STATISTICAL ANALYSIS ON HOUSING DATABASE
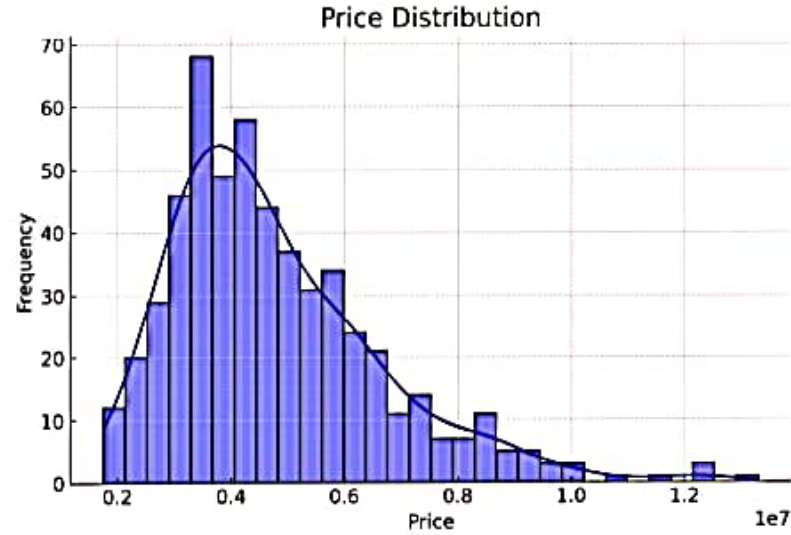
By Suman raj

Figure 1

2.2 Correlations

- **Price** has a strong positive correlation with:
    - **Area** ($\rho = 0.54$)
    - **Number of bathrooms** ($\rho = 0.51$)
    - **Number of stories** ($\rho = 0.41$)
- **Price** is weakly correlated with **parking spaces** ($\rho = 0.26$).
- **Visualization:**
    - A heatmap (Figure 2) reveals these relationships clearly, helping to identify the key features influencing price.
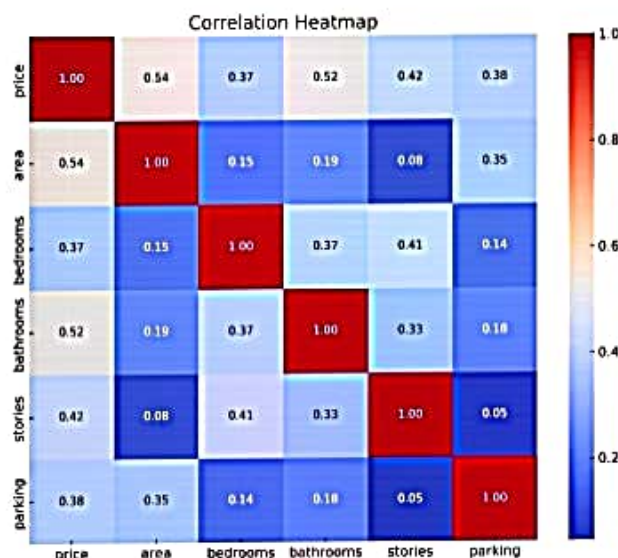

Figure 2

2.3 Influence of Furnishing Status on Price

- **Furnished houses** have the highest median price compared to semi-furnished and unfurnished properties.
- This trend suggests that furnishing status significantly impacts buyer preferences and property valuation.
- **Visualization:**
    - A boxplot (Figure 3) shows the distribution of prices for each furnishing status category, emphasizing the premium associated with furnished houses.
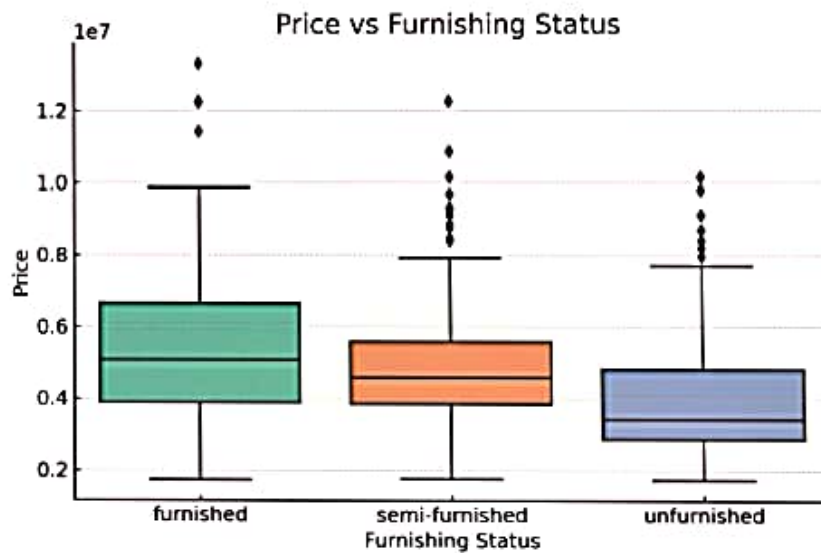
Figure 3

2.4 Mainroad Access and Furnishing Status

- Houses with **mainroad access** are predominantly furnished or semi-furnished, indicating a possible association between accessibility and property features.

- **Visualization:**

  - A countplot (Figure 4) highlights the frequency distribution of mainroad access across different furnishing statuses.
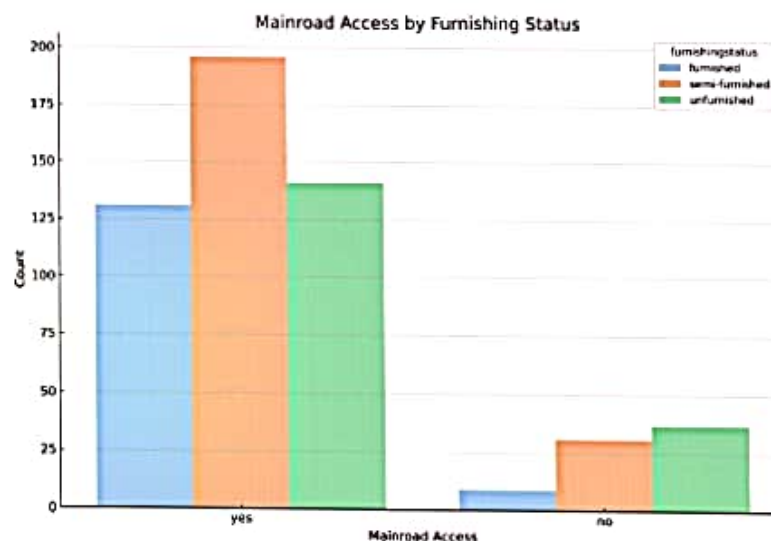

Figure 4

2.5 Pairwise Relationships Among Numerical Features

- Significant trends observed include:

4

  - Larger areas are generally associated with higher prices.
  - Properties with more bathrooms and stories tend to be more expensive.

- **Visualization:**

  - A pairplot (Figure 5) showcases these relationships, providing a holistic view of how numerical features interact.

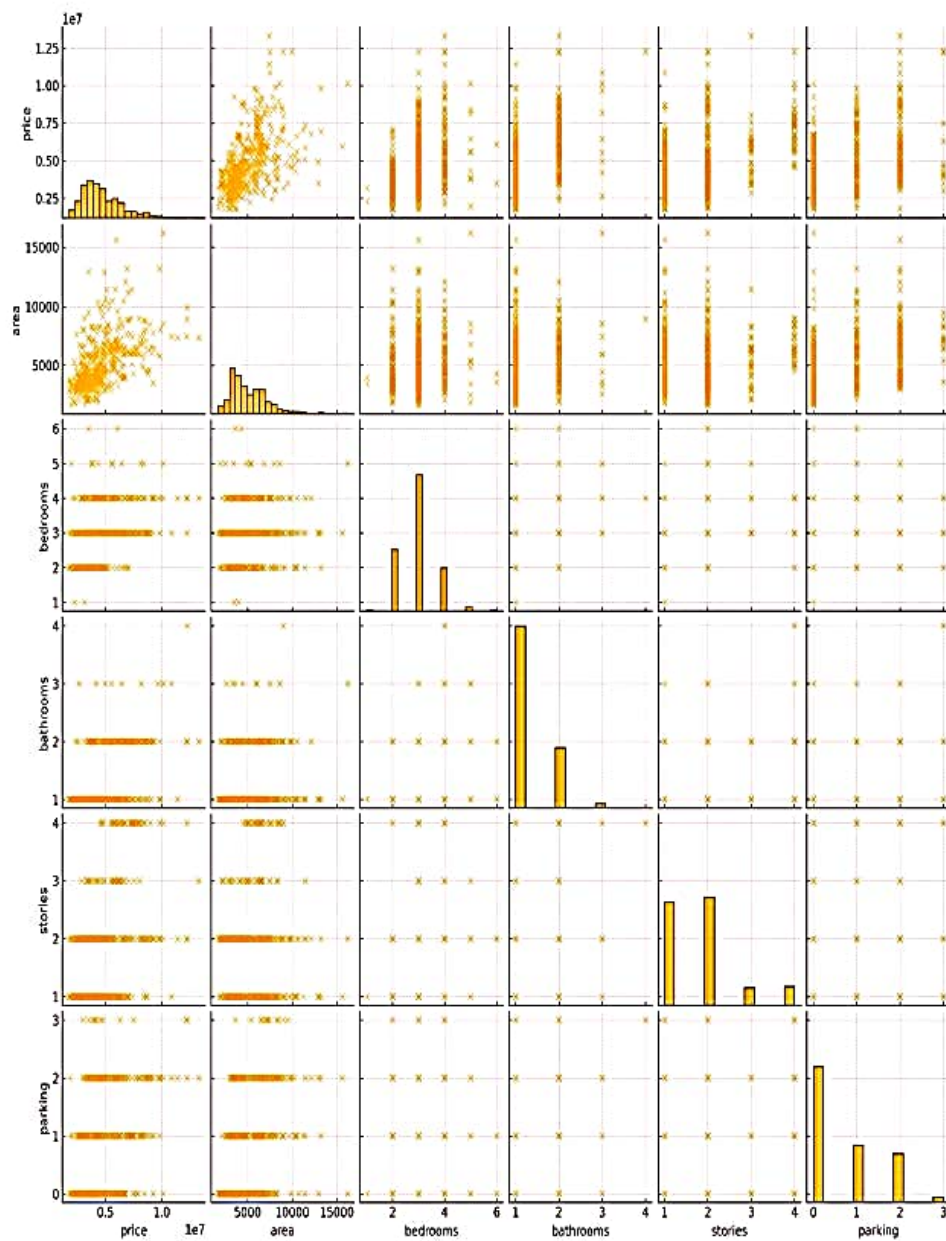   This graph is quite big; hence, a whole page would be needed.

Figure 5

2.6 Distribution of Categorical Features

- **Mainroad Access:** The majority of houses have mainroad access.

- **Guestroom Availability:** Guestrooms are less common.

- **Basement:** Most properties lack a basement.

- **Air Conditioning:** More than half of the properties are equipped with air conditioning.

- **Preferred Area:** Properties in preferred areas show a higher frequency.

- **Visualization:**

   - Individual bar plots (Figures 6-10) for each feature illustrate these distributions clearly.
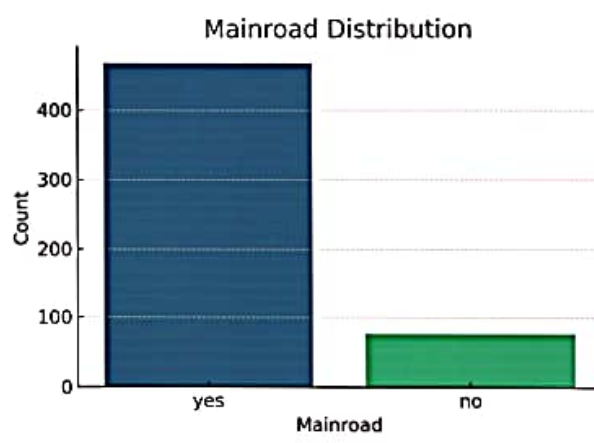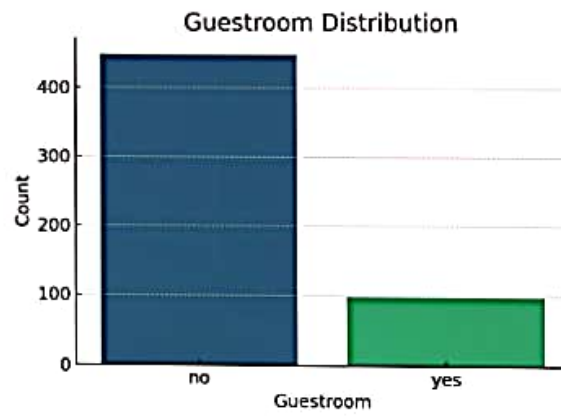

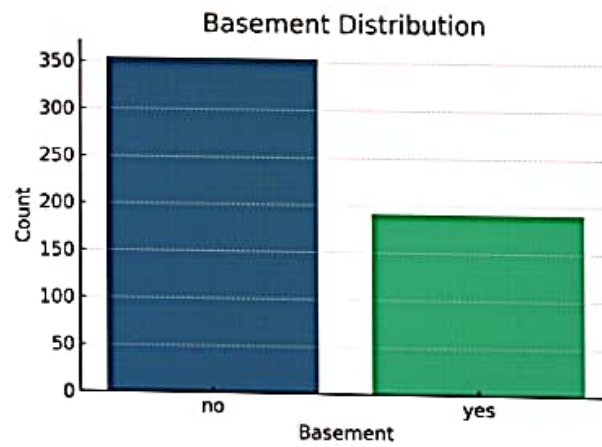Mainroad Distribution
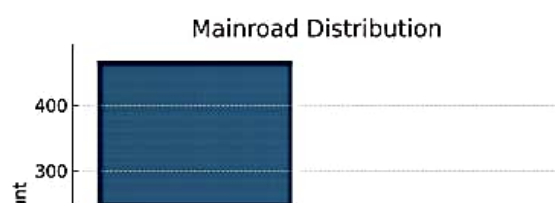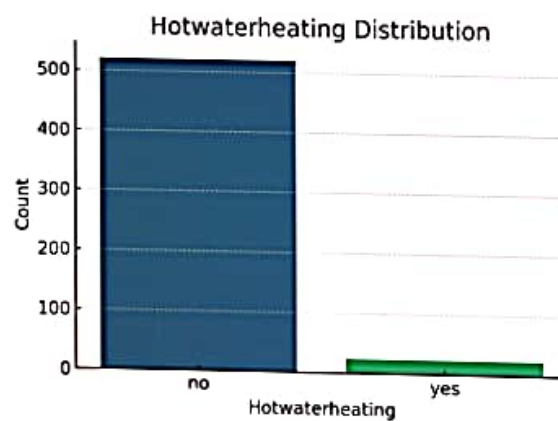
Figure 6

Figure 7

Figure 8

Figure 9

2.7 Price vs Area

- A positive linear relationship exists between **price** and **area**, with outliers observed in high-price properties.
- This trend underscores the significant role area plays in determining property value.
- **Visualization:**
  - A scatterplot (Figure 11) illustrates this relationship effectively, highlighting clusters and outliers.

Figure 11

2.8 Stories and Price

- Properties with more stories are generally more expensive, reflecting their potential as premium multi-story houses.
- **Visualization:**
  - A bar plot (Figure 12) demonstrates this trend.

Figure 12

3. Recommendations

For Buyers

- Consider properties in preferred areas or with air conditioning for better resale value.
- Invest in houses with larger areas and more bathrooms as they strongly influence pricing.

For Sellers

- Highlight features like mainroad access and furnishing status in marketing materials.
- Upgrading to a furnished status or adding air conditioning could increase property value.

For Real Estate Agents

- Focus on properties with multiple stories and larger areas to target high-budget buyers.
- Leverage the dataset to recommend optimal pricing based on features like furnishing status and area.

4. Appendices

Data Visualizations

- **Figure 1:** Price Distribution Histogram

3. Recommendations

For Buyers

- Consider properties in preferred areas or with air conditioning for better resale value.
- Invest in houses with larger areas and more bathrooms as they strongly influence pricing.

For Sellers

- Highlight features like mainroad access and furnishing status in marketing materials.
- Upgrading to a furnished status or adding air conditioning could increase property value.

For Real Estate Agents

- Focus on properties with multiple stories and larger areas to target high-budget buyers.
- Leverage the dataset to recommend optimal pricing based on features like furnishing status and area.

4. Appendices

Data Visualizations

- **Figure 1:** Price Distribution Histogram
- **Figure 2:** Correlation Heatmap
- **Figure 3:** Boxplot - Price vs Furnishing Status
- **Figure 4:** Countplot - Mainroad Access by Furnishing Status
- **Figure 5:** Pairplot of Numerical Features
- **Figures 6-10:** Bar Plots for Categorical Features
- **Figure 11:** Scatterplot - Price vs Area
- **Figure 12:** Bar Plot - Stories vs Price

Code Used for Analysis

Refer to the script provided in the "housing_data_analysis" file.

5. Conclusion

The analysis reveals critical factors influencing property prices, including area, furnishing status, and stories. These insights can guide buyers, sellers, and real estate professionals in making informed decisions. Further exploration could involve advanced modeling techniques for price prediction.

Visualizations such as histograms, scatterplots, and heatmaps were instrumental in identifying patterns and correlations, providing a robust foundation for the recommendations made in this report.