Capturing the True Value of Real Estate in Seattle Using Property Characteristics: A Machine Learning Approach*

Location and Age Prove Key Determinants in Calculating Intrinsic Value

Sameeck Bhatia

December 1, 2024

TO-DO: Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi faucibus ipsum dolor, in tincidunt justo tempor sit amet proin non turpis a orci imperdiet pretium. Aenean aliquam eros eu consectetur cursus. Proin vel tempor ante vestibulum felis nisi, ultricies sed porttitor in, viverra nec augue. Sed convallis a tortor in pretium. Vivamus metus ipsum, faucibus vitae turpis sed, sodales auctor dolor sed a orci id orci pulvinar tincidunt sed ut tellus. Morbi venenatis tempor arcu, a auctor dolor congue sit amet. Nam auctor tincidunt ligula rutrum aliquam. Sed facilisis at nulla et posuere.

Table of contents

*Code and data are available at https://github.com/SameeckBhatia/Seattle-Real-Estate

1 Introduction

Seattle's real estate market has experienced significant fluctuations since the start of the millennium. Despite these booms and busts, property prices in Seattle remain among the highest in the United States. Demand has steadily increased, and even during periods of lower demand, the market continues to attract buyers. At current price levels, more potential homebuyers are seeking reliable information to find the best deals. With advancements in technology and the increasing availability of data, tools for assessing property values have become essential. One critical tool is an estimate of a property's fair value, which can help buyers determine whether they are overpaying or underpaying relative to the market.

This paper focuses on developing an accurate and accessible property valuation model tailored to Seattle's housing market. The model incorporates current market data to provide fair value estimates for individual properties. The goal is to inform readers about the key factors driving real estate prices in Seattle while empowering them with a tool to make better home-buying decisions. By applying machine learning and statistical techniques, the model aims to ensure accuracy and reflect the intrinsic value of properties rather than merely their market prices. Existing valuation tools, such as Zillow's "Zestimate" and Redfin's proprietary estimates, have limitations. These models are often closed-source, leaving potential users unable to access the methodology or even the results without cost. This creates a significant gap in the availability of open, free, and transparent valuation tools. The absence of such tools denies buyers the advantage of comprehensive, unbiased information to assess property values independently. To address this gap, data was collected from Redfin's semi-public dataset on current Seattle property listings. The analysis was conducted using the R programming language for data cleaning, testing, model creation, and result interpretation.

The primary estimand of this paper is the intrinsic value of a property. This value is determined using a linear regression model applied to observational data collected across Seattle. The analysis aims to identify the true value of a property that a buyer should consider paying and a seller should consider accepting. Additionally, it examines the main factors influencing a property's intrinsic value and allows for comparison to market values.

Preliminary findings indicate that the model effectively predicts property prices, though deviations occur at extreme price ranges. These discrepancies suggest that additional factors, such as high-end amenities and neighborhood characteristics, play a significant role in influencing property value. Understanding these dynamics is important in incorporating socio-economic and structural variables into property valuation models, ultimately improving their accuracy and real-world applicability. The paper is structured as follows: Section 2 introduces the dataset and variables. Section 3 outlines the model design and its significance. Section 4 presents the model's predictions and actionable insights for buyers. Finally, Section 5 explores potential applications of the model in other cities, factors influencing property prices, and the generalizability of the model.

2 Data

2.1 Overview

The data used in this paper comes from Redfin, a real estate brokerage and mortgage company, and represents current property prices across Seattle. While the data was collected from Redfin (2024), similar data could have been sourced from competitors like Zillow and Realtor.com, with Zillow being the largest of the three. If data from Zillow or Realtor.com had been used, the listings might vary slightly since each platform likely features different sellers and listings. However, Redfin was chosen as the source because it is the only brokerage firm that allows public downloads of its listings, helping to avoid violations of data extraction policies.

The dataset includes only properties within Seattle's official city boundaries and covers condos, townhouses, and single-family homes currently on the market. The raw data contains 27 variables, 18 of which have been included in the cleaned dataset. These include variables such as MLS number, number of bedrooms, neighborhood, and geographic coordinates. Additionally, some variables were constructed specifically for this analysis, such as `half_bath`, `property_age`, and `price_sqft`. The first two were added to enhance the dataset and valuation model, while `price_sqft` was created as an alternate response variable. Further details about these variables can be found in Appendix A. The data was analyzed using R (R Core Team 2023) and the tidyverse (Wickham et al. 2019) package, while visualizations have been created using tidyverse.

2.2 Summary Statistics

Table 1

Statistic	N	Mean	St. Dev.	Min	Median	Max
beds	1,361	2.69	1.42	0	3	12
baths	1,361	2.03	1.02	0	2	13
sqft	1,361	1,706.78	1,192.46	223	1,373	13,710
year_built	1,361	1,987.55	37.51	1,890	2,002	2,024
days_on_market	1,361	64.31	65.10	1	49	878
hoa_month	1,361	323.15	623.58	0	13	10,281
price	1,361	1,107,431.00	1,575,578.00	199,000	775,000	39,950,000
half_bath	1,361	0.34	0.47	0	0	1
property_age	1,361	36.45	37.51	0	22	134
price_sqft	1,361	631.41	319.11	205.09	590.71	8,567.45

Table 1 presents summary statistics for all original and derived numeric variables in the dataset. The mean number of bedrooms is 2.69, with a median of 3, while the mean number of bathrooms is 2.03, with a median of 2. This indicates that the typical listing has around three bedrooms and two bathrooms, commonly seen in townhouses and single-family homes. The average property size is approximately 1,710 square feet, with a median of 1,373 square feet, suggesting a positive skew in property size due to a few larger homes in the dataset. The average time a property remains on the market is around 64 days (just over 2 months), with a maximum of 878 days (nearly 2.5 years), indicating low demand in Seattle's real estate market, especially since the data contains only active listings. The mean and median property prices are \$1,107,431 and \$775,000, respectively, while the mean and median price per square foot are \$631.41 and \$590.71, respectively. These figures highlight Seattle as one of the most expensive residential markets in the United States.

2.3 Measurement

In the United States, buyers and sellers have the freedom to select the real estate agent or brokerage firm they wish to work with for transactions. Consequently, agents often represent multiple listings within their region or city. Agents receive detailed information on each property from real estate appraisers, who measure variables such as the number of bedrooms, bathrooms, and square footage (National Association of Realtors 2024). These measurement practices, except for price, are strictly regulated to ensure accuracy for all stakeholders.

For property prices, appraisers typically estimate values based on the prices of recently sold comparable properties and the specific characteristics of the property. This valuation process is less regulated, as it serves primarily as a reference point for buyers and sellers. Real estate agents may gather price estimates from multiple appraisers to calculate an average. The prices observed in the data, although guided by these values, are ultimately set by the seller. All this information is uploaded to the Multiple Listing Service (MLS), a private database accessible only to agents and brokerage firms via subscription fees (Bankrate 2024). However, U.S. laws allow companies like Redfin, Trulia, and Zillow to extract and share MLS data with the public, fostering competition and transparency.

2.4 Outcome Variables

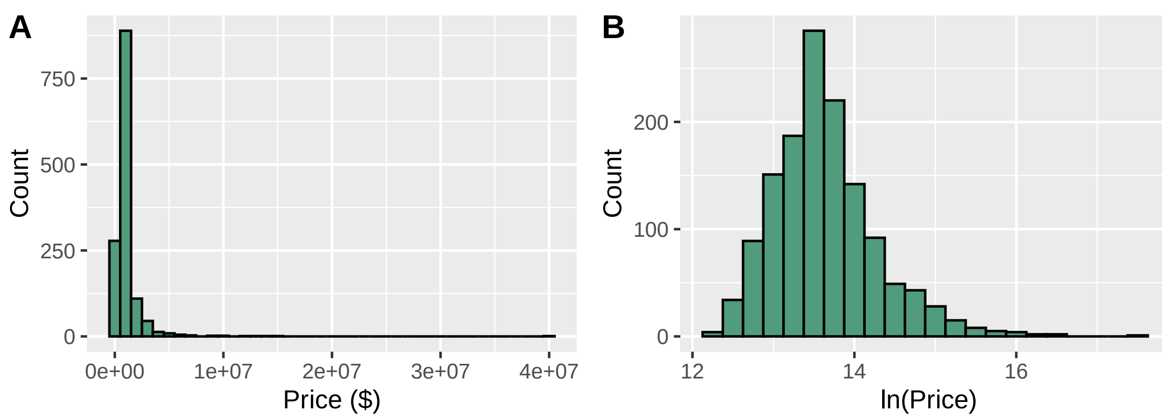

Figure 1: Heavily Skewed Property Prices and Skewed Still for Logarithmic Prices

The primary goal of this paper is to estimate the outcome variable, price, using a valuation model. This variable represents the market value of active listings at the time of data collection, focusing exclusively on properties located in Seattle. Figure 1 provides insights into the distribution of property prices. Plot A shows that property prices are highly skewed, with a maximum value near $40,000,000 and several listings exceeding $5,000,000. To better understand the distribution, Plot B presents the logarithmic transformation of property prices.

While the transformed distribution is less extreme, it remains skewed, indicating the presence of properties with exceptionally high valuations in the Seattle real estate market.

2.5 Predictor Variables

Property Type (`property_type`): This categorical variable identifies the type of property for each observation, based on classifications set by the MLS. The three main property types included in the analysis are "Condo/Co-op," "Single Family," and "Townhouse." These classifications are critical for valuation, as different property types possess distinct features that influence their market value.

Number of Bedrooms (`beds`): This numeric variable represents the count of full bedrooms in a property, as measured by appraisers. It is a key factor in the analysis, as more bedrooms often correlate with greater living space and the potential to accommodate larger households.

TO-DO: Number of Bathrooms (`baths`): This numeric variable represents the count of full bathrooms in a property, as measured by appraisers. It is a key factor in the analysis, as more bathrooms often correlate with greater living space and the potential to accommodate larger households.

Property Size (`sqft`): Measured in square feet, this variable reflects the total size of the property and is determined by appraisers. As the United States predominantly uses the imperial system, square footage is a standard unit. Larger properties generally hold more value, making this a significant predictor in property valuation.

Property Age (`property_age`): This derived variable calculates the age of a property in years, based on the difference between the year of data collection and the `year_built`. Age is a vital consideration in determining price, as newer properties are often smaller due to rising construction costs and increasingly strict zoning regulations.

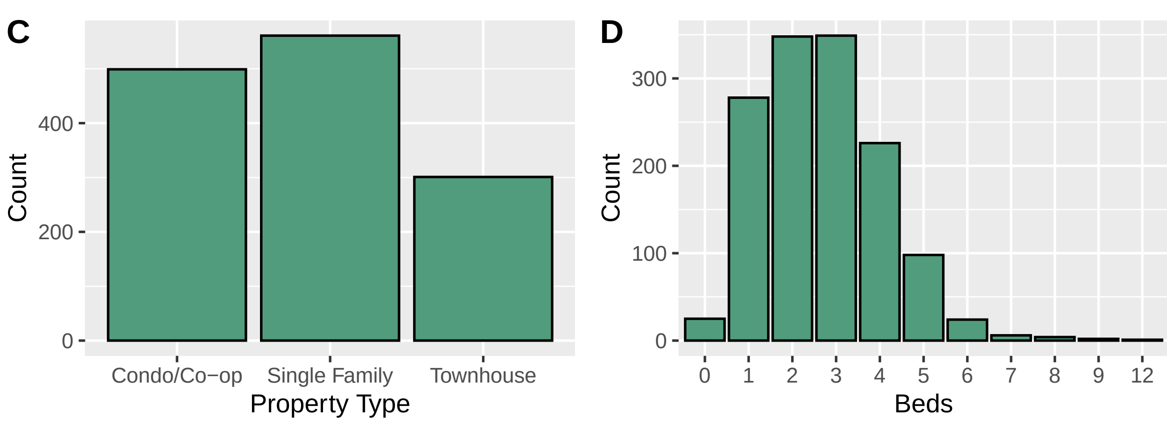

Figure 2: Single Family Homes and 2 or 3 Bedrooms Properties Most Listing on the Market

Figure 2 represents the distributions of property types and number of bedrooms. Plot C shows that single-family homes are the most common, followed closely by condos, with townhouses being the least prevalent. This likely reflects the mix of apartments and houses present in Seattle's real estate market. Plot D highlights that properties with 2 or 3 bedrooms are most frequent, while some properties even have north of 7 bedrooms. This is likely because condos typically feature 1-2 bedrooms, whereas single-family homes and townhouses commonly have 3-4 bedrooms.

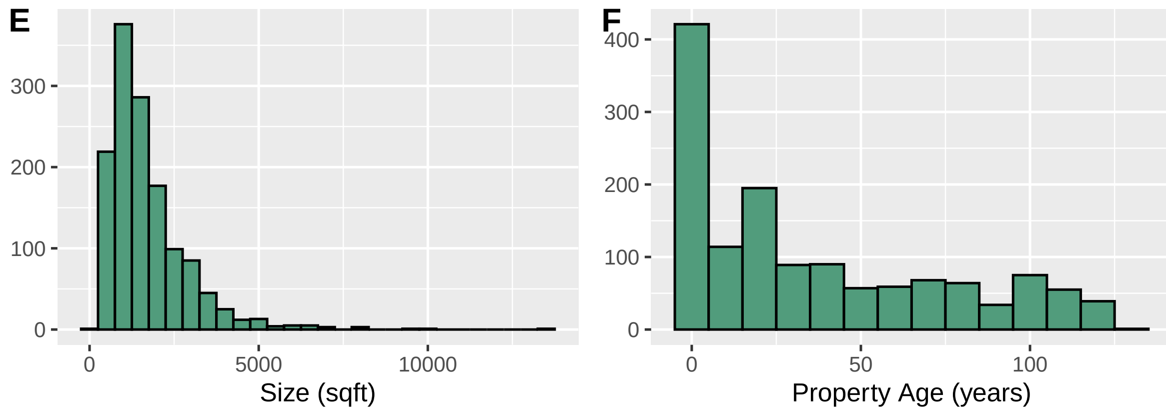

Figure 3: Skewed Distribution of Sizes and Most Properties Built Over the Last Decade

Figure 3 displays the distributions of property size and age. Plot E shows that property size is positively skewed, with most properties under 2,500 square feet and a peak between 700 and 1,200 square feet. Properties exceeding 5,000 square feet likely represent luxury homes or mansions. Plot F indicates that property age is also positively skewed, with the most common properties being less than 10 years old. The second most common group is 20-30 years old, likely reflecting construction surges before the 2008 financial crisis and real estate market downturn.

3 Model

3.1 Overview

The model uses a multiple linear regression approach to estimate property values in Seattle. It incorporates features like the number of bedrooms (`beds`), square footage (`sqft`), homeowner association fees (`hoa_month`), and an interaction term (`sqft` × `beds`) to predict prices. These characteristics were selected for their significance in influencing property valuations. The final model was chosen after thorough validation, achieving a good balance of accuracy and interpretability.

3.2 Setup

The linear regression model can be represented as:

$$\text{price} = \beta_0 + \beta_1 \cdot \text{beds} + \beta_2 \cdot \text{baths} + \beta_3 \cdot \text{sqft} + \beta_4 \cdot \text{hoa\_month} + \beta_5 \cdot (\text{sqft} \cdot \text{beds})$$

, where the coefficients are described as:

β_0: Intercept

The predicted value of `price` when all predictors (beds, sqft, hoa_month, sqft $\cdot$ beds) are zero.

β_1: Coefficient for `beds`

The effect of adding one more bedroom on price, holding other factors constant (except for the interaction term).

β_2: Coefficient for `baths`

The effect of adding one more bathroom on price, holding other factors constant.

β_3: Coefficient for `sqft`

The effect of increasing square footage by one unit on price, holding other factors constant (except for the interaction term).

β_4: Coefficient for `hoa_month`

The effect of a one-unit increase in monthly homeowner's association fees on price, holding other factors constant.

β_5: Interaction between `sqft` and `beds`

Represents how the relationship between sqft and `price` changes depending on the number of bedrooms.

3.3 Limitations

This model has a few limitations that should be considered when interpreting its predictions. First, it is trained on cross-sectional data rather than longitudinal data. As a result, the model is designed to provide accurate valuations for properties in the near future (typically less than a year), assuming minimal price fluctuations. Significant market changes over time would reduce the model's accuracy, requiring frequent updates with new cross-sectional data to maintain reliability. Additionally, the model is specifically tailored to Seattle's real estate market. While it might perform adequately in nearby cities with similar market characteristics, it is unlikely to generalize well to cities in other states, such as Los Angeles or New York, due to differing property attributes and pricing dynamics in those regions.

3.4 Justification

The four features included in the model are `beds`, `sqft`, `hoa_month`, and `sqft:beds`, as they were the most significant in influencing the model's property valuations. The beds feature was selected because the number of bedrooms is a key determinant of a property's utility and appeal to buyers, directly impacting its market value. The sqft feature, representing the total square footage, is a fundamental metric for assessing a property's size and, consequently, its worth. The hoa_month feature accounts for monthly homeowner association fees, which can significantly affect the affordability and desirability of properties, particularly in condominiums or communities with shared amenities. Finally, the interaction term sqft:beds captures the relationship between the size of the property and the number of bedrooms, highlighting how the distribution of space impacts valuation.

Figure 4 represents the causal relationships between the variables analyzed in the regression model. It illustrates that property price is influenced by the number of bedrooms, bathrooms, square footage, and HOA fees. Square footage itself depends on the number of bedrooms and bathrooms, while HOA fees are linked to square footage. The directed arrows capture the assumed causal pathways, emphasizing that the effects of bedrooms and bathrooms on price are partly mediated through square footage.

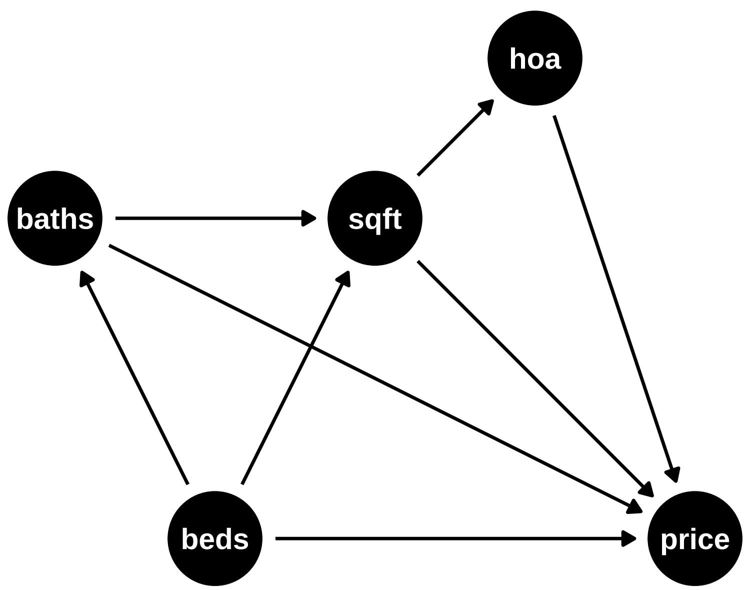

Figure 4: Multiple Causal Relationships Among Model Variables

3.5 Interpretation

Table 2 shows the values of the model's coefficients mentioned in the setup. The intercept of −66,054.23 suggests that, without any additional predictors, property prices would begin

Table 2

	Final Model
(Intercept)	−66 054.23
	(34 398.40)
beds	55 091.77
	(13 863.34)
baths	41 740.69
	(20 967.02)
sqft	592.23
	(22.87)
hoa_month	66.39
	(17.85)
beds × sqft	−38.66
	(3.33)
Num.Obs.	952
R2	0.717
R2 Adj.	0.716
RMSE	331 563.51

at this value, though it has limited standalone interpretation. The coefficient for `beds` is 55,091.77, meaning that, holding other factors constant, each additional bedroom is associated with an addition in price. However, this effect is modified by the interaction term `beds` × `sqft`, which has a negative coefficient of −38.66, indicating that larger properties have diminishing prices. The coefficient for `baths` is 41,740.69, meaning that for each additional bathroom, the predicted value of the price increases by around $41,740.69, holding all other variables constant. This suggests that bathrooms have a strong positive impact on the value of the property. The `sqft` variable has a strong positive impact (592.23 per additional square foot), highlighting size as a key driver of value. The `hoa_month` coefficient (66.39) shows a modest positive association with price, possibly reflecting higher costs in premium communities. With an adjusted R^2 of 0.716 and an RMSE of 331,563.51, the model has a decent fit.

3.6 Validation

The linear model was created using R (R Core Team 2023) to fit the data and generate predictions, while the MLMetrics package (Yan 2024) was used to evaluate performance. A train-test split was created with the rsample package (Frick et al. 2024), with the model trained on the 70% of the data and validated using out-of-sample testing on the remaining 30%. Key evaluation metrics included the Root Mean Square Error (RMSE) and R^2 The model achieved an RMSE of 331,564, indicating the average prediction error in dollar terms, and R^2

of approximately 0.717, reflecting the proportion of variance in property prices explained by the model. Further diagnostics and information are provided in Appendix B.

3.7 Alternate Models

Several alternative models were considered before selecting the final one. The first, a "full model," included all variables in the cleaned dataset, achieving an RMSE of 333,693.77 and an R^2 score of 0.714. While this model used all available information, some variables were statistically insignificant. A reduced model was then tested by retaining only significant variables, resulting in an RMSE of 333,736.71 and an R^2 score of 0.713. Though concise, it performed similarly to the full model. The final model, incorporating an interaction term, outperformed both alternatives with an RMSE of 331 563.51 and an R^2 of 0.717.

4 Results

4.1 Higher Prices Associated with Higher Prediction Errors

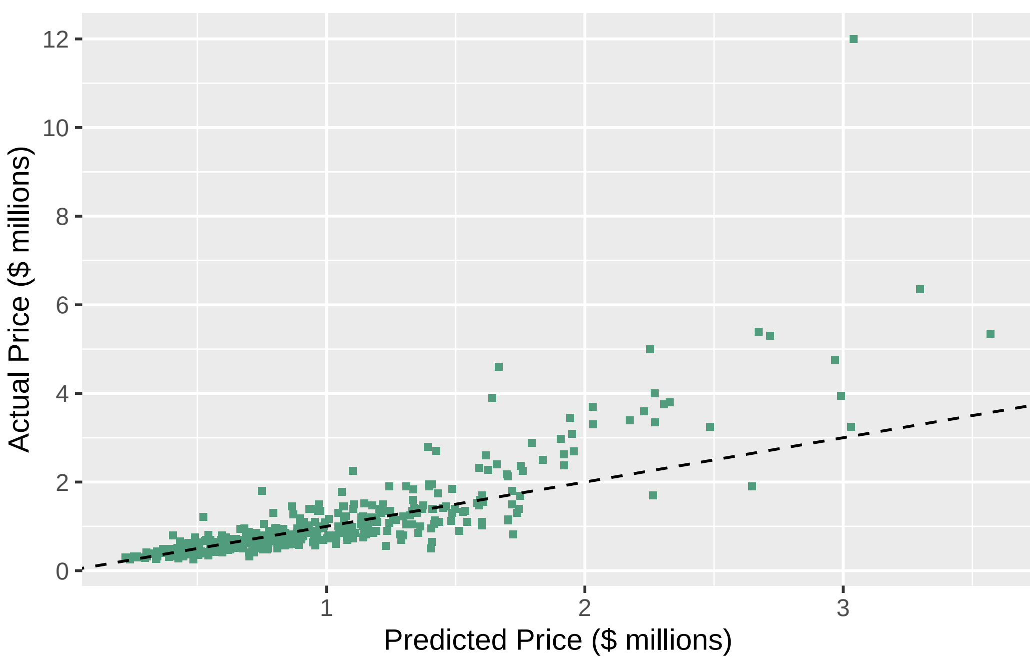

Figure 5: Large Uncertainty in the Predicted Values of Luxury Homes

Figure 5 shows the model-predicted prices alongside the actual price observations from the data. The dashed line represents the line of equality, where the predicted price matches the actual property price. Some data points deviate from this line, particularly at lower values, which is expected since no model can perfectly capture all variations in the data. In this case, the deviations may result from missing factors that could enhance the property valuation model. However, when property values exceed approximately 2 million dollars, the actual prices deviate significantly from the line of equality. This indicates higher prediction errors, likely due to a combination of internal and external factors. Internal factors might include additional amenities (e.g., pools, large gardens, or extra rooms) or premium furnishings (e.g., expensive flooring or tiling), which increase lot size or property desirability, driving up prices. External factors may include the property's location in high-income neighborhoods or proximity to desirable features such as beaches, shopping malls, or tourist attractions. These influences are analyzed further in Figure 8.

4.2 Similar Distributions to Market Price from the Model

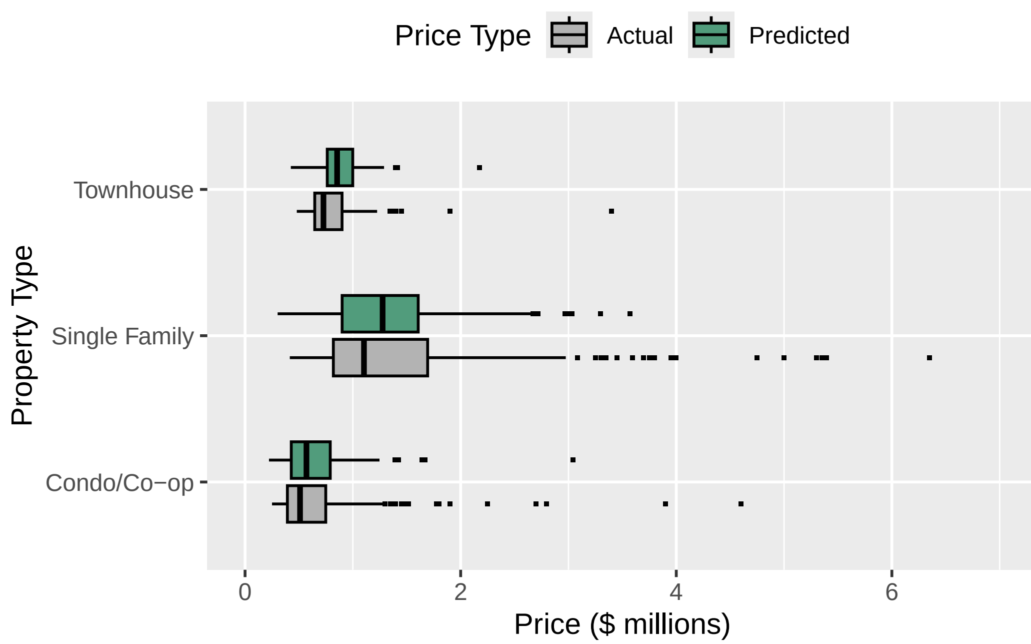

Figure 6: Single Family Homes Valued the Highest Among All Property Types

Figure 6 illustrates the distributions of actual and predicted prices based on MLS-classified property types. Condos have the lowest median prices, followed by townhouses, with single-family homes being the most expensive. This pricing trend is primarily influenced by the size of each property type, which plays a significant role in determining value. The predicted and actual distributions for all property types show similar shapes, although there are differences in their centers and outliers. The model predicts slightly higher intrinsic values for townhouses and condos compared to the actual prices, whereas there is a notable difference in the median price for single-family homes. This discrepancy is likely due to the real estate market being in a cooling phase, with prices remaining relatively stable over the past six months. Additionally, while market prices are subject to short-term fluctuations, intrinsic values tend to remain more consistent.

4.3 Influence of the Time Effect on Property Valuation

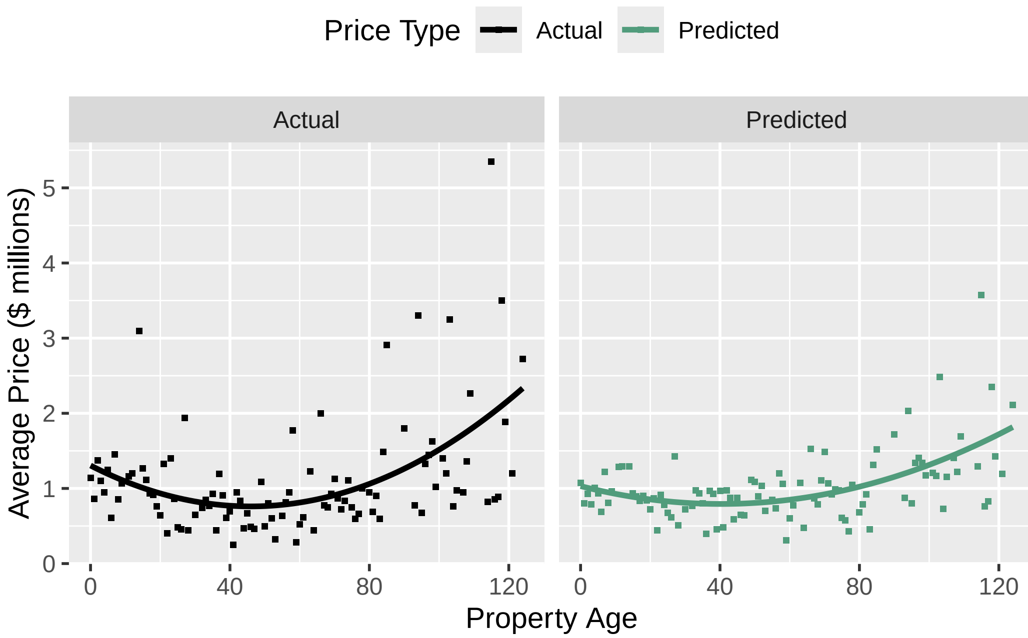

Figure 7: 40 Year Old Homes Among the Lowest in Actual and Predicted Values

Figure 7 highlights the trend between property age and the predicted and actual average prices. The relationship can be divided into three stages. First, the average price for brand-new properties starts at around 1 million dollars, followed by a gradual decrease as the properties age. This decline is expected, as older properties typically lose value due to higher upkeep costs and reduced desirability. Next, the average price bottoms out when properties reach around 40 years of age. This likely represents a threshold where property age significantly affects market value (for actual prices) or true value (for predicted prices). In the third stage, average prices increase for properties over 50 years old, continuing to rise for homes built a century ago. While this may seem counterintuitive given the earlier trend, it likely reflects the premium placed on vintage properties. Such homes often have historical significance, unique architectural features, or are built on highly desirable land, all of which enhance their perceived and intrinsic value.

4.4 Noticeable Clusters for High-End Properties

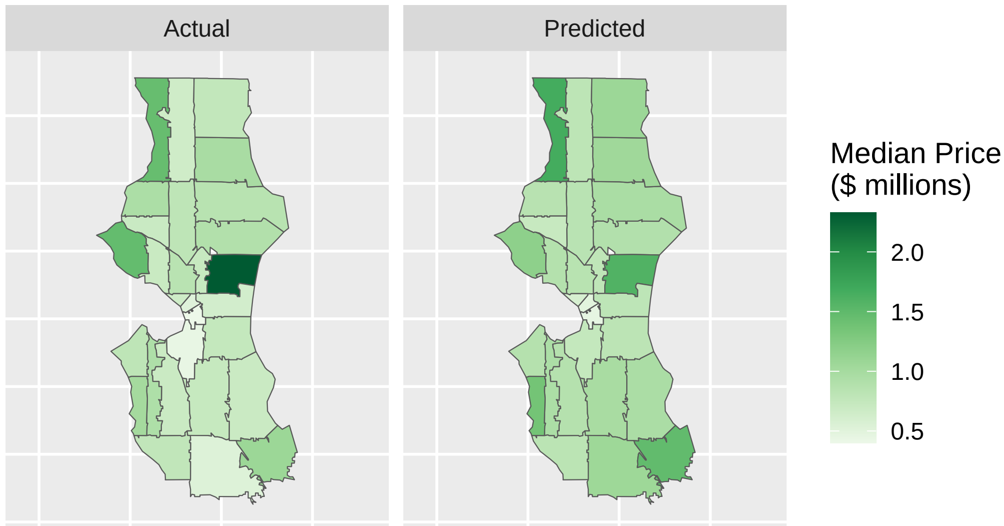

Figure 8: Highest Predicted Listing Prices in Seattle's Largest Gated Communities

Figure 8 presents the spatial distribution of median property values across Seattle, categorized by ZIP code. Two distinct patterns emerge from both maps: first, two ZIP codes stand out with the highest median listing prices; second, property values decrease the farther south they are in the city. The model captures these patterns well, showing similar clusters. The two ZIP codes with the darkest shades correspond to the Broadmoor and Broadview neighborhoods, which are 85- and 1,100-acre gated communities with affluent residents and amenities such as golf courses. These communities share features like higher incomes, better security, and a higher standard of living, enabling them to support significantly higher property values. Conversely, neighborhoods in southern Seattle tend to have lower incomes, higher crime rates, and less development, and evidently lower property values. These factors collectively explain the disparities in property values across the city. These disparities reflect broader socio-economic trends in urban areas, where wealthier neighborhoods often attract more investment in infrastructure, public services, and amenities, driving property values higher.

5 Discussion

5.1 Examining Racial Bias in Real Estate Price Estimates

TO-DO: Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed eget purus eleifend, vehicula dolor vitae, aliquet nibh. Curabitur ac leo non nunc ultrices dignissim. Phasellus faucibus elit id nulla pretium volutpat. Proin ullamcorper, nulla sit amet tempor faucibus, mi ligula interdum urna, eu condimentum arcu augue nec leo. Curabitur vulputate imperdiet nulla eu finibus. Vivamus sem lectus, commodo in lectus quis, posuere laoreet quam. Etiam nunc libero, faucibus nec porta a, dignissim a nisi. Phasellus laoreet elementum turpis a vestibulum. Quisque in urna varius, lobortis eros eu, blandit sapien. In lobortis orci id nunc facilisis, non tempor nisl ultricies. Sed pellentesque tortor cursus ipsum fringilla interdum. Cras lacinia tempor placerat. Proin eu elit elit. Sed molestie dolor et posuere sodales.

Fusce aliquam tempus tincidunt. Nulla viverra non nibh consectetur blandit. Proin malesuada, nunc et accumsan pellentesque, odio felis molestie augue, nec gravida lectus ipsum ut lectus. Donec at malesuada purus. Cras ultrices vitae metus eget aliquet. Aliquam sollicitudin elit placerat, porta mauris in, semper felis. Nulla vestibulum dolor sed tempor efficitur. In bibendum libero ex, in sollicitudin arcu maximus porttitor. Pellentesque ac sapien sit amet magna ornare auctor. In hac habitasse platea dictumst. Cras ut urna dapibus elit mattis aliquam non sed augue. Suspendisse lacinia ornare consectetur. Quisque ultricies tortor non ullamcorper porta. Mauris ac leo quis lacus tincidunt cursus ut id felis.

5.2 Using Alternative Algorithms for Nonlinear Price Trends

TO-DO: Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed eget purus eleifend, vehicula dolor vitae, aliquet nibh. Curabitur ac leo non nunc ultrices dignissim. Phasellus faucibus elit id nulla pretium volutpat. Proin ullamcorper, nulla sit amet tempor faucibus, mi ligula interdum urna, eu condimentum arcu augue nec leo. Curabitur vulputate imperdiet nulla eu finibus. Vivamus sem lectus, commodo in lectus quis, posuere laoreet quam. Etiam nunc libero, faucibus nec porta a, dignissim a nisi. Phasellus laoreet elementum turpis a vestibulum. Quisque in urna varius, lobortis eros eu, blandit sapien. In lobortis orci id nunc facilisis, non tempor nisl ultricies. Sed pellentesque tortor cursus ipsum fringilla interdum. Cras lacinia tempor placerat. Proin eu elit elit. Sed molestie dolor et posuere sodales.

Fusce aliquam tempus tincidunt. Nulla viverra non nibh consectetur blandit. Proin malesuada, nunc et accumsan pellentesque, odio felis molestie augue, nec gravida lectus ipsum ut lectus. Donec at malesuada purus. Cras ultrices vitae metus eget aliquet. Aliquam sollicitudin elit placerat, porta mauris in, semper felis. Nulla vestibulum dolor sed tempor efficitur. In bibendum libero ex, in sollicitudin arcu maximus porttitor. Pellentesque ac sapien sit amet magna ornare auctor. In hac habitasse platea dictumst. Cras ut urna dapibus elit mattis aliquam non sed augue. Suspendisse lacinia ornare consectetur. Quisque ultricies tortor non ullamcorper porta. Mauris ac leo quis lacus tincidunt cursus ut id felis.

5.3 Including External Factors into Property Valuation Models

TO-DO: Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed eget purus eleifend, vehicula dolor vitae, aliquet nibh. Curabitur ac leo non nunc ultrices dignissim. Phasellus faucibus elit id nulla pretium volutpat. Proin ullamcorper, nulla sit amet tempor faucibus, mi ligula interdum urna, eu condimentum arcu augue nec leo. Curabitur vulputate imperdiet nulla eu finibus. Vivamus sem lectus, commodo in lectus quis, posuere laoreet quam. Etiam nunc libero, faucibus nec porta a, dignissim a nisi. Phasellus laoreet elementum turpis a vestibulum. Quisque in urna varius, lobortis eros eu, blandit sapien. In lobortis orci id nunc facilisis, non tempor nisl ultricies. Sed pellentesque tortor cursus ipsum fringilla interdum. Cras lacinia tempor placerat. Proin eu elit elit. Sed molestie dolor et posuere sodales.

Fusce aliquam tempus tincidunt. Nulla viverra non nibh consectetur blandit. Proin malesuada, nunc et accumsan pellentesque, odio felis molestie augue, nec gravida lectus ipsum ut lectus. Donec at malesuada purus. Cras ultrices vitae metus eget aliquet. Aliquam sollicitudin elit placerat, porta mauris in, semper felis. Nulla vestibulum dolor sed tempor efficitur. In bibendum libero ex, in sollicitudin arcu maximus porttitor. Pellentesque ac sapien sit amet magna ornare auctor. In hac habitasse platea dictumst. Cras ut urna dapibus elit mattis aliquam non sed augue. Suspendisse lacinia ornare consectetur. Quisque ultricies tortor non ullamcorper porta. Mauris ac leo quis lacus tincidunt cursus ut id felis.

Appendix

A Additional Data Details

A.1 Raw Data Dictionary

Variable	Description
SALE TYPE	Type of sale (e.g., MLS listing, new construction home, new construction plan).
SOLD DATE	Date when the property was sold.
PROPERTY TYPE	Type of property (e.g., condo, single-family house, townhouse).
ADDRESS	Full address of the property.
CITY	City where the property is located.
STATE OR PROVINCE	State or province where the property is located.
ZIP OR POSTAL CODE	ZIP code of the property location.
PRICE	Sale price of the property.
BEDS	Number of bedrooms in the property.
BATHS	Number of bathrooms in the property.
LOCATION	Neighbourhood of the property.
SQUARE FEET	Total square footage of the property.
LOT SIZE	Lot size in square feet
YEAR BUILT	Year the property was built.
DAYS ON MARKET	Number of days the property has been listed on the market.
$/SQUARE FEET	Price per square foot of the property.
HOA/MONTH	Monthly Homeowners Association (HOA) fee, if applicable.
STATUS	Current status of the property (e.g., sold, pending, active).
NEXT OPEN HOUSE START TIME	Start time of the next scheduled open house, if available.
NEXT OPEN HOUSE END TIME	End time of the next scheduled open house, if available.
URL	URL to additional property details.
SOURCE	Source of the property data (e.g., MLS, Zillow).
MLS#	Multiple Listing Service (MLS) identification number for the property.
FAVORITE	Indicates whether the property is marked as a favorite (e.g., Y/N).
INTERESTED	Indicates whether the user has expressed interest in the property (e.g., Y/N).
LATITUDE	Latitude of the property location for geospatial analysis.
LONGITUDE	Longitude of the property location for geospatial analysis.

Variable	Description

B Additional Model Details

B.1 Model Diagnostics

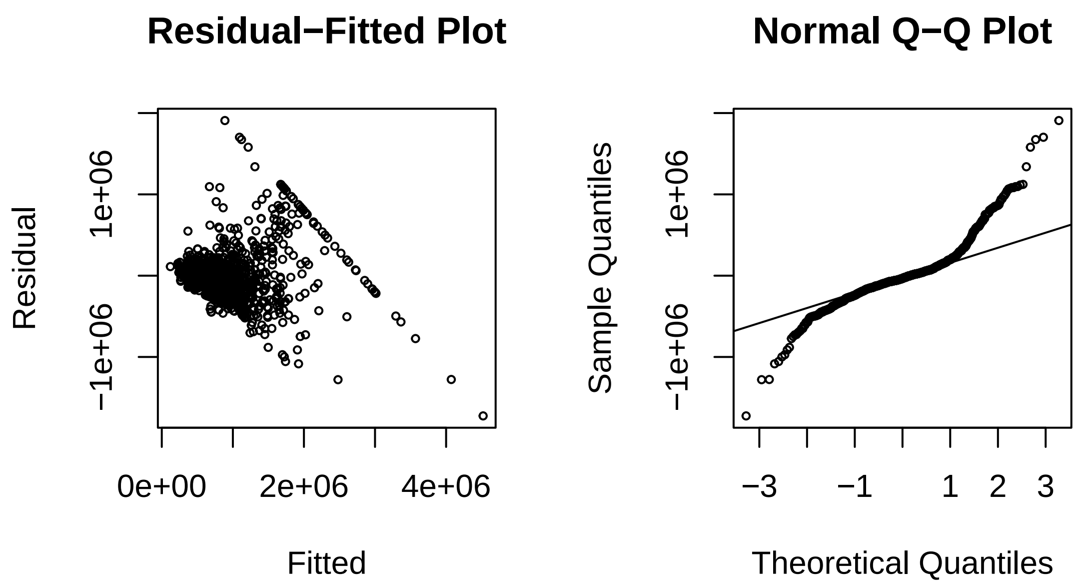

Figure 9: Visible Violations of Constant Variance and Normality Assumptions

TO-DO: Figure 9 ipsum dolor sit amet, consectetur adipiscing elit. Duis vitae pretium velit. Etiam fermentum enim ac venenatis fringilla. Nulla imperdiet metus ac ligula ultricies, at cursus eros varius. Curabitur commodo pretium ornare. Pellentesque ultrices sapien ut lectus viverra placerat. Mauris nec interdum tortor. Cras posuere turpis sed mi blandit dictum. Etiam accumsan ipsum quis sem ornare cursus. Aenean ligula mauris, vulputate a volutpat dapibus, consectetur a lacus. Nullam ornare placerat risus, in luctus urna. Ut molestie dignissim turpis, eget tincidunt arcu. Vivamus auctor fermentum lectus, eu vulputate tortor volutpat et. Aenean aliquam nulla eget felis faucibus faucibus ut nec nisi.

B.2 Alternate Model Summaries

TO-DO: Table 4 ipsum dolor sit amet, consectetur adipiscing elit. Duis vitae pretium velit. Etiam fermentum enim ac venenatis fringilla. Nulla imperdiet metus ac ligula ultricies, at cursus eros varius. Curabitur commodo pretium ornare. Pellentesque ultrices sapien ut lectus viverra placerat. Mauris nec interdum tortor. Cras posuere turpis sed mi blandit dictum. Etiam accumsan ipsum quis sem ornare cursus. Aenean ligula mauris, vulputate a volutpat dapibus, consectetur a lacus. Nullam ornare placerat risus, in luctus urna. Ut molestie dignissim turpis, eget tincidunt arcu. Vivamus auctor fermentum lectus, eu vulputate tortor volutpat et. Aenean aliquam nulla eget felis faucibus faucibus ut nec nisi.

C Surveys, Sampling, and Observational Data

TO-DO: Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce eu hendrerit mauris, in sodales mi. Aenean eget urna rutrum, tristique lorem vel, maximus ipsum. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nunc eget rutrum est. Donec tempor elit non nulla cursus, in consectetur est ullamcorper. Nam vitae porttitor felis, tempor maximus leo. Aenean ultricies molestie facilisis. Donec eget gravida eros, sit amet ornare erat. Cras a vehicula nibh. Nunc ac ligula malesuada, bibendum ante eu, luctus felis. Proin in tellus eu diam aliquam ultricies.

Duis sit amet augue sagittis, commodo eros sollicitudin, congue libero. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Maecenas convallis erat velit, eu consequat massa posuere et. Nullam tempor purus sit amet quam lacinia euismod. Praesent sagittis risus sit amet odio porttitor, eget pellentesque turpis luctus. Proin feugiat convallis diam, id vulputate mauris. Nunc facilisis quam nec tristique pellentesque. Sed laoreet tortor at ipsum faucibus iaculis. Sed cursus convallis molestie. Aliquam et accumsan tortor. Morbi volutpat rutrum elit sed fermentum. In vestibulum placerat eros, et consectetur lacus fringilla a. Mauris elementum velit vel eros placerat, eget commodo nisi convallis. Sed ac nisi mi. Nullam malesuada auctor semper.

Vestibulum at tortor ipsum. Quisque a elit ut odio suscipit fringilla ac ut nunc. Vivamus eget neque vel est iaculis sollicitudin in et turpis. Pellentesque quis risus ullamcorper metus porta viverra. In hac habitasse platea dictumst. In faucibus, sapien sed placerat semper, sapien lectus placerat leo, sit amet tempus risus dolor non erat. Morbi ac urna venenatis, pellentesque odio vel, iaculis tellus. Etiam tincidunt fringilla suscipit. Donec fermentum sollicitudin metus, quis tempor leo elementum eu.

Table 4

	Full	Reduced
(Intercept)	276 849.796	268 032.931
	(34 207.507)	(28 306.760)
property_typeSingle Family	198 743.750	198 184.223
	(42 548.641)	(37 982.409)
property_typeTownhouse	−50 299.334	−52 422.284
	(38 201.911)	(36 368.931)
beds	617.997	
	(15 837.696)	
baths	−10 142.810	
	(22 621.590)	
half_bath	102 491.848	105 609.130
	(26 985.796)	(26 118.559)
sqft	407.408	400.046
	(19.601)	(12.175)
days_on_market	−670.417	−663.538
	(163.646)	(162.787)
hoa_month	131.176	131.373
	(21.511)	(21.487)
property_age	−2415.871	−2367.412
	(358.420)	(344.105)
Num.Obs.	952	952
R2	0.714	0.713
R2 Adj.	0.711	0.711
AIC	26 938.7	26 934.9
BIC	26 992.1	26 978.7
Log.Lik.	−13 458.345	−13 458.468
RMSE	333 693.77	333 736.71

References

Bankrate. 2024. "What Is the MLS? Multiple Listing Service, Explained." https://www.bankrate.com/real-estate/mls-multiple-listing-service/.

Frick, Hannah, Fanny Chow, Max Kuhn, Michael Mahoney, Julia Silge, and Hadley Wickham. 2024. *Rsample: General Resampling Infrastructure*. https://CRAN.R-project.org/package=rsample.

National Association of Realtors. 2024. "What to Know about the Appraisal Process." https://www.nar.realtor/magazine/tools/client-education/handouts-for-buyers/what-to-know-about-the-appraisal-process.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Redfin. 2024. "Seattle, WA homes for sale & real estate." https://www.redfin.com/city/16163/WA/Seattle.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Yan, Yachen. 2024. *MLmetrics: Machine Learning Evaluation Metrics*. https://CRAN.R-project.org/package=MLmetrics.