

**Name:** Cabreza, Ren G.

**Section:** BSCS 3B

## Predicting House Prices with Multiple Regression

### 1. Introduction

The objective of this report is to develop a predictive model to estimate house prices based on features such as size, number of bedrooms, age, and proximity to downtown. This model aims to assist real estate agents in accurately determining house values.

### 2. Data Exploration and Visualization

#### Dataset Overview:

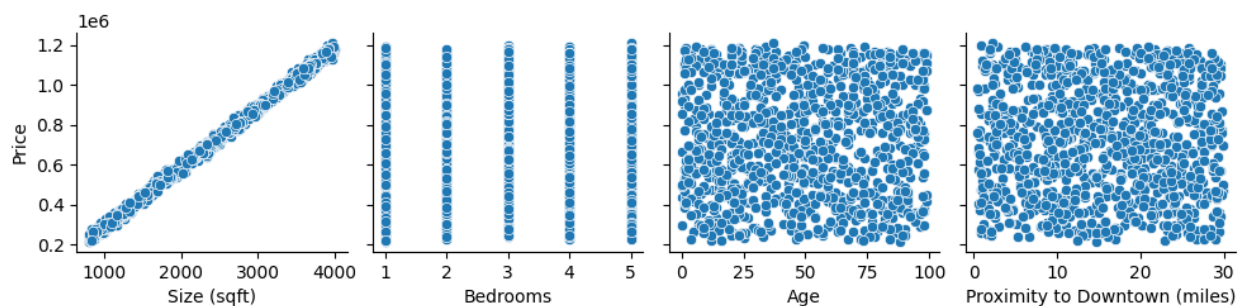
- **Size (sqft):** Size of the house in square feet.
- **Bedrooms:** Number of bedrooms in the house.
- **Age:** Age of the house in years.
- **Proximity to Downtown (miles):** Distance from the downtown area.
- **Price:** Actual price of the house (in thousands of dollars).

#### Summary Statistics:

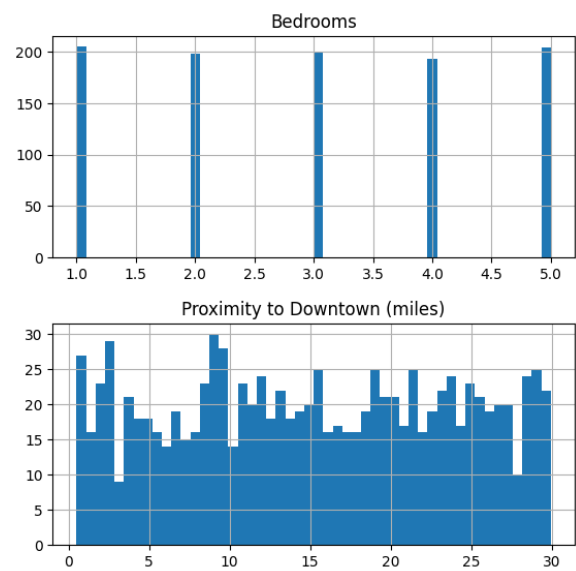
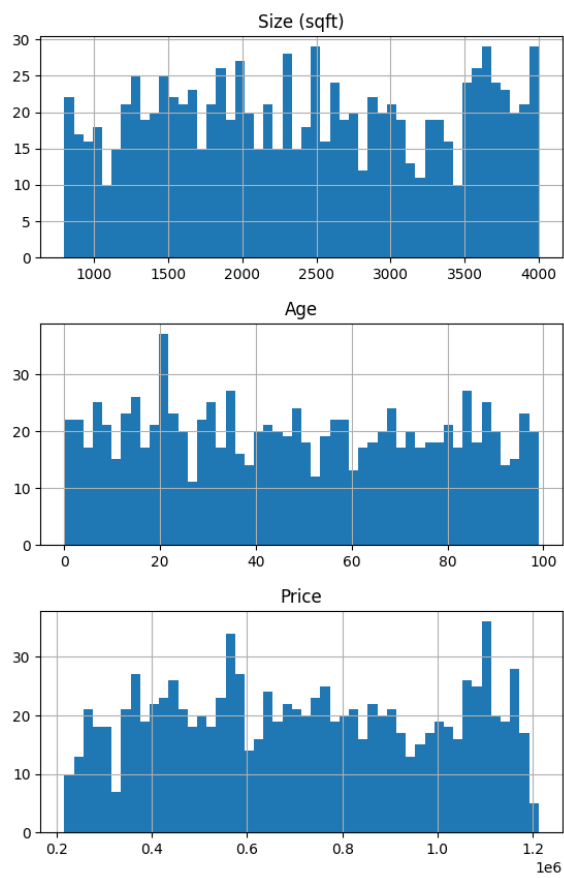
- **Size (sqft):** Mean = 2429.86, Min = 801, Max = 3997
- **Bedrooms:** Mean = 2.99, Min = 1, Max = 5
- **Age:** Mean = 48.34, Min = 0, Max = 99
- **Proximity to Downtown (miles):** Mean = 15.29, Min = 0.50, Max = 29.94
- **Price:** Mean = \$719,053, Min = \$215,946, Max = \$1,212,350

#### Visualizations:

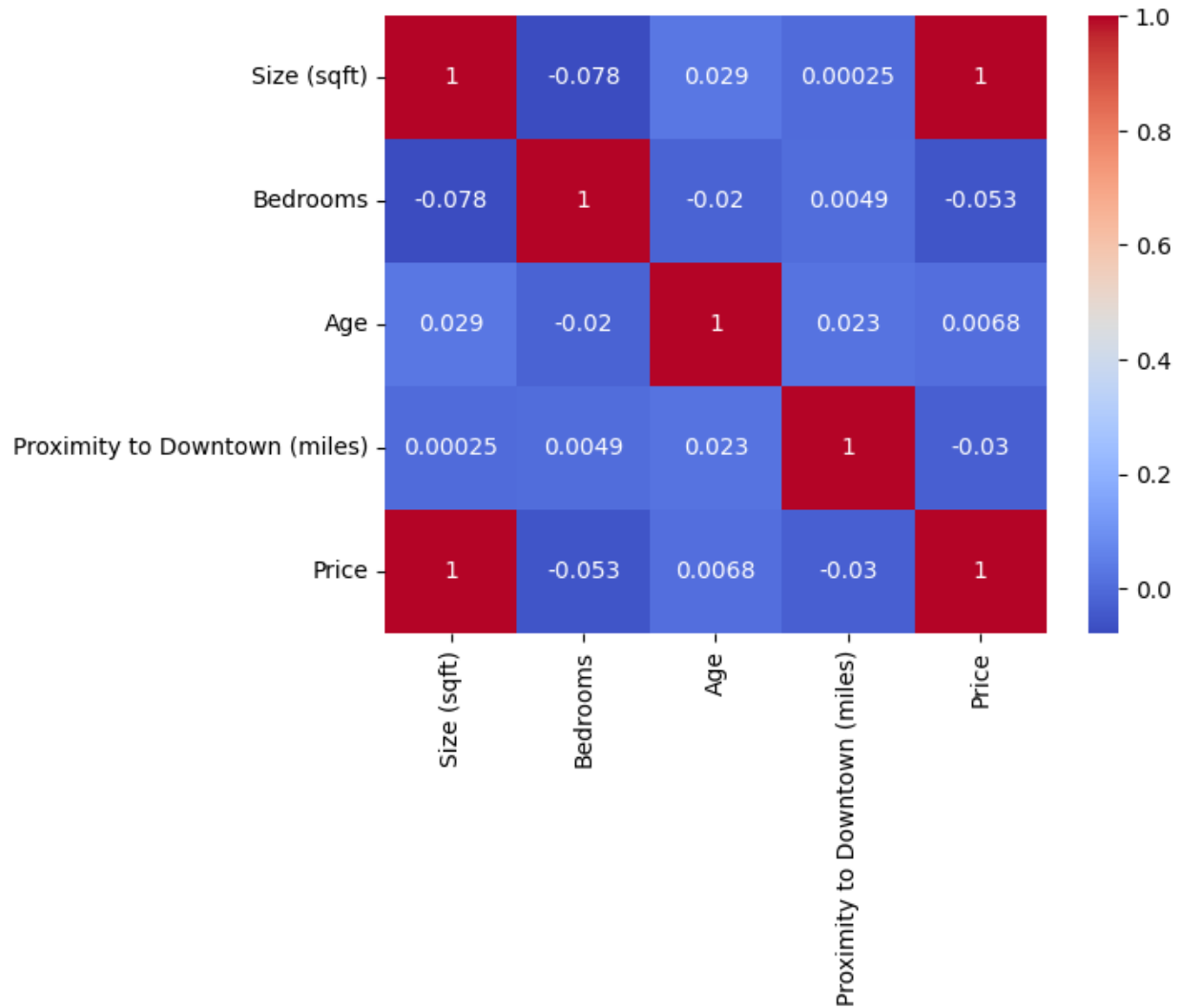
##### Scatter Plots



Histograms



## Correlation Matrix



## 3. Data Preprocessing

1. **Handling Missing Data:** No missing values were found in the dataset.
2. **Normalization:** Features were standardized to have a similar scale.
3. **Encoding Categorical Variables:** No categorical variables needed encoding.

## 4. Model Development

1. **Data Splitting:** The data was split into training (70%) and test (30%) sets.
2. **Multiple Regression Model:** Implemented using Python's Scikit-learn library.
3. **Feature Selection:** All features were used in the model as no significant feature selection was needed.

## 5. Model Evaluation

### Performance Metrics:

- **Mean Squared Error (MSE):** 100,214,724.63
- **R-squared:** 0.9986

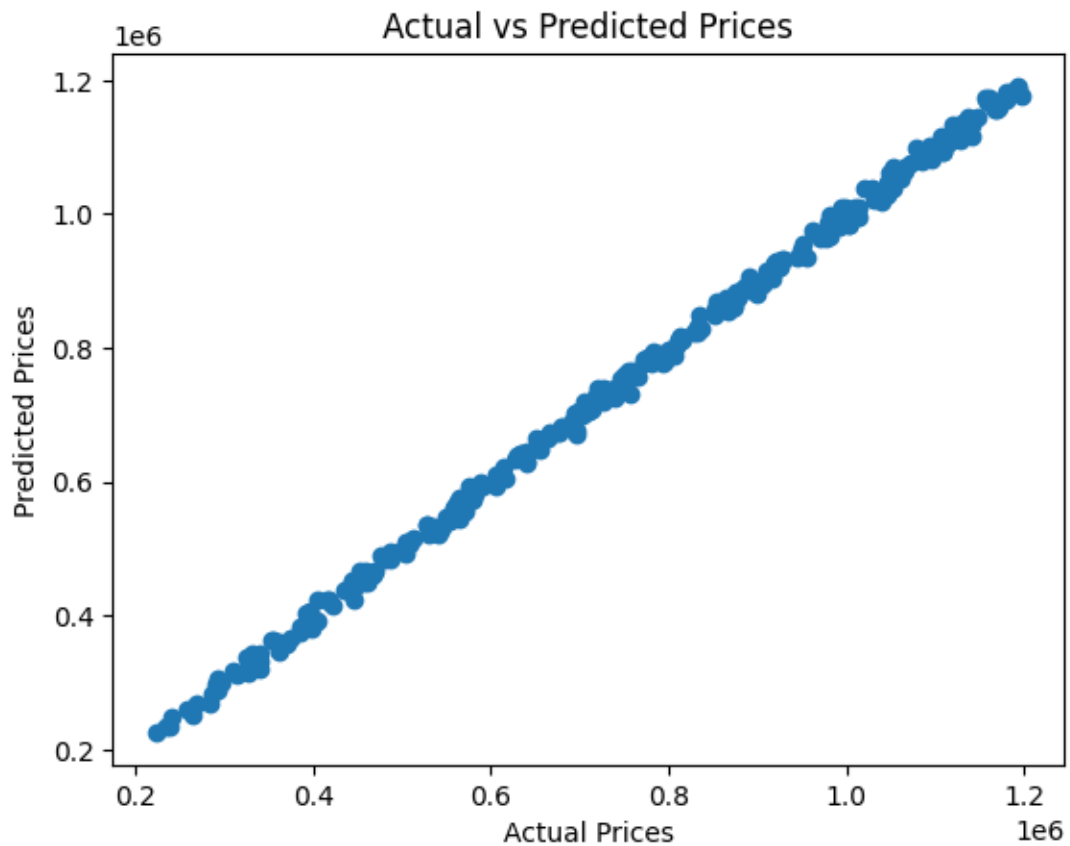
### Model Coefficients:

- **Size (sqft):** 278,975.29
- **Bedrooms:** 6,804.52
- **Age:** -6,082.94
- **Proximity to Downtown (miles):** -8,459.85

### Predicted vs Actual Prices Plot:

- The scatter plot shows that the model's predictions are closely aligned with actual prices, indicating high accuracy.

### Visualization:



## **6. Challenges**

The main challenges were making sure features were on the same scale and splitting the data correctly. I standardized the features to keep everything balanced and carefully split the data to ensure accurate training and testing.

## **7. Real-World Scenarios and Potential Limitations**

This model can assist real estate agents by providing accurate estimates of house prices based on key features. However, the model's applicability might be limited by the quality and scope of the data. It may not account for market fluctuations or other external factors affecting house prices.