What affects the prices of Airbnb Listings in Toronto?

Ayina Anyachebelu, Natalia Joseph & Leila Ashtaryeh

FNCE 237, Professor Michael Roberts

December 3, 2020

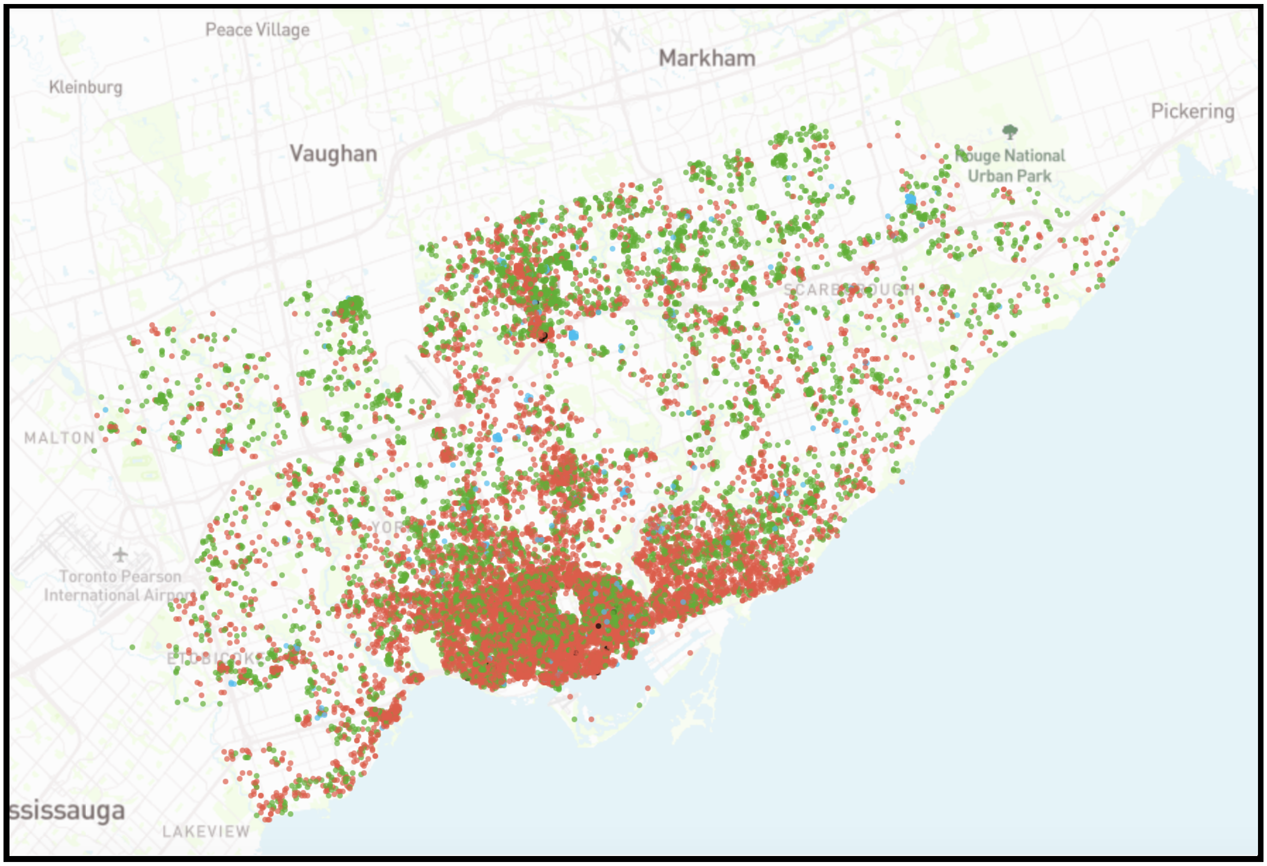


Table of Contents

Background

Toronto is the capital city of the Canadian province of Ontario. In 2016, the recorded census population of the proper city was 2.7 million people, with the greater Toronto area representing 6.4 million people. It is the most populous city in Canada and the fourth most populous city in North America. The city attracts over 43 million tourists annually, visiting to experience the popular art galleries, parks, urban multiculturalism, shopping, and entertainment. As a result of the large number of visitors to the city, Airbnb listings in Toronto are highly demanded, and the supply to meet such demand also exists. The large market for short-term rentals in Toronto is what drew our exploration to the city in addition to its tourist attractions and cultural amenities.

Moreover, the city boasts a thriving housing market. There are several neighborhoods in Toronto that tout exorbitant housing prices in addition to access to a variety of amenities; this data will allow us to deeply explore our research question. Each neighborhood in Toronto offers a different set of resources, such as nearby parks, restaurants, farmer's markets, museums, and other tourist attractions. We consider these to be external amenities when evaluating Airbnb listings. Meanwhile, amenities within the listing include line items such as the number of bathrooms, in-unit laundry machines, and other "perks" that certain hosts may offer to increase consumers' willingness to pay.

Airbnb listings for an entire home in Toronto make up 0.8% of the city's private dwellings (CBC, 2019). Experts believe that this percentage is "more than enough" to have a significant impact on housing prices in the city, and it does not even encompass the large number of condominium or apartment listings (CBC, 2019). In April of 2019, there were more than 9,500 Airbnb listings in Toronto; a large portion of these listings are located in Toronto's waterfront communities, representing one of the top five Airbnb locations in Canada. This is explained by the concentration of new condominium development and proximity to tourist attractions in the area.

Research Questions and Hypotheses

Our main research question is: What characteristics affect the prices of Airbnb listings in Toronto and to what extent? To answer this question, we have developed a set of 6 auxiliary research questions based on different groups of covariates that we expect to influence listing prices. Each area of investigation corresponds to a hypothesis about the effect of the corresponding characteristics on price; these are the alternative hypotheses. The null hypothesis is that these variables do not significantly influence price, meaning, in a multiple regression analysis, the coefficient estimates are not significantly different from zero.

1. How does geographic location affect listing price?

   The location of Airbnb listings is expected to have a significant effect on prices. In fact, we expect that geographic location will have the largest effect on prices in comparison to other categories because it is correlated with external amenities such as attractions. We hypothesize that neighborhoods near Toronto's central business district will command the highest prices for Airbnb listings. Specifically, the waterfront community is hypothesized to have the highest prices given its proximity to the downtown area as well as waterfront activities and views.

2. How does property type affect listing price?

   Property type is hypothesized to have a significant effect on price as well. We expect unique property types (ex: treehouse, cottage) will command higher listing prices. Moreover, it is hypothesized that certain property types are more valuable than others in certain neighborhoods, meaning the effect of property type on price may not be consistent across the investigation. For example, houses may be more valuable in neighborhoods near the central business district because the area tends to be more densely populated, so an entire spacious home is more rare.

3. How does listing capacity and amount of space affect listing price?

It is expected that listings of the entire property will command the highest prices compared to listings of private rooms and shared rooms (corresponding to the ‘room type’ variable). We use room type as a proxy for the amount of space available in the listing. Furthermore, we hypothesize that listing the entire space as opposed to a private or shared room will be more valuable when the property type is a house as opposed to an apartment. This is because we expect houses to have a larger total square footage compared to apartments. We also predict that private rooms will be more valuable if they are in a hotel (which includes bed and breakfasts). This is backed by our assumption that guests value privacy and safety from others very highly and, therefore, are willing to pay more. Guests are also highly familiar with the amenities that are associated with typical hotel stays, potentially making them willing to pay more to stay in a hotel listing.

4. How do a host’s characteristics affect listing price?

   There are a number of variables that speak to a host’s characteristics, including greeting, verification, response rate, tenure, and superhost status. Greetings, tenure, and superhost status indicate a higher quality of service with the listing and, as such, we expect them to positively impact the price of listings. On the other hand, host verification is not expected to significantly impact price because unverified hosts likely advertise similar prices to verified hosts, all else equal. In addition, we hypothesize that hosts with lower response rates have higher prices because they are less likely to actively monitor the Airbnb market, which would prompt them to adjust prices to remain competitive.

5. How do booking constraints affect listing prices?

   Hosts have the option to enable instant bookings from guests, meaning guests do not need to converse with the host prior to making the reservation. It is anticipated that hosts with instantly bookable listings use this feature as a tactic to lock in customers whose listing search is driven by price; as such, we expect a negative correlation between price and instantly bookable listings. Hosts are also able to set a minimum length of stay for guests.

Similarly, we expect listings with longer minimum stay restrictions to be offered at a lower price as a result of economies of scale used to incentivize long-term guests.

6. How do amenities affect listing prices?

   There are 20 different types of internal amenities included in our investigation (listed within the data dictionary in Appendix II). We hypothesize that having any of these amenities while controlling for others should command a higher price. We expect the value of outdoor space, balconies, and access to nature, and good views to be more pronounced in neighborhoods near the central business district because of the dense population. We also expect that a hot tub/sauna/pool or barbeque would be more valuable if the listing has outdoor space. Additionally, we hypothesize that breakfast amenities will have a larger impact on price if cooking basics are not provided and if the property type is a hotel. Gyms are expected to have a larger effect on price if the property type is a house because other property types (apartment, hotel) are more likely to have gyms already, so guests in those types of listings would see this as an expectation.

Methodology

Our data was sourced from *Insideairbnb*, an organization that scrapes Airbnb listings from various global cities. We downloaded our dataset on the 29th of September 2020, and it contains the 19,339 live listings on the site at that time. Additionally, we used a GeoJSON file from the City of Toronto's Open Data Portal that includes geospatial data for the 140 neighbourhoods in Toronto. This data will be merged with our Airbnb dataset for geographical analysis and mapping using python's geopandas package. One limitation of our data is that we only have access to the asking-price and not the actual prices that guests paid to stay at the property. Inexperienced property owners can set prices that are very low or high compared to the average nightly prices paid by prior guests. In reality, guests could have paid much less than the listing price after negotiating with the host. This limits our research because we do not obtain an accurate sense of how guests actually value the various listings. However, we will be able to obtain an overall sense of what prices properties are listed for.

Before delving into the exploration of characteristics influencing price, we will examine the overall Airbnb market in Toronto for background information and an understanding of the dataset. Through this preliminary data exploration, we will investigate how listing prices and reviews have changed over time in Toronto's Airbnb market.

To examine the relationship between our 6 categories of interest and price, we will use a hedonic pricing model. The hedonic price theory assumes that a commodity, such as a house, can be viewed as an aggregation of individual components or attributes (Griliches, 1971). It assumes that consumers buy products whose components or attributes maximize their utility functions (Rosen, 1974). The hedonic model is most commonly applied to the housing market, which explains its relevance to this study where we hope to use several attributes to predict Airbnb listing prices. The model involves regressing prices for properties against the characteristics hypothesized to be determinants of the price. We will use a multiple regression analysis to create a hedonic pricing model with the log of listing price as our dependent variable in order to determine the effect of various pertaining to our research questions.

As previously mentioned, our research is divided into 6 categories that will be included in our model: geographic location, property type, listing capacity and space, host characteristics, booking constraints, and amenities. To explore the influence of geographic location, we will use dummy variables for the 140 neighbourhoods in Toronto. For the property type group, we will separate listings into apartments, houses, and "other type" categories where "other type" includes unique listings such as tree houses or cabins. In order to explore the effect of capacity and amount of space on price, we will use the number of possible people accommodated by the listing as well as room type (entire space, private room, or shared room). To investigate host characteristics, we look at variables including the host's verification status, the host's superhost status, the host's response rate, the amount of days the host has been registered on Airbnb, and if the host greets the guest upon arrival. In order to explore booking constraints, we will use the variables for 'instantly bookable' listings as well as the minimum nights required to stay at the listing. Finally, under amenities, we will explore the presence of 20 variables: air conditioning, balcony or patio, bed linen, television, cooking basics (such as kitchenette, refrigerator, or microwave), major appliances (such as washer, dryer, or dishwasher), gym, parking, outdoor space, internet, private entrance, fire safety, a barbeque, a hot tub/sauna/pool, breakfast, proximity to nature and good views, a fireplace, and laptop-friendly workspace.

Preliminary Data Exploration

To obtain a background of the Airbnb market in Toronto, we begin our analysis with a preliminary exploration of prices, time series data, and review data. This process will ultimately help in understanding our data and the final predictive model.

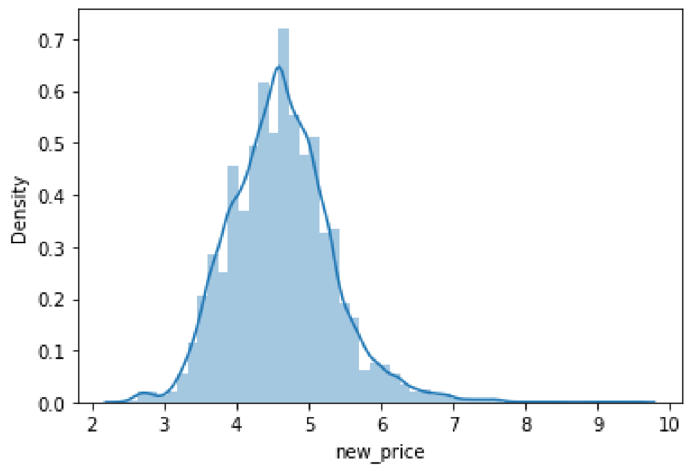
Price Distribution

The price of Airbnb listings in Toronto is our dependent variable in this analysis. As such, our preliminary data analysis will begin by looking at the properties and summary statistics of price in the dataset. The summary statistics below indicate a wide range of prices listed for Airbnbs in Toronto. As a highly populated city in North America, high prices are expected. The minimum price, on the other hand, is surprising; twelve dollars for a one-night stay at any property in Toronto is unexpected. This might be attributed to very low-cost accommodation such as hostels. Furthermore, comparing the mean price of $141.27 to the median price of $100.00 indicates that that distribution of prices is right-skewed. The mean is much higher than the median price because of high-priced outliers in the data, also seen by the exorbitant maximum price per night. Based on this skew, examining the distribution of the log price is more valuable than the un-transformed price. Once the log price is used, the distribution of Airbnb prices appears to be relatively normal.

Summary statistics:

Maximum price per night	$13,164.00
Minimum price per night	$12.00
Mean price per night	$141.27
Median price per night	$100.00

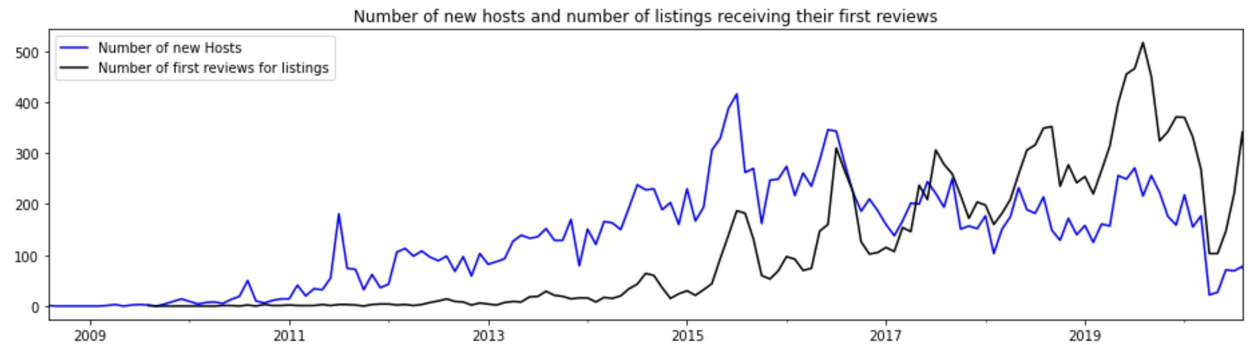
Log price distribution:



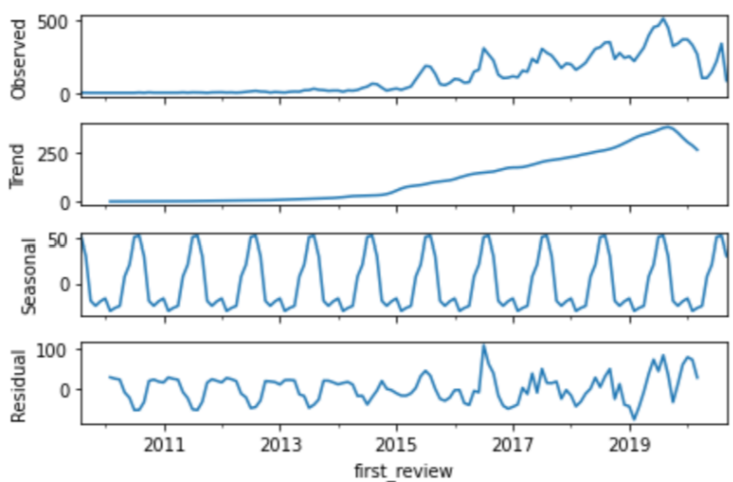
Time Series

The dataset provides information on Airbnb listings starting on August 8, 2008. The most recent listing date is September 7, 2020. Following, there are reviews provided from August 20, 2009 to September 10, 2020. This gives a large dataset with an adequate number of observations to conduct a meaningful analysis of the factors determining price.

Over time, the Airbnb market in Toronto has grown. This is demonstrated by the increasing number of cumulative Airbnb hosts and number of new hosts from 2009 to 2020. Additionally, the number of first reviews has increased; this indicates that a property has hosted its first guest, which gives valuable insight into demand for new listings. It is important to note that Airbnb requires all guests and hosts to leave reviews. The following time series plot illustrates this growth. A significant drop in the number of new hosts and first reviews for listings is seen in 2020, which we attribute to the coronavirus pandemic, which has significantly impacted the travel and hospitality industries as a whole.

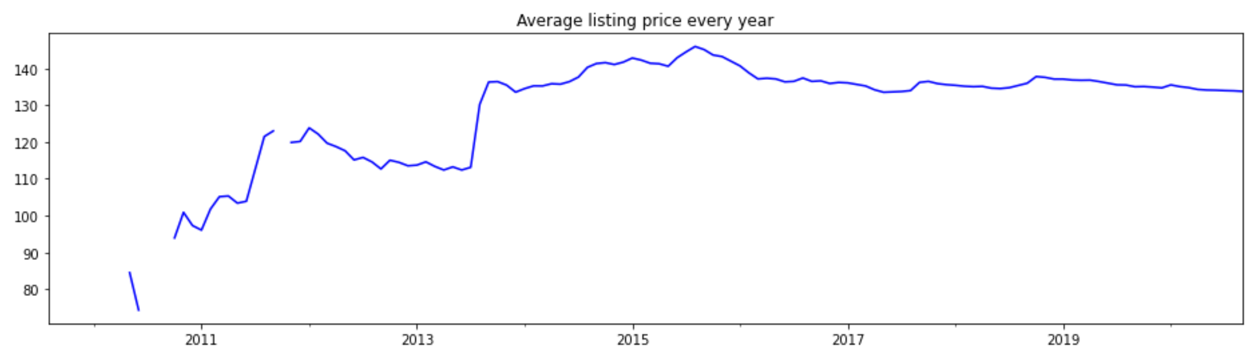


Looking further into the first review variable through the decomposition plots below, the plot corroborates the growing trend in Airbnb reviews from our first analysis and also shows an element of seasonality. This tells us that listings tend to receive their first reviews on a seasonal basis, which likely coincides with when travelling to Toronto is most likely.



Furthermore, we explore how prices have changed over time. The following series of boxplots shows the change in price of listings receiving their first review over time. Looking only at the price of listings receiving their first review each year gives an accurate picture of the price of Airbnbs when they are first listed, which is particularly important given that hosts can update their listing prices at any time. Again, the natural log transformation of price is used to improve the interpretability of the graph. Surprisingly, we see little-to-no increase in price over time in terms of the median listing price per night. Nonetheless, the number of outlying points that demonstrate higher-priced listings increases over time, with the exception of 2020, which is likely due to the impacts of the coronavirus pandemic.

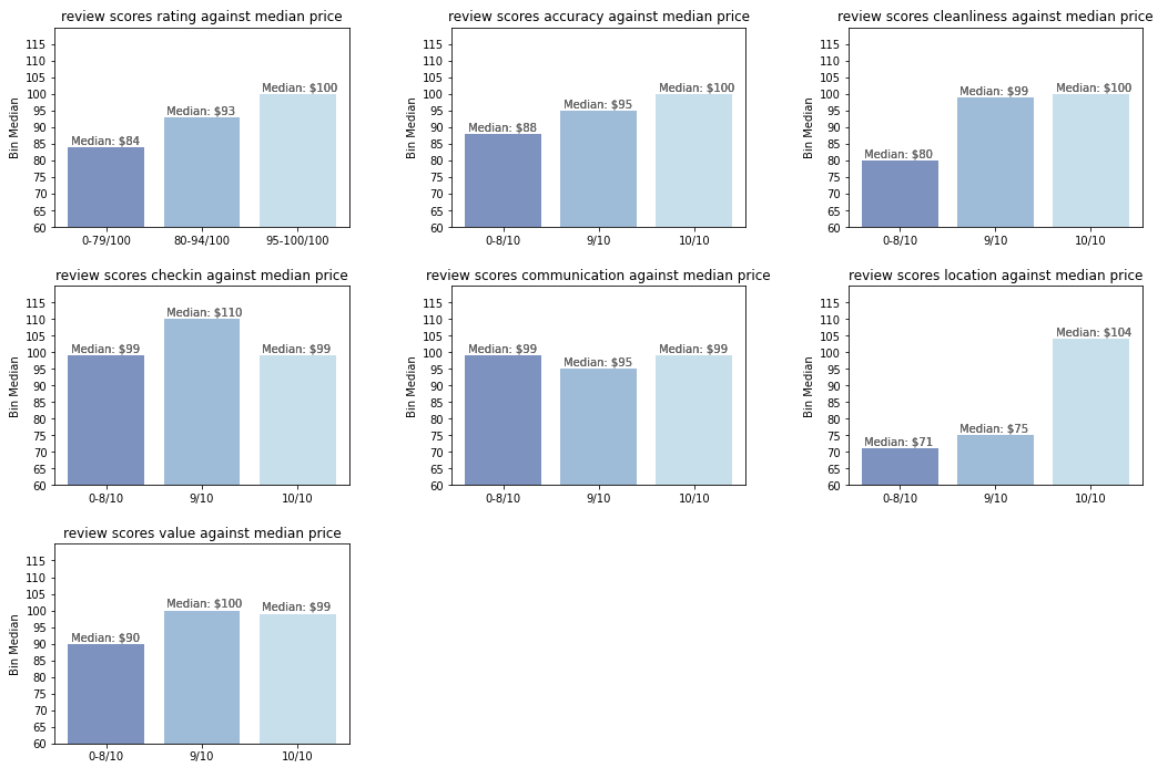
Subsequently, the plot below illustrates the average listing price over time. There is a steep increase in average price in mid-2013 and average price appears to be relatively stable thereafter.



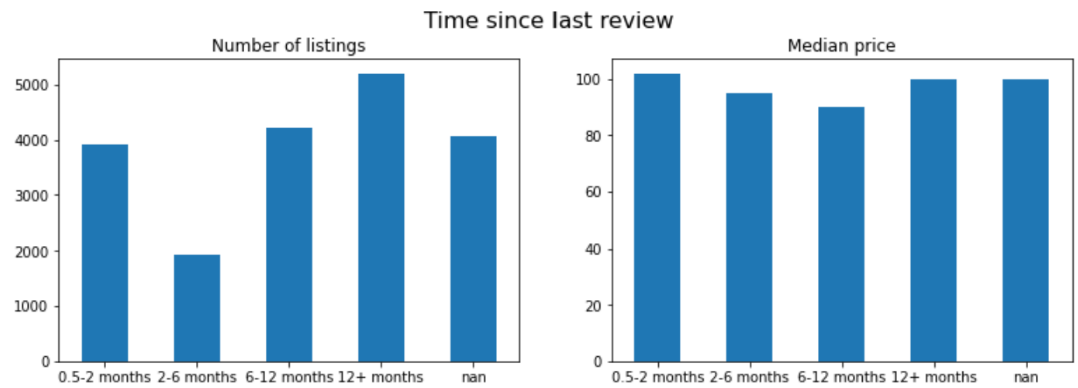
Reviews

Host and guest reviews are an important part of Airbnb's platform. Both the host and the guest are required to leave reviews in order to better inform future users. As such, there is review data on the overall rating, accuracy, cleanliness, check-in experience, communication, location, and value for money. To examine the potential relationships between review data and listing price, a series of bar graphs were used. It was found that review scores - which are values between one and ten or a percentage value - are highly concentrated at the high end of the spectrum. As such, the review data scores are bucketed into three bins for each review score.

The bar graphs below show the median prices for each rating bin. Higher scores do not always correspond with higher median prices. We attribute these anomalies to guests perceiving very little difference between a score of 9/10 compared to 10/10 for some of these rating categories. In addition, the price differentials between the two score buckets are relatively small in most cases. The largest difference in prices by review score is for the location category whereby the difference between a 9/10 and a 10/10 review score equates to a $29 difference in median price. Because each review characteristic has similar price distribution, we will only include the review scores rating, which is the overall rating, as a control to limit collinearity.



Finally, we examined the time since the last review seen in the table below. We noticed that about 20% of observations were missing, which we believe are listings that have not yet been reviewed. Only about 30% of the listings had been reviewed during the last 6 months. In terms of the price difference, median price does not appear to be greatly different depending on the time since the last review. However, the difference was statistically significant. For this reason, we include time since last review as a control in our model.

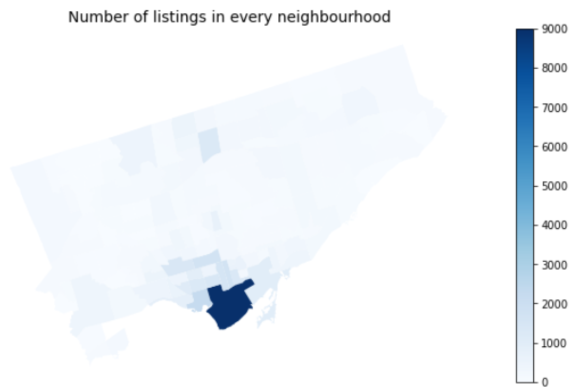


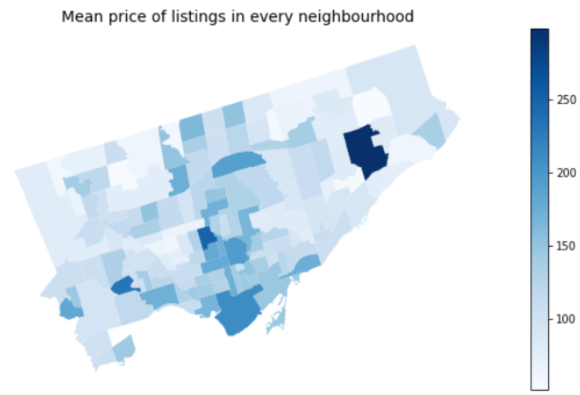
Exploration of the 6 Categories of Interest

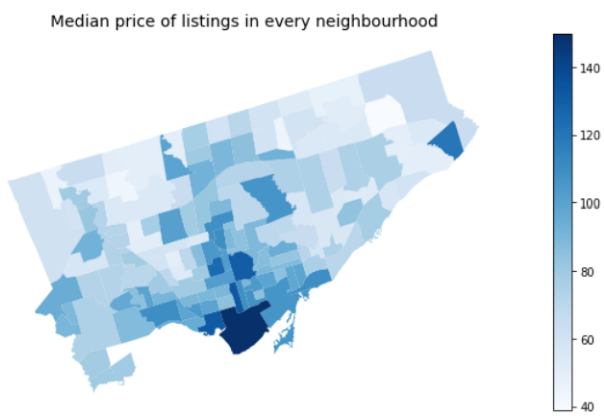
In this section, we will explore the six categories of interest outlined in our hypothesis. This includes geographic locations and neighborhoods, property type, listing capacity and amount of space (room type), host characteristics, booking constraints, and internal amenities.

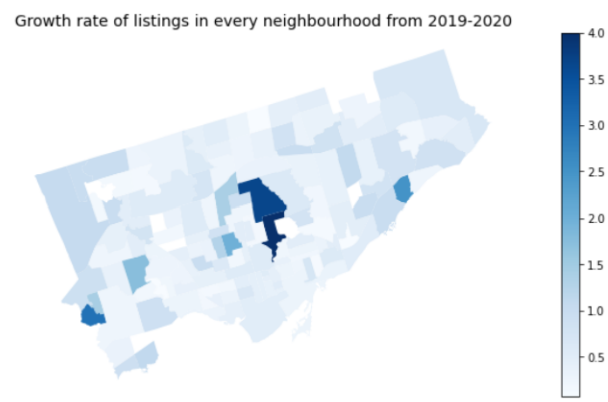
Geographical Location

For our geographic analysis, we investigated how the neighbourhood in which the listing was located affected price. The results are displayed visually in the 4 heat maps below. In the first map, it is clear that the neighborhoods with the greatest number of listings are concentrated in the Waterfront Communities and downtown area of Toronto, which is closest to the financial district and lake shore. In the second map, where the mean listing price is shown, the most expensive neighborhoods are more surprising, with the highest mean price located in Woburn (north east of the city center), instead of in the general downtown area. The median price map is closer to what was expected, with the highest prices being at the Waterfront Communities and closest to the downtown core, where many tourist attractions are located. The last map shows the growth rate in the total number of listings by neighborhood from 2019 to 2020. It appears as though up-and-coming, less saturated neighbourhoods had the most growth, especially Leaside-Bennington neighborhood, which received 4 new first-time reviews in 2020.



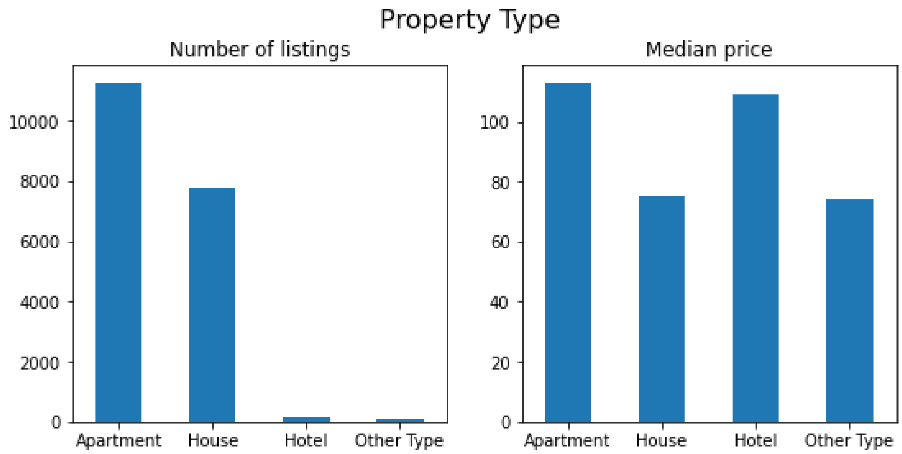






Property Type

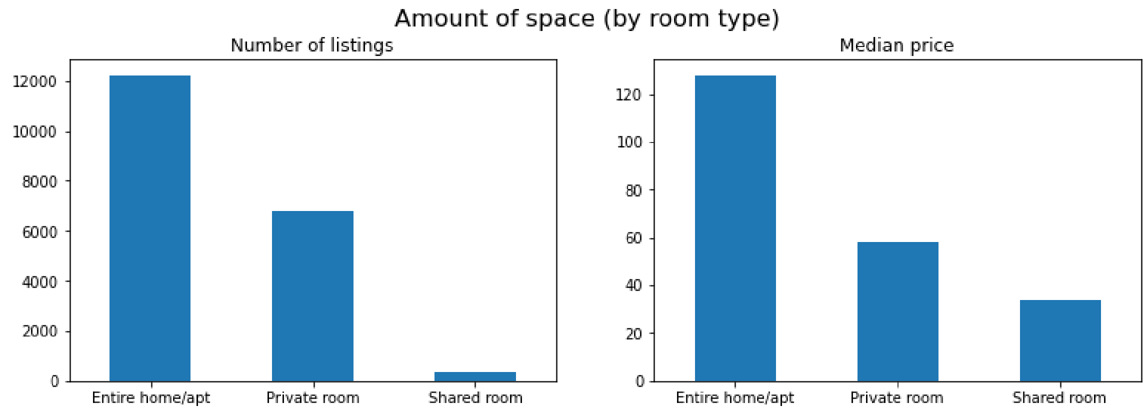
We also find that the type of listing influences the median price. In the following bar charts, the number of listings and median price are distributed by property type. With Toronto being a highly populated urban metropolis, it is not surprising that the majority of Airbnb listings are apartments. Additionally, apartments have the highest median price of all the property types, only slightly greater than hotels. Hotels listed on Airbnb are not mass-market chain hotels; they are boutique or lifestyle hotels or bed and breakfasts with the appropriate business licenses, differentiating this listing type from others on Airbnb. It is possible that these hotels are amongst the higher-priced property types because hosts are legally responsible for business licenses and property management.



Additionally, when we examine this data, there are signs that there may be a relationship between the type of listing and the listing's location. For example, it is more common for apartment listings to be in certain neighborhoods that are more densely populated, whereas houses are more commonly located in residential neighborhoods outside the city center. This indicates that we may want to use an interaction term between property type and neighborhood in our final model to account for this relationship. Given that the most expensive neighborhoods by median price are centralized in the downtown area and the median price of apartments is highest (see bar chart below), we extrapolate that an interaction between the two variables is present. This is further explored in the findings section from our hedonic pricing model.

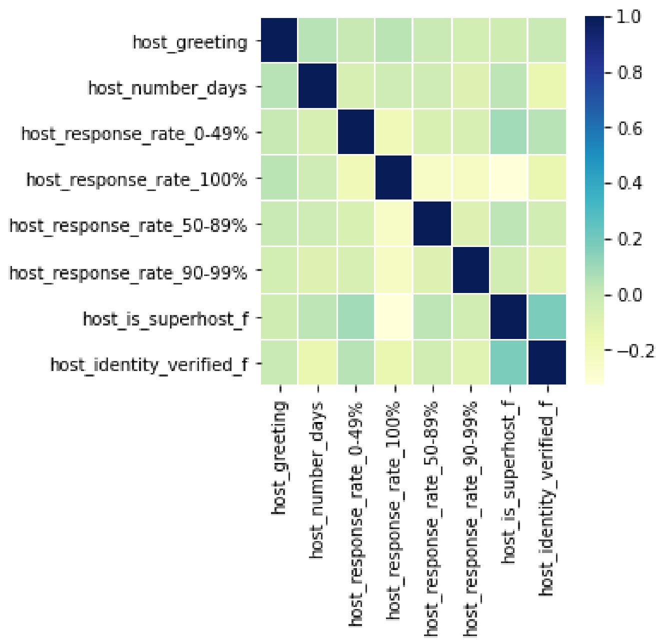
Listing Capacity and Space

In our study, we use the room type as a proxy for the amount of space available in each listing (ie. square footage). In Toronto, most listings are advertised as the entire home or apartment. This is evidently more attractive to guests as we assume privacy and safety are important concerns for guests. This logic follows with the number of private rooms and shared rooms listed as well as their corresponding median prices. As expected, it is clear that an entire home or apartment commands the highest price relative to a private room or shared room. Interestingly, there is also the highest number of listings for an entire home or apartment compared to private and shared rooms. Given the high variance in number of listings for each room type, it will be important to evaluate the significance of our findings for the impact of room type on price in our final model.

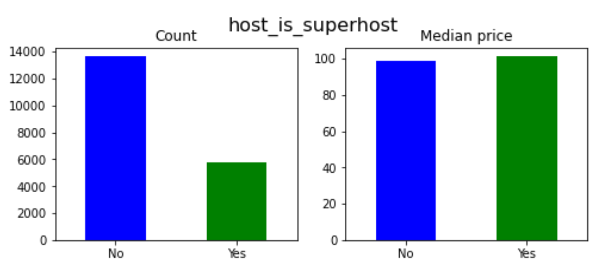


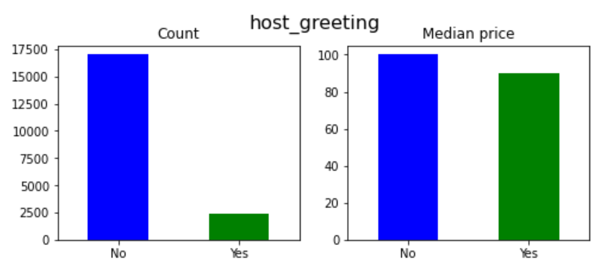
<u>Host Characteristics</u>

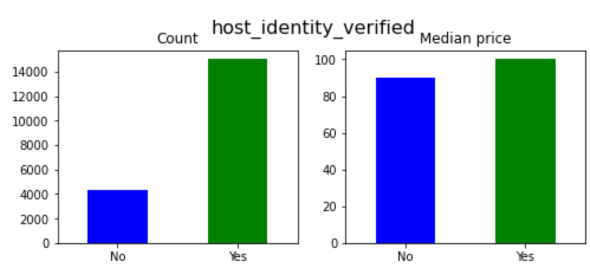
In the Airbnb dataset, there are almost 20,000 hosts with a minimum tenure of 22 days and maximum tenure of 4435 days (equivalent to around 12 years). The mean number of days being a host is 1626 days (around 4.5 years). We will use the host data, which is comprised of 5 unique variables, to examine the effects of host characteristics on Airbnb listing price. There is little correlation between these variables themselves. Nonetheless, the indicator variables for superhosts and host verification appear to be slightly correlated. This is expected based on our understanding of Airbnb; superhosts meet a set of requirements demonstrating their success with a significant number of guests. It follows that these hosts are likely to be verified, meaning they have provided a series of identification documents to confirm their identity. Given the success and track-record of superhosts, it makes sense that these hosts are also verified.



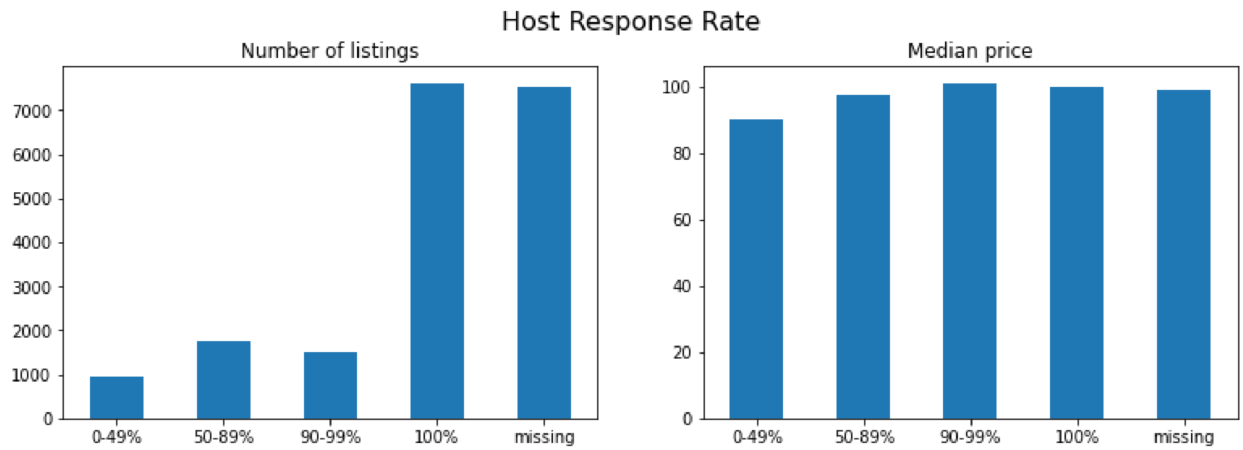
The following bar charts demonstrate the relationships between host characteristics and median listing price. They also indicate the likelihood of hosts being superhosts, including a host greeting, and verifying their identity, shown by the listing counts. By looking at the graphs, we see only small differences in median price for the host characteristics, possibly indicating that host characteristics might not influence price as greatly as we predicted. This is not sufficient to make an initial analysis of their significance in impacting price. They will be further examined in the final regression model.



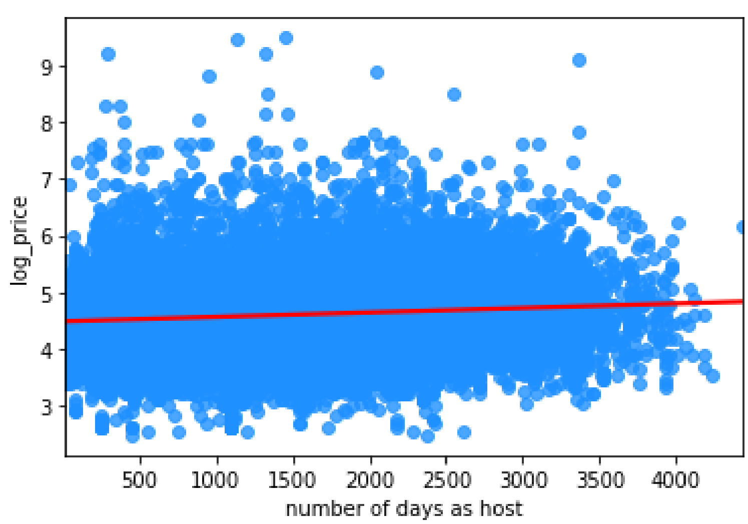




The bar charts on host response rates below show that while most hosts' response rates are missing (because a guest probably has not reached out to them yet), of the host's who do have a response rate, they tend to have a response rate of 100%. Similarly, the data on host response rates show very little difference in median price. The most apparent difference is for response rates below 50%.

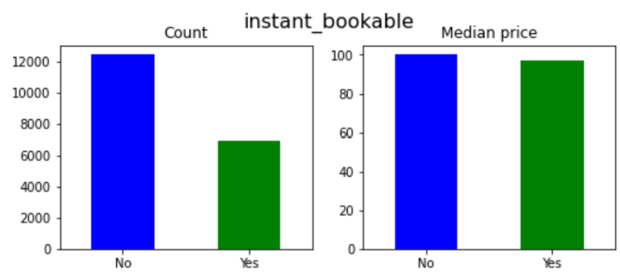


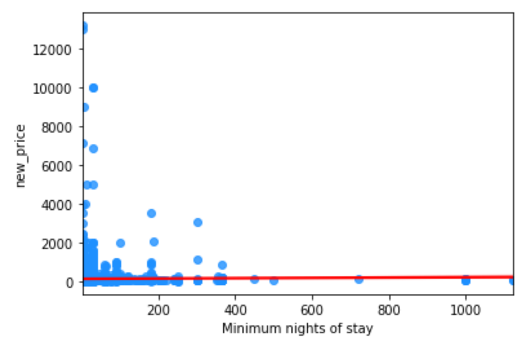
Lastly, the scatter plot below shows the relationship between the number of days a host has been using Airbnb and the log price of their listings. The fit line in the graph shows a slight increase in log price as a host's tenure increases. When the raw price is inputted into our regression model, we can now expect there to be a positive relationship between number of days as a host and listing price.



Booking Constraints

There are two potential booking constraints that hosts may choose to use when setting up their listings. This includes the ability for the listing to be instantly booked and the ability to set a minimum number of nights for the booking. We initially expected both of these variables to be negatively correlated with price. From our preliminary analysis, the median price for instantly bookable listings is $97, compared to standard listings with a standard booking process and a median listing price of $100. This confirms our hypothesis, but its significance is yet to be determined.



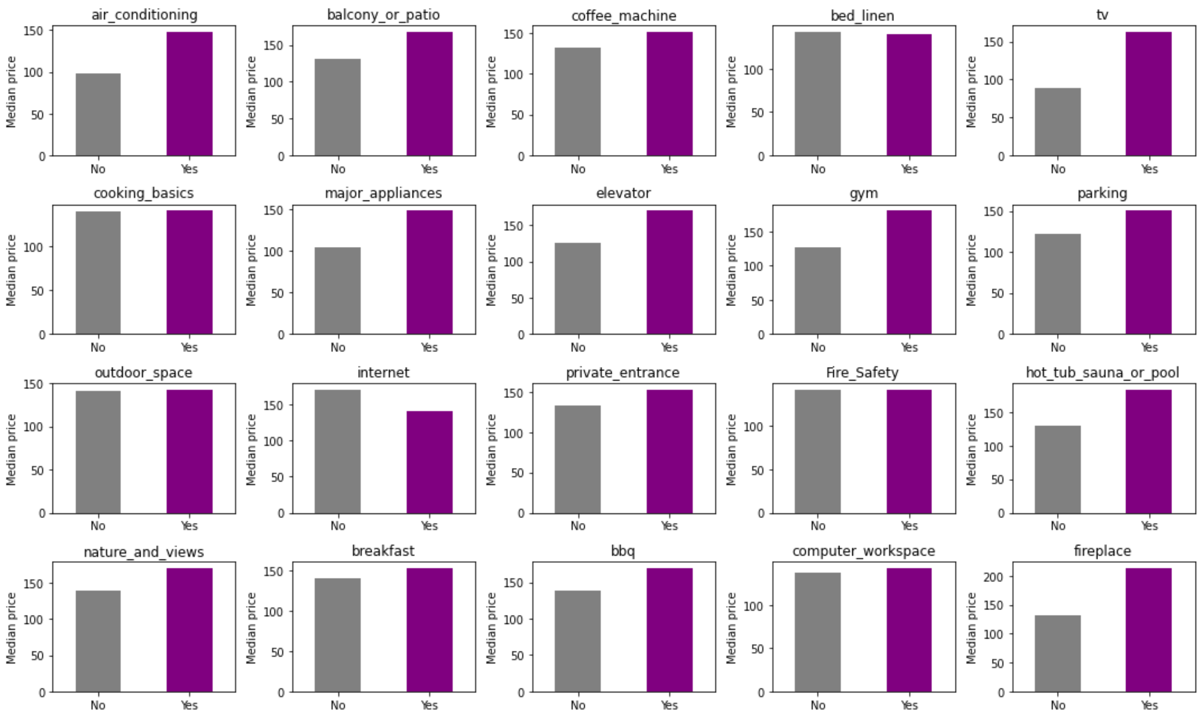


For listings with a minimum night stay, the scatterplot above does not visibly reveal a relationship between minimum nights and price though the relationship appears to be very mildly positive.

Amenities

We looked at 20 internal amenities in relation to the median listing price. These amenities are binary variables, meaning that a listing either has the amenity or does not. From these bar plots, it is difficult to determine whether or not the amenity has a significant impact on listing price. Nonetheless, notable differences in price are present given the addition of a balcony, a gym, major appliances, TV, hot tub/pool, a fireplace, and air conditioning. For these amenities, there is a positive relationship between their presence in the listing and the median listing price. Conversely, there appears to be a negative relationship between internet access and median price. This is unexpected, but further examination may reveal this is a result of confounding factors.

Additionally, after examining the correlation between the amenities, we decided to drop the coffee_machine and elevator variables from our model in order to decrease collinearity.



Findings from Hedonic Pricing Model

Our hedonic pricing models include a regression pertaining to the 6 categories of interest as well as review variables for control. The regression model including interaction terms had an adjusted R-squared of 0.585, meaning it explains 58.5% of the variability of the log of listing price around its mean. Furthermore, this model has an intercept of 3.51 which is the natural log of price for a shared room in an Apartment in Alderwood, where the host is a superhost, is verified, and offers instant booking.

<u>How does geographic location affect listing price?</u>

As we hypothesized, geographic location has a significant effect on the listing prices for Airbnbs. Table 1 shows the coefficients of the neighbourhoods with the five highest and three lowest coefficients after controlling for all the other categories of interest as well as review variables. The coefficients are relative to Alderwood, the neighbourhood that was dropped from the model. Similar to our data exploration findings, The Waterfront Communities neighborhood commands the highest prices. Specifically, having a listing in the Waterfront instead of Alderwood increased the price by 53.84%. Bay Street, University, and Church-Yonge, all neighbourhoods near downtown Toronto, had the next most expensive listings, followed by Casa Loma, one of the most touristic neighbourhoods in the city. On the other hand, Forest Hill South, Mount Olive Silverstone, Keelesdale-Eglinton West, Taylor Massey, and Clanton Park, which are all lower-income communities within the city, had listings with the lowest prices relative to Alderwood. It should be noted that while these communities had the lowest coefficients, their p-values were not statistically significant.

Table 1

Neighbourhood	**Coefficient**	**SE**
Waterfront Communities	0.5384	0.092
Bay Street	0.4965	0.099
University	0.4607	0.114
Church-Yonge	0.4426	0.099
Casa Loma	0.4016	0.105
Forest Hill South	-0.2866	0.220
Mount Olive Silverstone	-0.2363	0.127
Keelesdale-Eglinton West	-0.1861	0.112
Taylor Massey	-0.1617	0.116
Clanton Park	-0.1557	0.102
const	3.5108	0.099

How does property type affect listing price?

The coefficients and corresponding standard errors for variables pertaining to property type are found in Table 2. Only hotels commanded a log price that was significantly greater than that of apartments, with a 35.17% price premium. Contrary to our predictions, unique listings falling in the “other type” property category did not have the highest prices. While the other type category has a negative coefficient, it is not significant. The unique property type prices are likely being balanced out by high-end resorts and inexpensive campers and cabins placed in this category.

We hypothesized that certain property types would be more valuable in certain neighbourhoods. Specifically, we predicted that a house would be more valuable in downton neighbourhoods because this area is more densely populated and less spacious in general. We tested this hypothesis by interacting the house variable with 5 major downton neighbourhoods whose coefficients and standard errors can be found in Table 2. While only Bathurst Manor and Annex neighborhoods produced significantly different coefficients, their results corroborated our predictions. The premium in Bathurst Manor was especially large. Conditional on being located in Bathurst Manor, having a house over an apartment on average increases the price by 47.68%.

Table 2

Property Type	Coefficient	SE
Hotel	0.3517	0.165
House	-0.0204	0.014
Other Type	-0.0583	0.060
House*Yonge_St.Clair	0.1881	0.148
House*Yonge_Church	0.0011	0.067
House*Annex	0.1235	0.041
House*Uni	-0.0391	0.069
House*Bathurst	0.4768	2.907
const	3.5108	0.099

<u>How does listing capacity and amount of space affect listing price?</u>

Table 3 shows the coefficients and corresponding standard errors for the capacity and room type variables. The categorical room type variables are relative to having a shared room. As we predicted, the number of people accommodated by the listing is positively correlated with the price. For every extra person the listing is able to accommodate, there is a 13.49% price premium. Additionally, having a private room or the entire place as opposed to a shared room comes with a price premium of 45.54% and 82.94% respectively.

We hypothesized that the value of the amount of space could depend on the property type. Specifically, we believed that listing the entire space as opposed to a specific room will be more valuable when the property type is a House or unique listing as opposed to an apartment. Based on the coefficients of our interaction terms, our predictions were correct because, conditional on having the entire place, having a house or "other" property type instead of an apartment brought about a 13.54% or 34.95% price premium respectively. These results are reasonable because it is more valuable to have an entire house or entire boat versus having an entire apartment. Finally, we also hypothesized that a private room would be more valuable in a hotel which turned out to be the case given the statistically significantly positive coefficient (0.34) of our interaction term.

Table 3

Capacity/Room type	Coefficient	SE
Accommodates	0.1349	0.002
Private Room	0.4554	0.027
Entire Place	0.8294	0.028
private_room*Hotel	0.3401	0.168
Entire*House	0.1354	0.017
Entire*Other_type	0.3495	0.145
const	3.5108	0.099

<u>How do a host's characteristics affect listing price?</u>

Table 4 shows the coefficients and corresponding standard errors of the host's characteristics variables. Of our 6 categories of interest, the host's characteristics is one of the categories with the least influence on price as seen by its coefficients that are comparatively small in magnitude. Consistent with our predictions, the host's greeting and the host's tenure were both significantly positively correlated with log price. However, the magnitude of the coefficient of the number of days as a host was minute at 0.000021. Also consistent with our predictions was the fact that being a verified host did not have a significant effect on price. Although we expected that being a superhost would command higher prices, a host that was not a superhost did not have prices that were significantly different from those of a superhost. In fact, the mean log price of non-superhosts was higher than that of the superhosts. Finally, our last variable is the host response rate is relative to an unknown response rate (a host who may not have been messaged yet). We predicted that host's with lower response rates might have higher prices because they are not frequently active on the site, so they do not keep their prices competitive. While this might have been partly true since hosts with a response rate of 0-49% commanded the highest prices relative to hosts with no response rate, hosts with a 100% response rate commanded the next highest prices. Surprisingly, hosts with a response rate of 90-99% on average had listings that were significantly set at lower prices than hosts without response rates, and hosts with response rates between 50 - 89% were not significantly different from hosts without response rates.

Table 4

Host Characteristic	Coefficient	SE
host_greeting	0.0277	0.011
host_number_days	0.000021	0.00000429
host_response_rate_0-49%	0.0767	0.016
host_response_rate_100%	0.0332	0.009
host_response_rate_50-89%	0.0117	0.013
host_response_rate_90-99%	-0.0315	0.015
host_is_superhost_f	0.0025	0.009
host_identity_verified_f	0.0191	0.009
const	3.5108	0.099

How do booking constraints affect listing prices?

Table 5 shows the coefficients and corresponding standard errors of the variables pertaining to booking constraints. Consistent with our predictions, the minimum nights of a stay has a negative relationship with price. Although this coefficient is significant, it is small in magnitude, as an extra minimum night only commands a 0.04% premium. We had hypothesized that hosts that permitted instant bookings would have lower prices. Nevertheless, listings that were not instantly bookable actually had a lower mean, but this difference is not statistically significant. Overall, like host's characteristics, booking constraints were one of the categories with the least influence on price.

Table 5

Constraint	Coefficient	SE
minimum_nights	-0.0004	9.07E-05
instant_bookable_f	-0.0082	0.007
const	3.5108	0.099

How do amenities affect listing prices?

Table 6 shows the coefficients and corresponding standard errors of 18 of our initial 20 amenities because two of them were dropped to reduce collinearity.

Table 6

Amenity	Coefficient	SE
air_conditioning	0.0881	0.011
balcony_or_patio	0.0413	0.009
bed_linen	-0.0245	0.008
tv	0.1251	0.009
cooking_basics	-0.1692	0.017
major_appliances	0.0145	0.01
gym	0.0711	0.011
parking	0.0646	0.008
outdoor_space	0.0079	0.013
internet	-0.1416	0.024
private_entrance	-0.0065	0.01
Fire_Safety	-0.0827	0.016
hot_tub_sauna_or_pool	0.0072	0.012
nature_and_views	-0.0202	0.015
breakfast	-0.008	0.051
bbq	0.0278	0.021
computer_workspace	-0.0271	0.008
fireplace	0.1444	0.011
const	3.5108	0.099

We had predicted that all amenities would bring about a price premium, but our model revealed otherwise. Firstly, we discovered that listings with major appliances, outdoor space, private entrances, hot tubs/saunas/pools, barbeques, nature/views, or breakfast did not have significantly different prices from those that did not have these amenities. Furthermore, our model shows that listings with provided bed linen, cooking basics, internet, fire safety and laptop-friendly workspaces had significantly lower prices than those that did not have these amenities. Of the amenities that commanded significantly higher prices, fireplaces had the highest coefficient, having a 14.44% premium on average. This was followed by having a television which had a 12.51% price premium.

Looking more deeply into the listings whose presence of amenities had significantly lower prices, we found this peculiarity usually occurred either because the listing was of a unique property type or the host did not feel the need to mention the amenity since it was implicit in the type of listing. For example, listings without a laptop-friendly workplace amenity were mostly entire houses/villas/apartments. The host most likely did not feel the need to report a workspace since renters would have the entire property. At the same time, special listings without computer workspaces, such as the boat with a $900 nightly rate, might not have this amenity. However, these unique listings can charge higher rates due to their scarcity. Another type of special listing without key amenities that were still set at similar prices to those with the amenities such as fire safety or internet include camper/RVs and tiny homes (similar to fancy cabins usually in nature). Guests looking for these special types of listings are likely to expect the absence of fire safety. Moreover, they usually stay at these properties in order to get away from society so lack of access to the internet would not be seen as a disadvantage and might actually be a plus.

Furthermore, listings with private entrances may have commanded lower prices because they are not advertised for listings where the guest is given the entire place because it is assumed that the host has a private entrance. Private entrances are most likely reported for listings where the host is still on the premises. These correlations to other variables are likely reasons for the insignificant coefficient.

Notably, most of the most expensive listings that did not have cooking basics were rooms in boutique hotels. These types of properties still command high prices because they traditionally lack this type of amenity but have typical hotel amenities that are common knowledge to renters.

Table 7 shows the coefficients and corresponding standard errors for interaction terms pertaining to amenities based on our hypotheses.

Table 7

Amenity Interactions	Coefficient	SE
outdoor*baystreet	-0.2769	0.123
outdoor*bathurst	-0.4258	0.182
outdoor*other_type	0.0197	0.141
outdoor*bbq	0.016	0.027
outdoor*hottub_pool	0.0985	0.032
nature/views*yonge-church	0.5457	0.207
balcony_or_patio*church-yonge	-0.1278	0.042
house*gym	0.0877	0.033
breakfast*hotel	-0.4695	0.085
breakfast*cooking	0.0503	0.052
const	3.5108	0.099

We expected the value of outdoor space, balconies, and access to nature and good views to be more pronounced in neighborhoods near the central business district because of the densely populated space. On the contrary, we found out that conditional on having outdoor space, being in the Bay Street or Bathurst Manor neighbourhoods which are near downtown brought about a statistically significant price discount. Surprisingly, conditional on being located in the downtown neighbourhood of Church-Yonge, having a balcony or patio commanded a statistically significant lower price, but having proximity to nature or good views had a statistically significant price premium (54.57%) in comparison to not having the amenity. The price premium due to the nature and views amenity may be largely driven by the sought-after views from high rise buildings in the city center.

While the interaction between having outdoor space and a barbeque was not significant, as we predicted, conditional on having outdoor space, having a hot tub/sauna/pool increases the price by 9.85% relative to not having the amenity. Consistent with our hypothesis, having a gym conditional on the property type being a house brought about a significant price premium.

Interestingly, conditional on being in a hotel, offering breakfast decreased the price of the listing by 46.95% compared to not offering breakfast. We discovered that this was a result of the fact that most hotels that offer breakfast that are listed on Airbnb are actually “bed and breakfasts”. These are small lodging properties usually in people’s houses that are not very expensive which corroborates the price discount shown in our model.

Conclusion

The principal aim of this paper was to explore the relationships between prices and different attributes of Airbnb listings in Toronto in order to answer our main research question: what characteristics affect the prices of Airbnb listings in Toronto and to what extent? To answer this question, we developed a set of 6 auxiliary questions based on different groups of variables that we believed would influence listing prices, namely geographic location, property type, capacity and space, host’s characteristics, booking constraints, and amenities. The hedonic pricing model was used to determine the effect of these variables on the price. Our model had an intercept of 3.51, which is the log price of a shared room in an Apartment in Alderwood, where the host is a superhost, is verified, and offers instant booking. The hedonic pricing model found that some aspects of all 6 categories had statistically significant effects on price.

Overall, geographic location and listings capacity (space, room type) were the most influential covariates on price, while host characteristics and booking constraints were least influential to listing price. Based on the model, it appears that guests greatly value the neighbourhood where their listing is located, especially the Waterfront Communities and other downtown neighbourhoods. Some of the interaction terms of the model reveal that some specific property types, amount of space, and amenities are more valuable conditional on the neighbourhood in which the listing is located. Accordingly, geographical amenities and venues might be more valuable than built-in amenities since many of the interior amenities did not significantly impact price positively when controlling for all other variables.

Nevertheless, it is difficult for the model to predict the importance of these built-in amenities since it is the host’s responsibility to report them. Some of these amenities may be implicit in the property type. While the presence of these amenities may be apparent to the people booking these listings, they will not be apparent in the model. One signal that implicit amenities may be affecting price is the fact that many hotel property types that may not have listed many amenities command higher prices. Moreover, some interior amenities are only significant when paired with other amenities such as a hot tub/pool and outdoor space.

With regards to capacity and space and property type, as predicted, the listings with higher capacity that offered the entire space as opposed to a single room were the most expensive. While hotels commanded the highest prices, surprisingly, unique listings (other property type category) did not significantly affect prices. It appears that our "other type" classification had too many varied property types for it to significantly influence price. As said above, host's characteristics had little effect on price, demonstrating that it might not be worth it for the host to become a superhost and hosts do not have a significant advantage the longer they have been on Airbnb. Finally, booking constraints either did not have a significant effect on price or produced a very small discount possibly as a trade off for the constraints.

Given the importance of location, examining different types of geographical amenities and venues surrounding Airbnb listings could be a promising avenue for further research . For example, using a Google API to find the distance of listings from certain attractions or transit would provide insight into these amenities' impact on price. Including this spatial analysis could create a stronger model to understand the relationship between attractions and price as well assist in predicting prices of new airbnb listings.

Data Sources

Inside Airbnb (2020). *Adding Data to The Debate.* http://insideairbnb.com/get-the-data.html

Toronto Open Data Portal (October 19, 2020). *Boundaries of City of Toronto Neighborhoods.* Social Development, Finance & Administration. https://open.toronto.ca/dataset/neighbourhoods/

Toronto Open Data Portal (October 19, 2020). *City Wards.* City Clerk's Office. https://open.toronto.ca/dataset/city-wards/

References

Nicole Brockbank (May 1, 2019). *Airbnbs have 'serious impact' on Toronto housing despite making up less than 1% of dwellings, expert says.* CBC News. https://www.cbc.ca/news/canada/toronto/airbnb-listing-data-toronto-1.5116941

Griliches, Zvi (2013). *Price Indexes and Quality Change: Studies in New Methods of Measurement*. Harvard University Press. www.hup.harvard.edu/catalog.php?isbn=9780674592582.

Rosen, Sherwin (1974). "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy*. https://econpapers.repec.org/article/ucpjpolec/v_3a82_3ay_3a1974_3ai_3a1_3ap_3a34-55.htm

Google Colaboratory

https://colab.research.google.com/drive/15WucZw2fdtyuEmiqbtrIlquVFOXwsNNa?usp=sharing

APPENDIX I: Data Dictionary

Variable Name	Description	Number of NA Values
host_since	Date that the host first joined Airbnb	0
host_response_rate	Percentage of messages to which the host responds	0
host_is_superhost	Whether or not the host is a "superhost," classified by Airbnb	0
host_listings_count	Number of listings that the host owns/manages	0
host_identity_verified	If the host has been verified with government ID	0
neighborhood_cleansed	Name of the neighborhood in Toronto where the listing is located	0
property_type	Type of property: house, apartment, hotel, or other type	0
room_type	Type of room within the listing ie. entire home, private room, shared room	0

accommodates	Number of people the listing accommodates	0
bath_num	Number of bathrooms available within the listing	47
bedrooms	Number of bedrooms available within the listing	1429
beds	Number of beds available	196
amenities	Uncleaned list of amenities	0
new_price	Nightly price to book listing (dependent variable)	0
minimum_nights	Minimum length of stay defined by host	0
maximum_nights	Maximum length of stay defined by host	0
availability_365	Number of nights available to be booked in the next 365 days	0
number_of_reviews	Number of reviews left by customers for listing	0
first_review	Date of the first review made for the listing	4064
last_review	Date of the most recent review	4064
time_since_first_review	Number of days since the first review	0
time_since_last_review	Number of days since the last review	0
review_scores_rating	Overall property score from 1 to 5 stars, averaged from individual reviews	0
review_scores_accuracy	Accuracy of a property's description from 1 to 5 stars, averaged from individual reviews	0
review_scores_cleanliness	Review score of the property's cleanliness from 1 to 5 stars	0
review_scores_checkin	Review score for check-in experience from 1 to 5 stars	0
review_scores_communication	Review score for the host's communication from 1 to 5 stars	0
review_scores_location	Review score for the property's location from 1 to 5 stars	0
review_scores_value	Review score for the perceived value of the property for the customer's money from 1 to 5 stars	0
instant_bookable	Whether or not the property can be instantly booked on Airbnb	0
host_number_days	Number of days since the host joined Airbnb	0

** The ‘amenities’ variable has been split into a number of dummy variables (float64) to describe the amenities available at each listing. These amenity variable names include: air_conditioning, balcony, bed_linen, tv, coffee_machine, cooking_basics, major_appliances, elevator, gym, parking, outdoor_space, host_greeting, hot_tub_sauna_or_pool, internet, long_term_stays, private_entrance, fire_safety, and essentials.