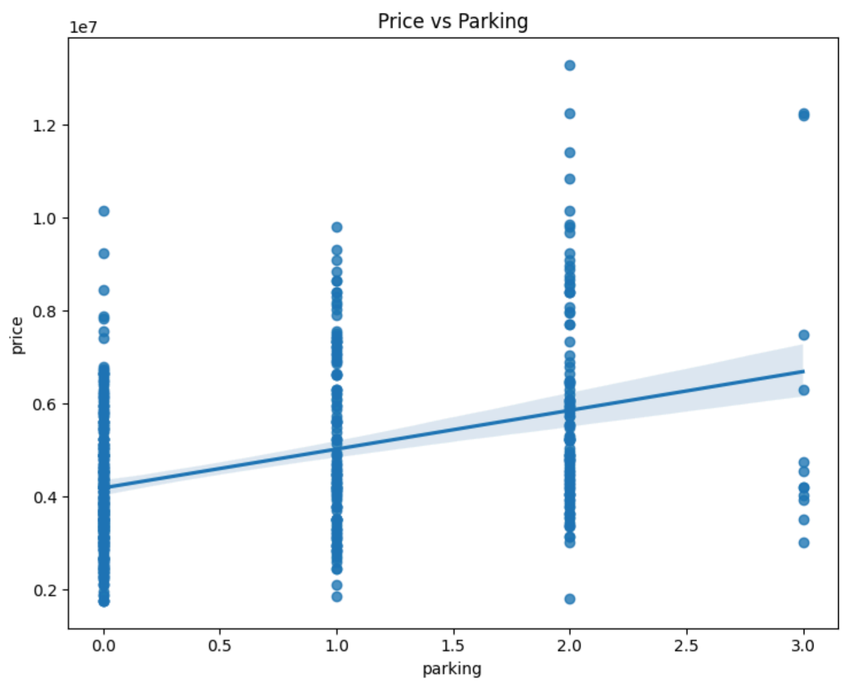

Correlation Matrix
price 1.00 0.54 0.37 0.52 0.42 0.30 0.26 0.19 0.09 0.45 0.38 0.33 0.30
area 0.54 1.00 0.15 0.19 0.08 0.29 0.14 0.05 -0.01 0.22 0.35 0.23 0.17
bedrooms 0.37 0.15 1.00 0.37 0.41 -0.01 0.08 0.10 0.05 0.16 0.14 0.08 0.12
bathrooms 0.52 0.19 0.37 1.00 0.33 0.04 0.13 0.10 0.07 0.19 0.18 0.06 0.14
stories 0.42 0.08 0.41 0.33 1.00 0.12 0.04 -0.17 0.02 0.29 0.05 0.04 0.10
mainroad 0.30 0.29 -0.01 0.04 0.12 1.00 0.09 0.04 -0.01 0.11 0.20 0.20 0.16
guestroom 0.26 0.14 0.08 0.13 0.04 0.09 1.00 0.37 -0.01 0.14 0.04 0.16 0.12
basement 0.19 0.05 0.10 0.10 -0.17 0.04 0.37 1.00 0.00 0.05 0.05 0.23 0.11
hotwaterheating 0.09 -0.01 0.05 0.07 0.02 -0.01 -0.01 0.00 1.00 -0.13 0.07 -0.06 0.03
airconditioning 0.45 0.22 0.16 0.19 0.29 0.11 0.14 0.05 -0.13 1.00 0.16 0.12 0.15
parking 0.38 0.35 0.14 0.18 0.05 0.20 0.04 0.05 0.07 0.16 1.00 0.09 0.18
prefarea 0.33 0.23 0.08 0.06 0.04 0.20 0.16 0.23 -0.06 0.12 0.09 1.00 0.11
furnishingstatus 0.30 0.17 0.12 0.14 0.10 0.16 0.12 0.11 0.03 0.15 0.18 0.11 1.00
price area bedrooms bathrooms stories mainroad guestroom basement hotwaterheating airconditioning parking prefarea furnishingstatus

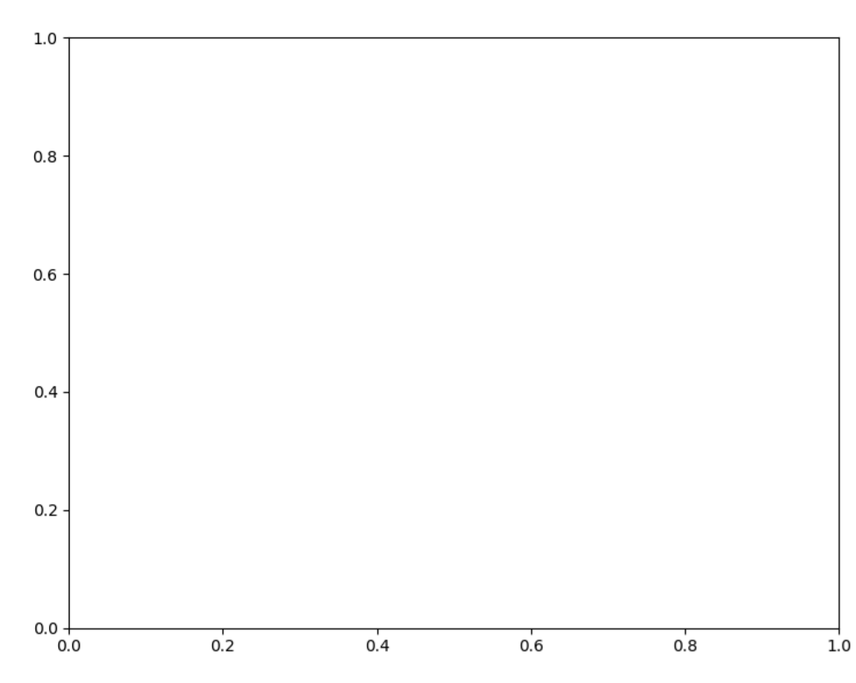

Price vs Area
1e7
price
area

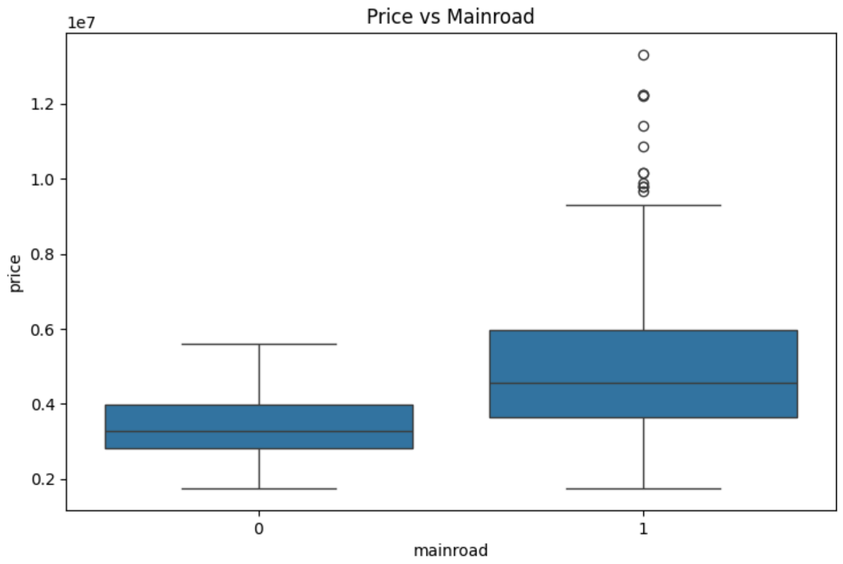

Price vs Bedrooms
1e7
price
bedrooms

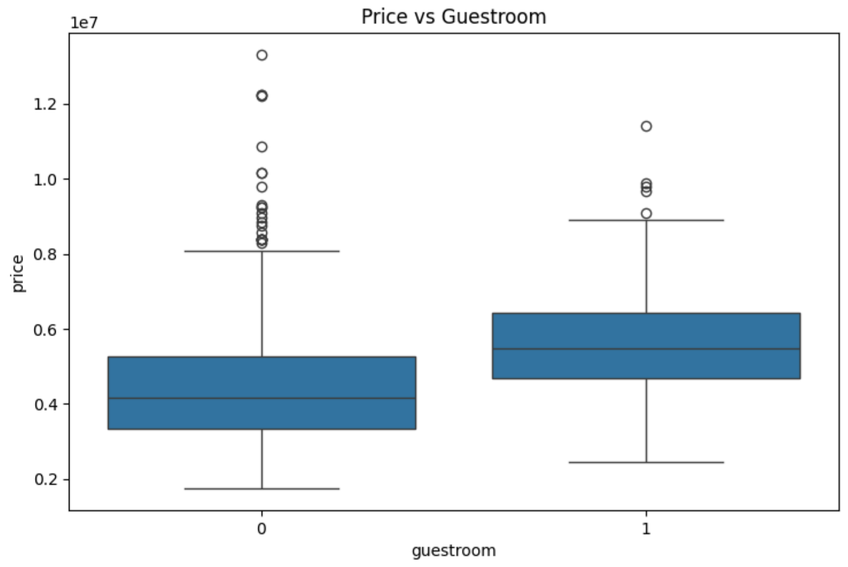

Price vs Bathrooms
1e7
price
bathrooms

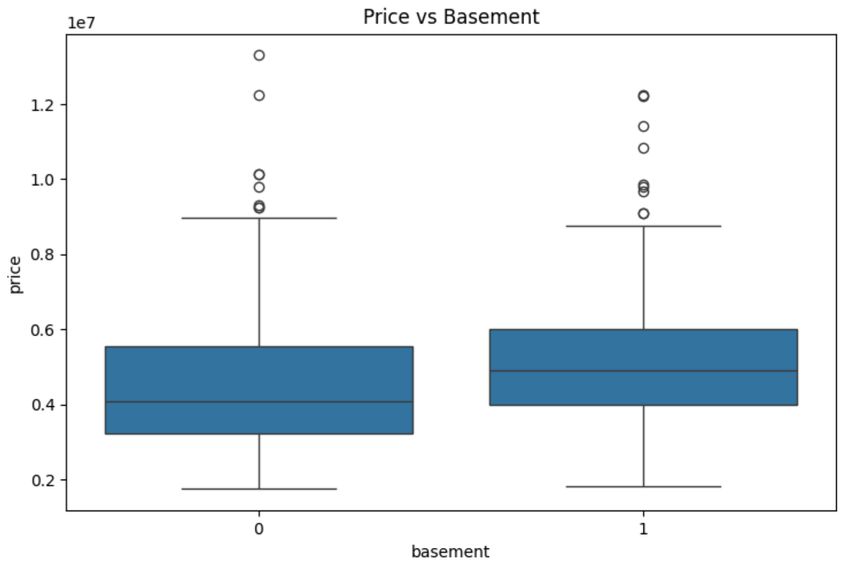

Price vs Stories
1e7
price
stories

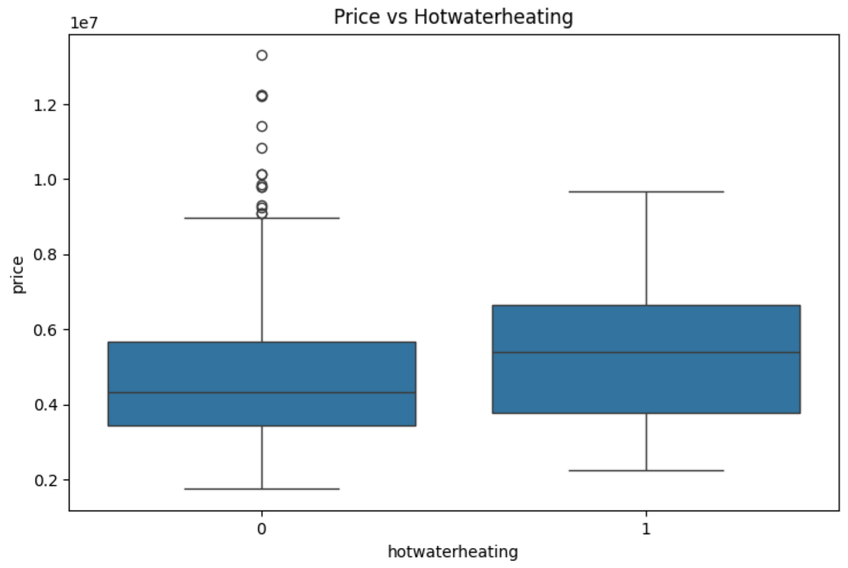

Price vs Parking
1e7
price
parking

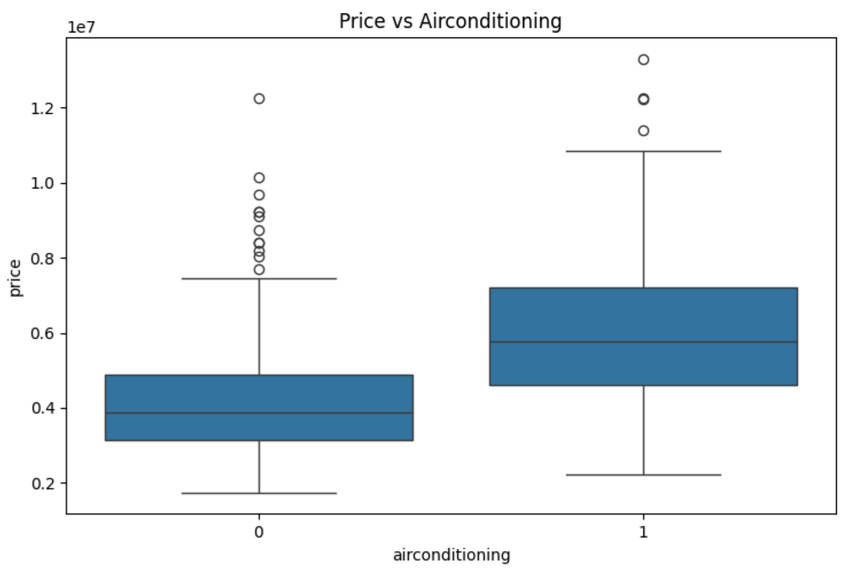


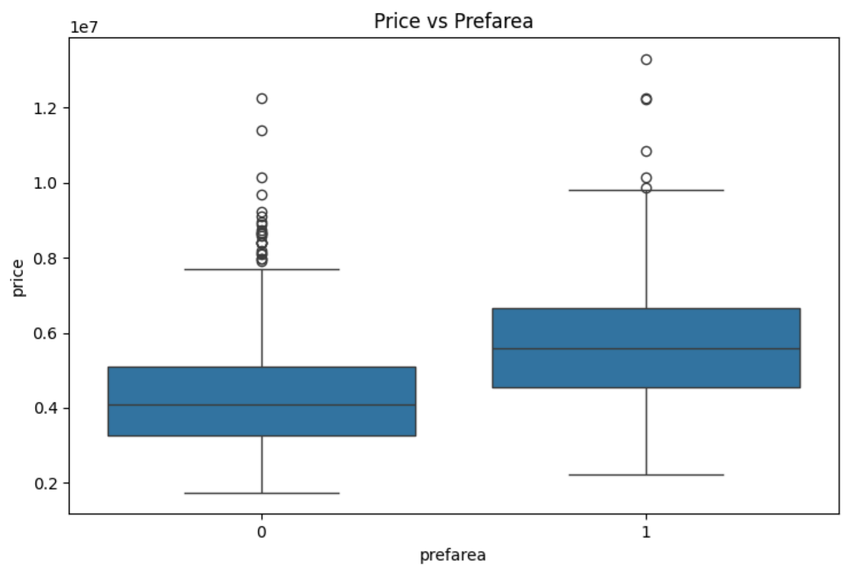

Price vs Mainroad
1e7
price
mainroad

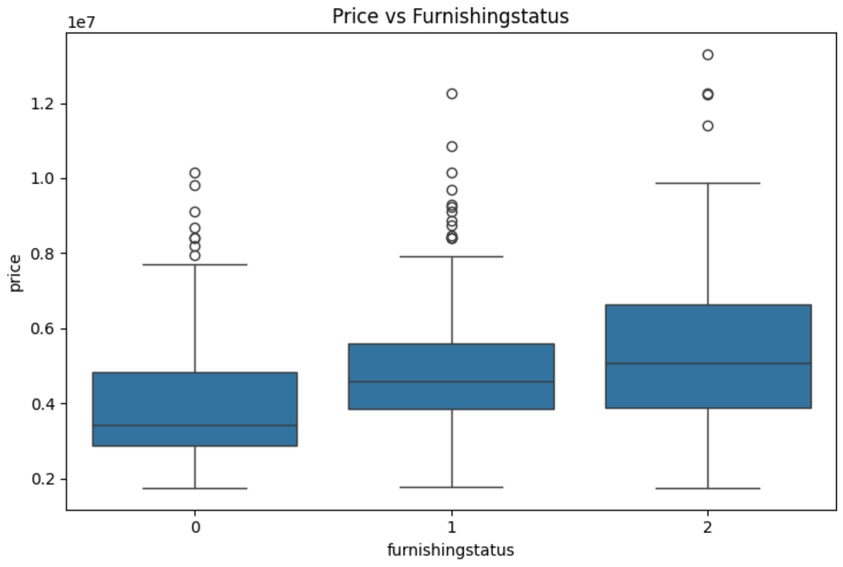

Price vs Guestroom
1e7
price
guestroom


Price vs Basement
1e7
price
basement


Price vs Hotwaterheating
1e7
price
hotwaterheating


Price vs Airconditioning
1e7
price
airconditioning


Price vs Prefarea
1e7
price
prefarea


Price vs Furnishingstatus
1e7
price
furnishingstatus