Impact of Reviews and Ratings on Airbnb Pricing in Canada: Analyzing Information Dynamics*

Alaina Hu and Abdul Ahad Qureshi

February 15, 2024

Airbnb has been emerging as a prominent player in the accommodation sector, providing travelers with a diverse range of lodging options. Existing research is primarily conducted on cities located in the United States and Europe, and few studies have focused on the market in Canada. This paper adds to the literature by analyzing the impact of information such as reviews and ratings on pricing strategies. Through data visualization, we find a negative relationship between the number of reviews of a listing and its price, a positive relationship between guest overall satisfication score and price, and no relationship between accuracy, cleanliness, location, and value ratings on. price.

Table of contents

*Code and data are available at: https://github.com/alainahu/airbnb_analysis. Replication on Social Science Reproduction platform available at: https://doi.org/10.48152/ssrp-yjrj-1d42

1 Introduction

In the changing landscape of modern travel, Airbnb has emerged as a transformative force, reshaping traditional notions of accommodation and hospitality. No longer are travelers confined to hotels; people have the choice to choose between various forms of accommodation and find the option most suitable for their needs. With Airbnb's vacation rentals, travelers have the option to gain access to more space, kitchens, home amenities, and lower cost (Guttentag 2016). Central to Airbnb's allure are the wealth of user-generated reviews and ratings, which serve as vital sources of information for prospective guests navigating a vast array of listings. These reviews not only offer insights into the quality and character of accommodations but also play a pivotal role in shaping consumer decisions. However, in this growing world of shared experiences, a fundamental question persists: What effect do these reviews and ratings have on Airbnb pricing strategies? This question lies at the heart of our research, as we delve into the relationship between information, consumer behavior, and pricing dynamics within the Airbnb marketplace in Canada. Our estimand for the pricing strategies of Airbnb units is the listed price for each rental unit since every Airbnb owner is able to set their own price. Through statistical analysis and data visualization, this paper endeavors to continue to explore the relationship between the explanatory variables of reviews and ratings with the response variable of Airbnb prices.

In an econometric analysis on Airbnb reviews and price in Boston, Lawani et al. (2019) find that reviews serve as a good proxy for rooms' quality and reviews affect both the host price and neighboring host price. Peng, Li, and Qin (2020) build a machine learning model and use natural language processing to predict Airbnb prices in nine cities across the United States, Australia, and Europe. Their research concludes that customer reviews, house features, and geographical data are all predictive factors for Airbnb rental prices. While there is an extensive amount of statistical and machine learning research focused on the relationship between Airbnb reviews and prices, this paper fills in the gap in literature by focusing on the information-price relationship within Canada, specifically the cities of Toronto and Vancouver. Existing research often focuses on cities in the United States or Europe. We are interested in seeing if the same relationship between reviews and price can be found when we zoom into Toronto and Vancouver. Regional differences should be examined: Ghosh, Jana, and Abedin (2023) find that predictability of Airbnb prices varies significantly across cities. For example, listing prices are the most predictable in Boston and least predictable in Chicago.

Our paper will follow a replication of research conducted by Laouénan and Rathelot (2022) on the effects of information on ethnic discrimination in the Airbnb market. We follow their paper to replicate the following aspects of the paper (1) the platform is effective in supplying useful information, so reviews and ratings impact the expectations of consumers and (2) there is a relationship between number of reviews and listed price. Data analysis for this reproduction is performed in `R` (R Core Team 2022), and additional help is provided by libraries such as `dplyr` (Wickham et al. 2023), `ggplot2` (Wickham 2016), `ggrepel` (Slowikowski 2024), `tidyverse` (Wickham et al. 2019), `kableExtra` (Zhu 2021), `knitr` (Xie 2023), `haven` (Wickham, Miller, and Smith 2023), 'readr' (Wickham, Hester, and Bryan 2024), `stargazer`(Hlavac 2022), `psych` (William Revelle 2024), `scales` (Wickham, Pedersen, and Seidel 2023), `RColorBrewer` (Neuwirth 2022), `gridExtra` (Auguie 2017), `here` (Müller 2020).

Through data visualization, we observe a negative relationship between number of reviews and listing price, meaning that more reviews on an Airbnb unit are associated with a lower price. We see a moderate positive relationship between Overall Guest Satisfaction scores and prices, meaning that higher guest satisfaction scores are related to higher prices. We find no relationship between the accuracy, cleanliness, location, and value ratings with price.

Our research paper begins with the Data section to visualize and further understand the measurement, source, methodology, and variables we are examining. Then, we go into Results of the relationships in the data. Finally, we include the Discussion of the findings, summarizing the takeaway and future of this research.

2 Data

2.1 Measurement

In order to analyze the relationship between reviews and ratings with the price, we need a method to measure these variables. Since we are replicating Laouénan and Rathelot's findings, we use their data set with our variables of interest. They collect their data through web-scraping of the Airbnb website, directly retrieving the number of reviews, rating scores, and the prices of rentals. This method leaves little room for measurement error of the variables of interest as consumers also receive the same market information. However, it is difficult to collect data for all Airbnb listings in the two cities as rentals enter and leave the market. More information on the methodology of data compliation is discussed in the Methodology section.

2.2 Source

The paper titled "Can Information Reduce Ethnic Discrimination? Evidence from Airbnb" (Laouénan and Rathelot 2022) published by the American Economic Association (AEA 2022) is the focus of our paper. We reproduce and investigate several areas of the original paper, using the same datasets for our analysis.

2.3 Methodology

The dataset was compiled from the publicly accessible listings on Airbnb's platform, capturing details visible on the initial listing page. (Laouénan and Rathelot 2022). The collection process was repeated every two to three weeks between June 2014 and June 2015 with an additional collection in November 2017, resulting in 21 total data collection waves. (Laouénan and Rathelot 2022). Following Laouénan and Rathelot's methodology, our analysis focuses on listings that received at least one review during the observation period, underlining the inclusion of only actively engaged listings. (Laouénan and Rathelot 2022) This criterion narrowed the original dataset from 663,090 to 220,939 listings. For our specific interest in the Canadian market, we further refined the dataset to include listings solely from Toronto and Vancouver, as detailed in Table 1.

We employ a methodological approach akin to Altonji and Pierret (2001), observing both the quantity and quality of information available about a property to potential guests, which could influence their decision-making process.(Altonji and Pierret 2001)

2.4 Features

The original research spanned 19 cities across North America and Europe, characterized by high listing volumes. Our study narrows this focus to Canadian cities, specifically Toronto and Vancouver, offering a localized perspective on the data while broadening the research question. The observations and listings for our research are detailed in Table 1

Figure 1 illustrates the broad spectrum of daily rental prices, highlighting significant price variability across listings. To mitigate outlier effects, we excluded the top and bottom 1% of the price range. The price distribution reveals a first quartile at $69, a median of $94, and a third quartile at $126 per night, indicating a skewed distribution with a mean price of $108.

Table 1: Number of Observations and Listings for Toronto and Vancouver. After filtering out observations with incomplete information regarding reviews and ratings, we are left with 87,059 observations. We continue the analysis with these observations.

City	Observations	Listings
Toronto	56,843	5,359
Vancouver	45,569	4,219

Our analysis emphasizes the importance of review quantity and quality in correlating information availability with listing prices. Utilizing the most recent rating for each property, which aggregates all received ratings over its Airbnb tenure, we noted a skewness towards higher ratings, a finding consistent with previous research by Fradkin, Grewel, and Holtz (2018)

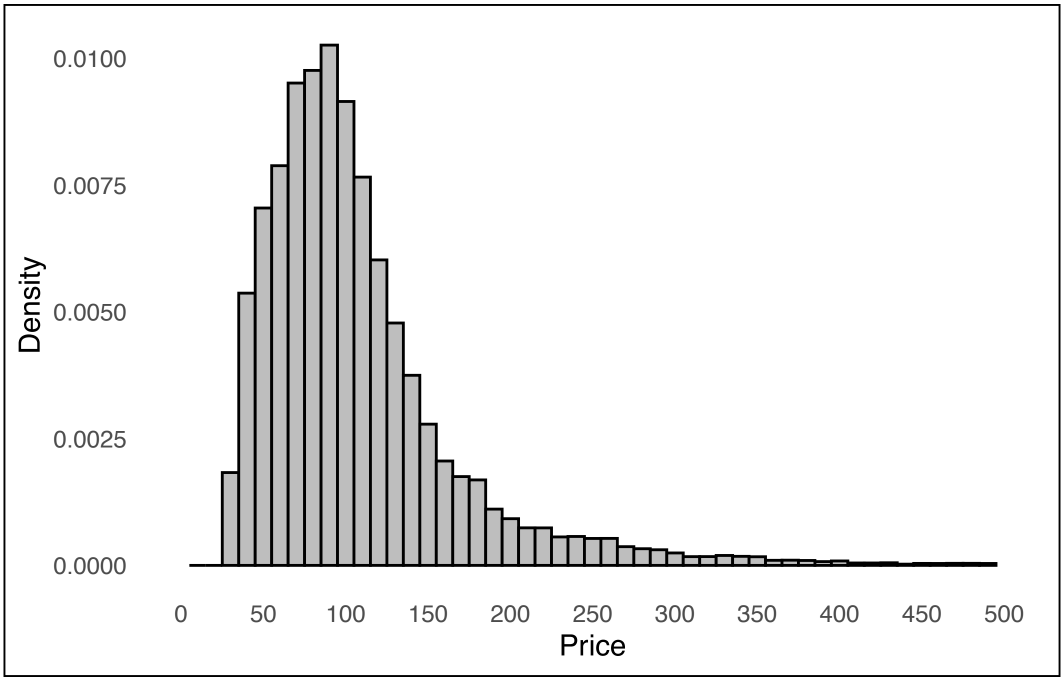

Figure 1: Distribution of Airbnb prices per night (including cleaning fees)

From the original dataset's 183 variables, we focused on 8 key variables deemed most relevant to our research question regarding information's impact on pricing:

1. Review: Represents the number of reviews received by the listing.It is the feedback provided by guests after their stay or experience.
2. Guest Satisfaction: A numerical rating indicating the overall satisfaction level of guests.
3. Accuracy Rating: Measured how well the description and images of a listing match the actual experience. Higher ratings suggest that guests found the property as advertised.
4. Cleanliness Rating: Assesses the cleanliness of the accommodation or service provided. It reflects guests' perceptions of the hygiene standards maintained at the property.
5. Location Rating: Rates the convenience, desirability, or attractiveness of the property's location, considering factors like proximity to tourist attractions, amenities, transport links, and the overall neighborhood.
6. Value Rating: Evaluates guests' perceptions of the worthiness of the service or accommodation relative to the price paid. It considers whether guests feel they received good value for their money.
7. New Price: Refers to the price in U.S dollars including cleaning fees and consequently due to the law of demand and supply.

8. Log Price: Relative change in prices due to availability of information

Table 2: Summary Statistics for the Reviews and Ratings

Variable	Mean	Median	Standard Deviation	Min	Max
Accuracy Rating	9.47	10	0.98	0	10
Cleanliness Rating	9.28	10	1.14	0	10
Guest Overall Satisfaction	93.47	95	7.78	20	100
Location Rating	9.48	10	0.95	0	10
Number of Reviews	14.90	7	23.86	1	566
Value Rating	9.29	9	1.00	0	10

Table 2 provides a statistical summary of various ratings and the number of reviews for Airbnb listings in Toronto and Vancouver. The table suggests that Airbnb listings in the study generally score very high across most rating categories, with a tendency for ratings to cluster around the higher end of the scale. The number of reviews, however, varies more widely than the other variables, which could be due to differences in listing popularity, duration on the platform, or other factors not specified by the ratings alone. This information, particularly the high ratings, could be influencing the pricing strategies of Airbnb hosts in these cities. We will test this intuition in this paper.

These variables were chosen for their direct relevance to our investigation into how informational transparency can affect Airbnb pricing strategies, with all variables being quantitative to facilitate our analysis.

3 Results

As we transition into a detailed examination of our findings, it is crucial to consider the graphical analyses that provide an empirical foundation for our interpretations. This section commences with an exploration of two scatter plots that render the relationship between the number of reviews, guest satisfaction scores, and the associated pricing of Airbnb listings in Toronto and Vancouver. These plots are instrumental in discerning the influence of consumer feedback and perceived satisfaction on the market value of the listings. They serve as a prelude to the violin plots, enabling us to contrast the discrete impact of reviews and satisfaction against the broader spectrum of ratings encompassing accuracy, cleanliness, and location. These visual representations provide a nuanced understanding of how different perceived attributes of a listing correlate with the pricing strategies adopted by hosts. As we delve into the results, these plots will serve to highlight the underlying trends and patterns that govern the pricing mechanisms across these urban markets, offering insights into the complex interplay between guest expectations and the economic imperatives of Airbnb hosts.

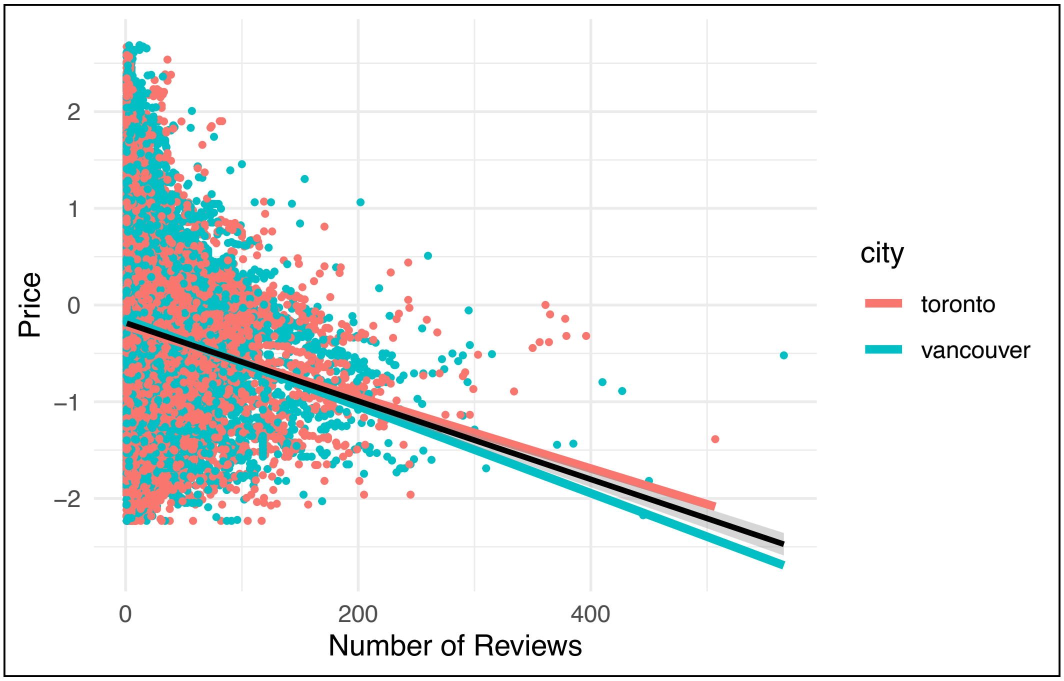

Figure 2: Plot of Listing Reviews and Price

Figure 2 appears to depict a scatter plot comparing the number of reviews to price for Airbnb listings in two cities: Toronto and Vancouver. The red dots represent listings in Toronto, while the blue dots represent listings in Vancouver. Two lines of best fit, one for each city, suggest a general trend in the data. The prices shown are deviations from a mean or median price, allowing for a more direct comparison of price trends relative to the average pricing in each city. From the graph, we can observe that there is a noticeable concentration of listings with fewer reviews, as evidenced by the dense clustering of points near the origin (where the number of reviews is low). This suggests that a large number of properties have not accumulated many reviews. As the number of reviews increases, there is a spread in the price range for both cities. However, the trend lines indicate that, on average, there is a slight negative relationship between the number of reviews and price, meaning that as the number of reviews increases, the price tends to decrease slightly. This could imply that more frequently reviewed (and possibly more frequently rented) properties are priced more competitively. The spread of prices in Toronto appears to be wider than in Vancouver, indicating more variability in how prices are set relative to the number of reviews in Toronto. Both cities have outliers with very high numbers of reviews compared to the general trend, which could represent properties that are exceptionally popular or have been listed for a long time. The graph helps us understand that while there is a general trend of price decreasing with more reviews, there are variations between the two cities. This could be due to differences in market dynamics, competition, or other local factors affecting Airbnb pricing strategies.

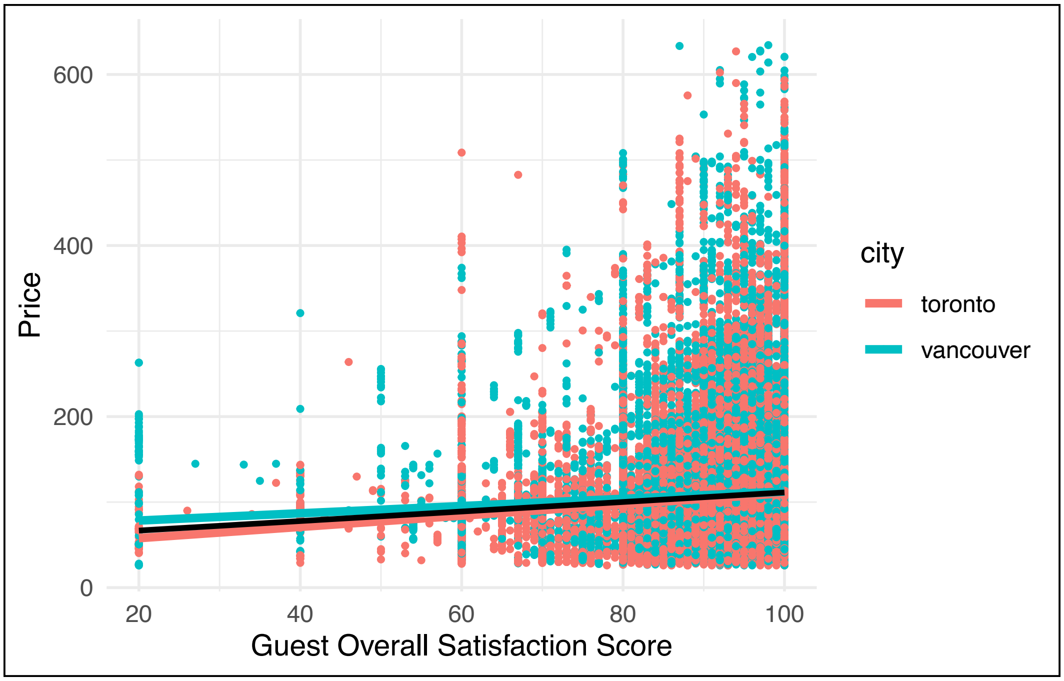

Figure 3: Plot of Guest Overall Satisfaction Score and Price

Figure 3 shows the relationship between guest overall satisfaction scores and the price of Airbnb listings in Toronto (represented by red dots) and Vancouver (represented by blue dots). The horizontal axis indicates the guest satisfaction score, which is on a scale of 0 to 100, while the vertical axis represents the price. The data points for both cities are densely packed at the higher end of the guest satisfaction score spectrum, suggesting that a large number of properties are highly rated by guests in terms of overall satisfaction. There's a significant spread in price at almost all levels of guest satisfaction scores, but particularly so at the higher end, indicating that a high satisfaction score does not necessarily correlate with a higher or lower price. This suggests that guests can find both high- and low-priced options among the well-rated properties. The trend lines for both cities appear to be flat, implying no strong correlation between the guest satisfaction score and the price. There is a slight upward trend which is more clear in Toronto but as the scores rise the prices in both cities converge to the same line of best fit. The slight upward trend indicates a weak relationship where as the satisfaction score increases, hosts respond by increasing prices or the more expensive listings have a higher guest satisfaction score. The distribution of prices in Toronto and Vancouver overlaps significantly, indicating that the overall satisfaction score's impact on price is similar in both cities. There are outliers, particularly at the lower end of the guest satisfaction score, where a few listings have high prices despite lower scores. Conversely, there are listings with high satisfaction scores that are priced low, possibly indicating good value. The graph supports the notion that while guest satisfaction is an important metric, it's not the sole determinant of price. Other factors likely play a significant role in pricing decisions, such as location, accuracy and value ratings. We will consider these factors next. The distribution of prices across the satisfaction score range highlights the diverse options available to guests in terms of price and satisfaction levels.

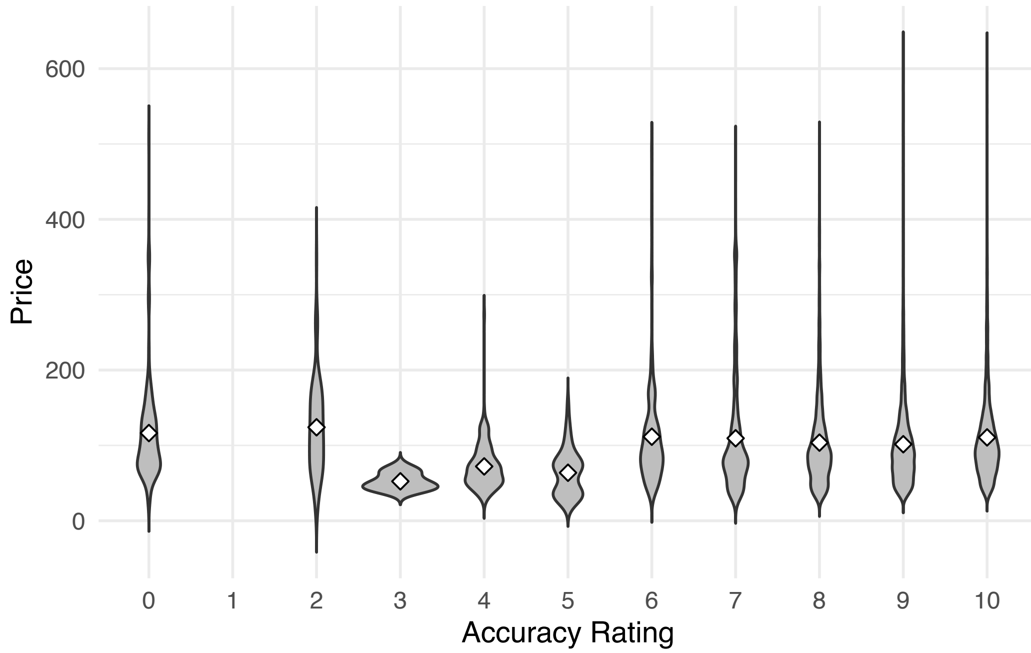

Figure 4: Plot of Accuracy Rating and Price

Figure 4 is a violin plot that displays the relationship between accuracy rating and price for Airbnb listings. The accuracy rating ranges from 0 to 10, and the price is displayed on the vertical axis. The width of each violin represents the density of data points at different prices for each accuracy rating. A wider section indicates a higher concentration of listings at that price range. The white diamond within each violin represents the mean price for that particular accuracy rating. Ratings from 4 to 10 show similar patterns with a concentration of prices at the lower end, indicating that most listings, regardless of accuracy rating, tend to be clustered around a lower price point. The plot for ratings of 0 and 1 is quite different, showing less data and possibly outliers, as these ratings are uncommon and may not provide a reliable indication of trends. As the accuracy rating increases, there doesn't appear to be a significant change in the concentration of lower-priced listings, suggesting that accuracy rating does not drastically affect the price. The violins for higher accuracy ratings seem to have longer tails extending to higher prices, possibly indicating that listings with higher accuracy ratings can command higher prices, although these are less common. The overall distribution suggests that most listings are priced in the lower to mid-range across all accuracy ratings, with a few listings at each rating level reaching higher price points. From this plot, it can be inferred that while there might be a slight trend of increased price with higher accuracy ratings, the effect is not pronounced, and the majority of listings are still concentrated in the lower price ranges regardless of the accuracy rating.

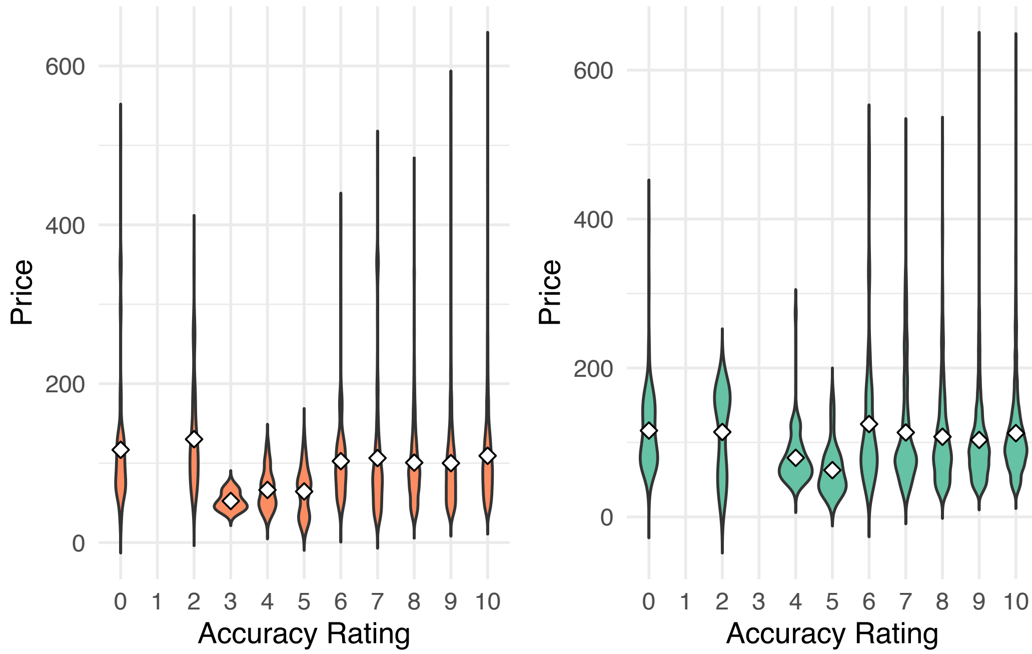

Figure 5: Plot of Accuracy Rating and Price in Toronto (left) and Vancouver (right)

Figure 5 presents two violin plots side by side, representing the relationship between accuracy rating and price for Airbnb listings in Toronto (left) and Vancouver (right). The accuracy rating is on a scale from 0 to 10, and the price is on the vertical axis.

Observations for Toronto (left plot):

Listings with the lowest accuracy ratings (0-2) are quite rare, as indicated by the narrow width of the violins. The median prices, shown by the white diamonds, are relatively stable across different accuracy ratings. There is a significant number of high-priced outliers, particularly in the 1 to 4 rating range, as indicated by the long upper tails. The bulk of the listings, especially those rated 5 and above, have a relatively concentrated price range, which is lower than the outliers.

Observations for Vancouver (right plot):

The distribution of prices in Vancouver appears more uniform across different accuracy ratings when compared to Toronto. The median prices in Vancouver are also fairly stable across ratings, but the price ranges (indicated by the width of the violins) are slightly more varied, particularly for ratings 5 through 10. There are fewer extreme price outliers in Vancouver than in Toronto, as seen by shorter upper tails in the plot. The listings in Vancouver with the highest accuracy ratings (9-10) show a tighter price concentration around the median, suggesting less price variation for highly accurate listings.

Comparing the two cities:

Toronto displays a greater presence of high-priced outliers across all accuracy ratings compared to Vancouver. Median prices do not seem to be heavily influenced by accuracy rating in either city. The price distribution is more varied in Toronto, with a broader range of prices at most ratings, whereas Vancouver shows a slightly tighter distribution. Both cities show a trend where listings with the highest accuracy ratings (8-10) tend to have less price variation.

These violin plots help in understanding the pricing strategy in relation to the perceived accuracy of listings in both cities. It appears that while the accuracy rating does influence the price to some extent, particularly at the higher end of the rating scale, it is not the sole factor determining pricing, as evidenced by the similar median prices across the ratings.

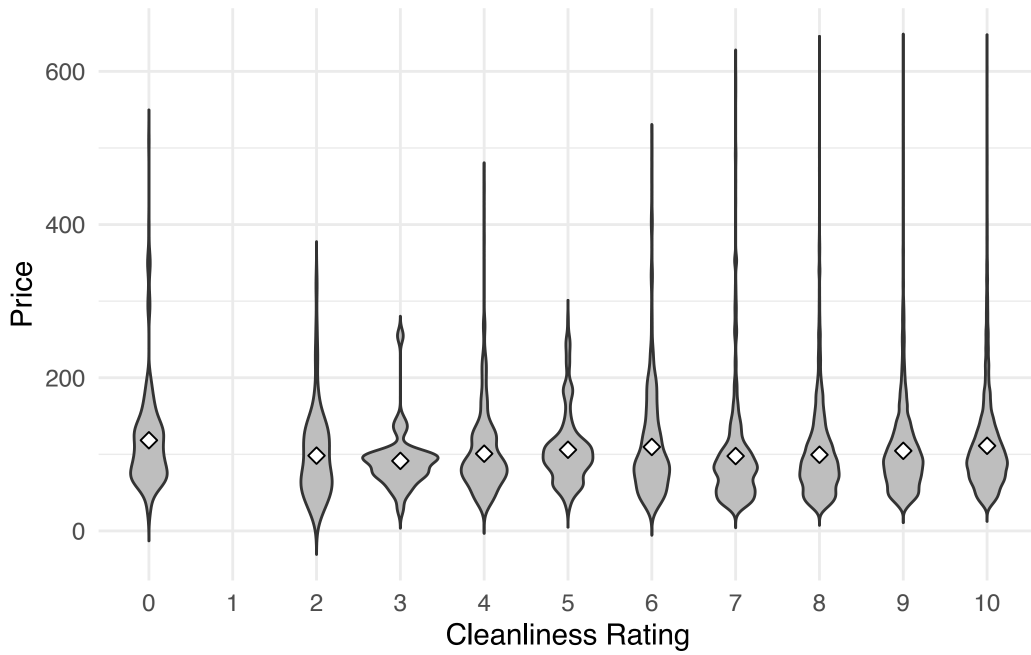

Figure 6: Plot of Cleanliness Rating and Price

Figure 6 illustrates the relationship between cleanliness rating and price for Airbnb listings.

The cleanliness ratings are on a scale from 0 to 10, and the prices are shown on the vertical axis. Similar to the previous graphs, the width of each violin indicates the density of listings at different price points for each cleanliness rating. Listings with the lowest cleanliness ratings (0-2) are less common, as shown by the narrower violins. As the cleanliness rating increases, the violins become wider at the bottom, indicating a higher concentration of lower-priced listings. There is a consistent presence of listings across all levels of cleanliness ratings, but there is a notable quantity of high-priced outliers, especially from ratings 1 to 4. From ratings 5 to 10, the distribution of prices is quite consistent, with a majority of listings concentrated at the lower end of the price spectrum. The violins' shapes suggest that there's a broad range of prices for listings with higher cleanliness ratings, but the median prices do not significantly increase with higher cleanliness ratings. The presence of outliers with high prices at almost all levels of cleanliness ratings indicates that some listings are priced much higher than the average, regardless of their cleanliness score. In general, this graph suggests that while cleanliness is an important factor for Airbnb listings, it may not have a strong direct influence on the price, as evidenced by the stable median prices across ratings. However, guests are likely to encounter a range of prices at any given level of cleanliness, including a number of premium-priced listings.

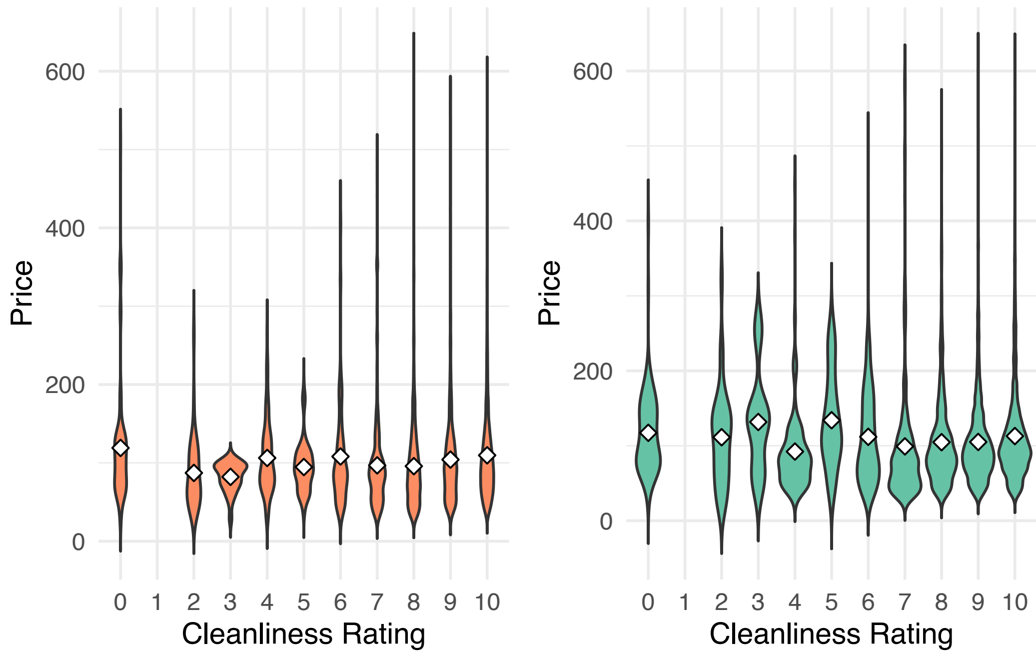

Figure 7: Plot of Cleanliness Rating and Price in Toronto (left) and Vancouver (right)

Figure 7 presents side-by-side violin plots for Toronto and Vancouver, showing the relationship between cleanliness ratings and price for Airbnb listings.

The graph presents side-by-side violin plots for two different cities, presumably Toronto on the left and Vancouver on the right, showing the relationship between cleanliness rating and price for Airbnb listings.

Toronto (left plot):

The violins for ratings 0 through 2 are narrow, suggesting fewer listings with low cleanliness ratings. There is a considerable number of listings at higher cleanliness ratings (7-10), as evidenced by the wider parts of the violins. The median prices across cleanliness ratings do not vary significantly, as indicated by the white diamonds remaining relatively consistent in the lower price range. Each rating level has some listings with very high prices, indicated by the long upper tails, especially for ratings 3 through 6.

Vancouver (right plot):

Listings with cleanliness ratings of 0 and 1 are also rare in Vancouver. There is a tighter concentration of listings around the median price for higher cleanliness ratings (8-10), suggesting less variability in price as cleanliness increases. The median prices appear consistent across the cleanliness ratings, similar to Toronto's trend. The long upper tails are present but slightly less pronounced than in Toronto's plot, indicating fewer extreme price outliers in Vancouver.

Comparing Toronto and Vancouver:

Both cities have a similar distribution of prices across different cleanliness ratings. The median prices are consistent across ratings in both cities, indicating that cleanliness rating has a limited impact on the median pricing of listings. Toronto shows slightly more price variability with longer tails at certain cleanliness ratings, which might suggest a broader range of property types or host pricing strategies. High cleanliness ratings do not necessarily command a much higher price in either city, given the median price points remain relatively stable.

These violin plots highlight the variation in price at different levels of cleanliness ratings and suggest that while cleanliness is an important factor in the guest experience, it might not be the primary driver of price for Airbnb listings. Other factors likely play a role in determining the price, such as location, size, and amenities of the listings.

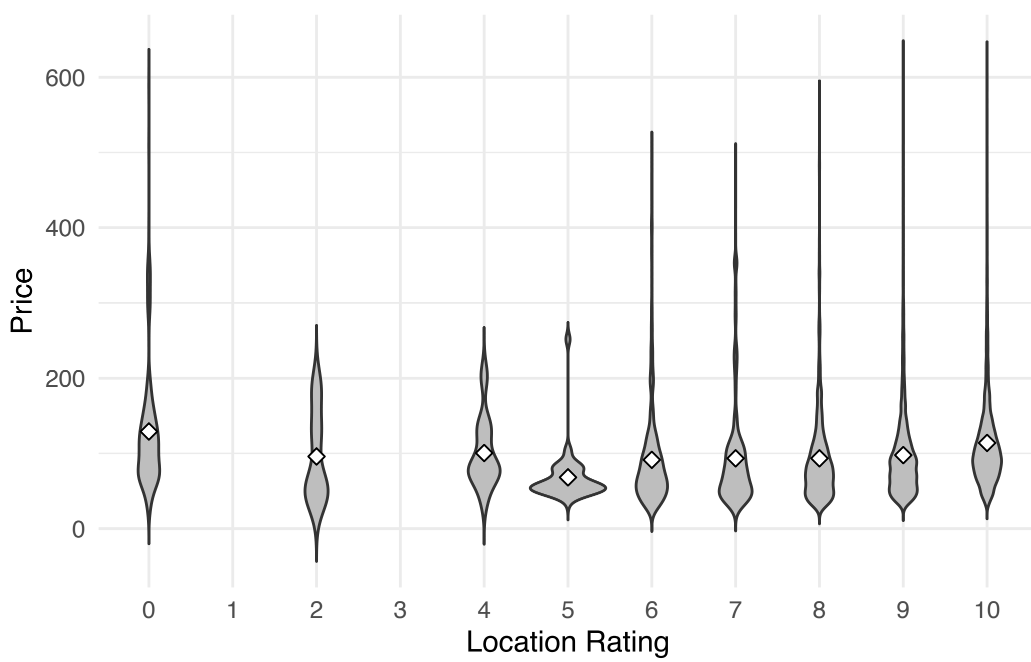

Figure 8: Plot of Location Rating and Price

Figure 8 is a violin plot that shows the relationship between location rating and price for Airbnb listings. The location ratings are on a scale from 0 to 10, with the price on the vertical axis. Listings with the lowest location ratings (0-2) appear to be less common, as indicated by the narrower violins. The width of the violins increases as the location rating improves, which suggests a higher density of listings at better-rated locations. The median prices do not show a significant upward trend with better location ratings; they remain relatively stable across the spectrum of ratings. Each rating level has listings with high prices but this is more pronounced for ratings 6 through 8. The broadest parts of the violins for higher ratings (9-10) indicate a higher concentration of listings around a central price range, which is not necessarily higher than the central price range of lower-rated locations. This plot suggests that while location is an important aspect of Airbnb listings, a higher location rating does not directly translate into a significantly higher price. There's a broad range of prices available at all levels of location rating, with the median price remaining fairly consistent. Listings at various locations, even those with the highest ratings, offer a range of prices, possibly accommodating a wide range of guest budgets.

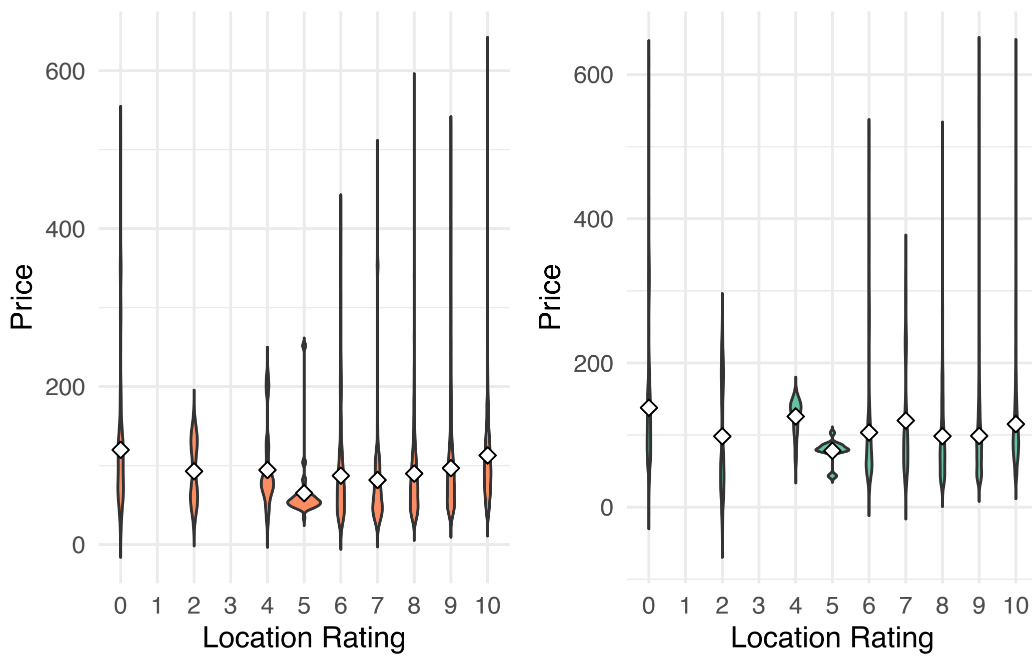

Figure 9: Plot of Location Rating and Price in Toronto (left) and Vancouver (right)

Figure 9 shows side-by-side violin plots for two different cities, representing Toronto on the left and Vancouver on the right. These plots illustrate the relationship between location rating and price for Airbnb listings.

For Toronto, listings with very low location ratings (0-2) are not common. The median prices across location ratings are relatively constant suggesting that location rating does not drastically alter the median price. There is significant variability in prices at almost all location ratings, particularly for the rating of 4. The broadest parts of the violins at higher ratings (7-10) suggest a larger number of listings clustered around the median price range.

Listings with the lowest location ratings are sparse in Vancouver, similar to Toronto. The median prices are again consistent across location ratings, reinforcing the observation that location rating does not have a strong impact on median pricing. Price variability is present but appears less pronounced than in Toronto, with shorter tails on the violins. The concentration of listings around the median price for higher location ratings (8-10) is tighter, indicating less price variability for well-located properties.

Both cities show a consistent median price across location ratings, suggesting other factors may influence the pricing more than location rating. Toronto exhibits a wider range of prices, especially at mid-level location ratings, while Vancouver's prices are more concentrated around the median. In both cities, higher location ratings do not necessarily command much higher prices, as evidenced by the stable median prices. These violin plots illustrate that while location is a key aspect for Airbnb listings, it doesn't strongly dictate the price. Properties with a wide range of location ratings offer varied prices, with no clear trend of increasing price with higher location ratings. This could indicate that guests prioritize different aspects of their stay beyond just location or that a good location is a baseline expectation for listings in these cities

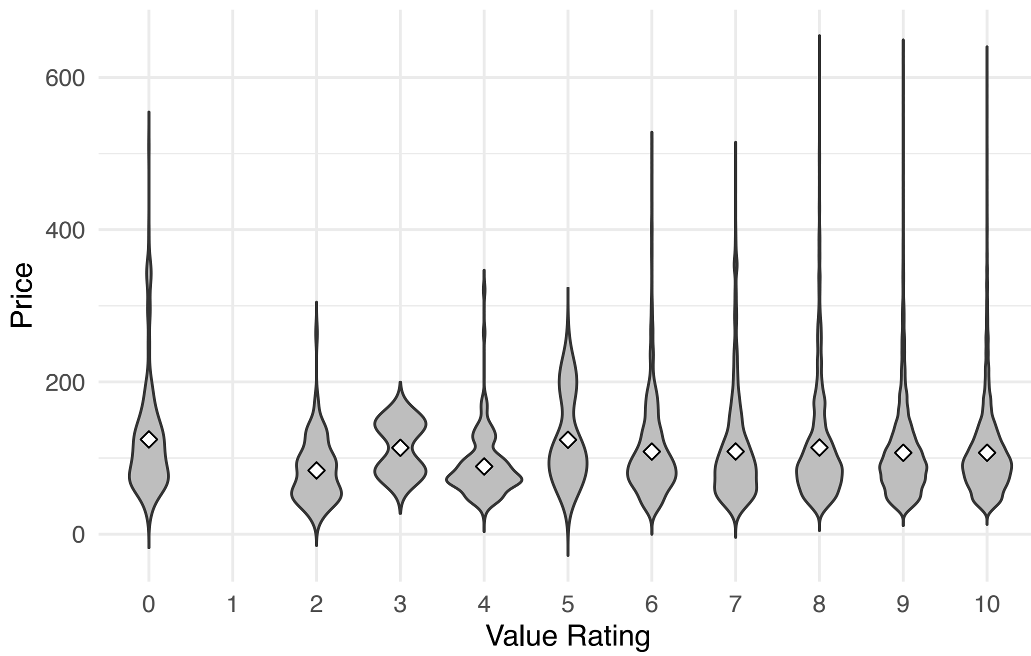

Figure 10: Plot of Value Rating and Price

Figure 10 displays a violin plot depicting the relationship between value rating and price for Airbnb listings. Value ratings are on a scale from 0 to 10, and the price is plotted on the vertical axis. The narrowest parts of the violins at the lower value ratings (0-2) indicate fewer listings, suggesting that few properties are perceived as offering poor value. Median prices, represented by the white diamonds, remain relatively consistent across different value ratings, implying that the perceived value does not drastically change the median price. The distribution of prices widens as the value rating increases, particularly from ratings 4 to 10, which shows a greater variability in the price at higher value ratings. Long upper tails across the violins suggest the presence of listings priced significantly higher than the median, regardless of value rating. The broad bases of the violins at higher ratings indicate a larger number of listings at lower price points, even for listings with high value ratings. The plot suggests that while listings across all value ratings have a wide range of prices, the perceived value offered by a listing does not necessarily correlate with higher prices. Guests can find Airbnb listings perceived to offer good value at both lower and higher price points.

```
Warning: Groups with fewer than two data points have been dropped.
Groups with fewer than two data points have been dropped.
```

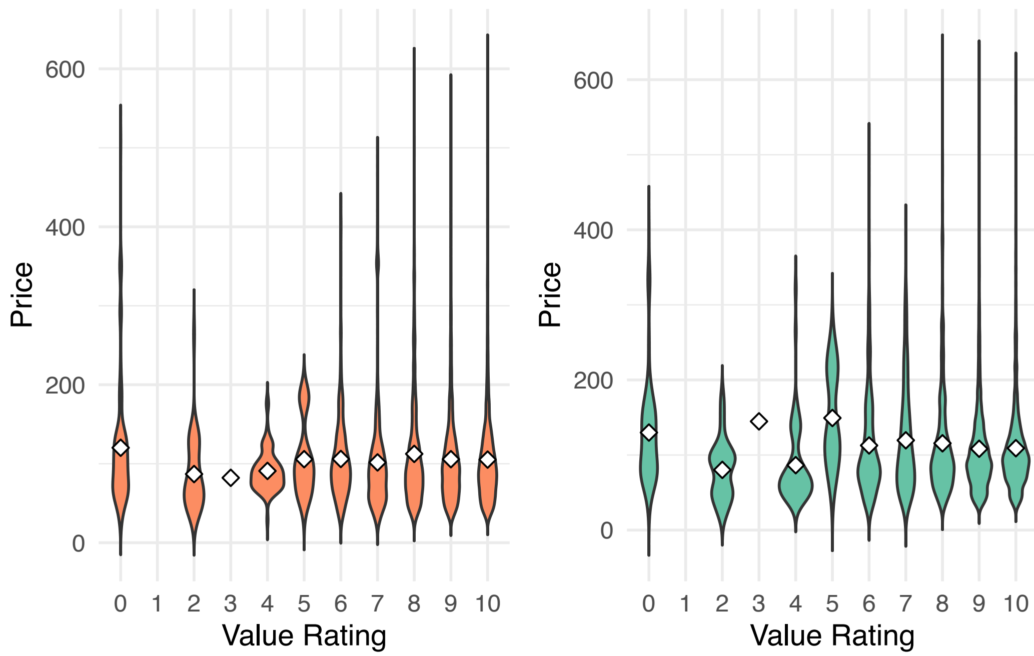

Figure 11: Plot of Value Rating and Price in Toronto (left) and Vancouver (right)

Figure 11 depicts two sets of violin plots, representing Airbnb listings from Toronto (left) and Vancouver (right), showing the distribution of prices across different value ratings.

In Toronto, Listings with the lowest value ratings (0-2) are not very common. The median prices do not drastically change across value ratings, suggesting that higher value ratings do not necessarily correspond with higher prices. There is considerable price variability at all levels of value ratings. The distribution of listings across value ratings seems fairly consistent, with a slight increase in density at the median price for higher ratings (8-10).

Similar to Toronto, lower value ratings (0-2) in Vancouver have fewer listings. Median prices are relatively stable across the value ratings, indicating that perceived value does not have a strong influence on the median listing price. The price variability in Vancouver appears slightly less pronounced compared to Toronto, with shorter tails, especially at higher value ratings. Higher value ratings in Vancouver (9-10) show a tighter concentration of listings around the median price, with less variability.

Both Toronto and Vancouver exhibit a similar pattern where the median price does not increase significantly with higher value ratings. Toronto has a wider range of prices for listings with mid to high value ratings, whereas Vancouver shows a somewhat more compact distribution. Listings in both cities with high value ratings (8-10) tend to cluster around a consistent median price, suggesting that guests value these listings for offering a good balance between cost and the quality of the experience. These plots suggest that guests in both cities can find listings that are perceived to offer good value across a broad range of prices. The concept of "value" in this context seems to encompass more than just a financial aspect, likely including other qualitative factors that guests consider important.

In conclusion, the results section has presented a multifaceted view of the Airbnb market in Toronto and Vancouver, weaving together the intricate threads of guest reviews, satisfaction scores, and various quality ratings. The scatter plots and violin plots collectively unravel the complex relationship between perceived quality attributes and pricing. Despite the intuitive expectation that higher ratings would command higher prices, our analysis reveals a more nuanced reality where prices do not uniformly increase with higher ratings. This suggests that guests may find well-rated listings across a spectrum of prices, reflecting a market that values accessibility and diversity. The data-driven insights gleaned from our analysis not only challenge conventional wisdom but also provide a robust framework for understanding the dynamics at play in the short-term rental market. Moving forward, these insights can inform both hosts and platform designers as they navigate the competitive landscape of urban Airbnb offerings.

4 Discussion

Our analysis aimed to understand the dynamics between Airbnb reviews, ratings, and listing prices in the Canadian cities of Toronto and Vancouver. Our results show some links between how customer feedback affects prices on the platform. Here, we discuss our findings, summarize the research, and consider future steps.

4.1 Key Findings

A key finding from our research is the negative relationship between the number of reviews a listing has and its price. This suggests that listings with a higher volume of reviews tend to be priced lower. One plausible explanation for this phenomenon could be that hosts with more reviews have successfully attracted more guests over time, possibly by offering more competitive pricing. Alternatively, it could indicate that guests are more inclined to choose and review more affordably priced accommodations, thereby increasing the review count of such listings. This finding challenges the intuitive expectation that a higher number of reviews, as an indicator for popularity or demand, would correlate with higher prices. Instead, it underscores the importance of pricing strategy in garnering guest expectation, interest, and feedback.

Our analysis also uncovers a slightly positive relationship between guest overall satisfaction scores and listing prices. This indicates that listings which manage to achieve higher satisfaction scores can command slightly higher prices, likely reflecting the premium that guests are willing to pay for a perceived higher quality or experience. This relationship aligns with expectations that positive guest experiences, as encapsulated by overall satisfaction scores, would be valued higher in the marketplace.

Our study finds no significant relationship between price and specific ratings such as accuracy, location, cleanliness, and value. This absence of correlation suggests that while these factors are crucial for guest satisfaction and decision-making, they do not directly influence the pricing of listings. It may be that these specific aspects of guest experience are considered baseline expectations, with deviations influencing guest choice but not necessarily allowing hosts to adjust prices upwards for excelling in these areas. Initially in Table 2, there is a left-skew of rating scores. As many listings generally score high in these ratings, the scores may not be a strong indicator of pricing strategy.

4.2 Weaknesses

A limitation of our study is the reliance solely on data visualization techniques. We relied on plotting the relationships between our independent variables (e.g., number of reviews, guest satisfaction scores) and the dependent variable (price), without the application of linear regression or the construction of predictive models. While these visualizations have offered valuable initial insights into potential trends within the data, they fall short of providing a detailed quantitative analysis that could confirm the strength, direction, and significance of these relationships.

The absence of model-based analysis prevents us from adjusting for potential confounds or exploring the interactive effects between different variables. As such, the relationships we report may be subject to underlying biases or influenced by variables not included in our visual explorations.

4.3 Further Steps

Our research paper replicates Laouénan and Rathelot's findings that show (1) the platform is effective in supplying useful information, so reviews and ratings impact the expectations of consumers and (2) there is a relationship between number of reviews and listed price. However, we leave out the focus on ethnic discrimination in this paper. A further step to take could be to perform an analysis on the impact of information on ethnic discrimination in the Airbnb market in Canada and fully replicate the original paper's findings. Researching the relationship between increased information on ethnic discrimination and lowering the gap in prices between majority and minority Airbnb hosts would be important because of Canada's growing minority population. Since 2017, over 50% of people in Toronto identified as a visible minority, making the city majority visible minority (Cole, Tulk, and Grzincic (2017)).

Not only can we expand the scope of our research, we can also deepen the complexity of our statistical analysis. To better understand the relationship between reviews and ratings on prices, we can conduct linear and multiple regression analyses or build models to incorporate a wider array of variables, such as location-specific factors and host characteristics. Further, integrating qualitative data, such as text analysis of review content, could provide a more nuanced understanding of how the qualitative aspects of guest feedback influence pricing decisions. By adopting these approaches, future research can bring us closer to better replicating the original findings and offer more definitive conclusions and recommendations for Airbnb hosts looking to optimize their pricing strategies in a competitive market in Canada.

References

AEA. 2022. *The American Economic Association Portal*. *The American Economic Association*. https://www.aeaweb.org/.

Altonji, Joseph G., and Charles R. Pierret. 2001. "Employer Learning and Statistical Discrimination." *The Quarterly Journal of Economics* 116 (1): 313–50. https://www.jstor.org/stable/2696451.

Auguie, Baptiste. 2017. *gridExtra: Miscellaneous Functions for "Grid" Graphics*. https://CRAN.R-project.org/package=gridExtra.

Cole, Matthew, Cameron Tulk, and Natasha Grzincic. 2017. "Toronto Is Now Majority Visible Minority. What about Your Neighbourhood?" https://www.thestar.com/news/gta/toronto-is-now-majority-visible-minority-what-about-your-neighbourhood/article_aee06d0f-bf38-5a81-a4b9-26acc63853f3.html.

Fradkin, Andrey, Elena Grewel, and David Holtz. 2018. "Reciprocity and Unveiling in Two-Sided Reputation Systems: Evidence from an Experiment on Airbnb." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2939064.

Ghosh, Indranil, Rabin K. Jana, and Mohammad Zoynul Abedin. 2023. "An Ensemble Machine Learning Framework for Airbnb Rental Price Modeling Without Using Amenity-Driven Features." *International Journal of Contemporary Hospitality Management*. https://doi.org/10.1108/IJCHM-05-2022-0562.

Guttentag, David. 2016. "Airbnb: Why Tourists Choose It and How They Use It." https://www.dg-research.com/airbnb-study.html.

Hlavac, Marek. 2022. *Stargazer: Well-Formatted Regression and Summary Statistics Tables*.

Bratislava, Slovakia: Social Policy Institute. https://CRAN.R-project.org/package=stargazer.

Laouénan, Morgane, and Roland Rathelot. 2022. "Can Information Reduce Ethnic Discrimination? Evidence from Airbnb." *American Economic Journal: Applied Economics* 14 (1): 107–32. https://doi.org/10.1257/app.20190188.

Lawani, Abdelaziz, Michael R. Reed, Tyler Mark, and Yuqing Zheng. 2019. "Reviews and Price on Online Platforms: Evidence from Sentiment Analysis of Airbnb Reviews in Boston." *Regional Science and Urban Economics*, 22–34. https://doi.org/10.1016/j.regsciurbeco.2018.11.003.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Neuwirth, Erich. 2022. *RColorBrewer: ColorBrewer Palettes*. https://CRAN.R-project.org/package=RColorBrewer.

Peng, Ningxin, Kangcheng Li, and Yiyuan Qin. 2020. "Leveraging Multi-Modality Data to Airbnb Price Prediction." *2020 2nd International Conference on Economic Management and Model Engineering*. https://ieeexplore.ieee.org/document/9434659.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Slowikowski, Kamil. 2024. *Ggrepel: Automatically Position Non-Overlapping Text Labels with 'Ggplot2'*. https://ggrepel.slowkow.com/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

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.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Wickham, Hadley, Evan Miller, and Danny Smith. 2023. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. https://CRAN.R-project.org/package=haven.

Wickham, Hadley, Thomas Lin Pedersen, and Dana Seidel. 2023. *Scales: Scale Functions for Visualization*. https://CRAN.R-project.org/package=scales.

William Revelle. 2024. *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois: Northwestern University. https://CRAN.R-project.org/package=psych.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. http://haozhu233.github.io/kableExtra/.