Individual Project Report
House Price Prediction Using Regression Techniques
Name: Varshini Sherapur Basavarajappa
Group: Group 8

1. Introduction:

This project focuses on addressing one of the most critical financial decisions individuals face: buying a home. Recognizing that house prices are influenced by numerous factors such as location, property size, and amenities, my work aims to tackle the challenges often associated with the housing market, including overpaying, lack of transparency, and difficulty in negotiation. To address these issues, I concentrated on developing a reliable House Price Prediction Model using regression techniques. By analyzing key housing features such as area, rooms, and amenities, my analysis provides actionable insights for buyers and sellers, enabling data-driven decision-making and enhancing market efficiency.
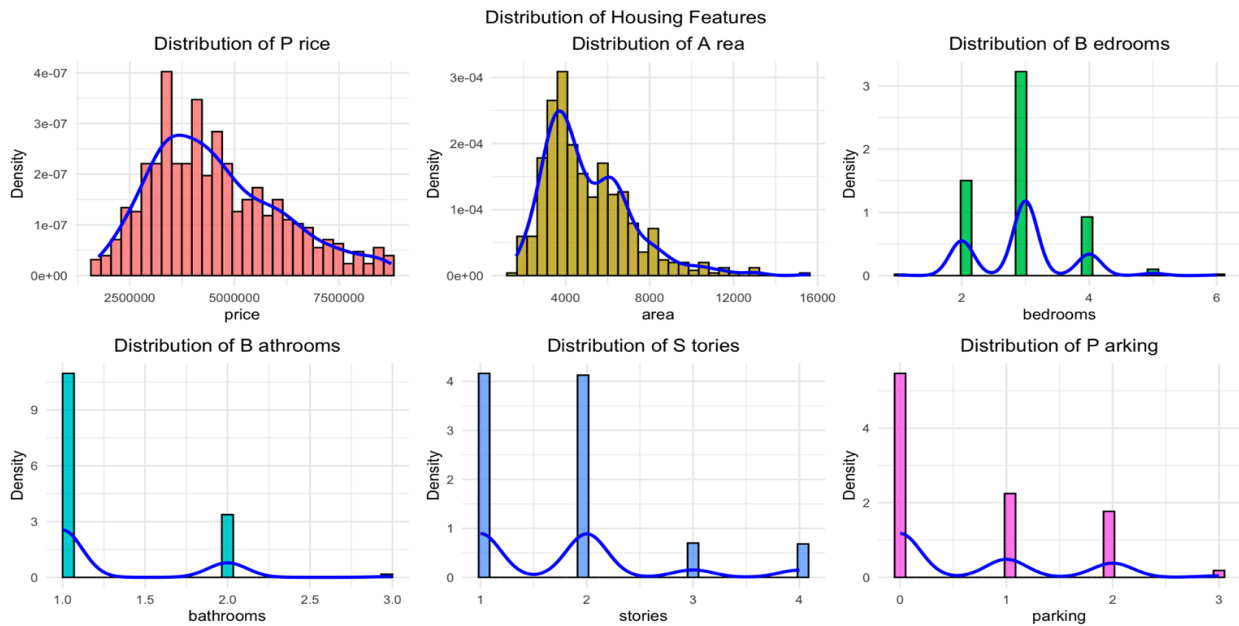
2. Exploratory Data Analysis (EDA)

As part of my analysis, I focused on visualizing data distributions to detect patterns and identify outliers, ensuring a comprehensive understanding of the dataset. By examining the relationships between housing features such as area, bedrooms, and price, I identified key trends and influential predictors. The distribution of prices, areas, and rooms revealed skewed patterns, prompting the need for log transformations to achieve normalization. I also utilized box plots to highlight categorical features, including 'Mainroad', 'Guestroom', and 'Air Conditioning', which significantly impacted price distributions. Additionally, through outlier analysis, I addressed extreme values in features like price and area, filtering these anomalies to improve the overall accuracy and reliability of the predictive model. This thorough approach laid a solid foundation for effective modeling and insights.
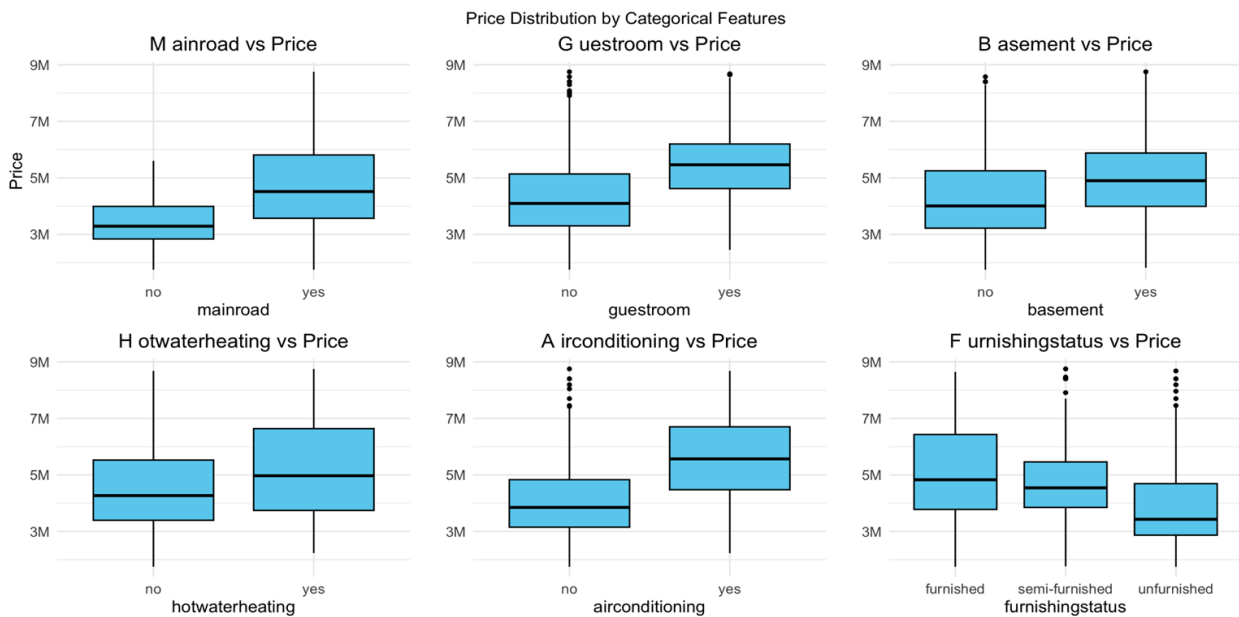
Visuals:

1. Data distributions (e.g., price, area, rooms):

- Visualized key housing features, including price, area, bedrooms, bathrooms, and parking.
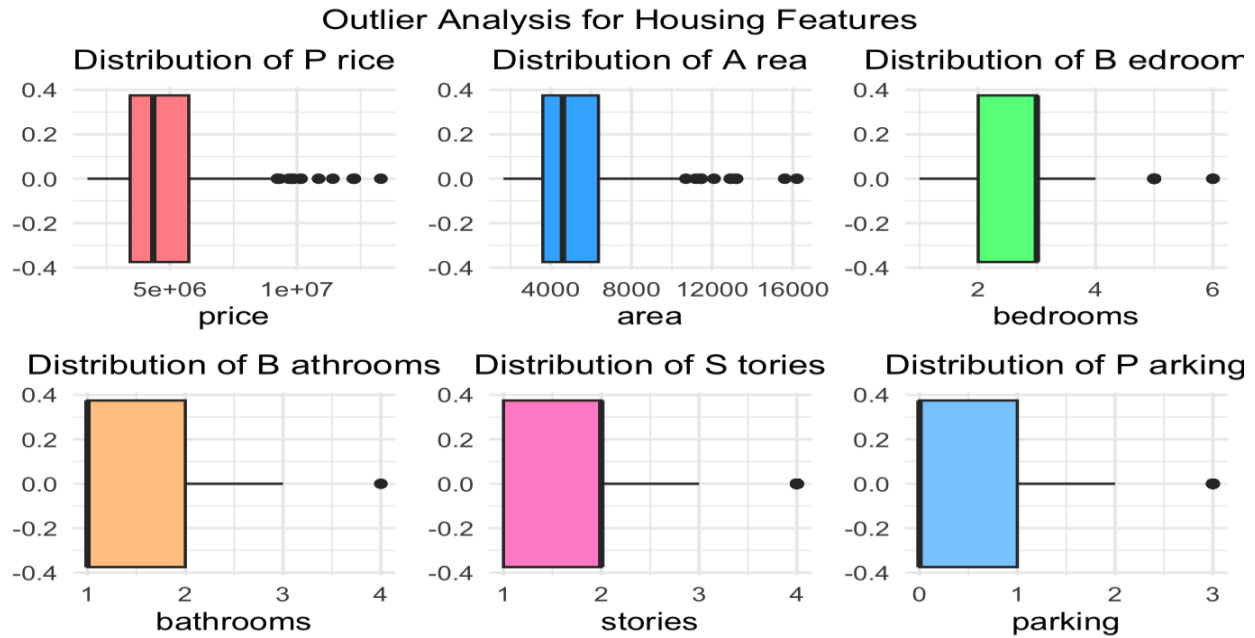
2. Box plots for categorical features:

Explored how features like Mainroad, Guestroom, and Air Conditioning impact price distributions.

3. Outlier Analysis:

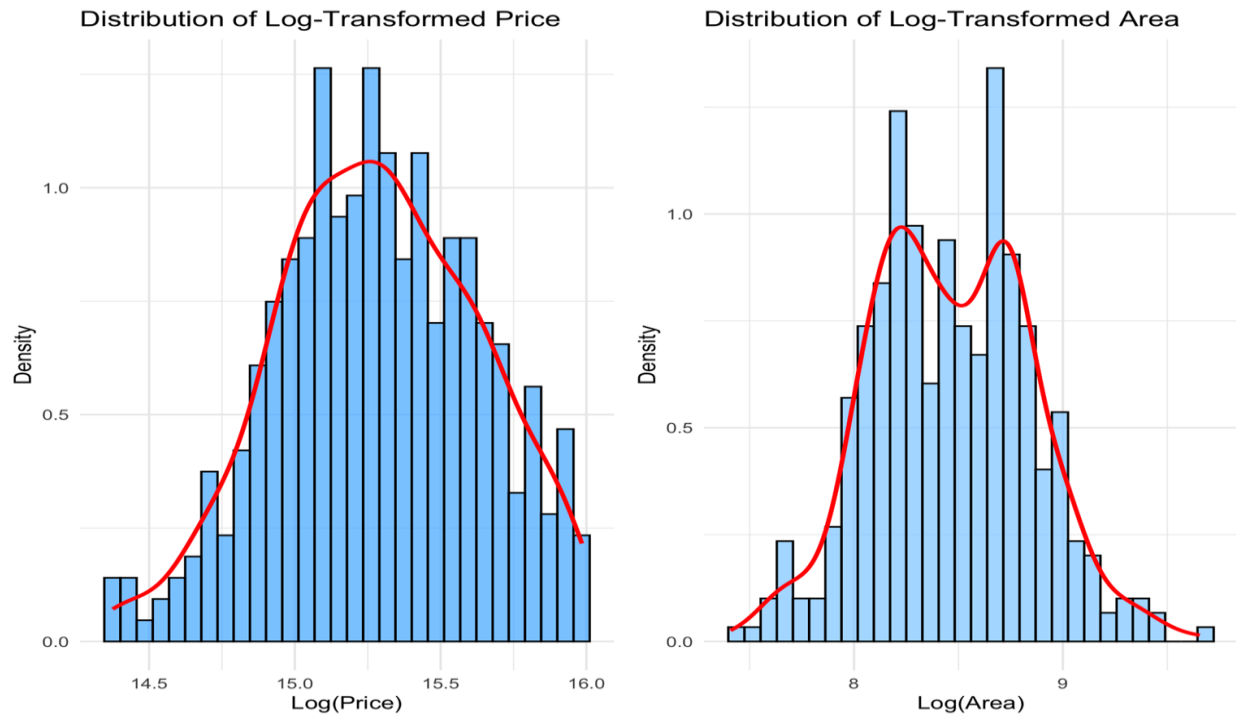- Identified outliers in numerical data (e.g., price, area, rooms).

Outlier Analysis for Housing Features

3. Data Transformation

I applied the Log transformations were applied to key features ('Price', 'Area') to normalize their distributions. This step ensures that regression assumptions of linearity and normality are met, improving the reliability of predictions.

Visuals:

1. Distribution of Log-Transformed Price and Area:

- These transformations improved normality and linearity for modeling.
- The log-transformed 'Price' and 'Area' exhibited near-normal distributions.

Distribution of Log-Transformed Price   Distribution of Log-Transformed Area

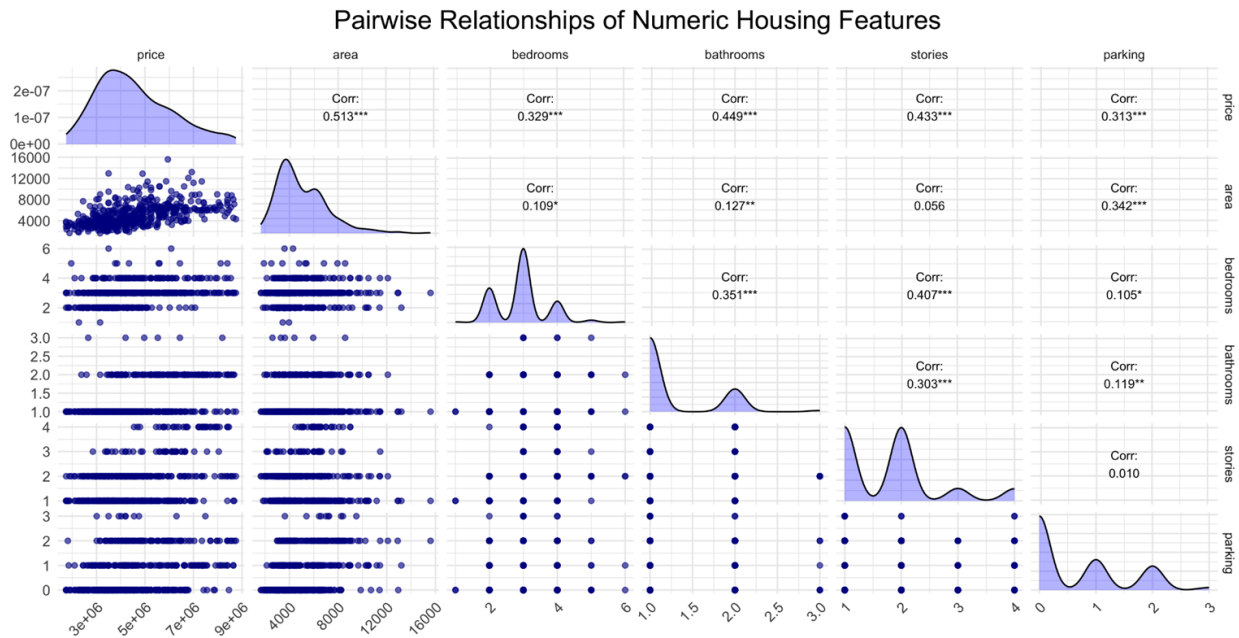## 4. Pairwise Relationships

Pairwise relationships between numeric features revealed:

- A strong correlation between 'Area' and 'Price' (Corr = 0.51).
- Moderate relationships between 'Bedrooms', 'Bathrooms', and 'Price'.
- Weak correlations for features like 'Parking', indicating limited influence.

Visuals:

Pairwise Correlation Matrix:

- Displayed significant relationships among housing features.

Pairwise Relationships of Numeric Housing Features

5. Conclusion

My contribution focused on EDA and Data Transformation. Through this process:

- Outliers were identified and handled.
- Key features influencing house prices were established.
- Transformed data to meet modeling assumptions.

This work highlights the importance of EDA in ensuring data readiness and extracting actionable insights for regression modeling.

6. References

1. C. R. Madhuri, G. Anuradha, and M. V. Pujitha, "House Price Prediction Using Regression Techniques: A Comparative Study," ICSSS, 2019.

2. Kaggle Dataset: Housing Price Prediction.