Predicting the Nightly Price of Denver's Airbnb in 2023:

Steps:

1. **Introduction and Problem Identification**
2. **Data Wrangling**
3. **Exploratory Data Analysis**
4. **Preprocessing And Training**
5. **Modeling**

1. Introduction and Problem Identification

Travel is one of the world's largest industries, and its approach has become commoditized. The travel industry has scaled by offering standardized accommodations in crowded hotel districts and frequently-visited landmarks and attractions. This one-size-fits-all approach has limited how much of the world a person can access, and as a result, guests are often left feeling like outsiders in the places they visit. Airbnb can help you with them.

In Denver, we are going to find a good way to predict the final price from our data like review scores, the number of bedrooms, and bathrooms. We should know how they affect to the final price.

Your property is never going to be expensive enough to satisfy your wishes, but it is never going to be cheap enough to satisfy a guest's wishes. The most common issue we hear is *'If I charge any more than $XX per night I won't get any bookings.'*

How can that data affect to the price? Or do they affect to the price at all?

We want to find the best model for predicting the price for Airbnb company.

- The purpose of the analysis: understanding the factors that influence Airbnb prices in Denver, or identifying patterns of all variables and Our analysis provides useful information for travelers and hosts in the city and also provides some best insights for Airbnb business.

- the factors that influence Airbnb prices in Denver, or identifying patterns of all variables and Our analysis provides useful information for travelers and hosts in the city and also provides some best insights for Airbnb business.

- This project involved exploring and cleaning a dataset to prepare it for analysis. The data exploration process involved identifying and understanding the characteristics of the data, such as the data types, missing values, and distributions of values. The data cleaning process involved identifying and addressing any issues or inconsistencies in the data, such as errors, missing values, or duplicate records and removing outliers.

- Through this process, we were able to identify and fix any issues with the data, and ensure that it was ready for further analysis. This is an important step in any data analysis project, as it allows us to work with high-quality data and avoid any potential biases or errors that could affect the results. The clean and prepared data can now be used to answer specific research.

- Once the data has been cleaned and prepared, now begin exploring and summarizing it with describing the data and creating visualizations, and identifying patterns and trends in the data. in exploring the data, may develop the relationships between different variables or the underlying causes of certain patterns or trends and other methods.

- using data visualization to explore and understand patterns in Airbnb data. We created various graphs and charts to visualize the data, and wrote observations and insights below each one to help us better understand the data and identify useful insights and patterns.

- Through this process, we were able to uncover trends and relationships in the data that would have been difficult to identify through raw data alone, for example factors affecting prices and availability. We found that minimum nights, number of reviews, and host listing count are important for determining prices, and that availability varies significantly across neighborhoods. Our analysis provides useful information for travelers and hosts in the city.

- The observations and insights we identified through this process will be useful for future analysis and decision-making related to Airbnb. And also, our analysis provides useful information for travelers and hosts in the city.

Source:

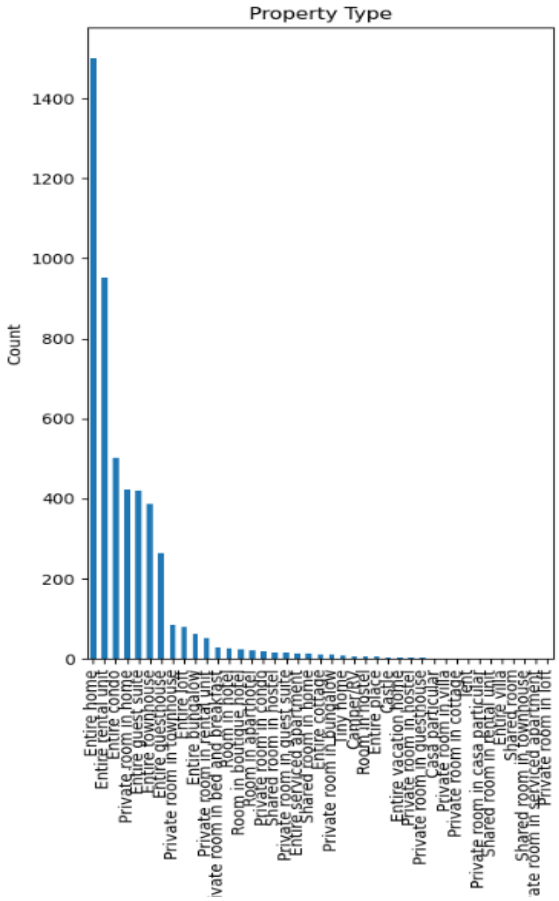
http://insideairbnb.com/get-the-data/

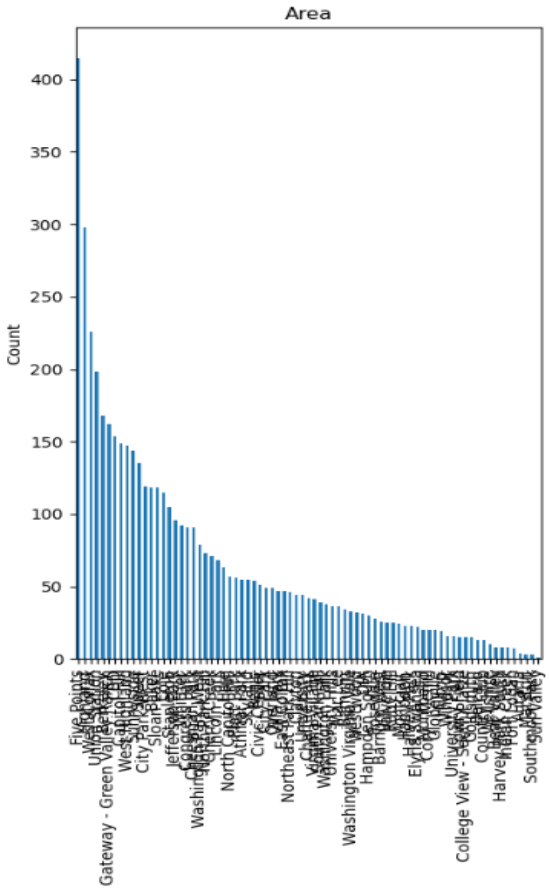
2. Data Wrangling

In this section, I tried to find the data that I need and I tried to have the summary of the data. After that I cleaned and found the distributions of the columns that I want to select as features to predict the price. Finally, I made the summary of the data. The details are listed below:

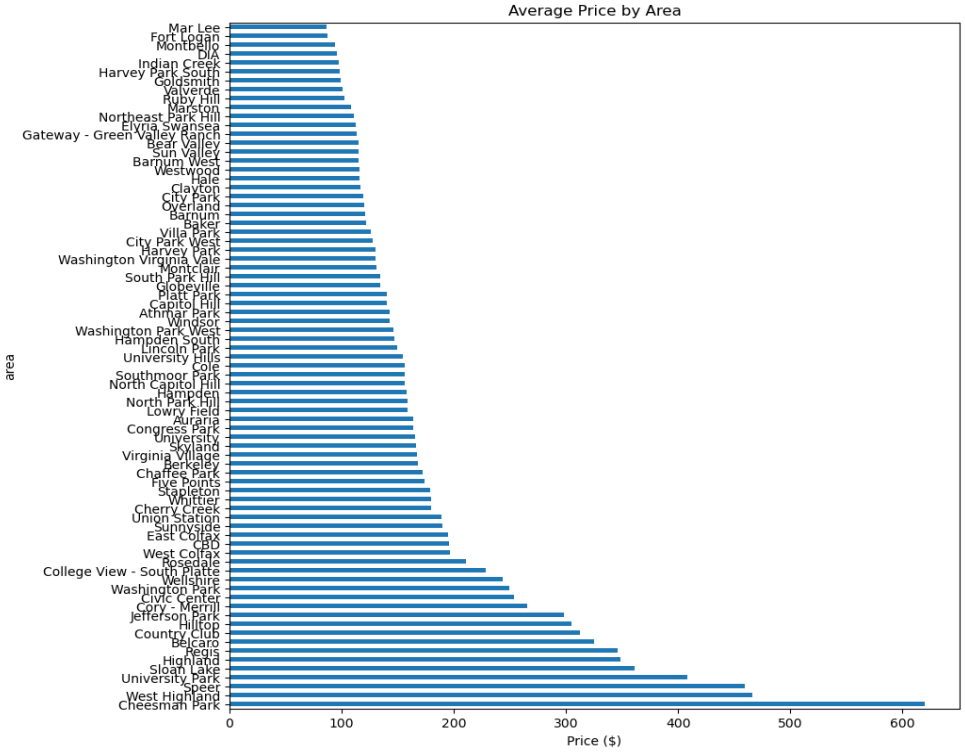
1. Find the Data that I Need Categorical Features from Raw Data
2. Number Of Missing Values by Column
3. Categorical Features

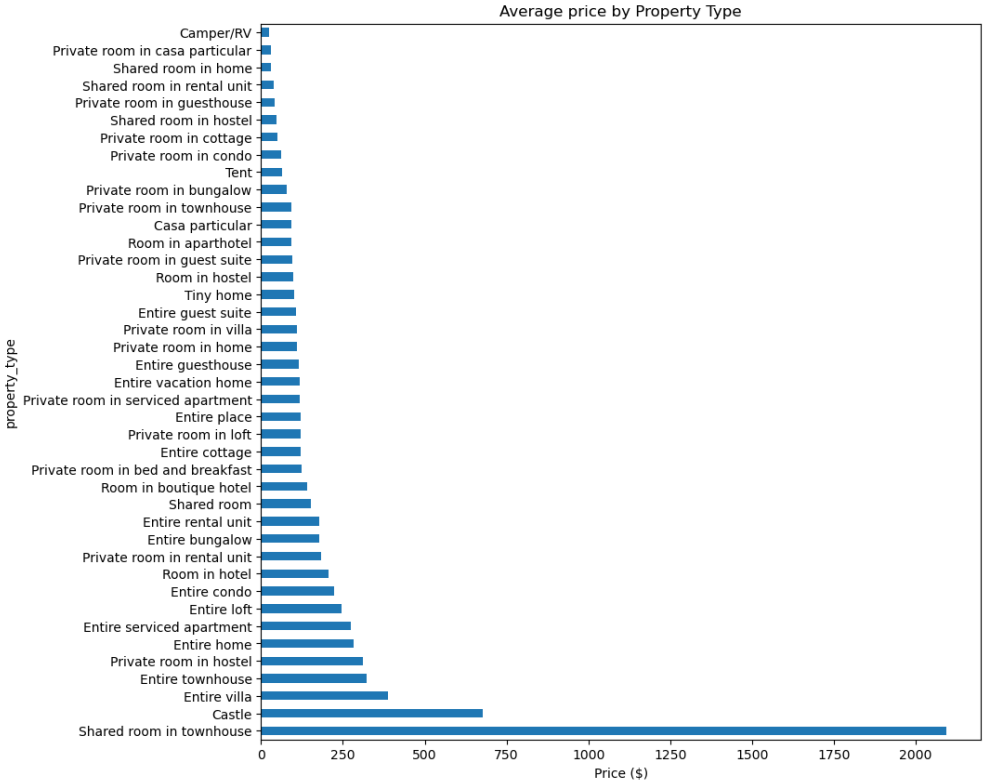
    1. Unique Home Type names
    2. Area
    3. Number of distinct regions
    4. Distribution by Property Type and Area





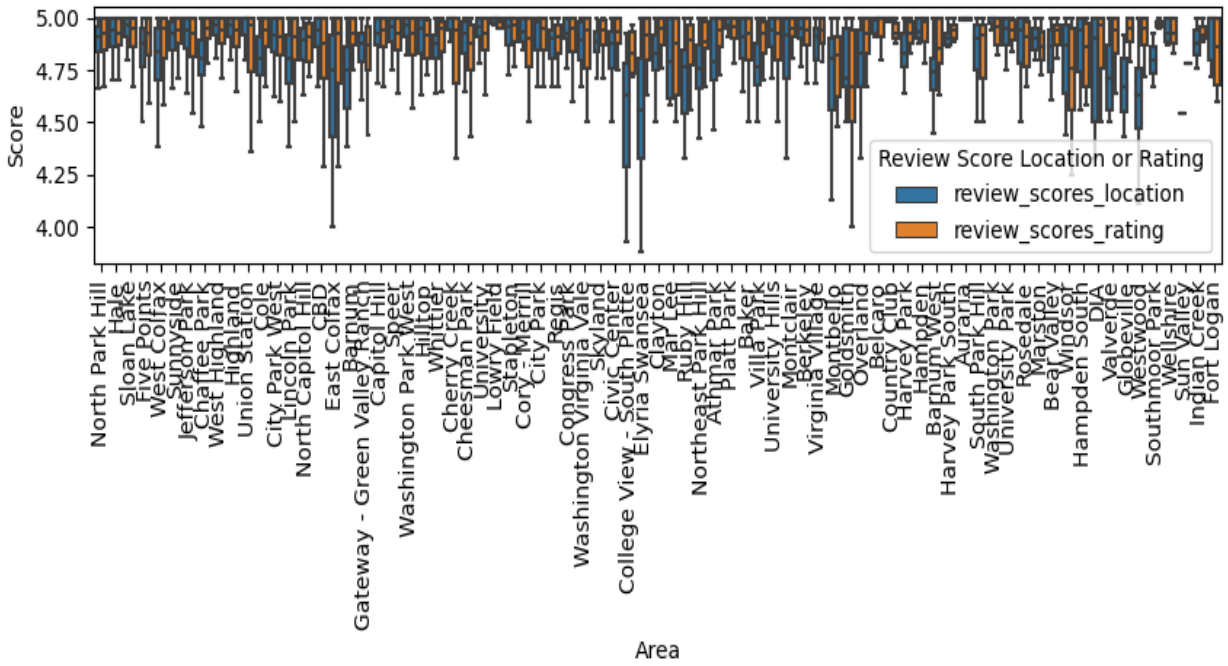
5. Distribution of Average of Price, Review Scores Rating and Review Score Location by Property Type and Area:
    a) Average of Price by Property Type and Area



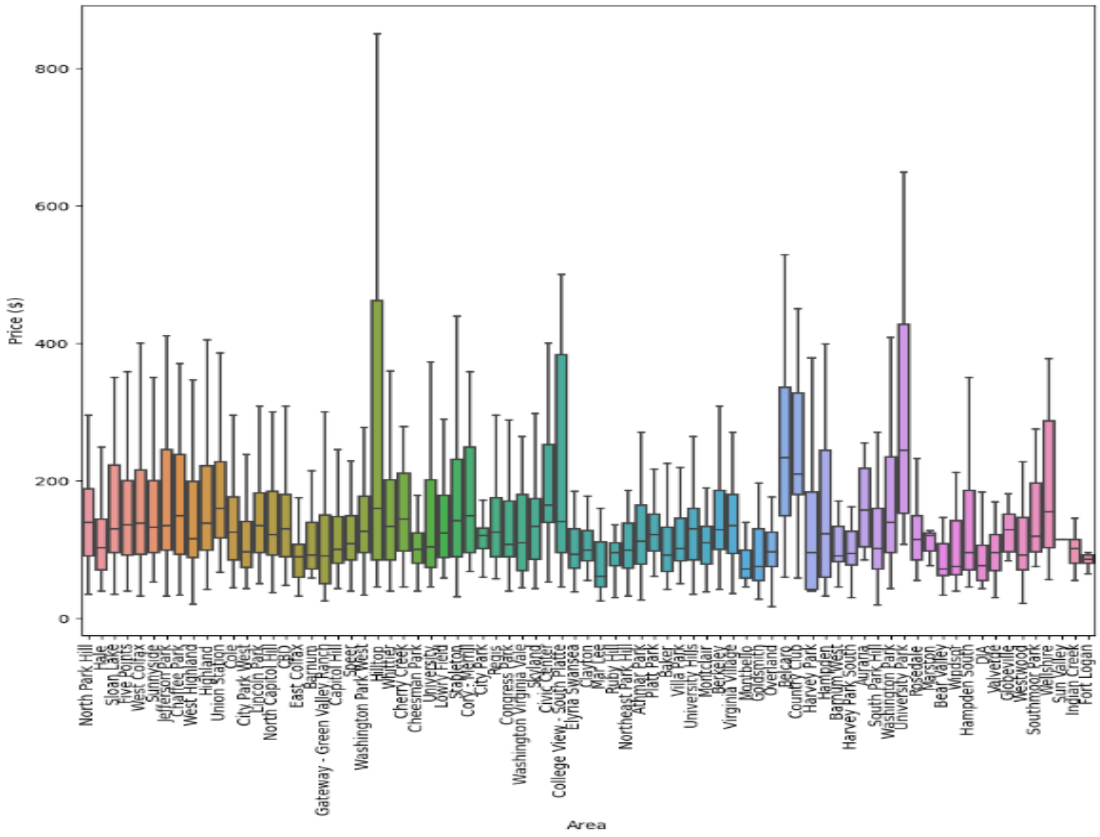


b) Average of Review Scores Location and Review Score Rating by Area

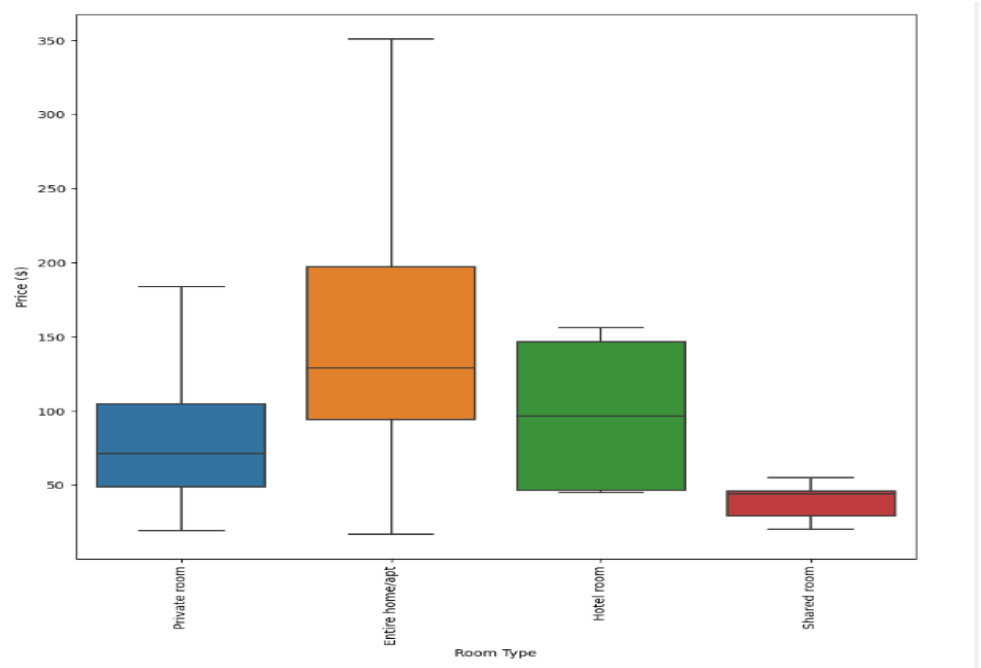
c) Distribution of Review Score Location, Review Score Rating and Price by Area



d) Distribution of Price by Area

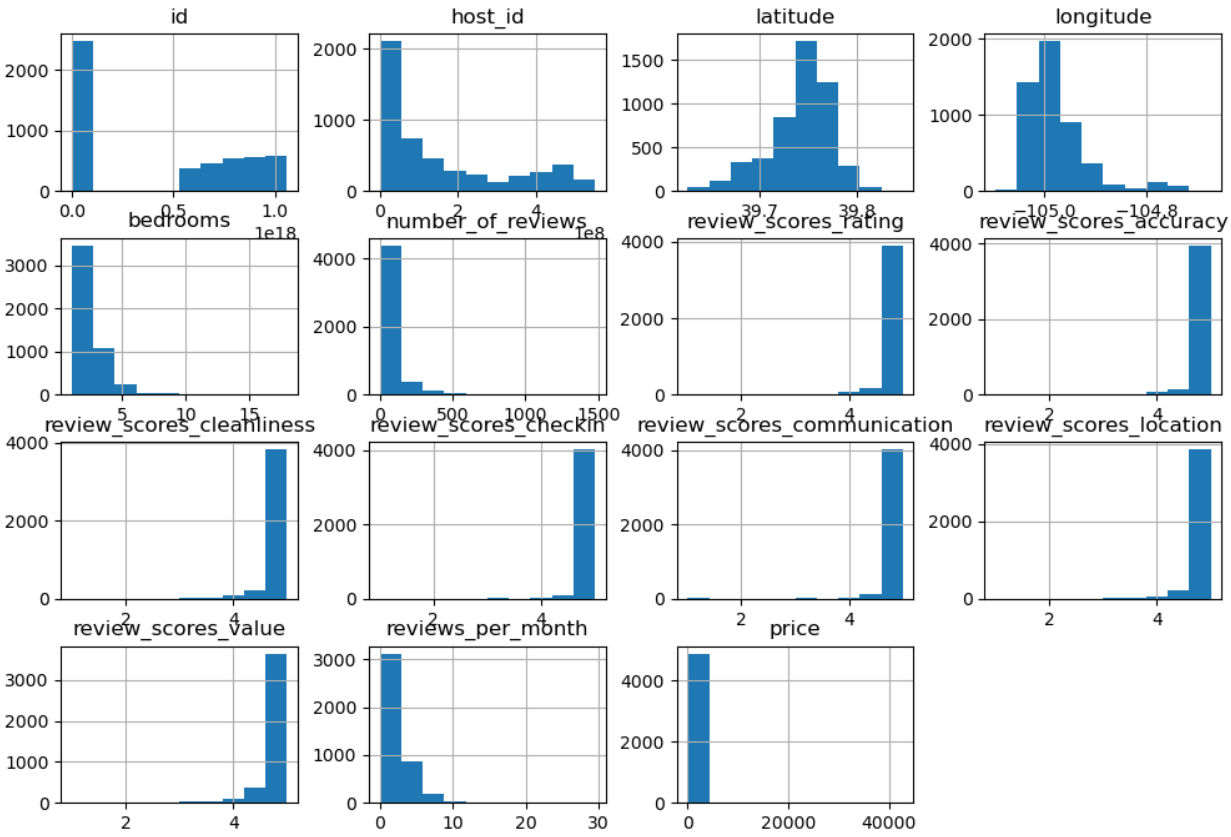


e) Distribution of Room Type by Area

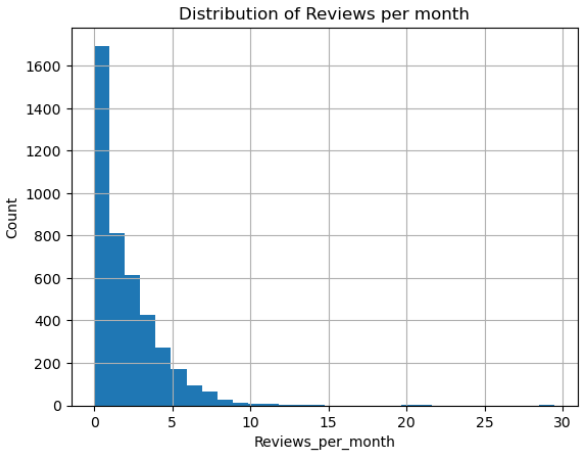


4. Numeric Features

1. Numeric Data Summary
2. Distributions Of Feature Value



a) Reviews per Month

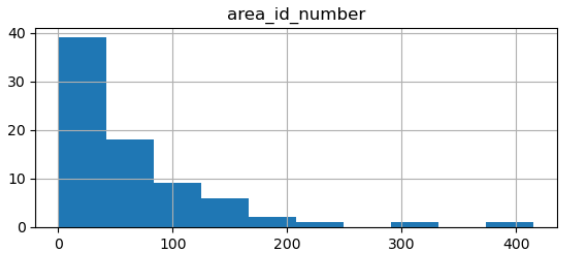


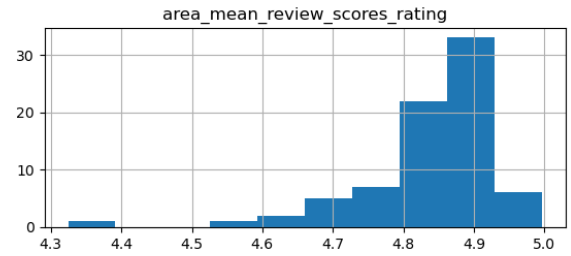
b) Price
c) Review Scores
d) Bedrooms and Bathrooms

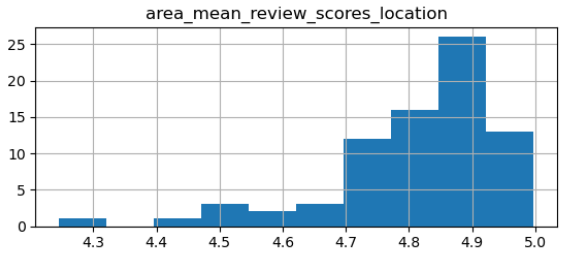
5. Derive Area-Wide Summary Statistics in Distinct Area

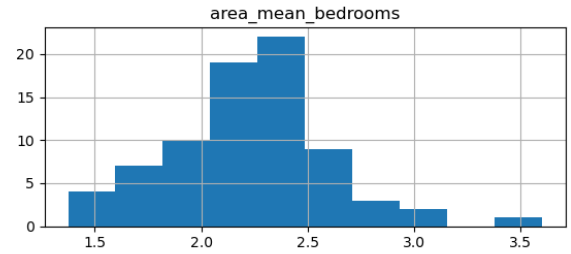
6. Drop Rows with no Price

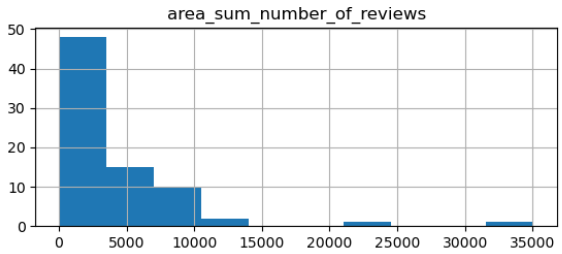
7. Review Summary Distribution

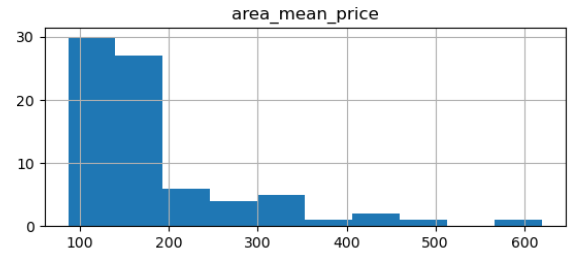












8. Add Area Population and Region Area to Summary

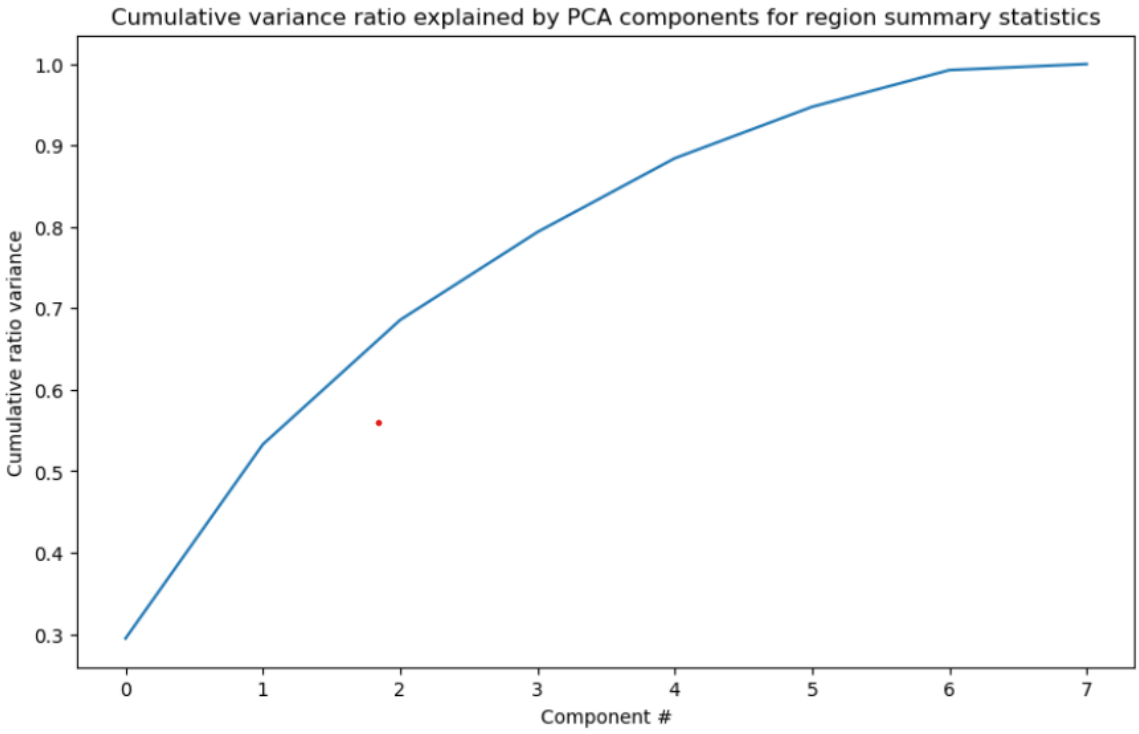
9. Target Feature (Price)

3. Exploratory Data Analysis

1. Top Areas by Order of Each of the Summary Statistics

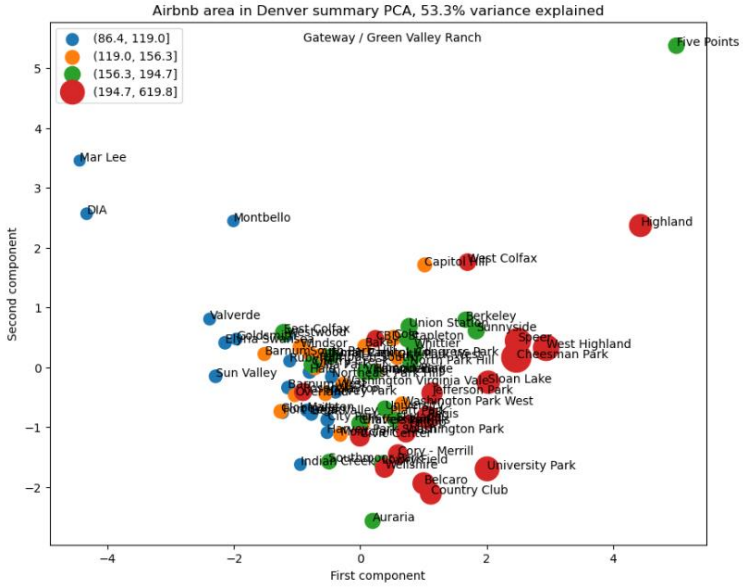
    1) Area review score rating
    2) Area review score location
    3) Id number
    4) Bedrooms
    5) Number of reviews
    6) Total population
    7) ID's Density

2. Visualizing High Dimensional Data
    1) Scale the data
    2) Calculate the PCA transformation

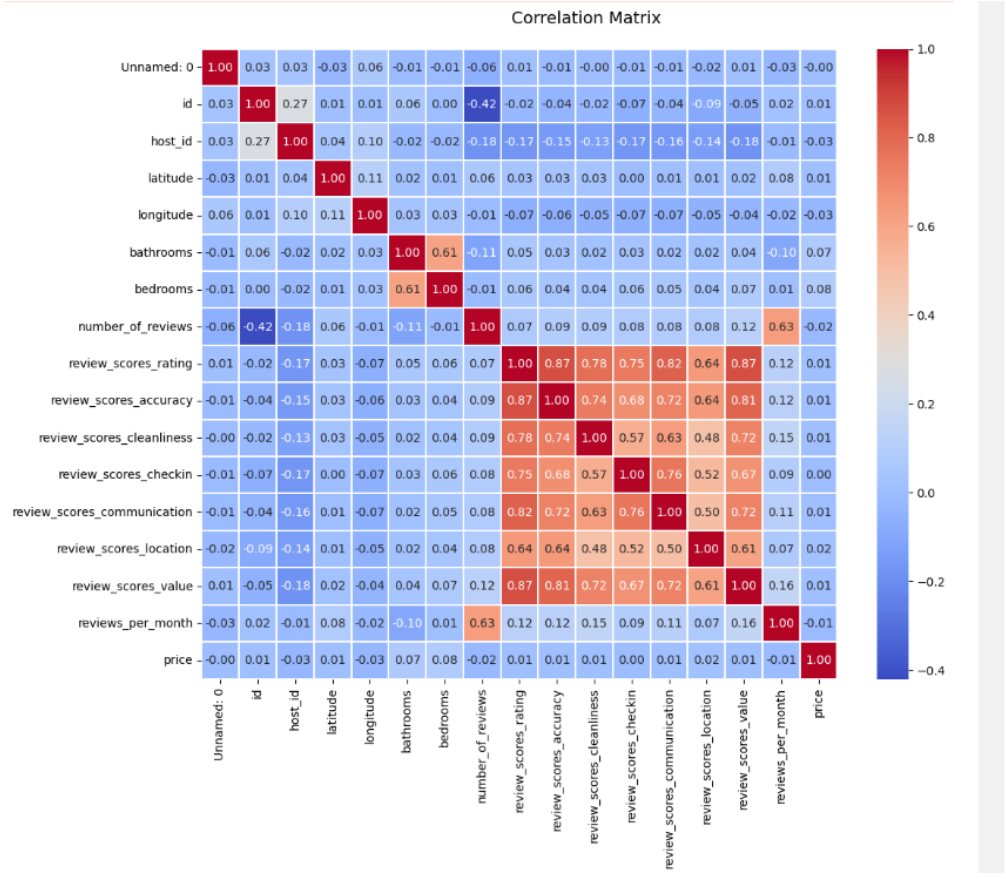


    a) Average Price by Area

b) Adding Average Price and Quartile to Scatter Plot

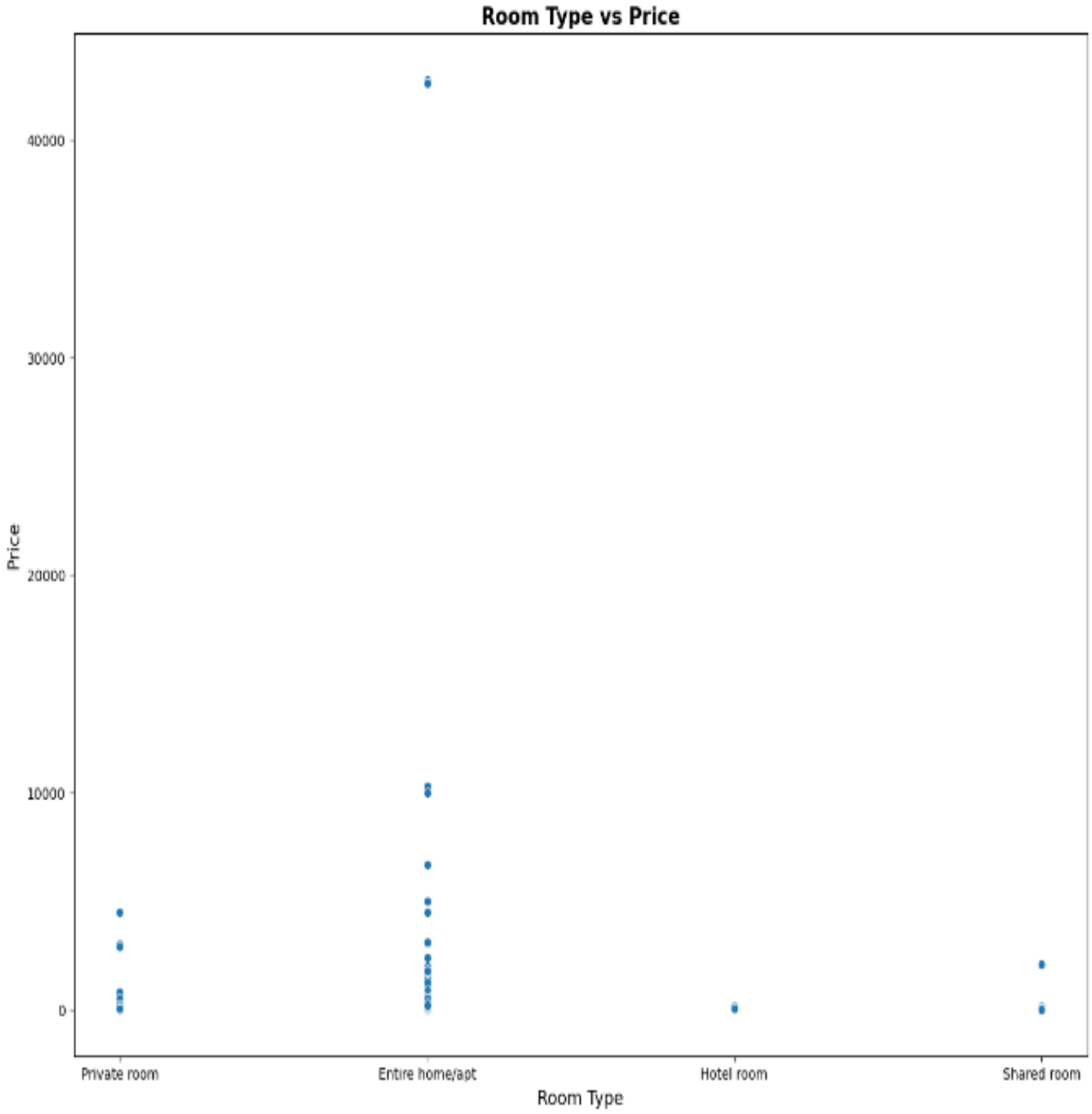


3) Feature Correlation Heatmap

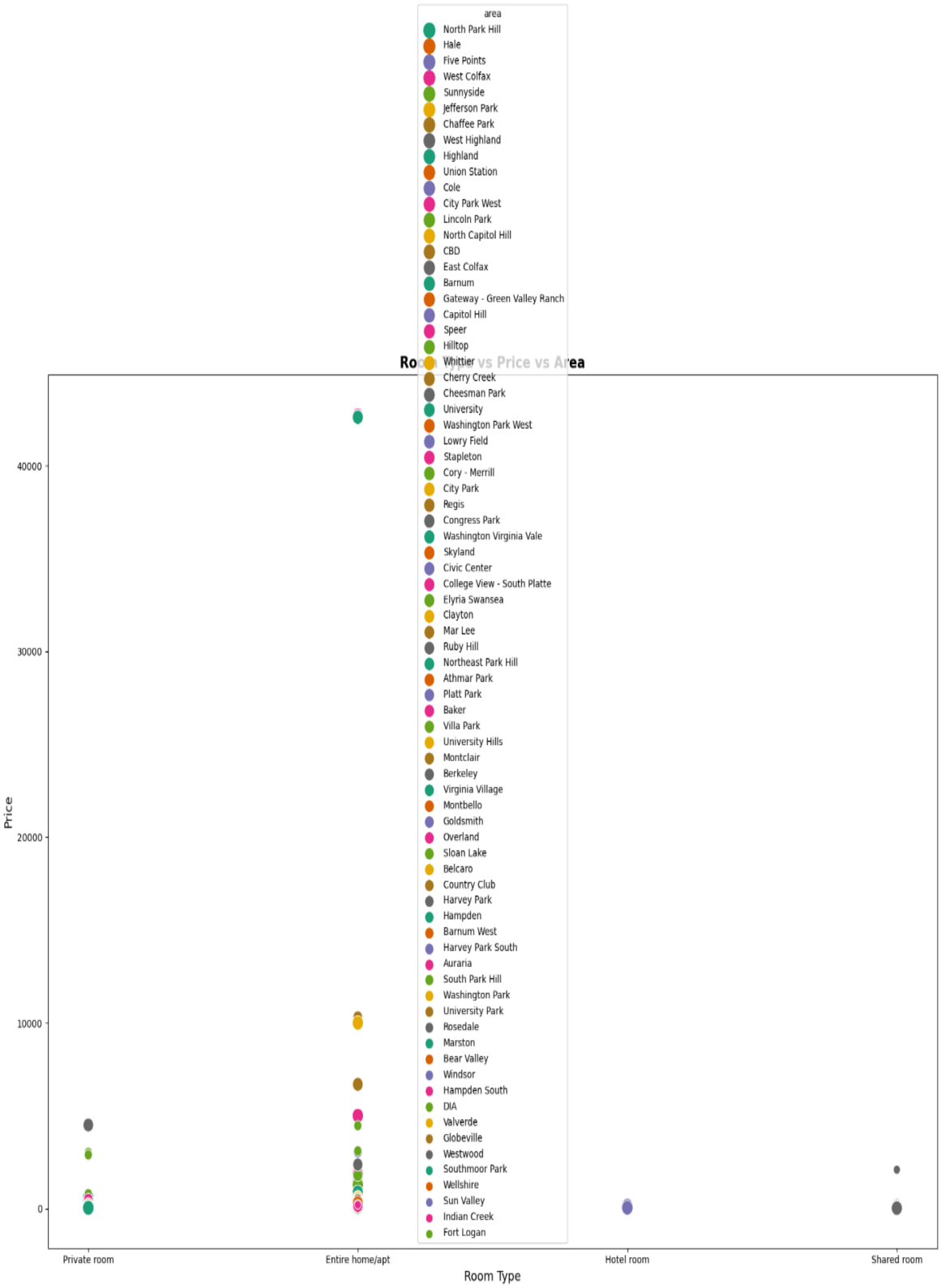


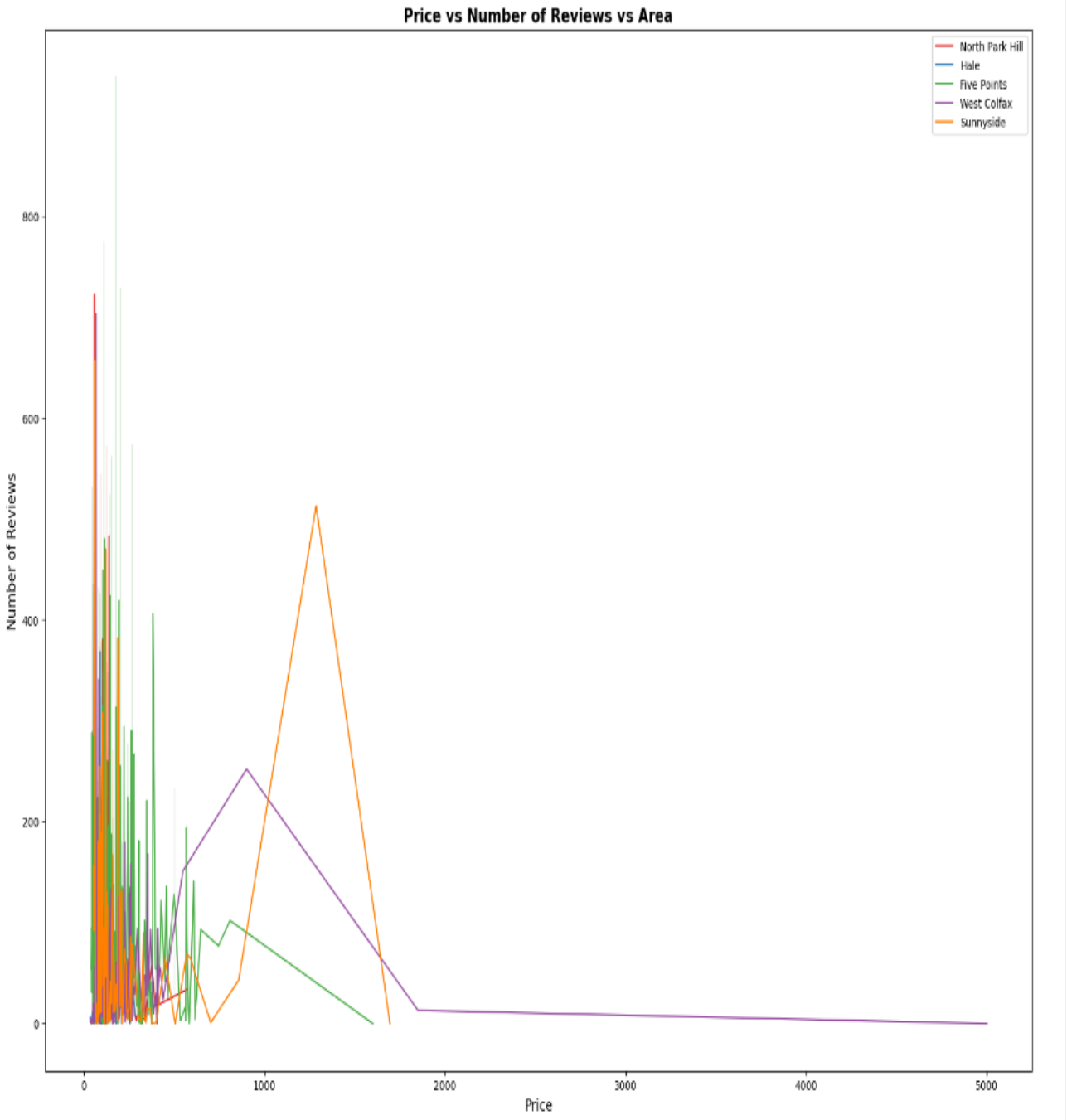
3. Explore Price with other Features

    1) Explore Price by Room Type

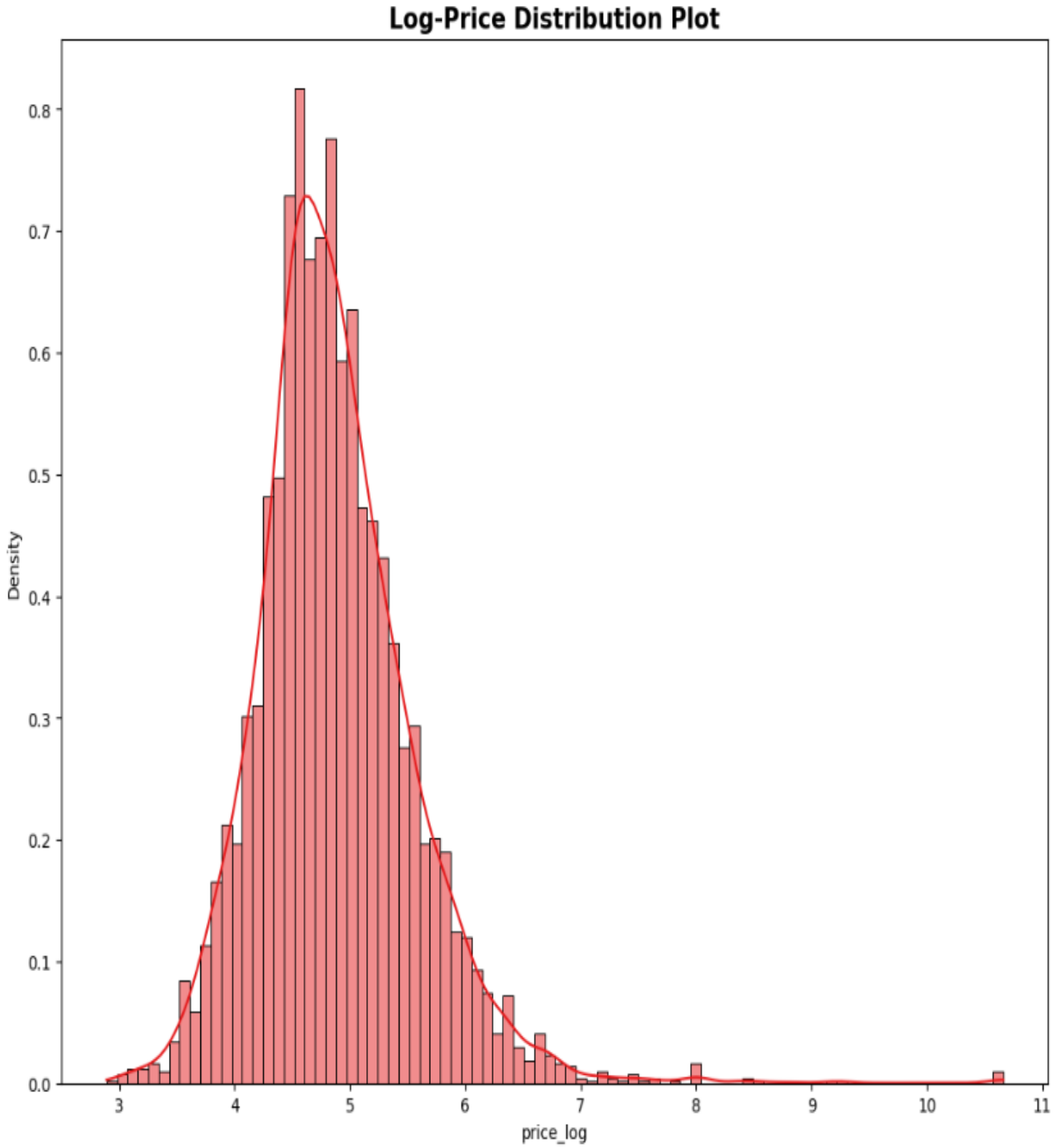


2) Explore Price and Room Type by Area

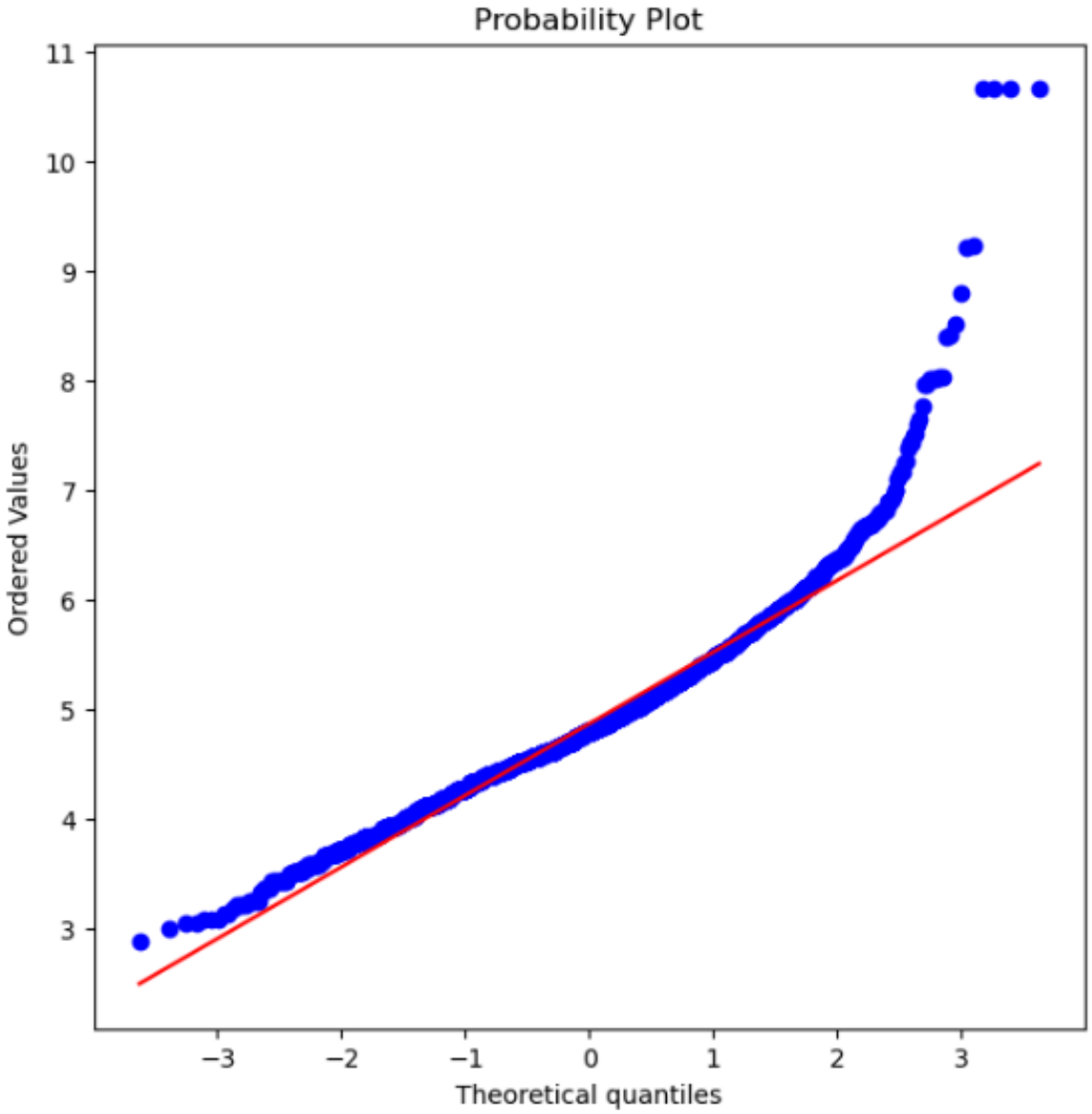



Price vs Number of Reviews vs Area
North Park Hill
Hale
Five Points
West Colfax
Sunnyside
Number of Reviews
0
200
400
600
800
0
1000
2000
3000
4000
5000
Price

3) Log Price Distribution Plot



4) Probability Plot

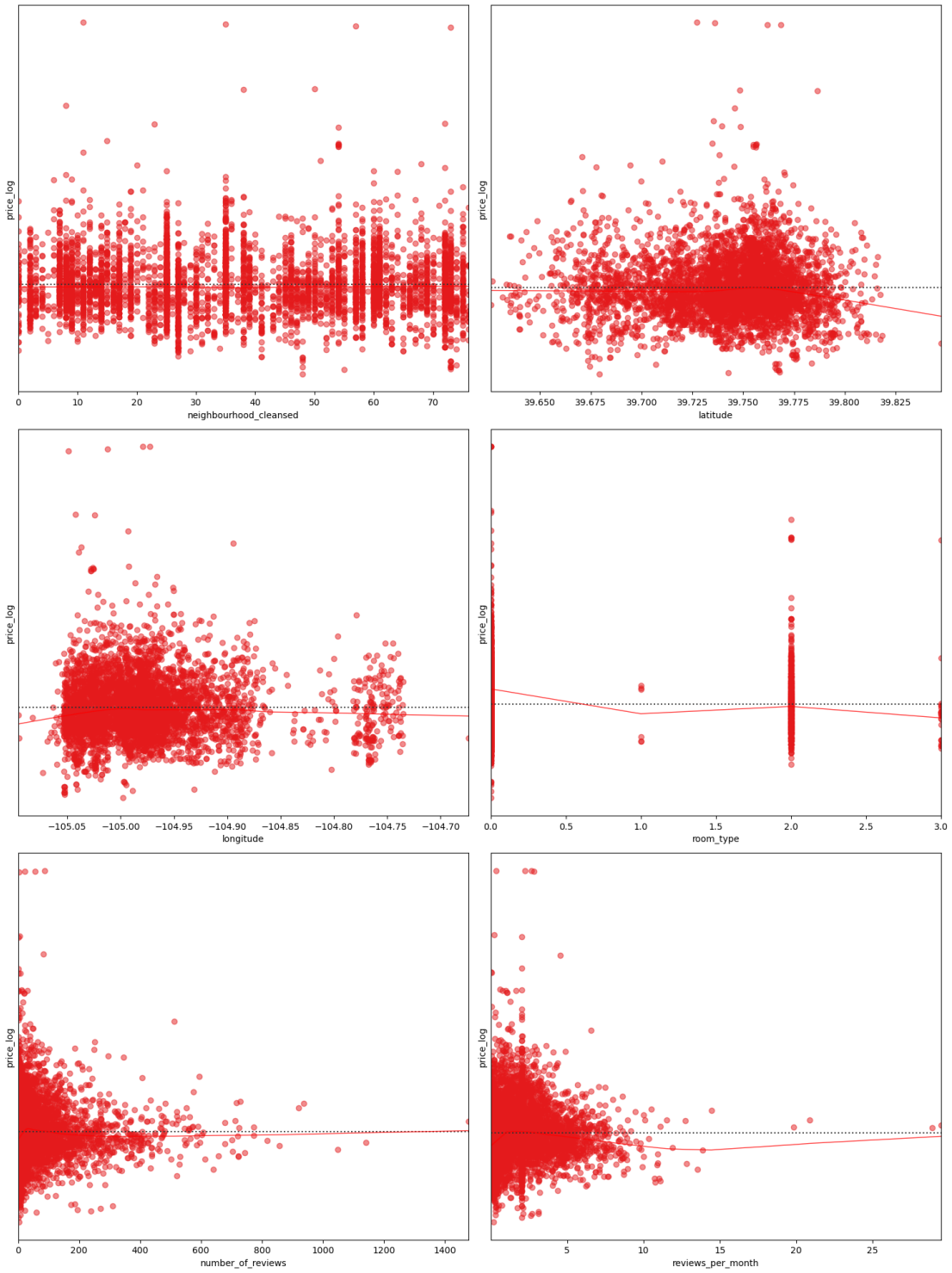


4. Preprocessing And Training

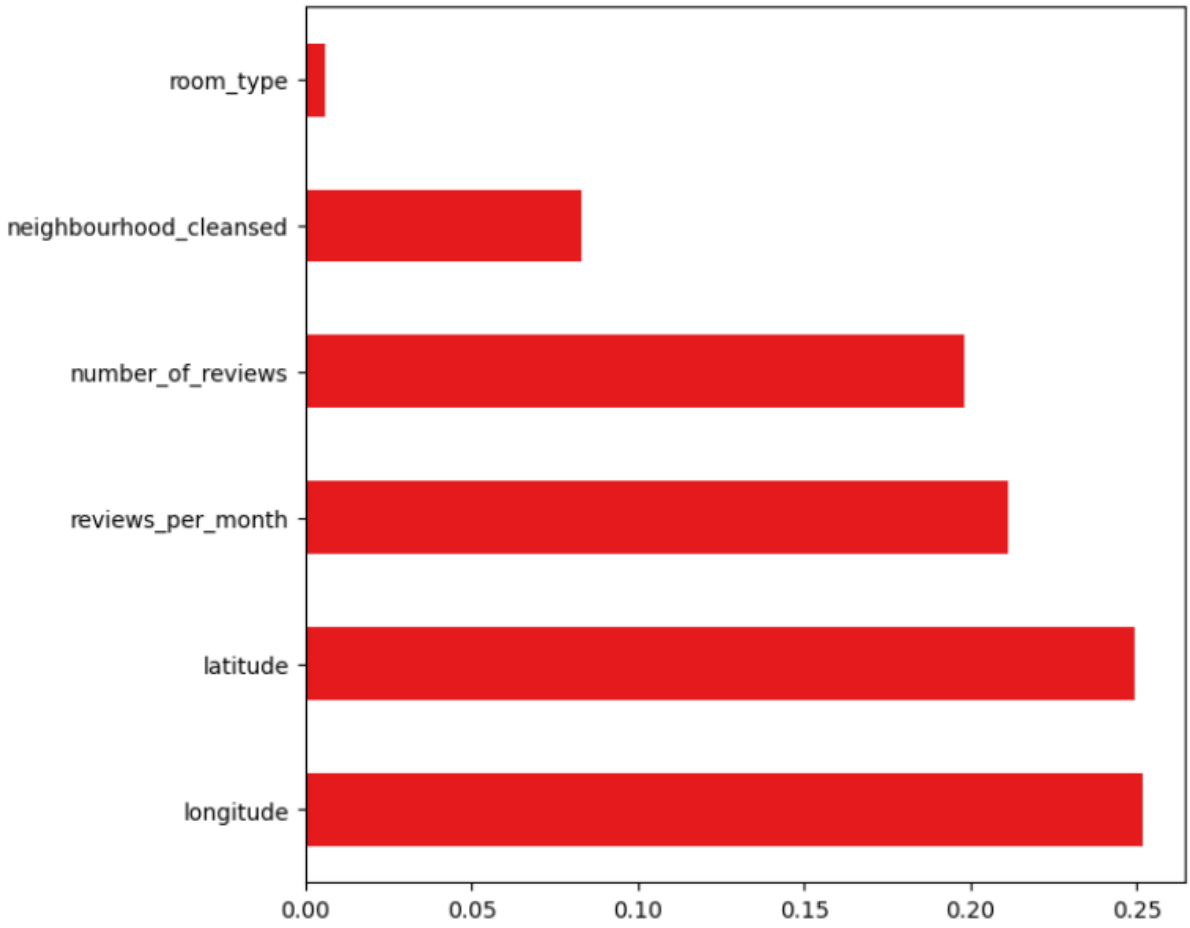
4. Create dummy features for room types
5. Standardize numeric features using a scaler
6. Train/Test Split

5. Modeling

1) Residual Plots



2) Feature Selection and Grid Search



3) Model Scenarios

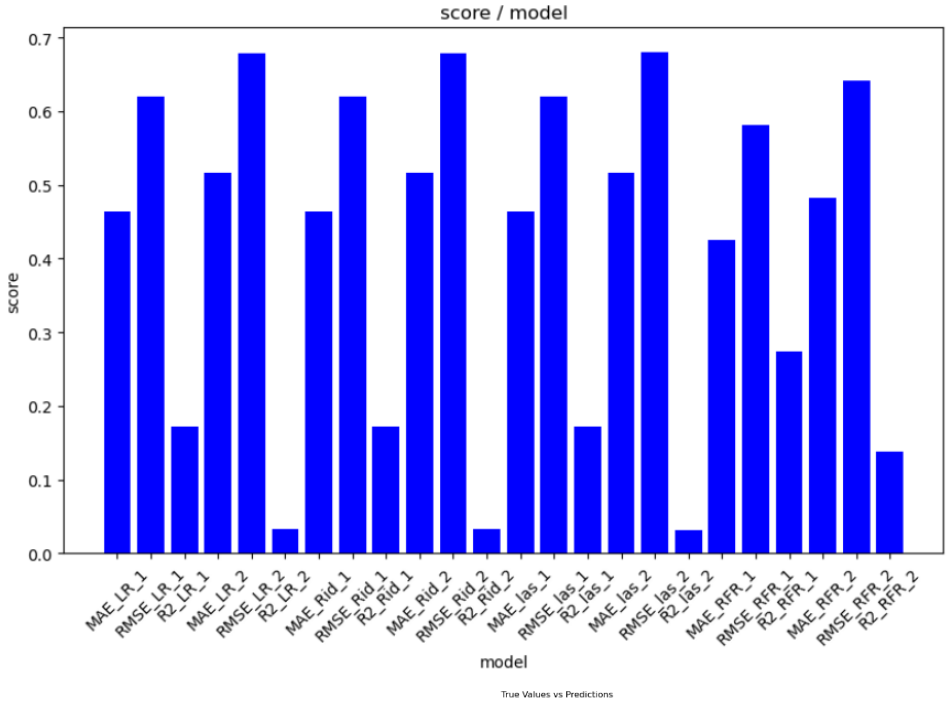
A. Scenario 1(With All Features)

a) K-Fold Cross Validation
b) Polynomial Transformation
c) Model Prediction

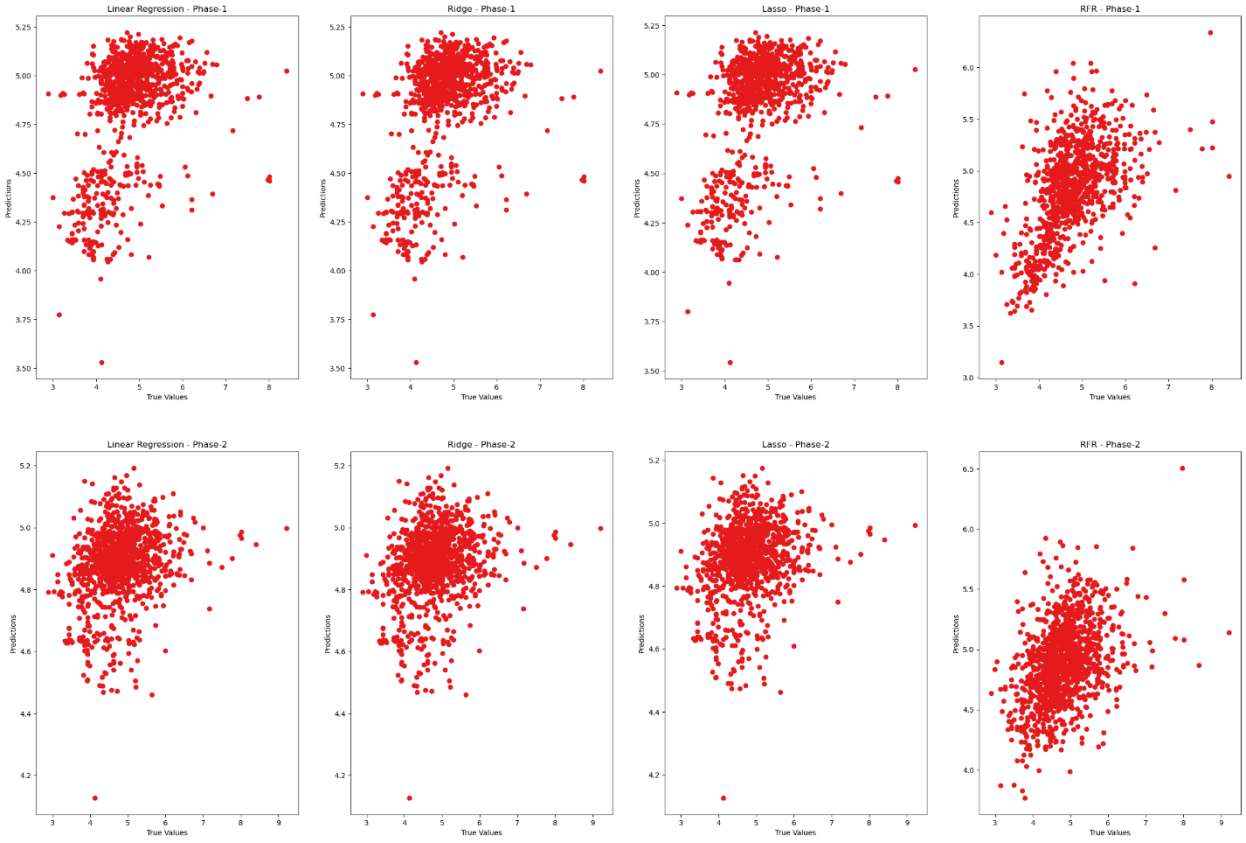
B. Scenario 2(Without All Features)

a) K-Fold Cross Validation
b) Polynomial Transformation
c) Model Prediction

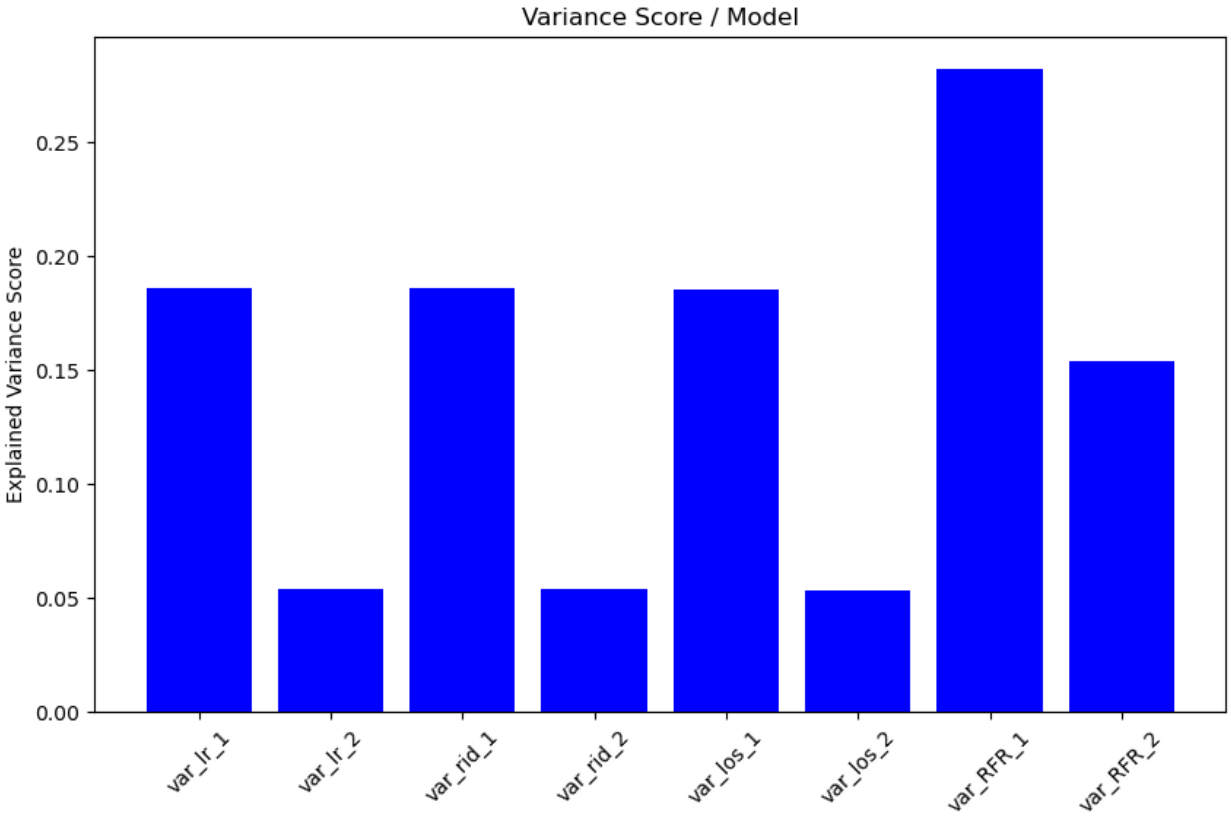
4) Model Comparison



True Values vs Predictions



5) Variance Comparision



6) Conclusion