

Comprehensive Analysis of Housing Market Trends in Ames, Iowa

Anthony Firkins

CSU-Global

24SA-MIS470-1: Data Science Foundations

Dr. Orenthio Goodwin

Contents	
Introduction	3
Analysis and Interpretation	4
Visual Analysis	5
Limitations of Current Data Source	6
Longitudinal Analysis:	6
Comparative Analysis:	6
Impact of Amenities and Home Features:	7
Community Feedback and Surveys:	7
Conclusion	7
Appendix A	9
Figure 1: Calculated Statistics	9
Figure 2: Histogram and Box Plot of Sale Prices (Training Dataset)	10
Figure 3: Histogram and Box Plot of Sale Prices (Testing Dataset)	11
Figure 4: Summary Statistics (Combined Dataset)	12
Figure 5: Linear Regression Model Output	13
Figure 6. Diagnostic Plots for Linear Regression Analysis of Ames Housing Data	14
References	15

Introduction

The real estate market serves as a mirror reflecting the broader economic health and prosperity of a community. By analyzing local housing data, stakeholders can gain valuable insights into economic trends, demographic shifts, and market demands. This report focuses on Ames, Iowa—a typical Midwestern city whose housing market exhibits characteristics unique to the region and indicative of broader national trends. By dissecting the various aspects of Ames' housing market, this study will provide a detailed portrait of the factors influencing property values and sales activities.

Utilizing the Ames Housing Dataset, which contains over 1,000 records of residential property sales, this analysis seeks to uncover the critical price determinants and understand the distribution of property values within the city. The dataset includes a wealth of variables such as area, quality, condition, and the presence of various amenities, which help paint a comprehensive picture of the local real estate landscape.

In this report, we employ various statistical tools and analytical techniques in R, a robust open-source statistical environment, to derive meaningful insights from the data. Our primary goal is to identify critical statistical measures such as the minimum, maximum, mean, and median sale prices and calculate the standard deviation to understand the variability within the market. Moreover, through graphical representations like histograms and box plots, we visually explore the distribution of sale prices, aiming to highlight underlying market dynamics and trends.

This comprehensive analysis is intended to provide a snapshot of the current state of Ames' real estate market and offers valuable insights for potential buyers, sellers, investors, and policymakers. By understanding the factors that drive property values and how they are distributed across the market, these stakeholders can make more informed decisions, contributing to more strategic planning and development within the community.

Analysis and Interpretation

All figures are in Appendix A: The initial exploration of the combined Ames Housing Dataset reveals that the sale price is essential for understanding the local real estate market dynamics. By employing R, a robust statistical computing environment, this analysis extracts several critical insights about the housing market in Ames, Iowa. The minimum sale price was recorded at \$34,900 at the lower end of the market spectrum. This price point indicates potentially entry-level or distressed properties, highlighting the most accessible segment of the market for buyers with limited budgets or investors looking for renovation opportunities.

Conversely, the maximum sale price soars to \$755,000, showcasing the presence of luxury or highly sought-after homes that cater to a wealthier clientele. This stark contrast in the price range underscores the economic diversity and the vast array of housing options available in Ames, catering to vastly different buyer groups. The mean sale price, approximately \$180,921.20, acts as a benchmark, providing a general sense of the market's pricing levels and helping potential buyers and investors gauge the typical costs they might encounter in this market.

The median sale price, standing at \$163,000, offers a central point that helps balance the skew caused by high-end property sales. This figure suggests that more than half of the homes in Ames are priced below this midpoint, offering a realistic view of what a median buyer might

expect to pay. Additionally, the calculated standard deviation of \$79,442.50 points to considerable variability in housing prices. Such variability can often be attributed to diverse factors, including the location of a property within Ames, its size, age, and the unique amenities it offers. These factors can significantly influence the desirability and, therefore, the price of homes, adding complexity to the market dynamics.

Visual Analysis

All figures are in Appendix A: The distribution of sale prices in Ames is further explored through detailed visual analysis using histograms and box plots for the training and combined datasets (figures 2-4). The histograms display a right-skewed distribution, indicating a higher concentration of properties at the lower end of the pricing spectrum and a gradual decline in frequency as prices increase. This skewness is typical of many residential real estate markets where many transactions occur at or below the median price range.

The accompanying box plots enhance this visualization by highlighting the median prices and the interquartile range, delineating the middle 50% of the data. These plots are particularly effective in illustrating the spread of the data and pinpointing outliers, which appear as points beyond the whiskers of the box plots. The presence of these outliers, particularly in the upper price range, corroborates the noted high variability and confirms the existence of premium properties that significantly exceed the average prices in Ames.

Together, these visual tools provide a comprehensive understanding of the housing market's structure in Ames. They reveal significant market segmentation, characterized by a substantial inventory of economically accessible homes contrasted by a smaller segment of premium properties that push the upper limits of the price range. This segmentation suggests that while the market is predominantly accessible to first-time homebuyers and middle-income

families, it also accommodates the luxury preferences of higher-income buyers. Thus, Ames offers diverse housing options, meeting the needs and preferences of a broad spectrum of residents and investors.

Limitations of Current Data Source

While detailed, the current analysis of Ames' housing market operates under certain limitations stemming from the nature of the data source used. To address these and enhance the robustness of future reports, several vital improvements are proposed:

Longitudinal Analysis:

Incorporating data spanning multiple years would provide a clearer understanding of trends. This could illuminate how housing prices and inventory levels respond to economic cycles, offering predictive insights into future market behaviors. A longitudinal perspective would help identify persistent trends or abrupt changes due to external shocks.

Comparative Analysis:

A deeper context could be established by comparing the Ames housing market with neighboring regions or similarly sized cities. This comparison would highlight Ames' unique characteristics or competitive advantages and provide a benchmark against which the performance of the Ames market can be evaluated.

Demographic and Economic Factors:

Integrating demographic and economic data would allow for a more nuanced analysis of the factors driving housing market trends. Correlating housing prices and sales volumes with variables such as median income, employment rates, and demographic shifts could elucidate why

specific patterns emerge in the data. This approach would link market dynamics directly to the socio-economic conditions of the area.

Impact of Amenities and Home Features:

Extending the analysis to consider the influence of specific home features and nearby amenities could yield valuable insights to buyers and developers. For instance, evaluating the impact of proximity to parks, schools, and public transportation or the value added by sustainable home features could help understand consumer preferences and plan future developments.

Community Feedback and Surveys:

Adding qualitative data through community feedback and surveys would provide a ground-level perspective often absent in quantitative analyses. Insights into residents' satisfaction, housing needs, and future aspirations could reveal personal factors and community sentiments influencing market dynamics. This feedback would make the analysis more comprehensive and rooted in the lived experiences of the community members.

These enhancements would address the current data source's limitations and enrich the analysis, making it more relevant and actionable for stakeholders across the Ames real estate landscape.

Conclusion

This detailed exploration of Ames' housing market underscores a narrative of diversity and inclusion within the housing sector. Using statistical and visual analysis tools has uncovered a market characterized by a wide array of pricing options, reflecting the broad spectrum of community needs and financial capabilities. The varied pricing, from modest starter homes to

luxurious estates, illustrates a housing landscape that is vibrant and dynamic, catering to a diverse demographic from first-time homebuyers to affluent investors.

These insights enhance transparency and facilitate informed decision-making for stakeholders across the real estate spectrum, including buyers, sellers, investors, and policymakers. By providing a clear and data-driven snapshot of the market conditions, this analysis empowers all participants to navigate the complexities of real estate with greater confidence and clarity.

Appendix A

Figure 1: Calculated Statistics

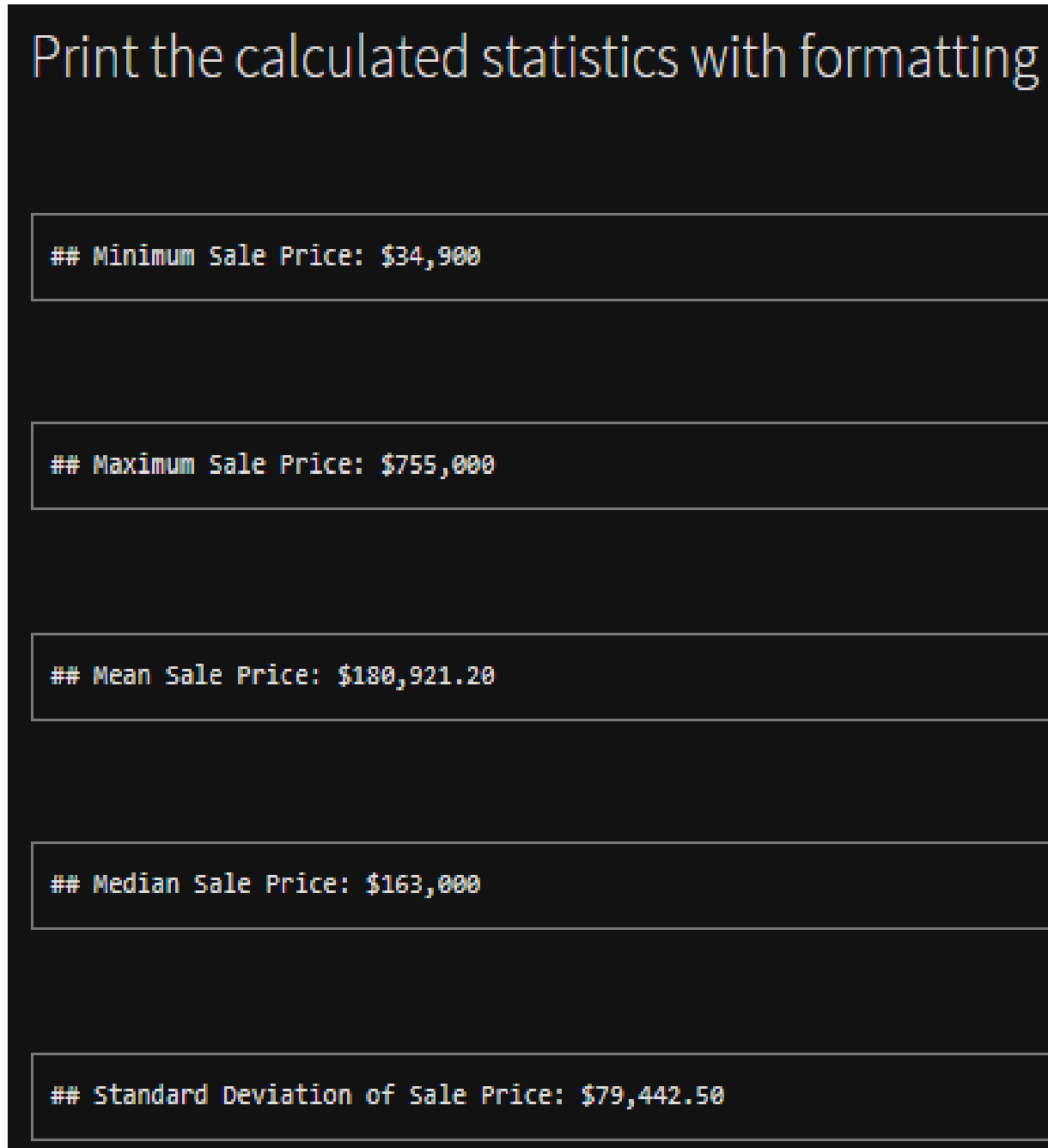


Figure 1 shows the calculated summary statistics for the combined dataset. The minimum, maximum, mean, median, and standard deviation of sale prices are displayed in U.S. currency.

Figure 2: Histogram and Box Plot of Sale Prices (Training Dataset)

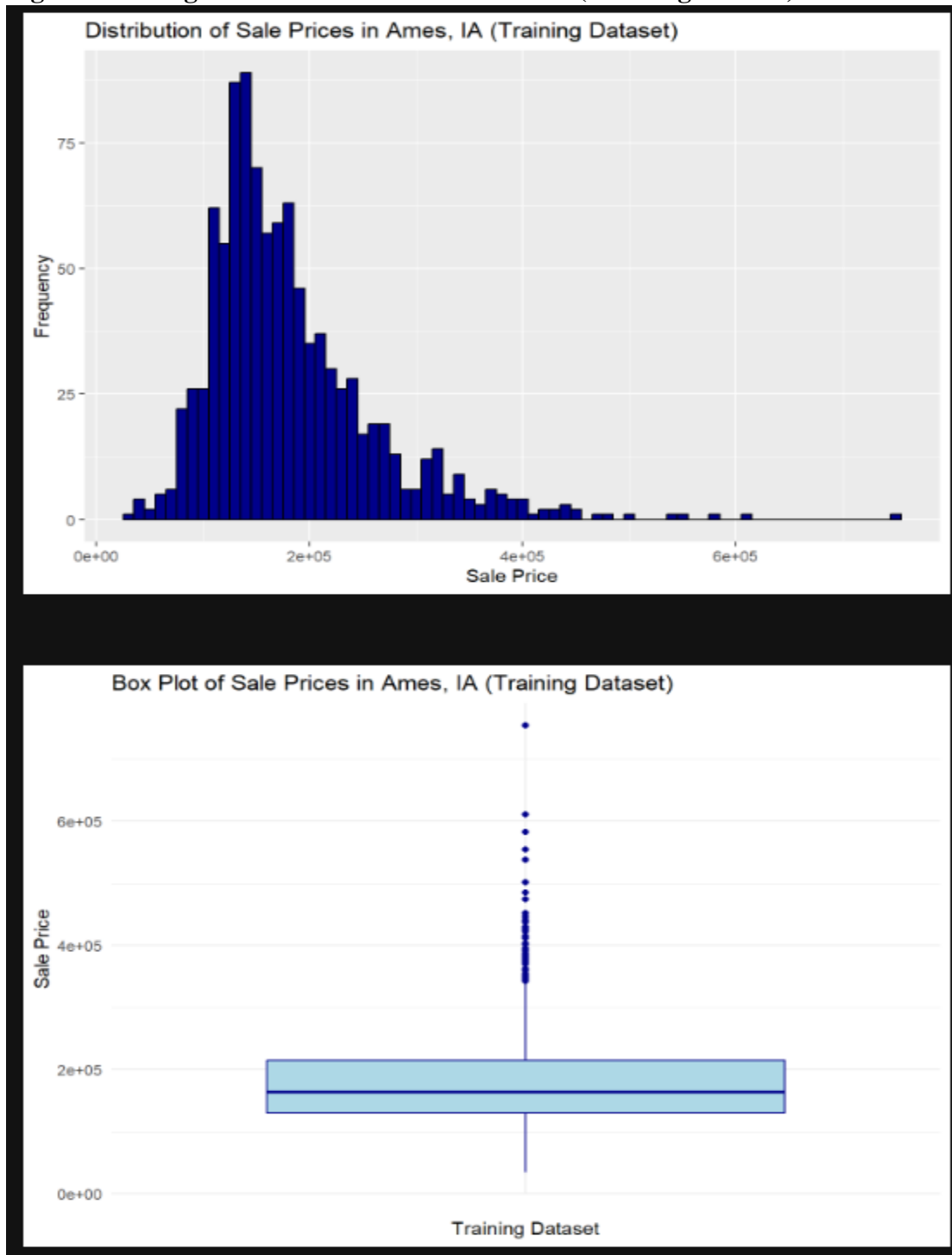


Figure 2 depicts the histogram and box plot for the training dataset. These visualizations show the distribution and the spread of sale prices, emphasizing lower-end and outlier properties.

Figure 3: Histogram and Box Plot of Sale Prices (Testing Dataset)

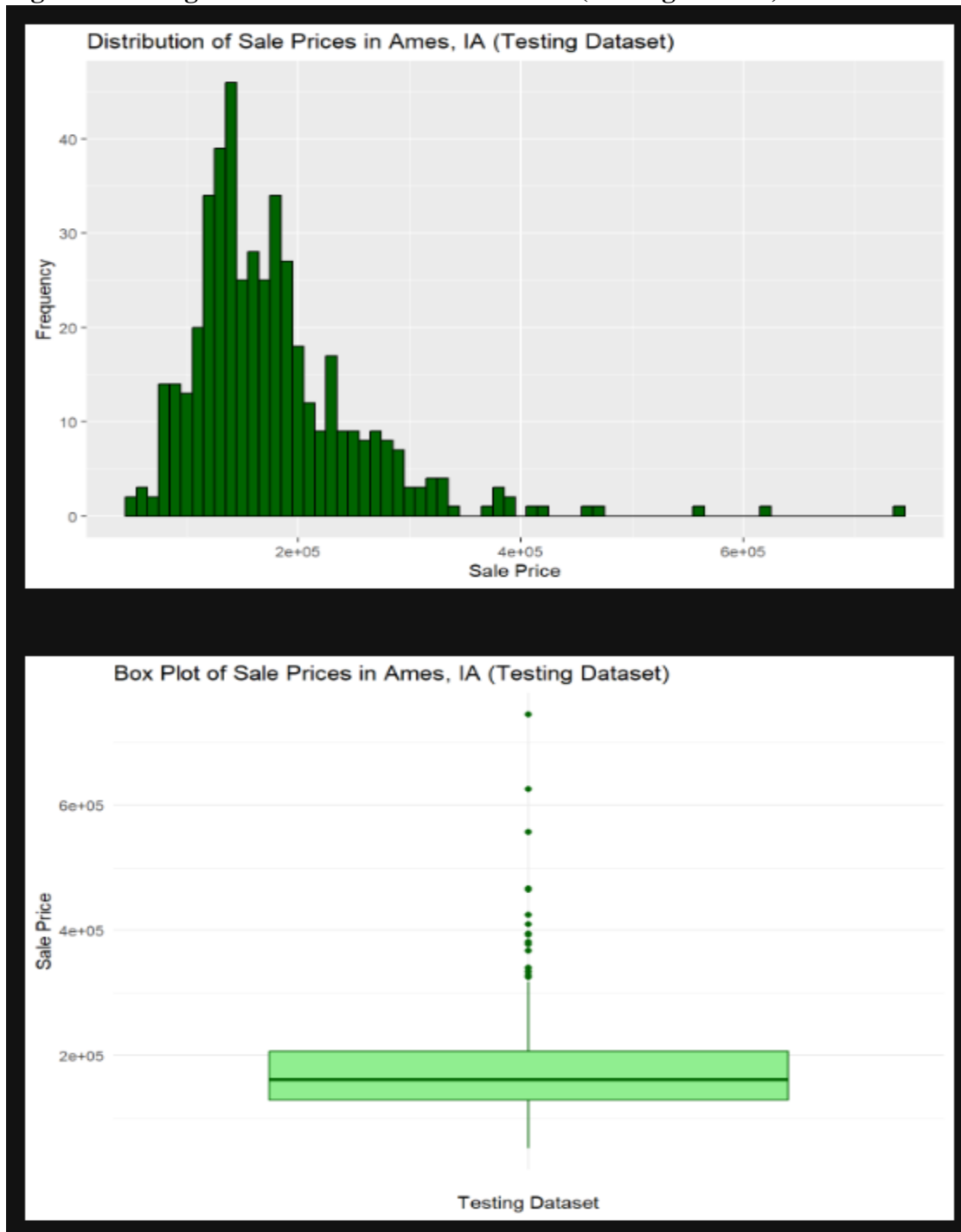


Figure 3 shows the distribution of sale prices within the testing dataset through histogram and box plot, indicating price variability and central tendency.

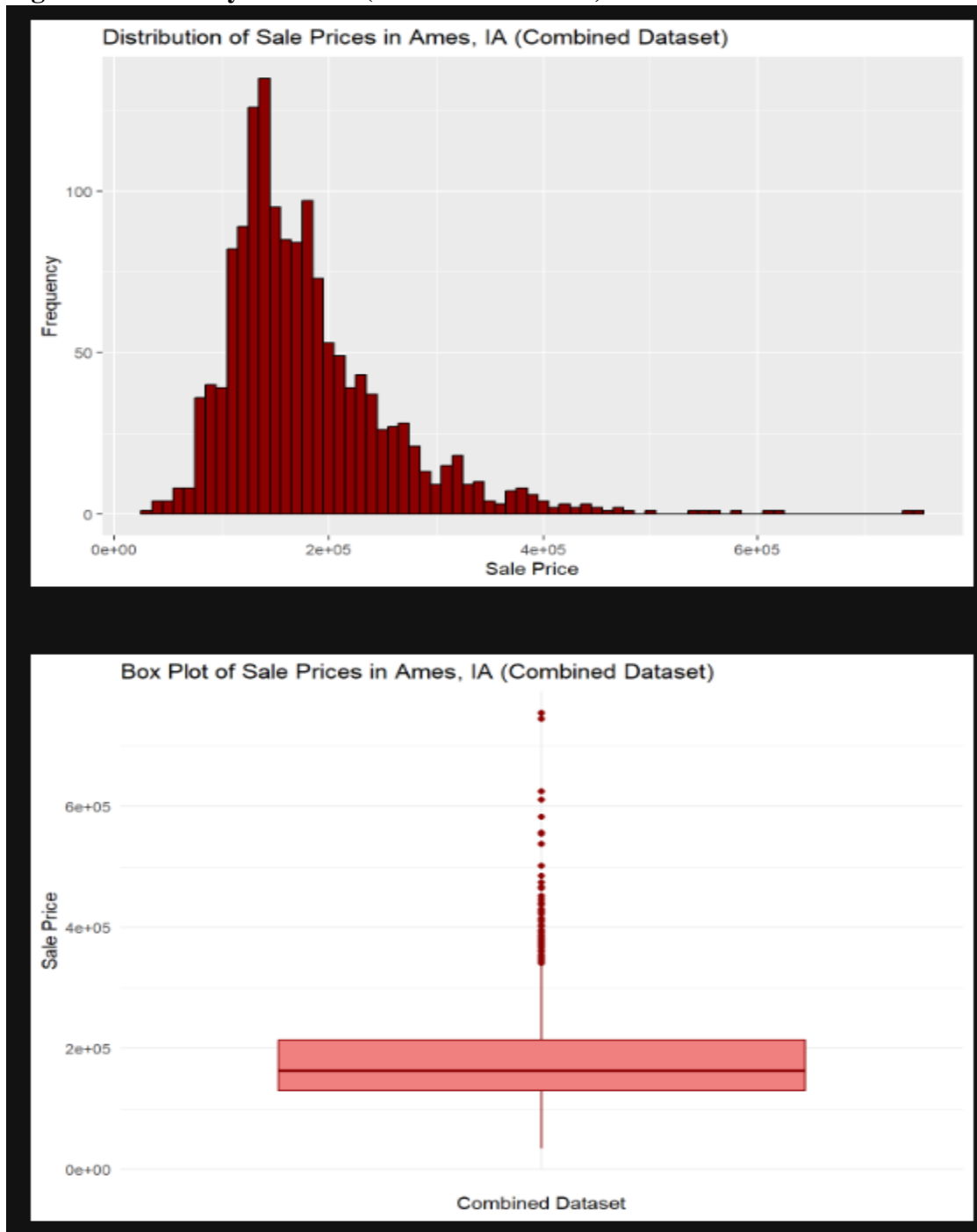
Figure 4: Summary Statistics (Combined Dataset)

Figure 3 provides summary statistics for the combined dataset, illustrating key measures such as quartiles, mean, and standard deviations across various housing attributes.

Figure 5: Linear Regression Model Output

```
##
## Call:
## lm(formula = SalePrice ~ ., data = training_data_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -378288 -17194  -1810   15486  235047
##
## Coefficients: (2 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.073e+06  1.864e+06  -0.576  0.56489
## Id            4.910e+00  4.290e+00   1.144  0.25281
## MSSubClass    -1.015e+02  3.716e+01  -2.732  0.00645 **
## LotFrontage   -7.293e+00  6.763e+01  -0.108  0.91416
## LotArea        7.206e-01  1.480e-01   4.868  1.37e-06 ***
## OverallQual    1.530e+04  1.634e+03   9.362  < 2e-16 ***
## OverallCond    6.498e+03  1.536e+03   4.231  2.62e-05 ***
## YearBuilt      4.289e+02  9.210e+01   4.657  3.80e-06 ***
## YearRemodAdd   1.706e+02  9.408e+01   1.813  0.07016 .
## MasVnrArea     3.279e+01  7.357e+00   4.458  9.56e-06 ***
## TotalBsmtSF    2.581e+01  4.405e+00   5.858  7.01e-09 ***
## GrLivArea      4.773e+01  6.250e+00   7.638  6.78e-14 ***
## FullBath       -2.674e+03  3.782e+03  -0.707  0.47974
## HalfBath        5.362e+02  3.379e+03   0.159  0.87394
## BedroomAbvGr   -1.497e+04  2.339e+03  -6.400  2.74e-10 ***
## KitchenAbvGr   -2.802e+04  7.166e+03  -3.911  0.00010 ***
## TotRmsAbvGrd    9.040e+03  1.659e+03   5.448  6.93e-08 ***
## Fireplaces      6.267e+03  2.359e+03   2.656  0.00807 **
## GarageYrBlt     -9.373e+01  9.898e+01  -0.947  0.34393
## GarageCars       5.174e+03  3.864e+03   1.339  0.18093
## GarageArea      4.690e+01  1.331e+01   3.524  0.00045 ***
## WoodDeckSF      2.016e+01  1.119e+01   1.802  0.07197 .
## OpenPorchSF     -2.212e-01  2.136e+01  -0.010  0.99174
## MoSold         -8.515e+02  4.709e+02  -1.808  0.07095 .
## YrSold          2.068e+00  9.300e+02   0.002  0.99823
## HasGarage              NA         NA      NA      NA
## HasLotFrontage        NA         NA      NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 33730 on 746 degrees of freedom
## Multiple R-squared:  0.8473, Adjusted R-squared:  0.8423
## F-statistic: 172.4 on 24 and 746 DF,  p-value: < 2.2e-16
```

Figure 5 displays the output from the linear regression model based on the training dataset. Key predictors of sale price are identified with coefficients, standard errors, and significance levels.

Figure 6. Diagnostic Plots for Linear Regression Analysis of Ames Housing Data

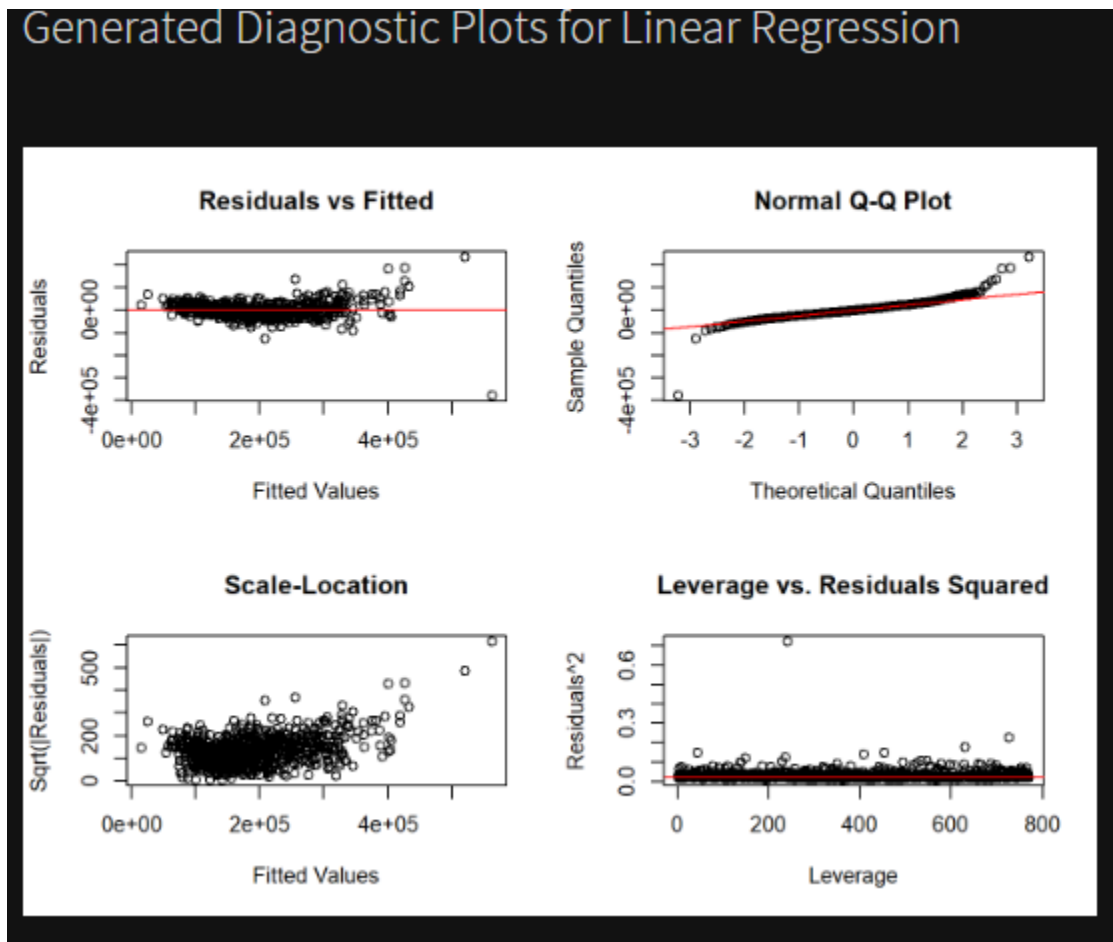


Figure 6 presents four diagnostic plots to assess the adequacy of a linear regression model predicting housing prices in Ames, Iowa. The plots include (1) Residuals versus Fitted Values, showing whether residuals have non-linear patterns, which would indicate non-linearity in the data; (2) Normal Q-Q Plot, assessing whether residuals are normally distributed; (3) Scale-Location (or Spread-Location) Plot, showing whether residuals are spread equally along the ranges of predictors (homoscedasticity), and (4) Leverage versus Residuals Squared, identifying influential cases that might affect the regression model significantly. The plots collectively help verify the assumptions of linear regression and identify potential anomalies or model deficiencies.

References

- Kramer, A., Warren, A., Maher, W., McLaughlin, M., Nelson, A., Payne, O., & Tsang, E.
(2021). Emerging Trends in Real Estate®, 2021 - ULI Americas. Urban Land Institute.
<https://americas.uli.org/wp-content/uploads/2021/02/Emerging-Trends-in-Real-Estate-United-States-and-Canada-2021.pdf>
- Nowak, A., & Smith, P. (2016). Textual analysis in real estate. *Journal of Applied Econometrics*, 32(4), 896–918. <https://doi.org/10.1002/jae.2550>