Valuing Real Estate in Seattle Using Property Characteristics: A Machine Learning Approach*

Subtitle

Sameeck Bhatia

November 26, 2024

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

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.

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.

Donec commodo diam non arcu facilisis, sit amet egestas massa consequat. Vestibulum semper congue libero, eu vestibulum mi cursus sed. Morbi ut justo quis ligula molestie mattis. In luctus, dolor ac posuere pretium, ex ex sollicitudin ligula, a varius libero nisi sed arcu. Nulla et bibendum nunc. Morbi sit amet lorem sit amet lectus bibendum tempus aliquam sit amet lectus. Sed ligula nulla, pulvinar quis ipsum a, condimentum egestas mi. Aenean libero tortor, vehicula vitae velit sit amet, eleifend semper elit. Nam non fringilla nisi. Vivamus quis consequat odio, et consectetur risus. Proin vulputate tristique mauris vitae interdum. Phasellus efficitur dictum nibh, id dictum nisl rutrum et.

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	733	600.02	745.73	0	510	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

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.
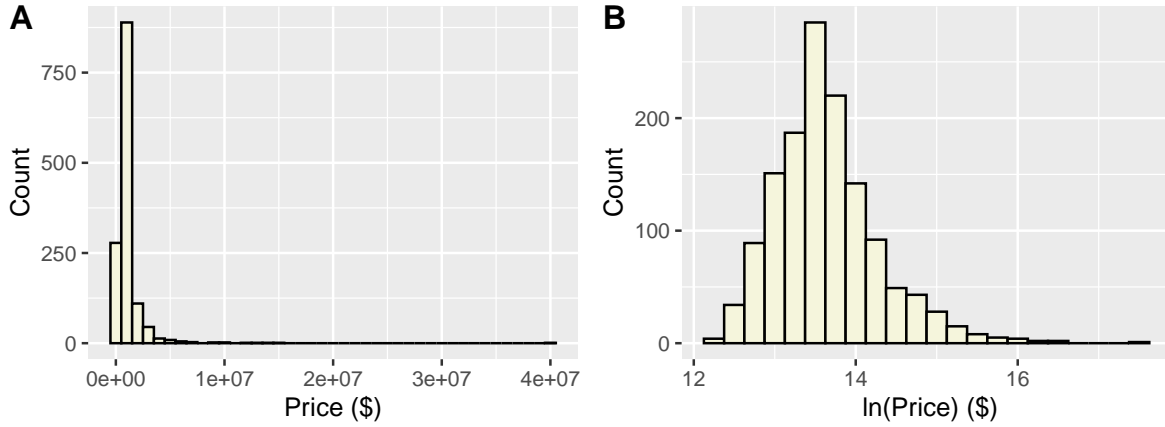
Figure 1: Distributions of Property Price and Logarithmic Property Price

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.

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 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.

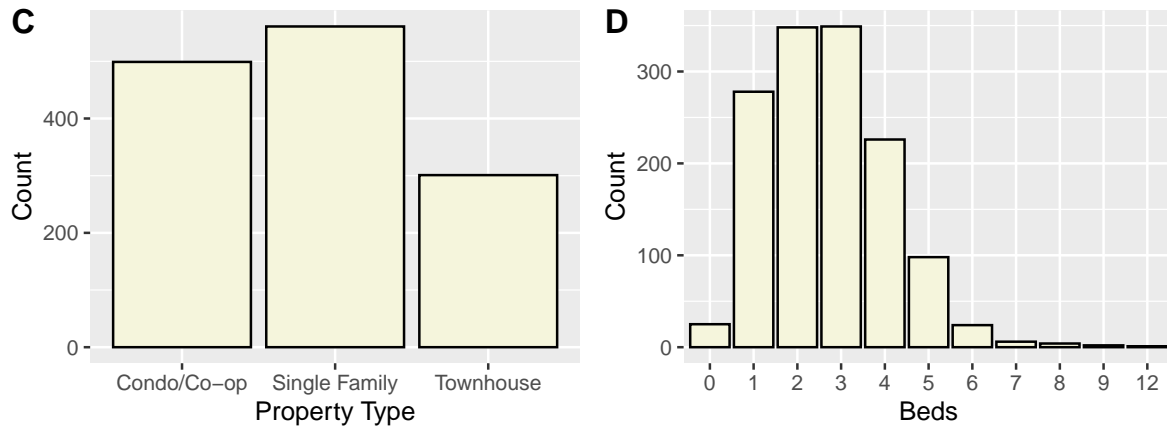Figure 2: Distributions of Property Type and Number of Bedrooms

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.

Figure 3: Distributions of Property Size and Property Age
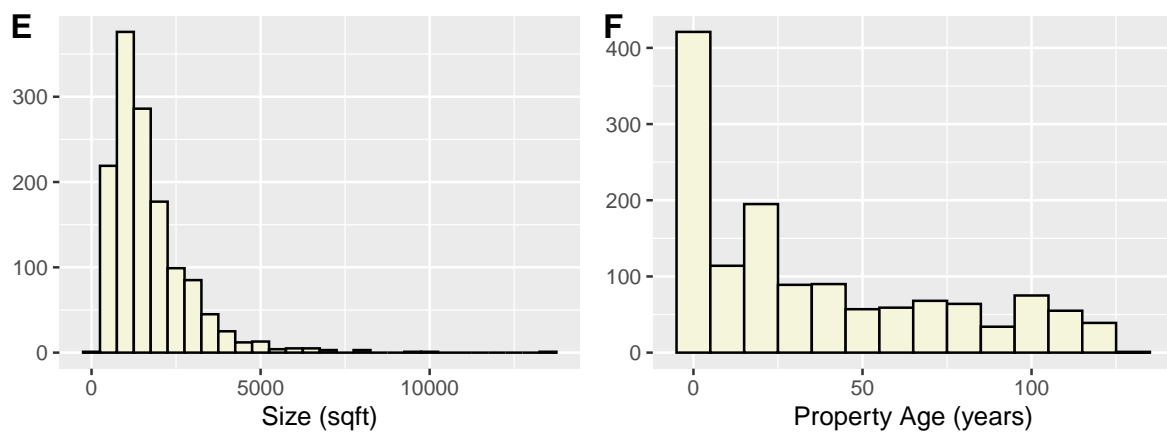
3 Model

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc euismod nisl sollicitudin blandit fermentum. Pellentesque molestie mi augue, a consectetur tellus commodo a. Sed dui lectus, sollicitudin ut sagittis eget, venenatis eget est. Proin neque massa, interdum et laoreet vulputate, cursus vitae mi. Ut scelerisque, dui vitae consequat fringilla, nunc felis pharetra dui, ac condimentum nibh ipsum non tortor. Vivamus dictum volutpat nisl. Quisque ullamcorper erat mi, ac tempus leo efficitur non. Nulla porta id est id tempus. Quisque pulvinar, dolor eget elementum tincidunt, est neque bibendum purus, quis sodales ipsum eros ac turpis.
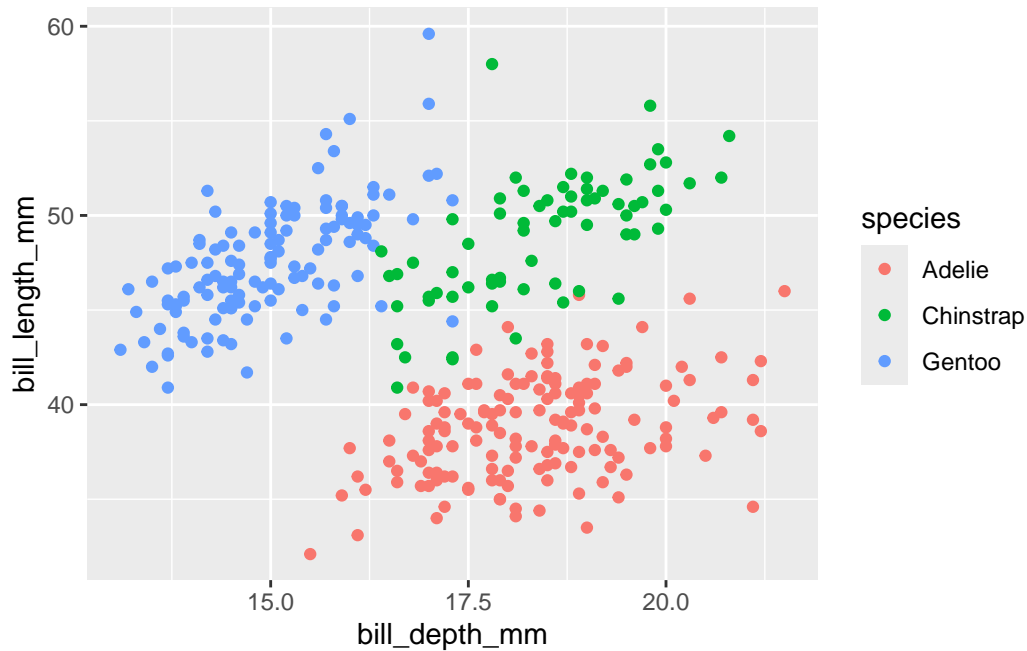
3.1 Model Setup

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc euismod nisl sollicitudin blandit fermentum. Pellentesque molestie mi augue, a consectetur tellus commodo a. Sed dui lectus, sollicitudin ut sagittis eget, venenatis eget est. Proin neque massa, interdum et laoreet vulputate, cursus vitae mi. Ut scelerisque, dui vitae consequat fringilla, nunc felis pharetra dui, ac condimentum nibh ipsum non tortor. Vivamus dictum volutpat nisl.

3.2 Model Justification

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc euismod nisl sollicitudin blandit fermentum. Pellentesque molestie mi augue, a consectetur tellus commodo a. Sed dui lectus, sollicitudin ut sagittis eget, venenatis eget est. Proin neque massa, interdum et laoreet vulputate, cursus vitae mi. Ut scelerisque, dui vitae consequat fringilla, nunc felis pharetra dui, ac condimentum nibh ipsum non tortor. Vivamus dictum volutpat nisl. Quisque ullamcorper erat mi, ac tempus leo efficitur non. Nulla porta id est id tempus. Quisque pulvinar, dolor eget elementum tincidunt, est neque bibendum purus, quis sodales ipsum eros ac turpis.
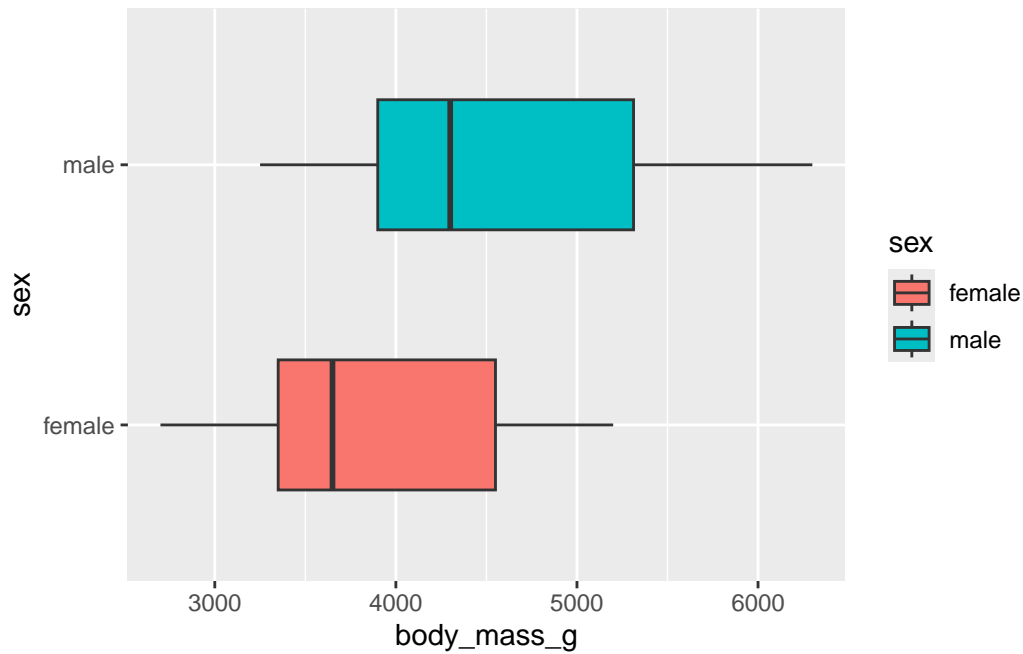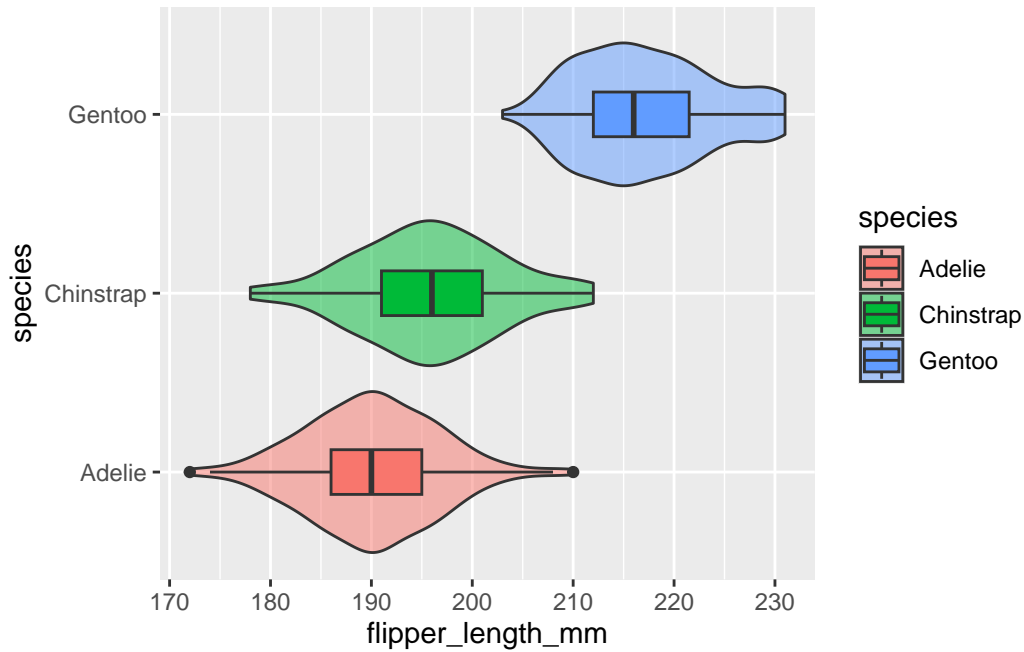
4 Results

4.1 Result 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis volutpat ut ante vestibulum sagittis. Etiam sollicitudin ligula neque, et sollicitudin erat iaculis eu. Pellentesque aliquam semper sapien, a faucibus nulla consequat vitae. Praesent in vulputate neque. Fusce fermentum tempor turpis, nec iaculis diam porta nec. Maecenas nec porta felis, at tincidunt ex. Mauris convallis tincidunt posuere. Quisque rutrum imperdiet turpis, quis bibendum est ultrices ut. Phasellus in tortor eu ligula pellentesque pellentesque. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Curabitur eleifend ac purus quis dapibus. Ut tincidunt posuere tortor, vitae venenatis nunc placerat eu. Praesent sit amet mi urna. Sed ultrices est et ligula posuere, eget molestie lorem porta. Quisque tempor, dolor porta semper aliquam, urna magna rhoncus sapien, at mattis metus nibh eleifend mi. Sed id bibendum urna, ut tempor ligula. Nulla facilisi. Nulla facilisi.

4.2 Result 2

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis volutpat ut ante vestibulum sagittis. Etiam sollicitudin ligula neque, et sollicitudin erat iaculis eu. Pellentesque aliquam semper sapien, a faucibus nulla consequat vitae. Praesent in vulputate neque. Fusce fermentum tempor turpis, nec iaculis diam porta nec. Maecenas nec porta felis, at tincidunt ex. Mauris convallis tincidunt posuere. Quisque rutrum imperdiet turpis, quis bibendum est ultrices ut. Phasellus in tortor eu ligula pellentesque pellentesque. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Curabitur eleifend ac purus quis dapibus. Ut tincidunt posuere tortor, vitae venenatis nunc placerat eu. Praesent sit amet mi urna. Sed ultrices est et ligula posuere, eget molestie lorem porta. Quisque tempor, dolor porta semper aliquam, urna magna rhoncus sapien, at mattis metus nibh eleifend mi. Sed id bibendum urna, ut tempor ligula. Nulla facilisi. Nulla facilisi.

4.3 Result 3

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis volutpat ut ante vestibulum sagittis. Etiam sollicitudin ligula neque, et sollicitudin erat iaculis eu. Pellentesque aliquam semper sapien, a faucibus nulla consequat vitae. Praesent in vulputate neque. Fusce fermentum tempor turpis, nec iaculis diam porta nec. Maecenas nec porta felis, at tincidunt ex. Mauris convallis tincidunt posuere. Quisque rutrum imperdiet turpis, quis bibendum est ultrices ut. Phasellus in tortor eu ligula pellentesque pellentesque. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Curabitur eleifend ac purus quis dapibus. Ut tincidunt posuere tortor, vitae venenatis nunc placerat eu. Praesent sit amet mi urna. Sed ultrices est et ligula posuere, eget molestie lorem porta. Quisque tempor, dolor porta semper aliquam, urna magna rhoncus sapien, at mattis metus nibh eleifend mi. Sed id bibendum urna, ut tempor ligula. Nulla facilisi. Nulla facilisi. Proin mollis quam ut rutrum vulputate. Praesent commodo.

5 Discussion

5.1 Discussion 1

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 Discussion 2

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 Discussion 3

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

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.

References

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

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. "Real Estate, Homes for Sale, MLS Listings, Agents." https://www.redfin.com.

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.