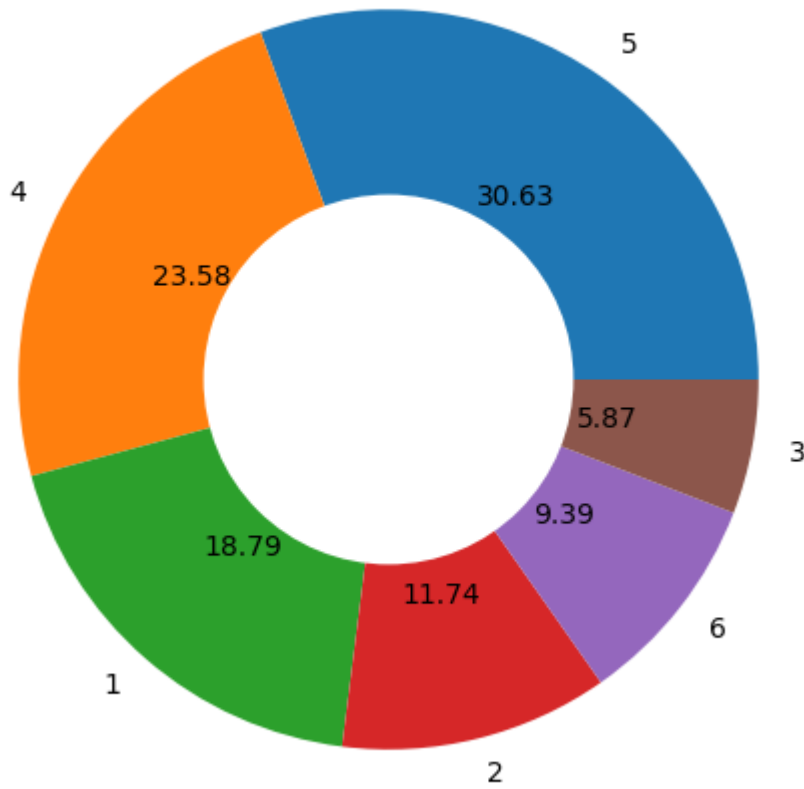
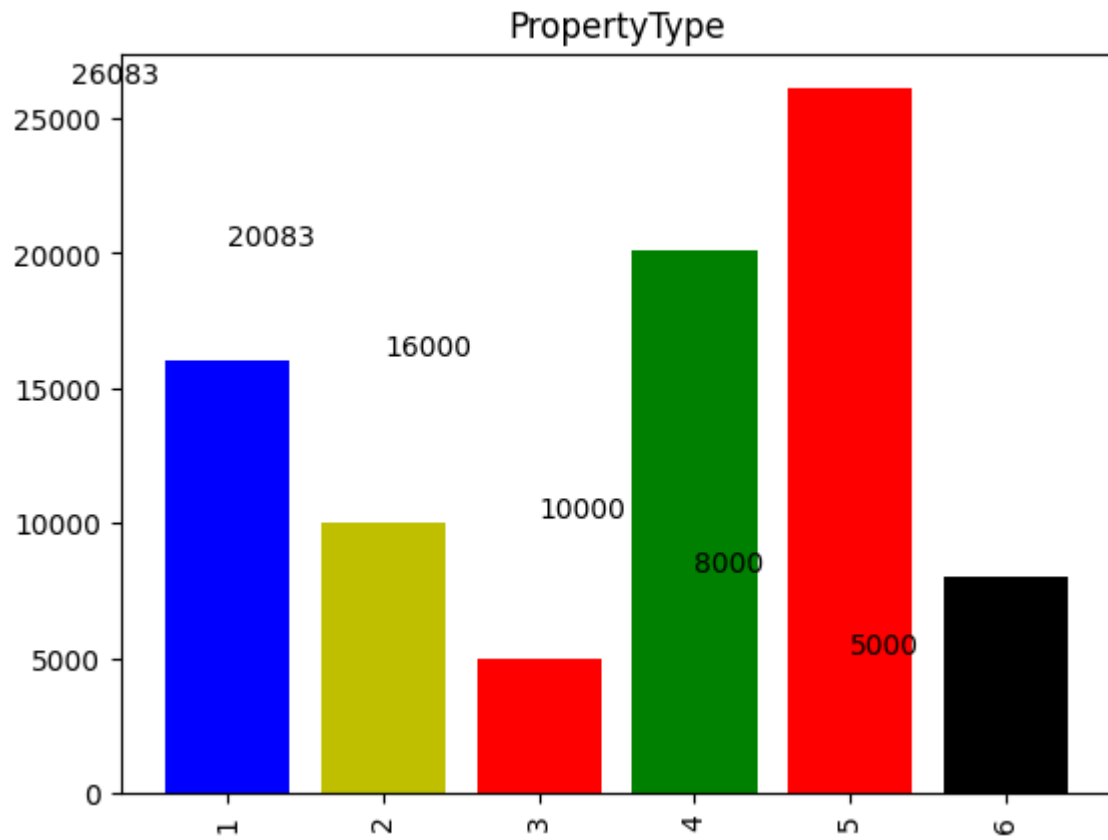


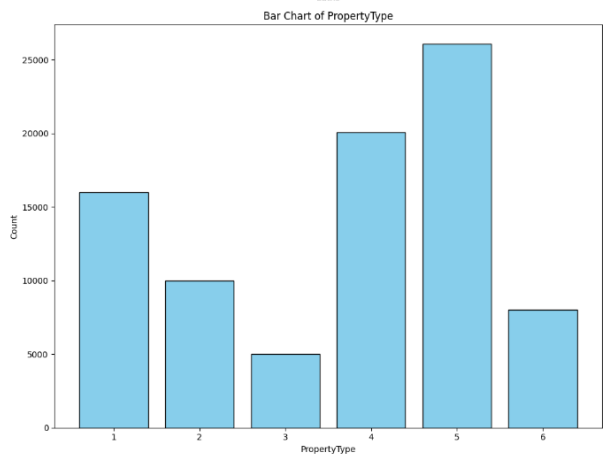
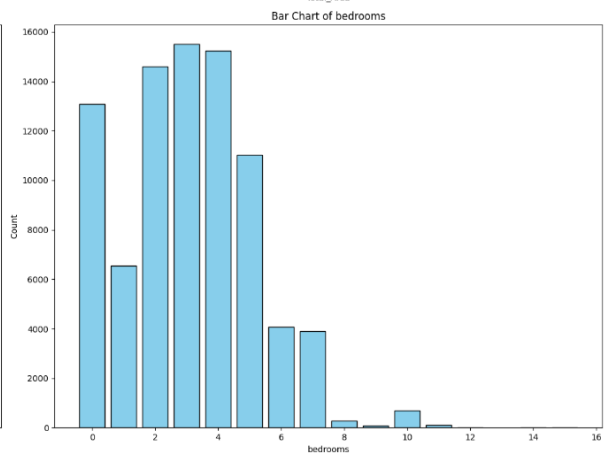
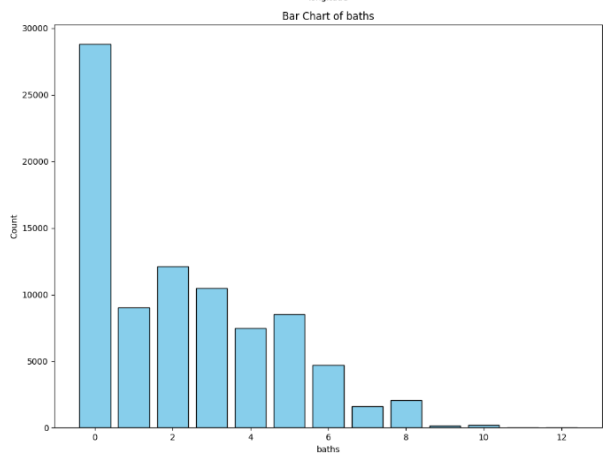
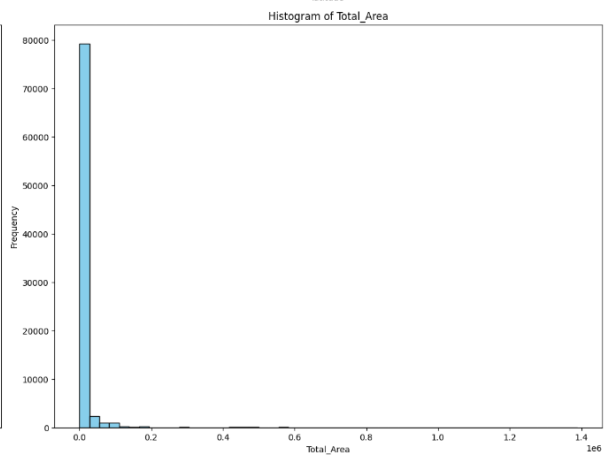
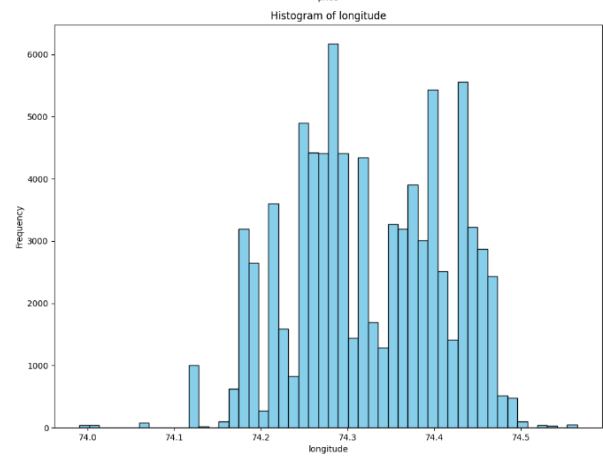
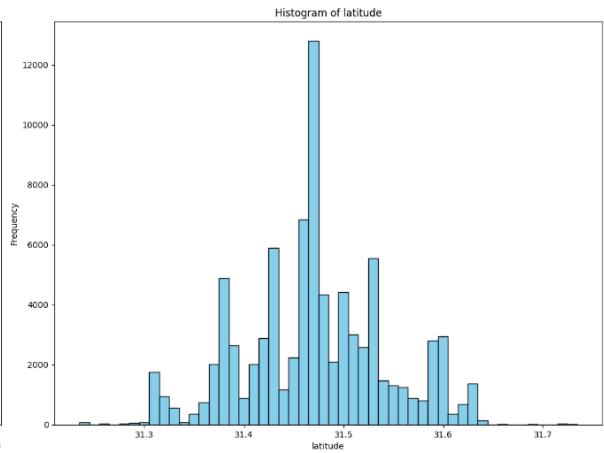
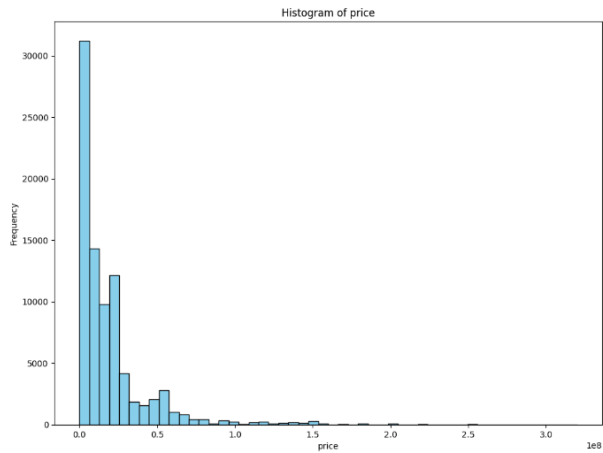
Univariate Analysis:

Property Type :

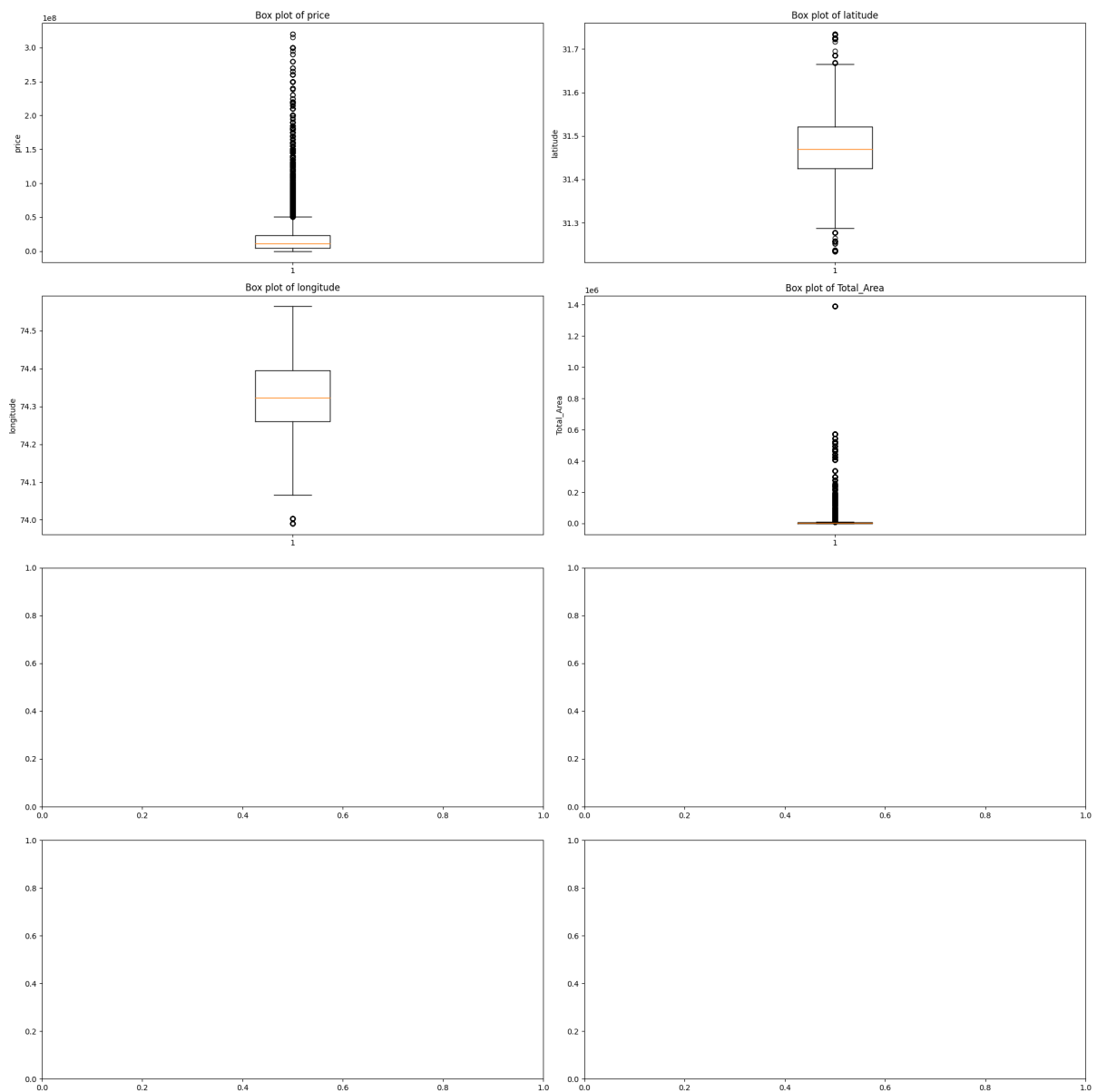




Histograms & Bar Charts : Plotting histogram for continuous variables and bar chart for discrete variables.



Box plots for all variables



Distribution, Central Tendency, and Variability :

Distribution, central tendency, and variability is given below:

1. Price:

Distribution: The distribution is right-skewed, meaning most of the properties are priced at the lower end with a few properties being significantly expensive.

Central Tendency:

Mean (Average): Approximately 19,905,990

Median (50th percentile): 11,000,000

Variability:

Range: 1 to 320,000,000

Standard Deviation: Approximately 27,068,300, indicating a wide spread in property prices.

2. Latitude:

Distribution: The distribution has multiple peaks, suggesting properties come from different regions within a confined geographical area.

Central Tendency:

Mean: Approximately 31.47

Median: 31.47

Variability:

Range: 31.23 to 31.73

Standard Deviation: 0.0726

3. Longitude:

Distribution: The values are clustered, indicating properties are primarily from specific regions.

Central Tendency:

Mean: Approximately 74.32

Median: 74.32

Variability:

Range: 73.99 to 74.56

Standard Deviation: 0.088

4. Baths:

Distribution: Most properties have between 0 to 4 bathrooms. The distribution decreases as the number of bathrooms increases.

Central Tendency:

Mean: Around 2.3

Median: 2

Variability:

Range: 0 to 12

Standard Deviation: 2.28

5. Bedrooms:

Distribution: Most properties have between 2 to 4 bedrooms.

Central Tendency:

Mean: Around 3.07

Median: 3

Variability:

Range: 0 to 15

Standard Deviation: 2.08

6. Total_Area:

Distribution: Right-skewed. Most properties have a smaller area, with a few having a significantly larger area.

Central Tendency:

Mean: Approximately 12,648 (likely square units)

Median: 2,722.5

Variability:

Range: 0 to 1,388,475

Standard Deviation: 54,733.37

7. PropertyType:

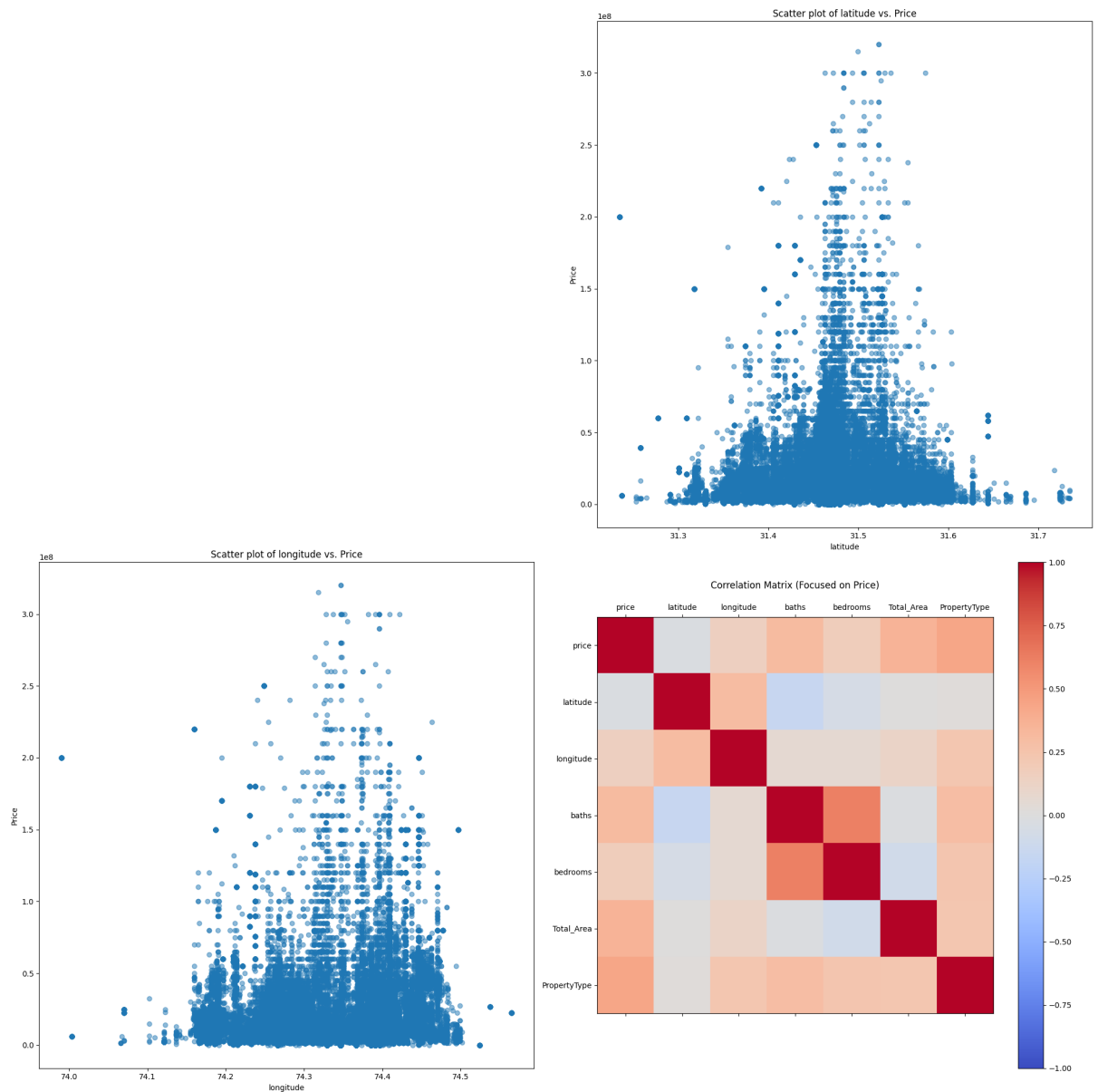
Distribution: The dataset contains properties of types 1 to 6, with differing counts for each type.

Central Tendency: Given it's a categorical variable, mode (most common value) is more relevant. Type 5 is the most common.

Variability: Types range from 1 to 6, with different counts for each type due to our oversampling strategy.

These metrics give a comprehensive understanding of each variable's characteristics. The central tendency provides a sense of the "average" or "middle" of the data, while the variability indicates how spread out the data is around this centre. The distribution gives a broader perspective on the shape and characteristics of the data.

Bivariate Analysis



Scatter Plots:

These plots visualize the relationship between the PropertyType (our main label class) and other continuous variables.

price vs. PropertyType: There's a noticeable trend where higher property prices are associated with specific property types.

latitude & longitude vs. PropertyType: The scatter plots hint at certain geographical areas being associated with particular property types. For instance, specific latitudes and longitudes are densely populated with certain property types.

Total_Area vs. PropertyType: Certain property types seem to be associated with larger areas, while others are clustered around smaller areas.

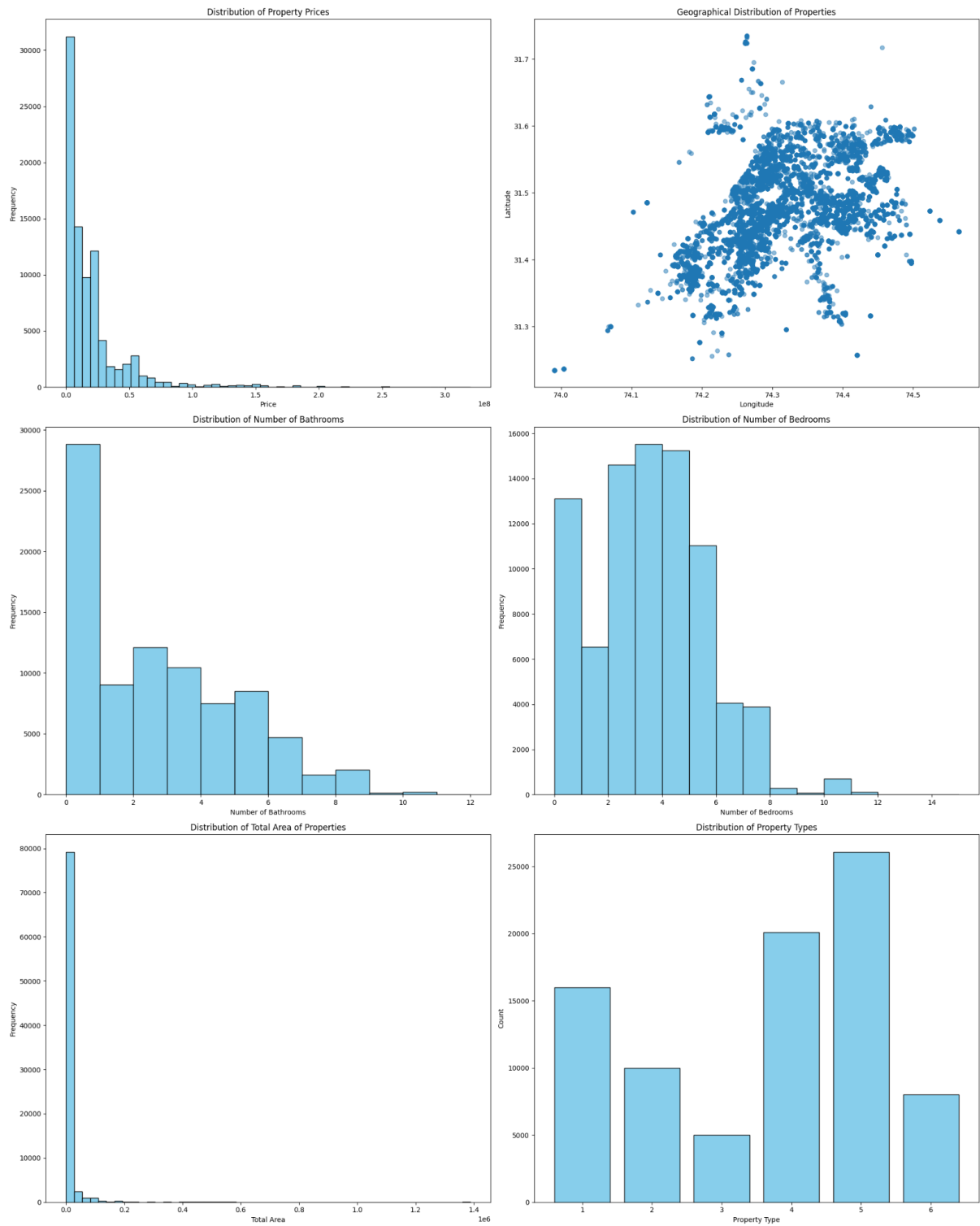
Correlation Matrix:

The correlation matrix provides a quantitative measure of the linear relationship between variables.

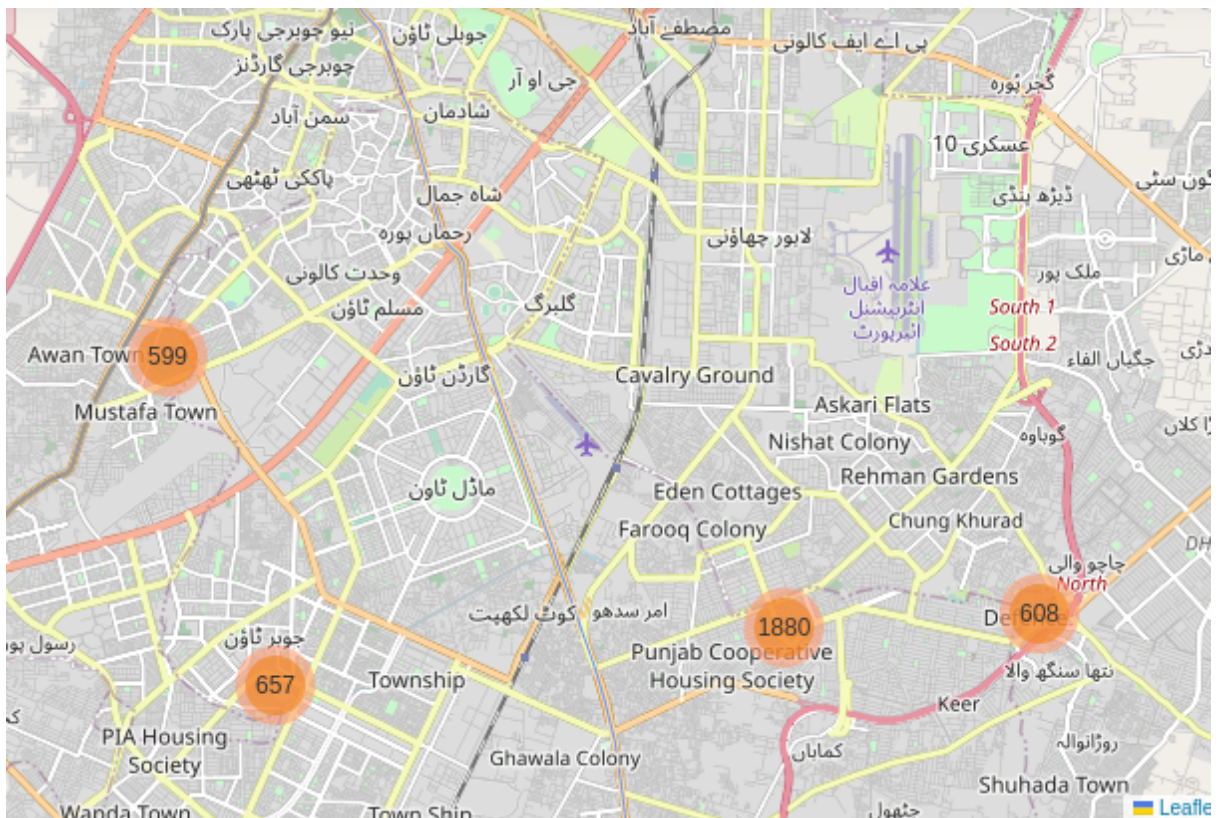
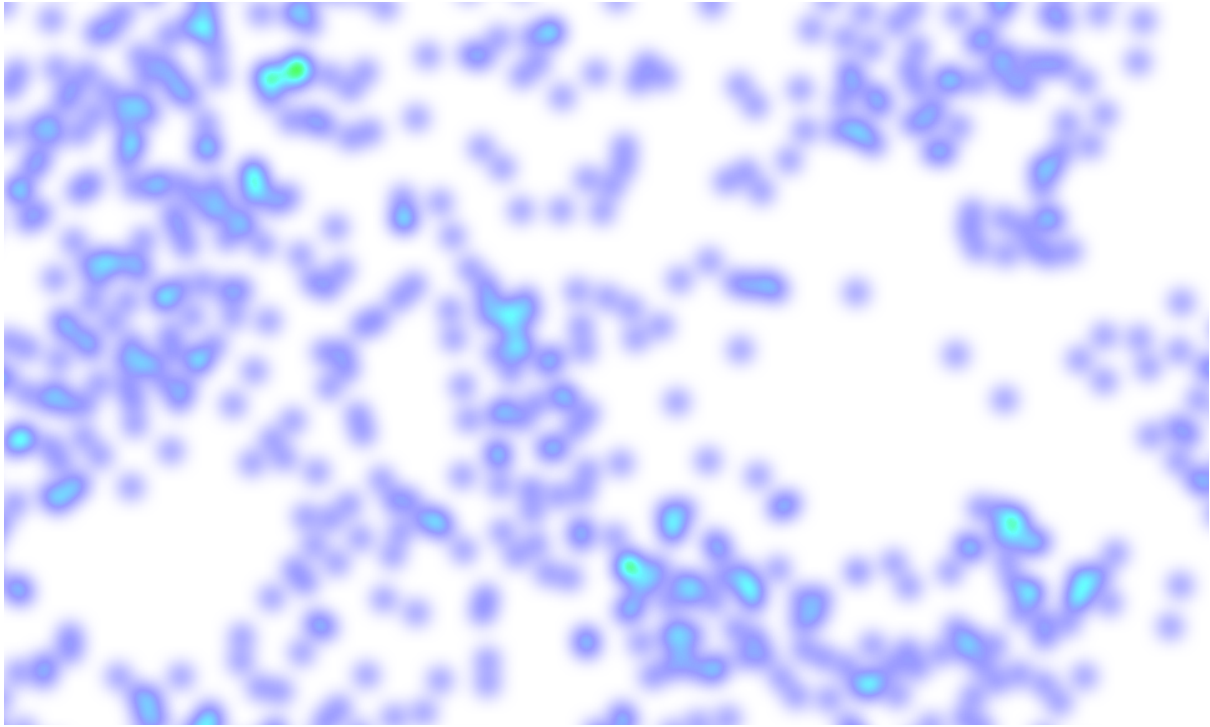
- **PropertyType & price:** A moderate positive correlation of 0.44 suggests that as the property type changes, the price tends to increase.
- **PropertyType & baths/bedrooms/Total_Area:** Positive correlations in the range of 0.23 to 0.29 indicate that certain property types tend to have more baths, bedrooms, and a larger total area.
- **PropertyType & longitude:** A positive correlation of 0.23 suggests a relationship between property location (longitude) and its type.
- **PropertyType & latitude:** A very weak correlation of 0.02, which might not be significant.

Descriptive Analysis:

Visualization



Longitude and Latitude map :



Descriptive Analysis:

1. price:

- **Description:** This represents the price of the property.
- **Insights:**
 - The property prices in the dataset vary widely.
 - Most properties are priced on the lower end, with a few properties having significantly higher prices.

2. latitude and longitude:

- **Description:** These represent the geographical coordinates of the property.
- **Insights:**
 - The properties are confined to a specific geographical region.
 - Certain latitudes and longitudes are densely populated, indicating specific popular areas or neighbourhoods.

3. baths:

- **Description:** This represents the number of bathrooms in the property.
- **Insights:**
 - Most properties have between 0 to 4 bathrooms.
 - A significant number of properties have no bathrooms, which could either be plots, commercial spaces, or missing data.

4. bedrooms:

- **Description:** This represents the number of bedrooms in the property.
- **Insights:**
 - The majority of properties have between 2 to 4 bedrooms.
 - There are some properties with a high number of bedrooms, which are rare.

5. Total_Area:

- **Description:** This represents the total area of the property.
- **Insights:**
 - Most properties have a smaller area, with a few properties having a much larger area.
 - The vast majority of properties have areas below 20,000 units, suggesting that larger properties are rare.

6. PropertyType:

- **Description:** This represents the type of property.
- **Insights:**
 - There are 6 different types of properties in the dataset.
 - Type 5 properties are the most common, but this distribution was adjusted based on our oversampling strategy.

Here are the visualizations that provide a descriptive analysis of the dataset:

Distribution of Property Prices:

- The histogram for property prices shows that most properties are on the lower end of the price spectrum. There's a sharp decline as the price increases, which is indicative of the right-skewed distribution.

Geographical Distribution of Properties:

- The scatter plot for latitude and longitude displays the concentration of properties. It's evident that there are specific regions with a higher density of properties.

Distribution of Number of Bathrooms:

- The majority of properties have between 0 to 4 bathrooms. There's a drop in frequency as the number of bathrooms increases.

Distribution of Number of Bedrooms:

- Most properties have between 1 to 5 bedrooms. Very few properties have more than 7 bedrooms.

Distribution of Total Area of Properties:

- The vast majority of properties have a total area on the lower end. The distribution is right-skewed with a few properties having a significantly larger area.

Distribution of Property Types:

- This bar chart showcases the number of properties for each type. It's evident that Type 5 properties dominate, followed by Types 4, 1, 2, 6, and 3.