Home Price Regression Analysis

Kameron Jackson

Contents

1. Introduction **1**

2. Data Summary **1**

3. Model Interpretation **2**

4. Diagnostic Plots **3**

5. Scenerio Analysis **7**

1. Introduction

This study aims to perform a comprehensive analysis of home prices, leveraging a dataset derived from 150 transactions in a local real estate market. Focusing on the response variable of home prices and two explanatory variables—square footage and the number of bathrooms— this analysis reveals intricate relationships within the data. Our objectives include visually exploring the dataset through scatterplots to identify patterns, fitting a multiple regression model to quantify these relationships, and evaluating the model's conditions and significance. By addressing these objectives, our analysis offers valuable insights for homeowners and real estate professionals navigating the dynamic landscape of housing transactions.

Scenario: A homeowner asked the realtor if she should spend $40,000 to convert a walk-in closet into a small bathroom in order to increase the sale price of her home. Analyze the home dataset, and determine whether the homeowner should remodel his current home according the analysis.

2. Data Summary

```
##
## Summary Statistics of Home Dataset
## ================================================
## Statistic      N   Mean   St. Dev.  Min   Max
## ------------------------------------------------
## Price         150 322.807 117.841   77    826
## Size          150  4.025   1.696   2.100 12.800
## Bathrooms     150  2.273   0.835     1     7
## Lot Size      150  7.827   3.239     1     21
## Median Income 150 155.867 128.760   31    668
## ------------------------------------------------
```

3. Model Interpretation

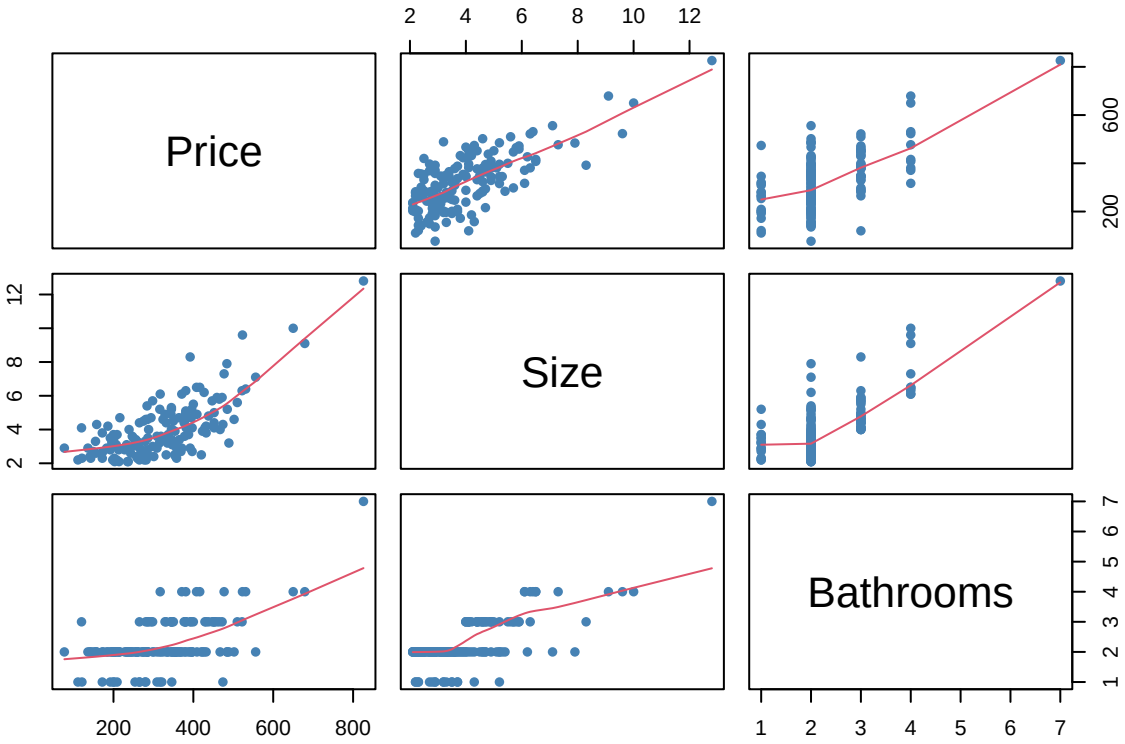
```
##
## Regression Results
## ================================================
##                          Dependent variable:
##                      ---------------------------
##                                 Price
## ------------------------------------------------
## Size                           45.161***
##                                 (5.782)
##
## Bathrooms                       14.794
##                                (11.747)
##
## Constant                       107.419***
##                                 (19.591)
##
## ------------------------------------------------
## Observations                      150
## R2                               0.534
## Adjusted R2                      0.527
## Residual Std. Error       81.031 (df = 147)
## F Statistic            84.060*** (df = 2; 147)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

In the context of our multiple linear regression model, an increase in size by one square foot corresponds to a price increase of $45,161, while the addition of each bathroom leads to a $14,791 increase in house price.

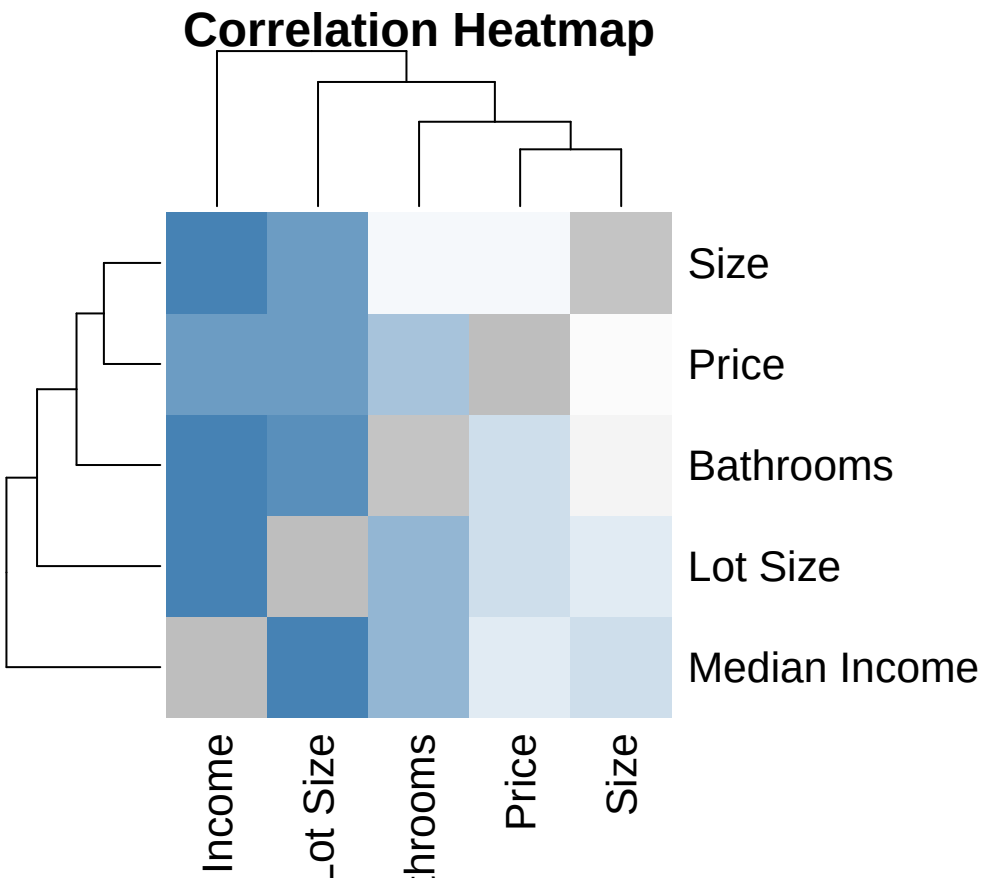
The decrease in adjusted R-squared aligns with the theory that one of the regressors may adversely affect the results. However, the overall R-squared values affirm the reliability of our dataset, explaining 53% of the linear variation. The F-statistic leads to rejecting the null hypothesis, which provides strong evidence that our model holds statistical significance, and enables a comprehensive analysis of home prices in an unpredictable market.

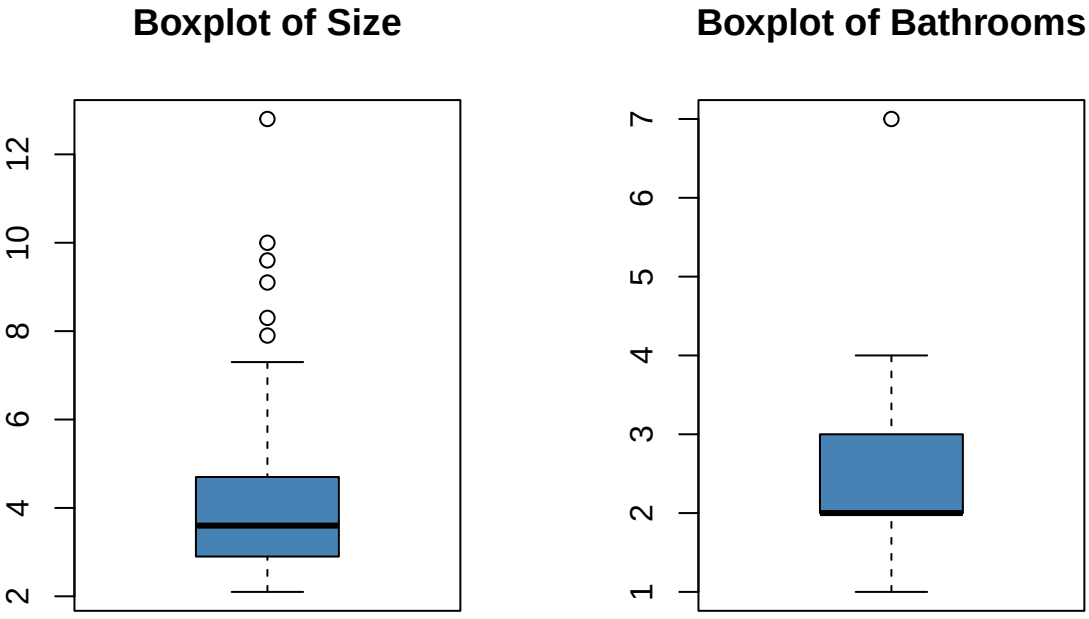
It is evident that the Size variable holds statistical significance, with a p-value < 0.01 denoted by '***' in the summary table. This indicates a meaningful relationship between the size of a house and its price. On the contrary, the corresponding p-value for the Bathroom variable is 0.21, a notably larger value. This higher p-value suggests a lack of statistical significance for the number of bathrooms in influencing house prices, therefore we can infer from this model that the number of bathrooms does not play a statistically significant role in explaining the variation in house prices. We will further explore the statistical significance of this model through data visualizations.

4. Diagnostic Plots

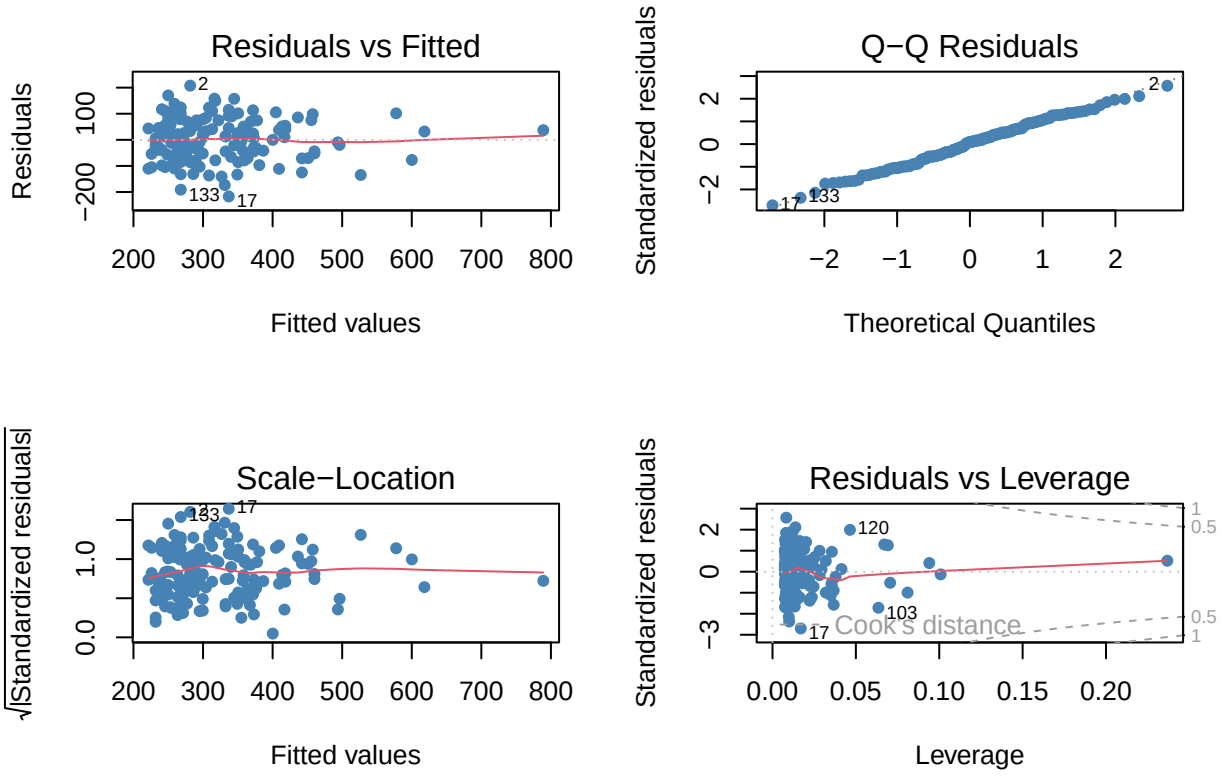


Summarizing the data with a scatter plot matrix reveals a clear positive linear relationship among home prices, size, and the number of bathrooms. This confirms that as house size and the amount of bathrooms increase, so does its price. The plots validate the linearity of these relationships, presenting a compelling case for utilizing multiple regression to investigate home pricing based on size and the number of bathrooms.



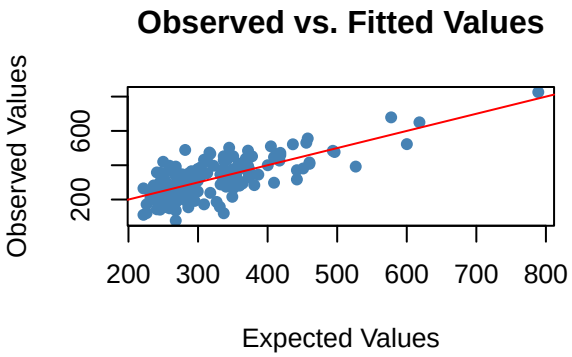


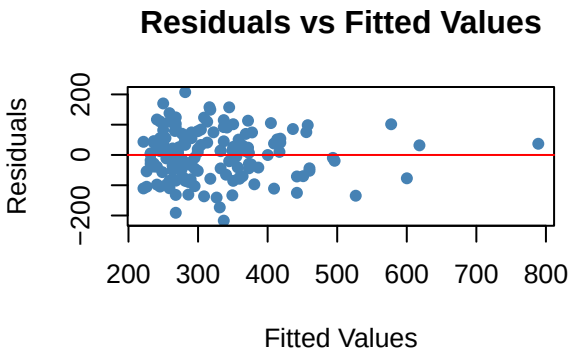
After viewing the boxplot visualizations, there are evident outliers for each explanatory variable - Size and Bathroom. Further investigation using statistical tests like a Leverage/Cook's distance analysis is necessary to assess their statistical significance and potential impact on the multiple regression model.

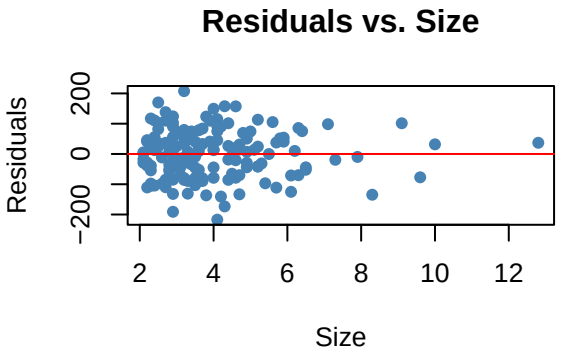



```
## NULL
```

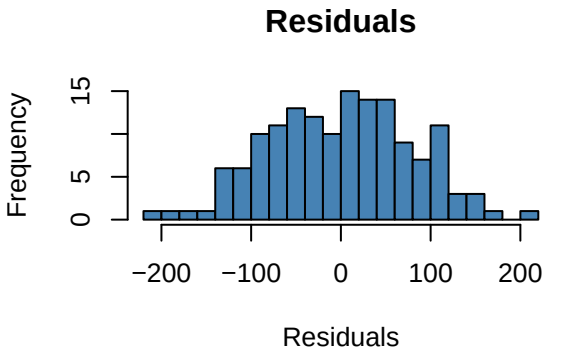
While a few consistent outliers are present in each visualization, they do not appear to significantly impact the overall trends of the graphs. The absence of points beyond Cook's threshold in the graph labeled 'Residuals vs Leverage' in the lower right-hand corner affirms the statistical significance of the data. This approach further reinforces the validity of our conclusions.

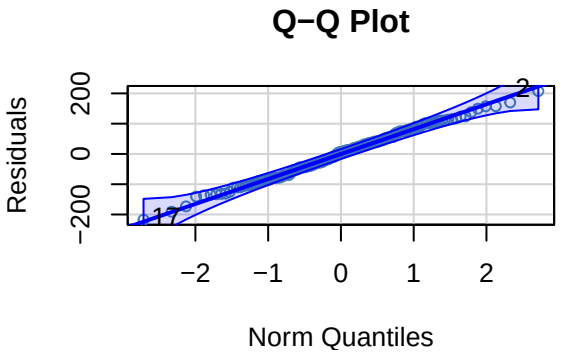







```
## [1] 17  2
```





After analyzing diagnostic plots for our multiple linear regression model predicting home prices based on Size and the number of Bathrooms, several positive observations emerge. The 'Observed vs Fitted' values plot reveals a strong positive linear relationship, affirming that our model effectively captures the overall trend of our hypothesis. Similarly, the 'Residuals vs Fitted' values plot and the 'Residuals vs Explanatory' values plots indicate ano correlation, with residuals evenly distributed around y=0.

This is further corroborated by the histogram of residuals, which shows an even distribution around 0, and the Normal QQ Plot displaying residuals along the line. In our case, the residuals closely align with the line, suggesting that our model has no systematic errors or biases.

This consistency across different levels of predictors reinforces the reliability of our model in predicting home prices, as seen by the even distribution of residuals (homoscedasticity). These features enhances the confidence in the model's overall performance, solidifying its ability to provide accurate predictions for home prices.

5. Scenerio Analysis

A homeowner asked the realtor if she should spend $40,000 to convert a walk-in closet into a small bathroom in order to increase the sale price of her home. What does your analysis indicate?

```
##
## ========================
##             2.5 %  97.5 %
## ------------------------
## (Intercept) 68.703 146.134
## Size        33.734 56.587
## Bathrooms   -8.421 38.009
## ------------------------
```

As seen based on the multiple linear regression mode, the increase in one bathroom will yield to a $14,794 increase in value of the hosue. Therefore, spending $40,000 to remodel does not justify the overall cost and benefit of turning the closet into a bathroom.

In 97.5% of cases, the actual unknown value of a house will likely fall within this interval. For example, when considering the price of a bathroom, the confidence interval suggests that the true cost will be between an $8,421 decrease in the home value and an $38,00 increase in home value. This range falls below the expense associated with replacing the small closet with a bathroom, which is $40,000. This indicates that the cost of the renovation and the increase in the house's value may not offer significant statistical evidence supporting the financial benefit for the homeowner. In essence, the statistical analysis suggests that the projected price of converting a walk-in-closet into a small bathroom may not be a justifiable expense of the renovation.