# Buy Vs Rent

Shahriar Nekouei

December 2024

## 1 Introduction

"Is buying better than renting?" This question has long puzzled economists due to its complexity and the multitude of variables that influence it. Conventional wisdom suggests buying is preferable because mortgage payments contribute to home-ownership, whereas rent is merely a temporary living expense. However, for young professionals at the beginning of their careers, with adventurous mindsets rather than a desire to settle down, renting may seem more appealing. Renting provides the flexibility to explore different areas to determine the most suitable location before making a long-term commitment.

In recent years, the affordability of housing has drastically diminished, forcing many to choose the most economically viable option available. According to CNBC, approximately 31 percent of Generation Z still live with their parents, unable to afford either buying or renting their own place. The situation is particularly acute in Colorado, where the cost of living exceeds the national average by 6 percent, and the state ranks as the fifth most expensive in the United States, according to RentCafe and CNBC.

This project seeks to identify the variables that significantly impact rent and mortgage prices in Colorado and to develop predictive models to determine whether renting or purchasing is the better option. To ensure relevance and focus, the analysis is confined to single-family homes, the most common type for first-time buyers. As defined by Rocket Mortgage, a single-family home is a free-standing residential structure intended for one household. The U.S. Census Bureau includes townhouses with individual laundry and air conditioning units in this category. The study focuses on three Colorado cities: Denver, Boulder, and Fort Collins, excluding condominiums and apartments due to their price volatility and lack of reliable datasets.

# 2 Methods and Data

For this project, the main analytical approach involves two separate time-series multilinear regression models: one for predicting purchase (mortgage) costs and the other for rent costs. The data used in this analysis was sourced from the Zillow Housing Database, the Federal Reserve Bank of St. Louis, and the U.S. Bureau of Labor Statistics. Due to data limitations, the analysis covers the period from August 2018 to August 2024.

## 2.1 Dependent Variables

- **Single Family Home Value Index:** Obtained from the Zillow database, this reflects the typical value of homes in the 35th to 65th percentile range.
- **Single Family Observed Rent Index:** Also sourced from Zillow, this represents the average of listed rents for single-family homes within the same percentile range.

## 2.2 Independent Variables

| **Independent Variables** | **Definition** | **Reason** |
|---|---|---|
| Monthly Inflation | A measure of change in price between one month and the previous month | Economical Factor |
| Unemployment Rate | The number of unemployed people as a percentage of labor force in Colorado | Economical Factor |
| 30 Year Mortgage Rate | A fixed rate of mortgage for 30 years | Economic Factor |
| Inflation Expectation Rate | The rate at which people, consumers, businesses, and investors expect prices to rise in the future | Economical Factor |
| Consumer Sentiment Index | An economic indicator that measures how optimistic consumers feel about their finances and the state of the economy | Economical Factor |
| Average Hourly Earning | The average amount people make per hour per month | Economical Factor |
| Labor Participation Rate | The percentage of people aged 16 and older who are in the labor force | Economical Factor |
| Heat Index | A time series dataset that aims to capture the balance of for-sale supply and demand in a given market. A higher number means the market is more tilted in favor of sellers | Market Factor |
| Number of Permits Authorized | New Private Housing Units Authorized by Building Permits: 1-Unit Structures for Colorado | Market Factor |
| Number of Listings | The count of unique listings that were active each month | Market Factor |
| Personal Saving Rate | The portion of personal income that is used either to provide funds to capital markets or to invest in real assets such as residences | Economical Factor |

For a more robust analysis and accurate comparisons between purchasing and renting, I used the following formula to calculate the monthly payment for a 30-year fixed mortgage. This calculation incorporates the Single-Family Home Value Index as the principal amount and the 30-Year Fixed Mortgage Rate as the interest rate:

$$M = P \times \frac{r_{\text{monthly}} \times (1 + r_{\text{monthly}})^N}{(1 + r_{\text{monthly}})^N - 1}$$

$$M = \text{Monthly Mortgage Payment}$$

$$r_{\text{down}} = \text{Down Payment Percentage}$$

$$P = \text{Principle Amount} = \text{Single Family Home Value Index} \times (1 - r_{\text{down}})$$

$$r_{\text{annual}} = \text{Annual Mortgage Rate (30 Year Fixed Mortgage Rate)}$$

$$r_{\text{monthly}} = \frac{r_{\text{annual}}}{12 \times 100}$$

$$N = \text{Total Number of Monthly Payments} = 30\,\text{years} \times 12 = 360$$

It is important to mention that the monthly mortgage payment variable would not include escrow which includes the property insurance and property tax.

# Equations

**Monthly Mortgage Payment:**

$$\begin{aligned}\text{Monthly Mortgage Payment} = {} & \beta_0 + \text{Monthly Inflation} \cdot \beta_1 + \text{Unemployment Rate} \cdot \beta_2 \\ & + \text{Mortgage Rate (Fixed 30 Year)} \cdot \beta_3 + \text{Inflation Expectation Rate} \cdot \beta_4 \\ & + \text{Consumer Sentiment Index} \cdot \beta_5 + \text{Average Hourly Earnings} \cdot \beta_6 \\ & + \text{Labor Participation Rate} \cdot \beta_7 + \text{Heat Index} \cdot \beta_8 \\ & + \text{Number of Permits Authorized} \cdot \beta_9 + \text{Number of Listings} \cdot \beta_{10} \\ & + \text{Personal Saving Rate} \cdot \beta_{11}\end{aligned}$$

**Observed Rent Index:**

$$\begin{aligned}\text{Observed Rent Index} = {} & \beta_0 + \text{Monthly Inflation} \cdot \beta_1 + \text{Unemployment Rate} \cdot \beta_2 \\ & + \text{Mortgage Rate (Fixed 30 Year)} \cdot \beta_3 + \text{Inflation Expectation Rate} \cdot \beta_4 \\ & + \text{Consumer Sentiment Index} \cdot \beta_5 + \text{Average Hourly Earnings} \cdot \beta_6 \\ & + \text{Labor Participation Rate} \cdot \beta_7 + \text{Heat Index} \cdot \beta_8 \\ & + \text{Number of Permits Authorized} \cdot \beta_9 + \text{Number of Listings} \cdot \beta_{10} \\ & + \text{Personal Saving Rate} \cdot \beta_{11}\end{aligned}$$

# 3 Transformation (Monthly Mortgage Payment)

Given that this project involves time-series analysis across three cities (Denver, Boulder, and Fort Collins), it was necessary to include three binary dummy variables representing these cities. These dummy variables allow the model to account for city-specific effects and variations over time.

To capture temporal dependencies inherent in the time-series data, I created a new variable, the ”Time Index,” which incorporates the chronological sequence of the data. This variable helps the model recognize and adjust for temporal trends and patterns in the dataset.

The dataset was divided into training and testing subsets, with 80 percent of the data allocated for model training and 20 percent reserved for validation. This split ensures that the model is trained effectively while preserving data for unbiased evaluation.

# Actual vs. Fitted Mortgage (Training Set)

Here is the plot comparing actual vs. fitted values for the mortgage model on the training set:

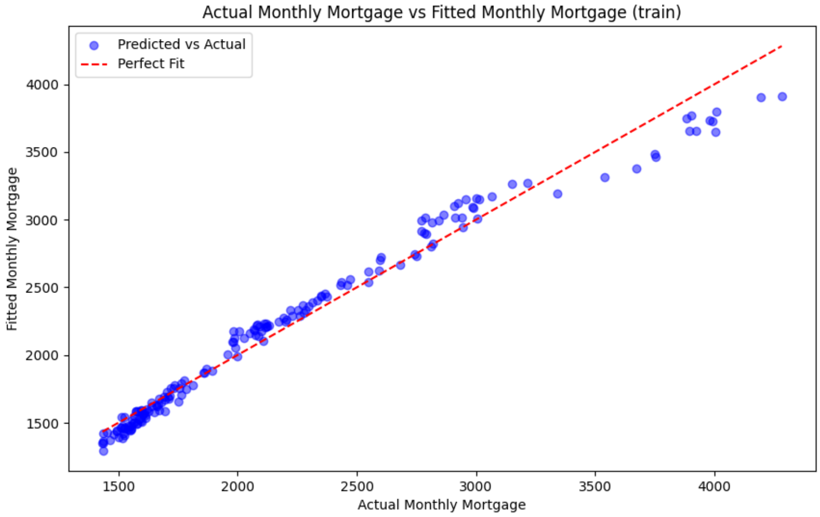


Figure 1: Actual vs. Fitted Values for the Mortgage Model (Training Set)

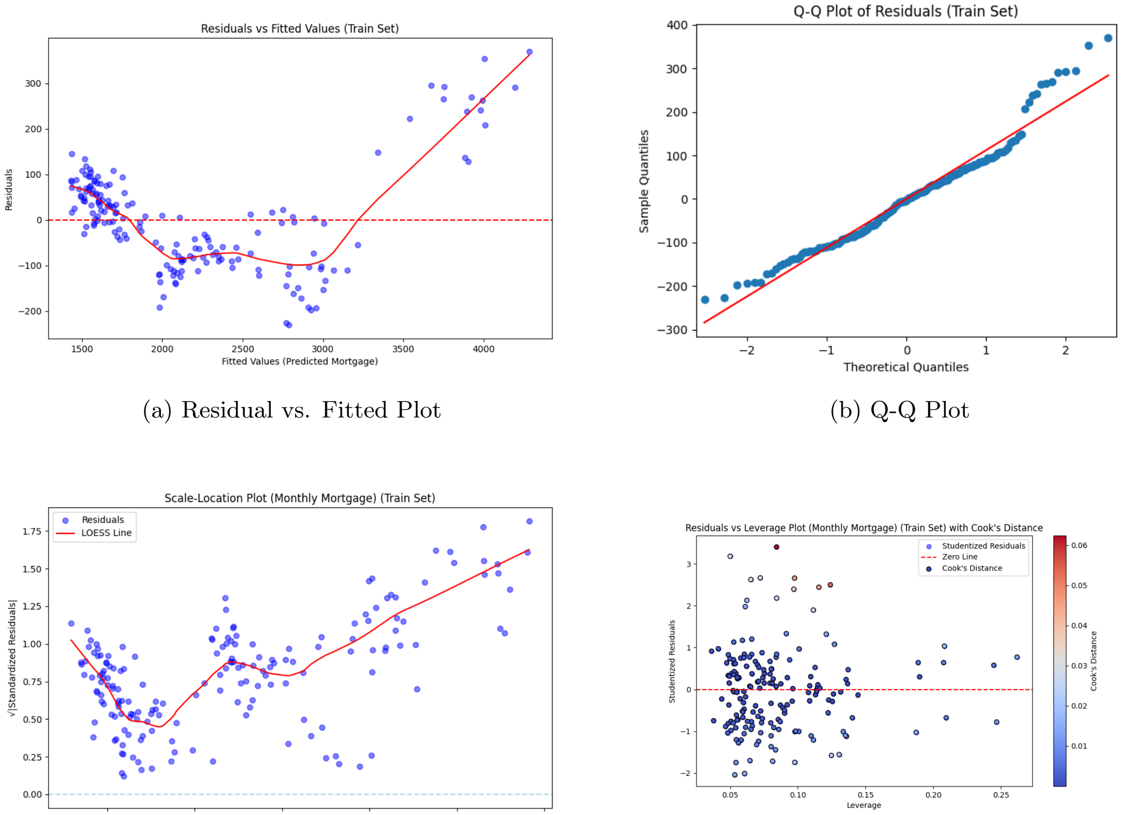
## 3.1 Assumption Check: Linear Regression

The actual vs. fitted value plot reveals deviations from linearity, particularly as the monthly mortgage payment increases. This indicates potential non-linearity in the model. Before proceeding with further analysis, it is essential to validate the assumptions of linear regression as well as the Gauss-Markov assumptions to ensure the model's reliability and interpretability.

### 3.1.1 Key Assumptions of Linear Regression

**Linearity**: The relationship between the predictors and the response variable should be linear. Any non-linear patterns in the residuals suggest a violation of this assumption. . **Homoscedasticity**: The variance of residuals should remain constant across all levels of the predicted values. If the residuals show a pattern or funnel shape, it indicates heteroskedasticity. **Independence of Errors**:The residuals (errors) should be independent of one another. This is often tested using the Durbin-Watson statistic or an autocorrelation plot. **Normality of Errors**: The residuals should follow a normal distribution, which can be assessed visually using Q-Q plots or quantitatively using normality tests like the Shapiro-Wilk test.

# Diagnostic Plots (Monthly Mortgage Payment)



(a) Residual vs. Fitted Plot

(b) Q-Q Plot

(c) Scale-Location Plot

(d) Residuals vs. Leverage Plot (with Cook's Distance)

Figure 2: Diagnostic Plots for the Mortgage Model: Residual vs. Fitted, Q-Q Plot, Scale-Location, and Residuals vs. Leverage.

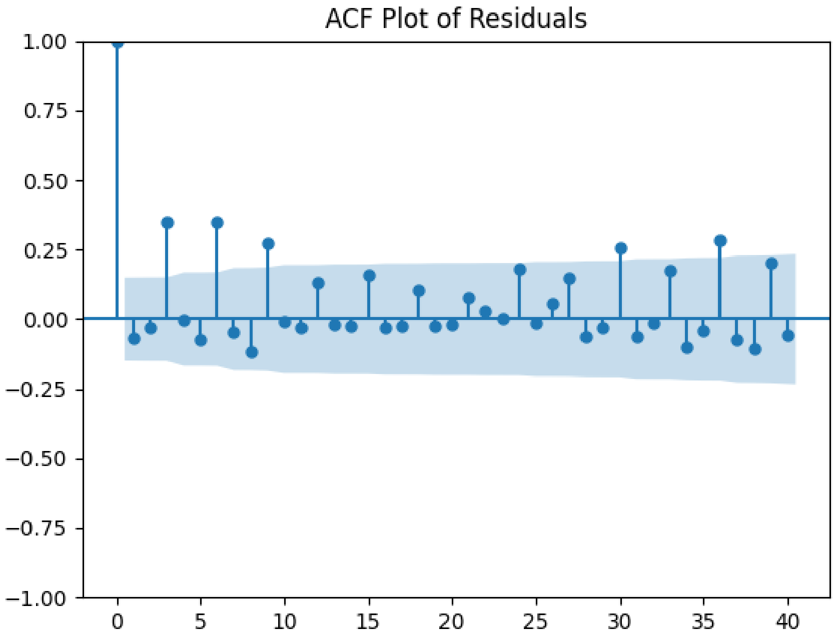


Figure 3: Autocorrelation Function (ACF) Plot

#### 3.1.2 Interpretation of Diagnostic Plots

The diagnostic plots indicate that all four assumptions of linear regression have been violated, necessitating corrective measures before proceeding with any analysis. Here's a detailed interpretation:

1. **Residual vs Fitted Values Plot**:This plot shows a clear pattern, suggesting a non-linear relationship between the predictors and the response variable.

2. **Scale-Location Plot**: The presence of a pattern indicates heteroskedasticity, meaning the variance of residuals is not constant across predicted values.

3. **Q-Q Plot**:The residuals deviate significantly from the reference line, suggesting they do not follow a normal distribution.

4. **Autocorrelation Function (ACF) Plot**: The ACF plot displays significant autocorrelation, meaning the residuals are not independent and are influenced by time-dependent patterns

These violations imply that the regression metrics derived from this initial model are biased and unreliable.

## 3.2 Initial Biased Model Metrics (Training Data)

| Metric | Value | Unit/Description |
|---|---:|---|
| Sum of Squared Errors (SSE) | 2,194,613.1316 | – |
| Sum of Squares for Regression (SSR) | 84,828,535.9319 | – |
| Total Sum of Squares (SST) | 87,023,149.0635 | – |
| Mean Squared Error (MSE) | 12,540.6465 | – |
| R-Squared ($R^2$) | 0.9748 | Proportion of variance explained |
| Adjusted R-Squared | 0.9724 | Adjusted for number of predictors |
| Mean Absolute Error (MAE) | 86.5101 | – |
| Mean Absolute Percentage Error (MAPE) | 3.7678 | Percent |

Table 2: Summary of Regression Metrics

## 3.3 Addressing the Four Assumptions of Linear Regression

To resolve the identified violations of linear regression assumptions, the following corrective measures were implemented:

### 3.3.1 Non-Linearity

- Applied quadratic terms (degree 2) to the predictors to capture the non-linear relationships between the predictors and the response variable.

### 3.3.2 Heteroskedasticity (Non-Constant Variance)

- Log-transformed the response variable (Monthly Mortgage Payment).

Log transformation stabilizes the variance, reducing heteroskedasticity. Additionally, lagged features of the response variable were introduced to account for potential temporal dependencies that could affect variance.

### 3.3.3 Autocorrelation (Lack of Independence of Errors)

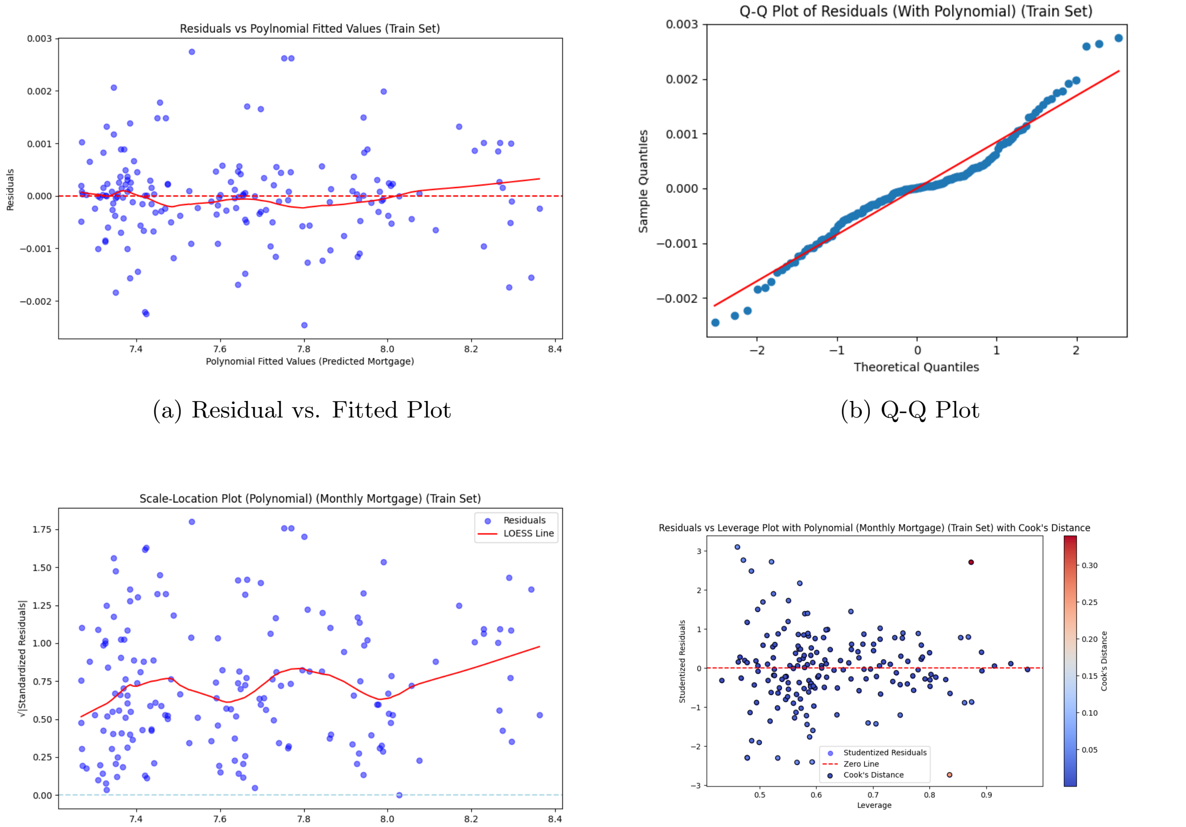
- Introduced lagged features of the response variable.

These lagged variables help model and account for time-dependent patterns in the residuals, thereby improving independence.

### 3.3.4 Non-Normality of Errors

- The combination of log transformation and adding lagged features often helps in making the residuals more normally distributed.

# Post Transformation Plots (Monthly Mortgage Payment)



(a) Residual vs. Fitted Plot

(b) Q-Q Plot

(c) Scale-Location Plot

(d) Residuals vs. Leverage Plot (with Cook's Distance)

Figure 4: Diagnostic Plots for the Mortgage Model: Residual vs. Fitted, Q-Q Plot, Scale-Location, and Residuals vs. Leverage.

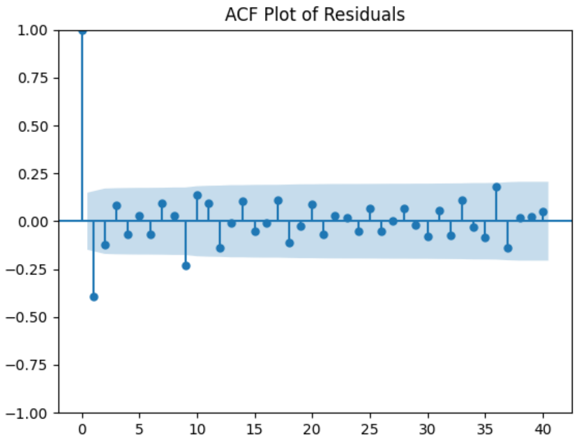


Figure 5: Autocorrelation Function (ACF) Plot

### 3.4 Post-Transformation Model Assessment

The transformations applied to the model successfully addressed the four assumptions of linear regression

- **Linearity**:The residual vs. fitted values plot confirms that non-linearity has been resolved. Residuals now scatter randomly without any discernible pattern.
- **Homoscedasticity**:The Scale-Location plot shows a more uniform spread of residuals, indicating that heteroskedasticity has been significantly reduced
- **Homoscedasticity**:Breusch-Pagan Test: The p-value (0.6966) is greater than 0.05, supporting the assumption of homoscedasticity.
- **Independence od Errors**:The ACF plot does not show significant signs of autocorrelation.
- **Independence od Errors**:The value of Durbin-Watson Statistic (1.97) is very close to the ideal value of 2, confirming that residuals are approximately independent.
- **Normality**:The p-value of Shapiro-Wilk Test (0.7441) is greater than 0.05, indicating that residuals follow a normal distribution

### 3.5 New Challenge: Multicollinearity

While the assumptions of linear regression have been addressed, the inclusion of polynomial terms and lagged features introduced multicollinearity,a strong interdependence among predictors. Given the economic nature of the predictors (e.g., inflation, unemployment), some degree of multicollinearity is inherent. However, the transformations exacerbated this issue.

### 3.6 Limitations of OLS/GLS

Multicollinearity violates the Gauss-Markov assumptions, rendering Ordinary Least Squares (OLS) and Generalized Least Squares (GLS) unreliable for this model.

## 4 Elastic Net Regression

To address multicollinearity, Elastic Net Regression was employed as an effective solution. Elastic Net combines the strengths of Ridge and Lasso regression techniques. By introducing a mixture of L1 and L2 regularization penalties, it balances the benefits of coefficient shrinkage (Ridge) and variable selection (Lasso). This approach not only reduces multicollinearity but also ensures a parsimonious model by selecting relevant predictors while shrinking less important coefficients, thereby enhancing overall model robustness and interpretability.

## Training Performance Metrics

| **Metric** | **Value** |
|---|---:|
| Mean Squared Error (MSE) | 0.00033 |
| R-Squared ($R^2$) | 0.9962 |
| Adjusted R-Squared | 0.9954 |
| Mean Absolute Error (MAE) | 0.0130 |
| Mean Absolute Percentage Error (MAPE) | 0.1682% |

Table 3: Summary of Model Performance Metrics

## Test Performance Metrics

| **Metric** | **Value** |
|---|---:|
| Mean Squared Error (MSE) | 0.0024 |
| R-Squared (R$^2$) | 0.8937 |
| Adjusted R-Squared | 0.6484 |
| Mean Absolute Error (MAE) | 0.0370 |
| Mean Absolute Percentage Error (MAPE) | 0.4536% |

Table 4: Summary of Model Performance Metrics

### 4.1 Interpretation of Elastics Net (Monthly Mortgage Payment)

The Elastic Net regression results highlight significant concerns with the model's ability to generalize effectively. While the training data metrics, including a high 2 R 2 of 0.9962 and a very low Mean Squared Error (MSE) of 0.00033, suggest excellent performance on the training dataset, the testing data reveals a sharp decline in performance. The 2 R 2 drops to 0.8937, and more critically, the Adjusted 2 R 2 plummets to 0.6484. This substantial decrease in Adjusted 2 R 2 , which accounts for the number of predictors, underscores the model's over-reliance on training data patterns and its failure to generalize to unseen data.

The presence of multicollinearity, even with the application of Elastic Net regularization, further complicates the model's performance. Elastic Net is designed to handle multicollinearity by combining Lasso and Ridge regression penalties, yet its limited effectiveness here suggests either insufficient regularization or data issues that amplify the problem. The persistence of correlated predictors likely contributes to the model's instability and overfitting, allowing it to capture noise and spurious relationships rather than robust patterns.

Overfitting is the primary issue, as evidenced by the stark performance gap between training and testing metrics. While the training data metrics reflect a nearly perfect fit, the testing metrics—such as the increase in MSE to 0.0024 and the rise in Mean Absolute Percentage Error (MAPE) from 0.1682 percent to 0.4536 percent clearly demonstrate that the model struggles to predict new data accurately. This overfitting indicates that the model is overly complex, capturing characteristic of the training data rather than the underlying structure of the problem.

# 5 Transformation (Rent)

## Actual vs. Fitted Rent (Training Set)

Here is the plot comparing actual vs. fitted values for the rent model on the training set:

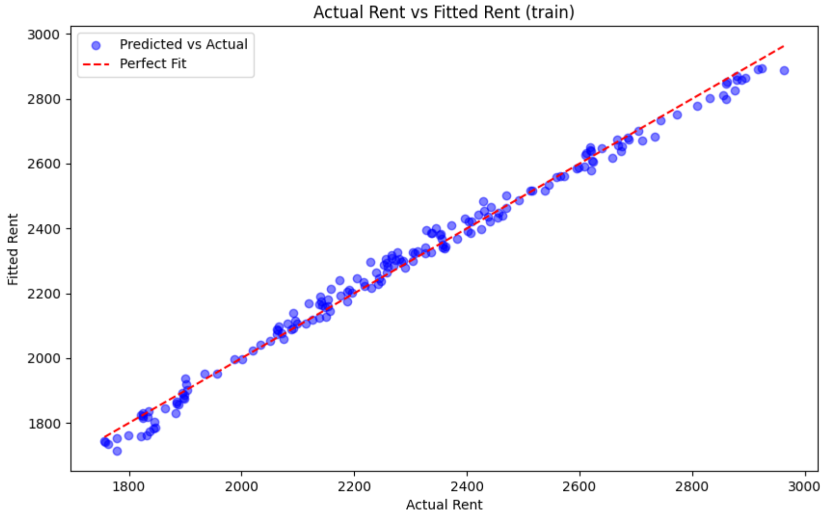


Figure 6: Actual vs. Fitted Values for the Rent Model (Training Set)

## 5.1 Assumption Check: Linear Regression

We will go through the same process as we did with Monthly Mortgage.

### 5.1.1 Non-Linearity

- Applied quadratic terms (degree 2) to the predictors to capture the non-linear relationships between the predictors and the response variable.

### 5.1.2 Heteroskedasticity (Non-Constant Variance)

- Log-transformed the response variable (Monthly Mortgage Payment).

Log transformation stabilizes the variance, reducing heteroskedasticity. Additionally, lagged features of the response variable were introduced to account for potential temporal dependencies that could affect variance.

### 5.1.3 Autocorrelation (Lack of Independence of Errors)

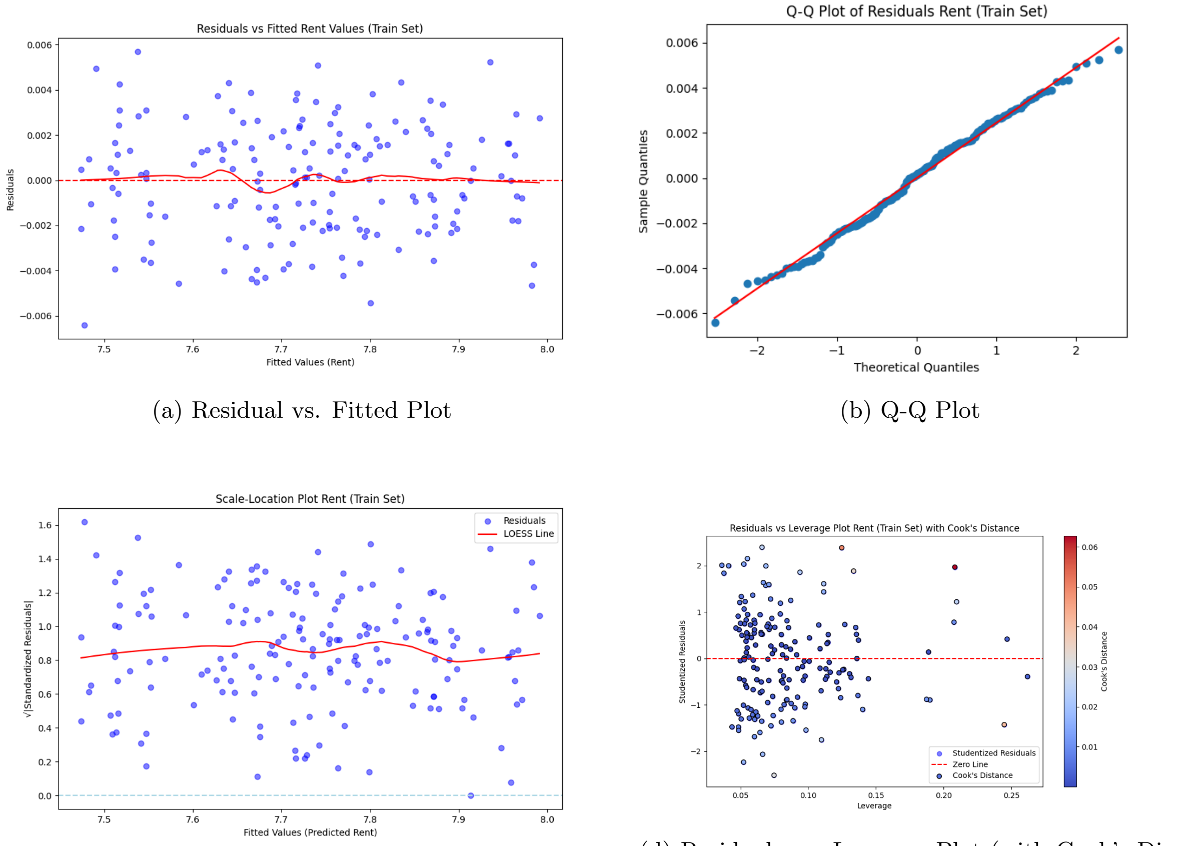
- Introduced lagged features of the response variable.

These lagged variables help model and account for time-dependent patterns in the residuals, thereby improving independence.

### 5.1.4 Non-Normality of Errors

- The combination of log transformation and adding lagged features often helps in making the residuals more normally distributed.

# Post Transformation Plots (Rent)



(a) Residual vs. Fitted Plot

(b) Q-Q Plot

(c) Scale-Location Plot

(d) Residuals vs. Leverage Plot (with Cook's Distance)

Figure 7: Diagnostic Plots for the Mortgage Model: Residual vs. Fitted, Q-Q Plot, Scale-Location, and Residuals vs. Leverage.

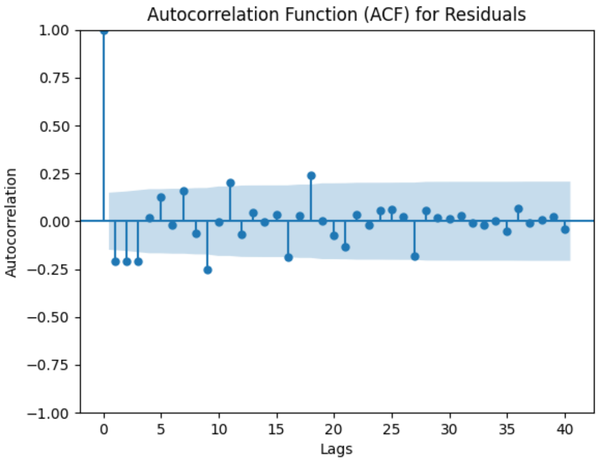


Figure 8: Autocorrelation Function (ACF) Plot

### 5.2 Post-Transformation Model Assessment (Rent)

The transformations applied to the model successfully addressed the four assumptions of linear regression:

- **Linearity**: The residual vs. fitted values plot confirms that non-linearity has been resolved. Residuals now scatter randomly without any discernible pattern.
- **Homoscedasticity**: The Scale-Location plot shows a more uniform spread of residuals, indicating that heteroskedasticity has been significantly reduced.
- **Normality of Residuals**: The residual distribution aligns closely with a normal distribution, as indicated by Q-Q plots, suggesting that normality is no longer a concern.
- **Autocorrelation**: The Durbin-Watson statistic of 2.4154, which is close to the ideal value of 2, confirms that autocorrelation is not a major issue in the model.

## Elastic Net Regression Results

| Metric | Value |
|---|---|
| Sum of Squared Errors (SSE) | 0.0097 |
| Sum of Squares for Regression (SSR) | 3.1043 |
| Total Sum of Squares (SST) | 3.114 |
| R-Squared ($R^2$) | 0.9969 |
| Adjusted R-Squared | 0.9966 |
| MAPE (%) | 0.0762 |

Table 5: Training Performance Metrics

| Metric | Value |
|---|---|
| Sum of Squared Errors (SSE) | 0.0125 |
| Sum of Squares for Regression (SSR) | 0.4504 |
| Total Sum of Squares (SST) | 0.4629 |
| R-Squared ($R^2$) | 0.9729 |
| Adjusted R-Squared | 0.9552 |
| MAPE (%) | 0.1655 |

Table 6: Testing Performance Metrics

### 5.3 Interpretation of Elastic Net Regression Results

The Elastic Net regression results demonstrate the following insights:

1. **Model Performance:** The high R-Squared values for both training ($R^2 = 0.9969$) and testing ($R^2 = 0.9729$) indicate that the model explains nearly all of the variability in the response variable. The low MAPE values for training (0.0762%) and testing (0.1655%) confirm high prediction accuracy with minimal percentage errors relative to the scale of the data.
2. **Overfitting Analysis:** The similarity between the training and testing metrics demonstrates that the model is not overfitting. Elastic Net regularization effectively balances model complexity and prediction accuracy, enabling good generalization to unseen data.
3. **Robustness to Multicollinearity:** Elastic Net successfully mitigates multicollinearity by penalizing large coefficients and retaining only the most influential predictors. This regularization ensures that the model remains robust and reliable, despite the inherent dependencies among predictors.

# Necessary Assumptions of Linear Regression for Different Models

| Model | Linearity | Independence of Errors | Homoscedasticity | Normality of Error |
|---|---|---|---|---|
| Arima | Yes | Yes | Yes | No (for prediction) |
| Sarima | Yes | Yes | Yes | No (for prediction) |
| Prophet | No | Yes | No | No |
| XGBoost | No | No | No | No |
| LSTM | No | No | No | No |

Table 7: Assumptions of Linear Regression for Different Models

# 6 XGBoost Model for Monthly Mortgage

XGBoost (Extreme Gradient Boosting) is a powerful machine learning model that excels in handling complex datasets, particularly time-series data.

One key advantage of XGBoost is that it does not require adherence to the four assumptions of linear regression, making it practical and easy to implement for datasets where these assumptions are violated.

## Feature Importance Plot

The following figure shows the feature importance derived from the XGBoost model:

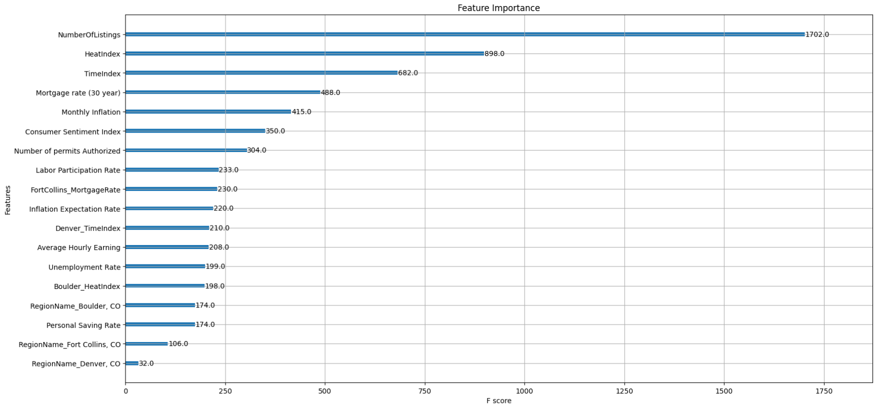


Figure 9: Feature Importance Plot for XGBoost Model

# XGBoost Model: Test Data Evaluation

## Performance Metrics (Test Data)

| **Metric** | **Value** |
|---|---:|
| Mean Squared Error (MSE) | 11,014.00 |
| Mean Absolute Error (MAE) | 84.77 |
| R-Squared ($R^2$) | 0.9610 |
| Adjusted R-Squared | 0.9329 |
| Mean Absolute Percentage Error (MAPE) | 2.41% |
| Sum of Squared Errors (SSE) | 484,616.13 |
| Sum of Squares for Regression (SSR) | 11,118,391.00 |
| Total Sum of Squares (SST) | 11,603,007.13 |

Table 8: Performance Metrics for XGBoost Model on Test Data

## Interpretation of Results

**1. Model Performance:**
The high R-squared value (0.9610) indicates that XGBoost captures most of the variability in the test data, showcasing its strong predictive power. The Adjusted R-squared (0.9329) suggests that the model performs well even when accounting for the number of predictors, with no significant signs of overfitting. The low MAPE (2.41%) indicates that the model's predictions are accurate relative to the scale of the data, with minimal percentage errors.

**2. Predictor Importance:**
The analysis highlights that the *number of listings per month* and *heat index* are the most influential predictors in determining housing prices. This result underscores the significance of housing supply dynamics (captured by the number of listings) and market conditions (reflected in the heat index) in shaping price trends.

**3. Overfitting Analysis:**
With a consistent R-squared and Adjusted R-squared, as well as a low error rate on test data, there is no indication of overfitting. The model generalizes well to unseen data, balancing accuracy and robustness effectively.

# 7 Conclusion

To streamline this report, which initially spanned over 40 pages, I have included only the models that delivered the most accurate and reliable results out of the 12 developed for this project.

The key findings are as follows:

1. **XGBoost** emerged as the most effective model for predicting monthly mortgage payments due to its superior performance and ability to handle complex relationships in the data.
2. The **multiple linear regression model with transformation procedures** was identified as the most robust approach for predicting rent prices.

While these models demonstrate strong predictive capabilities, their application to future forecasts would require simulating predictor variables such as inflation, mortgage rates, and unemployment. This can be achieved using **Monte Carlo simulations**; however, accurately modeling such predictors over a 4-6 year horizon is an intricate and extensive analytical challenge due to economic uncertainty and volatility.

Future work could focus on building a simulation framework for these predictors, which would enable these models to estimate rent prices and mortgage payments more dynamically. For the scope of this project, my objective was to establish reliable and well-performing models to predict rent and mortgage prices based on the available data.

A key insight from this analysis is that purchasing remains a favorable option when financially feasible. Even during periods of elevated mortgage rates, refinancing offers a strategic path forward when rates eventually decline.

## Limitations

The primary constraint in this project was data availability. Although I utilized up to 6 years of historical data, the time range was limited by the availability of certain predictors, which restricted the depth of the analysis.

# References

1. CNBC. (2024, January 11). High housing costs have kept 31% of Gen Z adults living at home. *CNBC*. Retrieved from https://www.cnbc.com/2024/01/11/high-housing-costs-have-kept-31percent-of-gen-z-adults-living-at-home.html

2. RentCafe. Cost of living calculator: Colorado. *RentCafe*. Retrieved from https://www.rentcafe.com/cost-of-living-calculator/us/co/

3. CNBC. (2024, July 12). America's 10 most expensive states to live in. *CNBC*. Retrieved from https://www.cnbc.com/2024/07/12/americas-10-most-expensive-states-to-live-in.html

4. Rocket Mortgage. (n.d.). What is a single-family home? *Rocket Mortgage*. Retrieved from https://www.rocketmortgage.com/learn/single-family-home

5. Keith, T. Z. (2019). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (3rd ed.). Routledge.

6. Vittinghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E. (2012). *Regression methods in biostatistics: Linear, logistic, survival, and repeated measures models* (2nd ed.). Springer.

7. Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press.

8. Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Routledge.

9. Huntington-Klein, N. (2022). *The effect: An introduction to research design and causality*. Routledge.

10. Bureau of Labor Statistics. *Data*. Retrieved from https://www.bls.gov.

11. Zillow Research. Zillow housing data. *Zillow*. Retrieved from https://www.zillow.com/research/data/.

12. Federal Reserve Bank of St. Louis. FRED economic data. *Federal Reserve Economic Data (FRED)*. Retrieved from https://fred.stlouisfed.org.