How to make a killing on Airbnb in New York City

- The causal relationship between Airbnb rental amenities & features and its business performance in the five boroughs of NYC

Bi Wu
Akshita Gupta

Final paper for Advanced Test, Analysis and Experimental Design

School of Professional Studies
New York University

December 13th, 2018

Abstract:

New York City is one of the prime tourist spots in the United States. Being a prominent city of the country, it often gets people from all around the world to come and experience the magic that it has to offer. While the hospitality industry has greatly benefitted from the city, a new peer-to-peer business model came into existence as the internet grew stronger. Now anyone with an empty house or room has been enabled to become a property renter and earn a few extra dollars by renting out their property for substantially lower prices than the big hoteliers. With this new model, many got into the business of temporarily renting out their spare spaces to visitors and has now led to extreme competition in the space. While it seems easy, sustaining your listing on Airbnb can be somewhat challenging since the business operates on a strict review system. If subsequent guests give the host a low rating, Airbnb can go so far as to blacklist the host. In an environment where fake lists are becoming common each day, it is extremely important for hosts to maintain their scores and be top hosts on Airbnb in order to continue getting more guests. Through this paper, we will be looking at potential areas that affect the overall business of Airbnb and areas that affect the status of a host to understand more clearly which areas hosts can work on to provide a valuable experience to guests.

Introduction:

Airbnb has been growing exponentially in the “lodging” business. The company has been giving big hotel chains like Hilton, Marriott and the like a run for their money, forcing these giants to step up their efforts to sustain in the market. Airbnb brought about the disruption in the hotel, or rather, lodging business which was long due, to end the monopoly that the hotel chains seemed to have had. With Airbnb’s model, anybody and everybody that had even a spare room could now be a hotelier.

While the business model has stood the test of time and has been gaining popularity, it is rather difficult for room owners or “renters” to consistently make money, and a huge component leading to this is the reviews. Since the company relies on individual renters to provide stay to customers, there is no consistency in experience and quality of stay that they can rely on (a major disadvantage to the business and a major advantage for hotel chains that stand for consistency in quality and experience).

New York City is one of the major markets for the company and understanding how the business can further prosper in the 5 boroughs is of high importance to the company. If you have a spare place in NYC, you probably have thought about getting into the Airbnb business. You can rent out a room for $2000 per month but if you were to rent it on Airbnb, you can charge $150 per night. Sounds lucrative, we know.

However, not every host on Airbnb is that fortunate. You can't charge as much as you want because there are too many competitors out there and there will always be cold nights your listing doesn't get any tenants. To make your listing profitable, there are lots of factors to take into consideration.

This is the reason why we wanted to study the data to help solve or at least understand this problem. We want to study how you can improve your listing on Airbnb and become a top host on the site. We wish to give recommendations on how to boost your listing value, improve your ratings, and become a superhost, which will not only help the list owner or "renter" but also the company at large. We also aim to find out the profitability based on ratings and reviews and group listings into boroughs to compare what has been working for one and how others can learn from these best practices and thus identify the business opportunities that Airbnb can explore from the housing market in New York City.

Bjørkelund, Burnett, & Nørvåg state that tourism and hospitality are "particularly information- and service-oriented industry" (as cited in Luo 2018, p. 6). They further argue that "the field involves customer-based service in a context where multiple factors, such as noise, nearby construction, weather, and even customer expectations, may impact customer experience and the evaluation of services"
(as cited in Luo 2018, p. 6).

In an industry that is growing and crumbling at the same time, our objective for this study is to come up with a solution for the company as well as the individual property renters known as "hosts", listed on Airbnb, to help grow business and generate more revenue for all involved. This study will also help related industries such as space rentals that can leverage these findings and boost their own businesses with our insights.

Hypothesis Development:

Our overall business objective is to find ways in which the company and the renters can maximize their dollar value and the quantitative and qualitative factors that will primarily enable us to come up with those recommendations are:

- Price
- Review score ratings
- Is Superhost or not

Through this study, we aim to examine the effect of other factors on these factors and vice versa to determine their relationship. The causal relationships determined by us are:

Causal relationship 1:

Objective qualities like Location, Room Type, Accommodates and room facilities have a direct and positive effect on the Price of the listing.

Through this causal relationship, we state that, although the price is a human factor, which is completely decided by the host, we want to find out if objective factors including Location, Apartment Type, Room Type, Accommodates and room facilities have a statistically significant impact on a listing's market price. We took different statistical approaches to input these four factors with the dependent variable "price" in R:

1) *oneway.test(price ~ neighbourhood_group_cleansed , data=list_1, var.equal=FALSE)*
2) *t.test(Private_rooms$price, Shared_rooms$price, var.equal = FALSE)*
3) *scatterplot(price~accommodates, data=list_1)*
4) *m1 <- lm(price~bathrooms+bedrooms+beds+as.factor(bed_type), data=list_1)*

Causal relationship 2:

An increase in review score ratings accuracy, cleanliness, check-in, communication, location, and value will lead to an increase in overall review score ratings.

Reasons: guests are asked to rate their overall experience, cleanliness, accuracy, value, communication, arrival and location of the room; Because overall rating is a reflection of all the other ratings combines, we think if other ratings are high, a guest will rate his/her overall experience high too.

This is represented in R as:

```
m_1 <- lm(review_scores_rating~review_scores_accuracy+review_scores_cleanliness+
review_scores_accuracy+review_scores_checkin
        +review_scores_communication+review_scores_location+review_scores_value,
data=list_1)
```

Causal relationship 3:

An increase in listing counts, listing price, minimum nights of stay, availability, number of reviews, review score ratings and reviews per month will increase the likelihood of a host becoming a superhost.

Superhosts are Airbnb renters that are all stars on Airbnb. They are people who have aced the art of providing a fantastic experience to customers and often stand out as the gold standards for other renters to look up to. We used logistic regression to test this hypothesis in R:

```
logistic_reg <-glm(host_is_superhost~
                neighbourhood_group_cleansed+room_type+instant_bookable
               +cancellation_policy, data=list_1, family="binomial")
```

To test the validity of these causal relationship guesses, we developed the following hypotheses:

Data and Variable Selection:

Our dataset obtained from the Airbnb data site [2] is the data for the five boroughs of New York, namely, Manhattan, Brooklyn, Queens, Staten Island, and Bronx. In this dataset, we were able to get the following variables:

1) Listing ID (listing_id) 2) Host ID (host_id) 3) Host Name (host_name) 4) Host response time (host_response_time) 5) Host Response Rate (host_response_rate) 6) Host is Superhost (host_is_superhost) 7) Host Listing Count (host_listings_count) 8) Neighbourhood Group Cleansed (neighbourhood_group_cleansed) 9)Zip Code (zip code) 10) Latitude (latitude) 11) Longitude (longitude) 12) Property Type (property_type) 13) Room Type (room_type) 14) Accommodates (accommodates) 15) Bathrooms (bathrooms) 16) Bedrooms (bedrooms) 17) Beds (beds) 18) Bed Type (bed_type) 19) Price (price) 20) Cleaning Fees (cleaning_fee) 21) Minimum nights to be stayed at the listing (minimum_nights) 22) Availability 30 (availability_30) 23) Number of reviews so far (number_of_reviews) 24) Review Score for rating (review_scores_rating) 25) Review Score for Accuracy (review_scores_accuracy) 26) Review score for cleanliness (review_scores_cleanliness) 27) Review score for check-in ease (review_scores_checkin) 28) Review score for communication - given by renter to host (review_scores_communication) 29) Review score for location of the listing (review_scores_location) 30) Review score for value - of the listing and overall experience (review_scores_value) 31) Instantly bookable (instant_bookable) 31) Cancellation Policy (cancellation_policy) 32) Number of reviews per month - indicative of how many times the listed was booked by renters (reviews_per_month).

The main units or variables that we will be studying for our analysis and find factors affecting the outcome of are:

- Price
- Review Score Ratings
- Host is Superhost

For each of the causal relationship assumption, we will be examining the following variables:

Causal relationship 1: Objective qualities like Location, Room Type, Accommodates have a direct and positive effect on the Price of the listing.

We will be studying the effects of the neighborhood the listing is in (neighbourhood_group_cleansed), type of room (room_type), and accommodates

(accommodates)on the price of the listing (price). we only study the effects of them on the price one at a time and not involving any other factors.

Causal relationship 2: An increase in review score ratings accuracy, cleanliness, check-in, communication, location, and value will lead to an increase in overall review score ratings.

For this assumption, we will be studying the effects of the review score given for accuracy (reiview_score_accuracy), review scores given for cleanliness (review_score_cleanliness), review scores that are given for ease of check-in (review_score_checkin), review scores given for communication (review_score_communication), review scores given for location of the listing (review_score_location), review scores given for value provided for the overall experience for the amount charged for the listing (review_score_value) on the overall review score received by the host for the listing (review_scores_rating).

While scores are given for the overall stay, there are individual categories that the host and the listing receive scores for. We wanted to see if a particular category of scores has a more significant impact in driving the overall score. For instance, while the check-in process might have been tedious and thus the host received lower ratings in that category, but the location was a prime location, which one of these factors would have a more significant impact on the overall score the listing and host receives.

Causal relationship 3: An increase in listing counts, listing price, minimum nights of stay, availability, number of reviews, review score ratings and reviews per month will increase the likelihood of a host becoming a superhost.

To test the validity of this assumption, we will be looking at the effects of the host's total listing count (host_listing_count), price of the listing (price), minimum nights that can be booked for the listing (minimum_nights), availability for 30 days (availabilty_30), total number of reviews the listing has received (number_of_reviews), overall review scores for the listing (review_scores_rating), number of reviews the listing receives per month which is indicative of the number of times the listing was rented out in a month (reviews_per_month) on the status of the host being a superhost or not (host_is_super_host).

Here, we wanted to study the effect of specific review and score related characteristics to study the effects of the status of the host. This is the reason why these specific factors have been considered.

EDA:

We performed exploratory data analysis on the data variables to come up with basic characteristics and number for our dataset.

Some of the Univariate Analysis were:

- Spread analysis
- Boxplot analysis
- Summary Statistics.

The overall summary statistics for our dataset can be seen in Table 1.
We looked at which borough had the maximum number of listings and the type of listing (room_type) which is shown in Figure 1 (a) and (b). The total number of listings in each borough can be seen in Figure 3. Manhattan has the highest number of listings at 23,052 listings and Staten Island has the least number of listings at 341 listings.

Figure 1 (a).

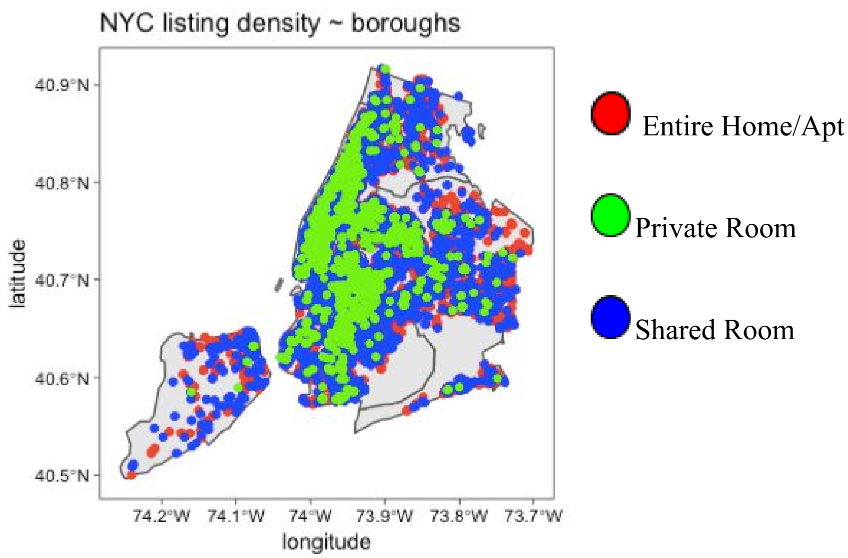
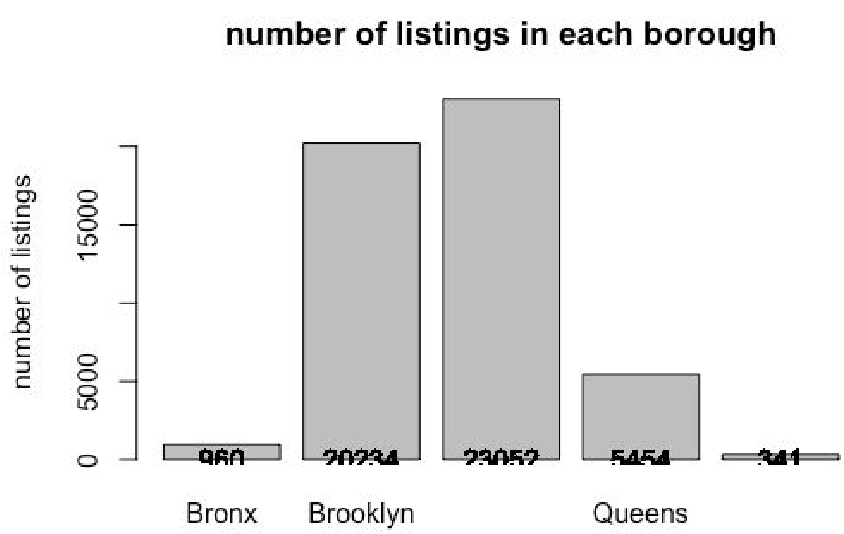
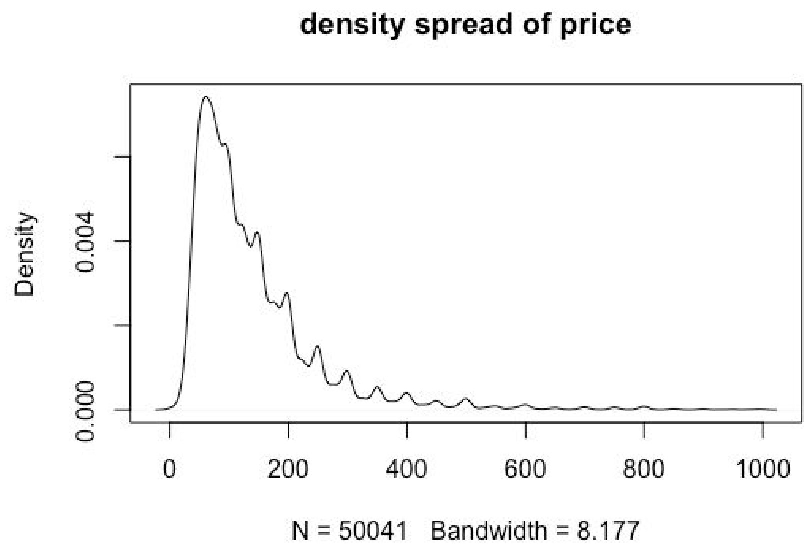


Figure 1(b)



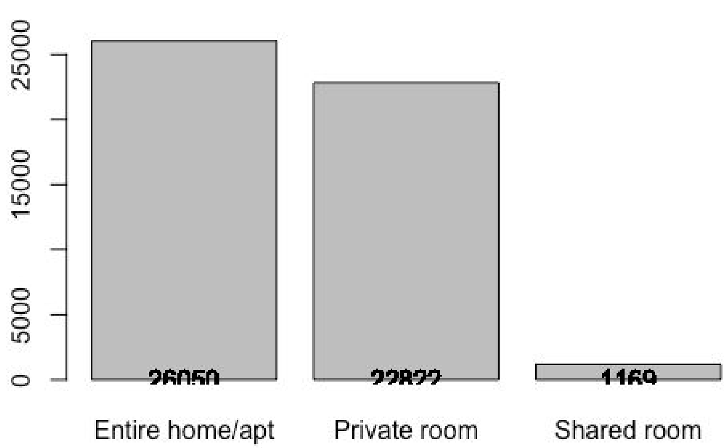
To look at the most common price range, a density spread of price was performed which is seen in Figure 2.

Figure 2:



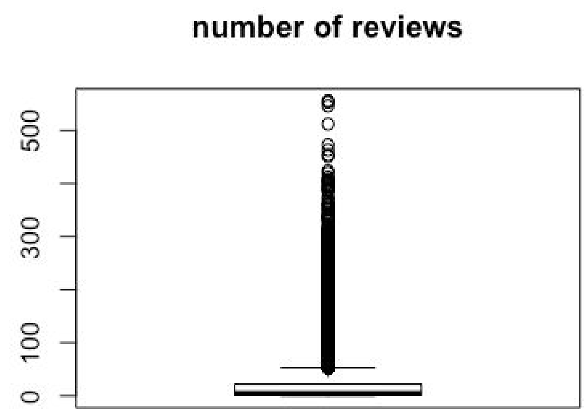
Here, we can see that the most listings in New York were between the $0 to $200 range. We assume that this is well in keeping with the general hotel prices of the city per night and in most cases, even lower.
For our analysis, we also looked which type of listings were most common for the New York region and saw that listings which were the entire apartments for renting were the highest in number across the five boroughs at 26,050 listings on Airbnb. The count can be seen in Figure 4.

Figure 4:



We created a boxplot for the total number of reviews of listings to find out that the number of reviews has a mean of 21 and it's positively skewed with a heavy tail towards the right. This can be seen in Figure 5.

Figure 5:



To study the overall review score ratings, we conducted the box plot analysis and a density spread and saw that the review ratings are skewed negative and with a heavy tail on the left (see Figure 6 (a)). Most review scores are between 80 to 100 and two spikes are seen on scores 99 and 95 which can be seen in Figure 6 (b).

Figure 6 (a)

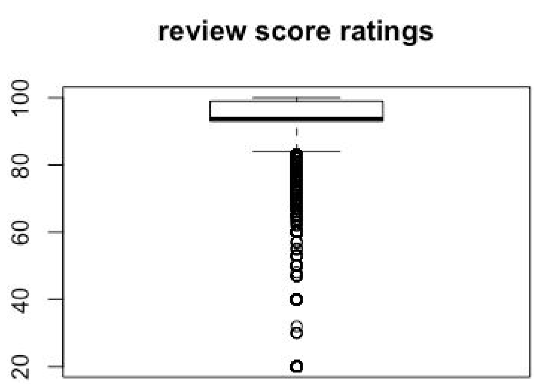
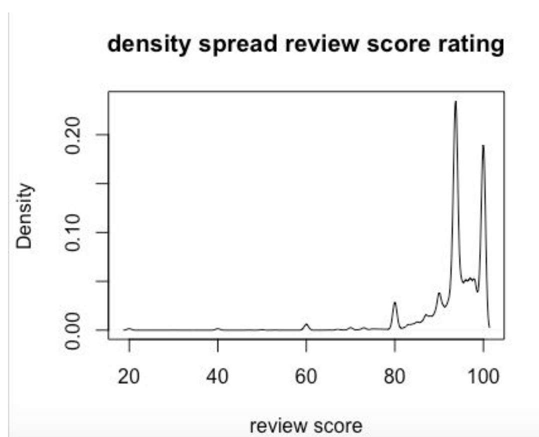


Figure 6 (b)

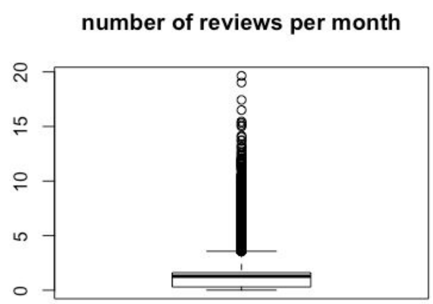


We performed an analysis on the number of reviews per month per listing. The results for the summary statistics can be seen in Table 1. The results of the number of reviews per listing can give a rough idea about how many unique bookings a listing receives per month. Considering that there are 30-31 days a month and that not all people who stay in an Airbnb actually rate the listing and the host, it is acceptable to get low numbers for this analysis. A box plot analysis of the same was conducted and the results can be seen in Figure 7.

Table 1:

Minimum	1st Quarter	Median	Mean	3rd Quarter	Maximum
0.010	0.300	1.280	1.402	1.610	19.640

Figure 7:



Multivariate tests helped us to test two variables against each other to support our hypothesis. We analyzed the price of listings with boroughs to see which borough has listings at a higher price. Manhattan is seen to have the highest average price. The least expensive listings are in the Bronx. According to Wachsmuth, while listings in Manhattan have a higher revenue per listing, Brooklyn has a faster growth rate (Wachsmuth, D. (2018). The high cost of short-term rentals in New York City. McGill University.). The results can be seen in Figure 9 (a) and Figure 9 (b).

Figure 9 (a)

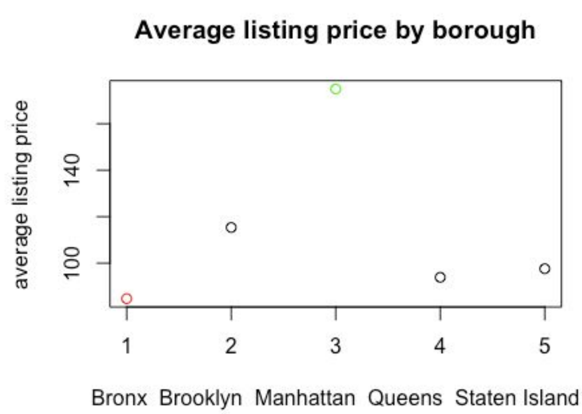
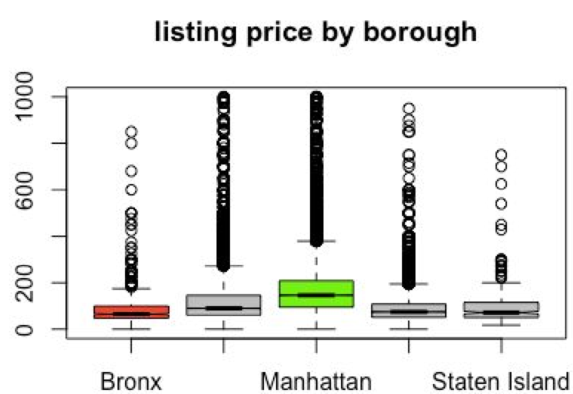
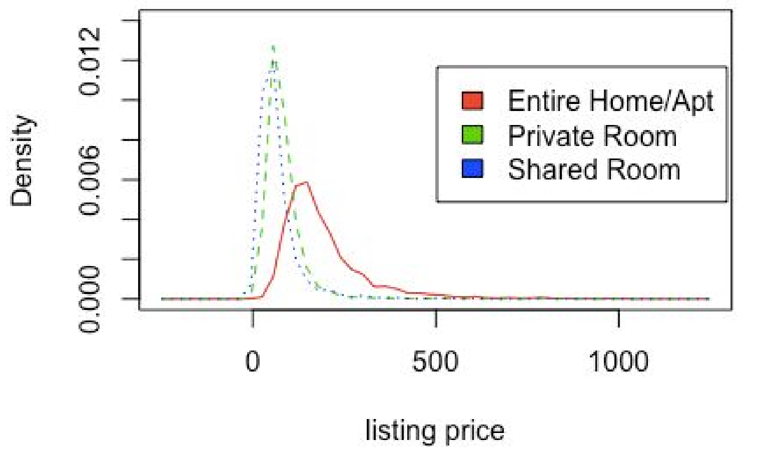


Figure 9 (b)



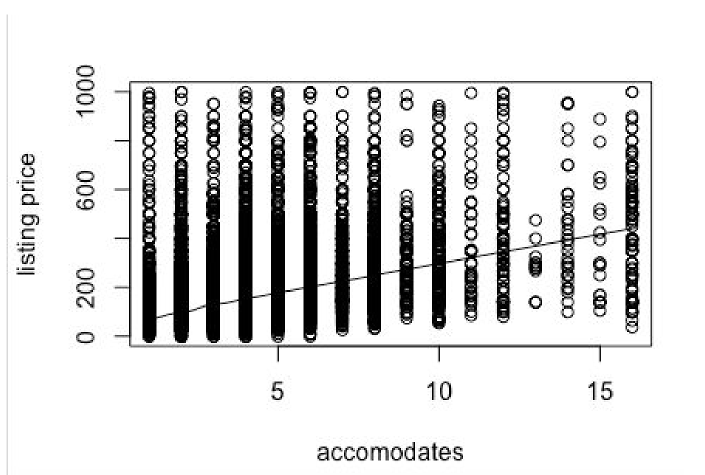
By analyzing listing price by room type, we found that entire apartment had the highest price per listing compared to other listing types. This can be said since the host rents out the entire unit to one renter and thus more is charged. The results can be seen in Figure 10.

Figure 10:



We also studied the effects of how many people the listing accommodates to see its effects on price for the listing. The causal relationship seen through our analysis states that as the number of “accommodates” increases, the price for the listing increases. This is also seen in the traditional hotel industry and since Airbnb is a new take on the traditional hospitality industry, it can be said that it still follows some of the rules of the former. It can also be assumed that the reason behind this could be that the listings that have the ability to accommodate more renters are bigger spaces and that is the reason why the price of the listings increase. The results of this analysis can be seen in Figure 11.

Figure 11:



To get the count for the number of superhosts by the boroughs, we performed cross tabulation and see that while Manhattan had the highest price per listing (Figure 9 (a) and 9 (b)), Brooklyn has the highest number of superhosts. To depict if the host has the status of a superhost or not, the dataset assigned the values “t” if the host is a superhost and “f” if the host is not a superhost. From the cross tabulation, we can see that the number of “t” values are highest for listings in Brooklyn while they are lowest in Staten Island. The results can be seen in Table 2.

Table 2:

Count (t or f)	Brooklyn	Queens	Manhattan	Bronx	Staten Island
	4	3	2	0	0
f	16891	4421	19952	774	251
t	3339	1030	3098	186	90

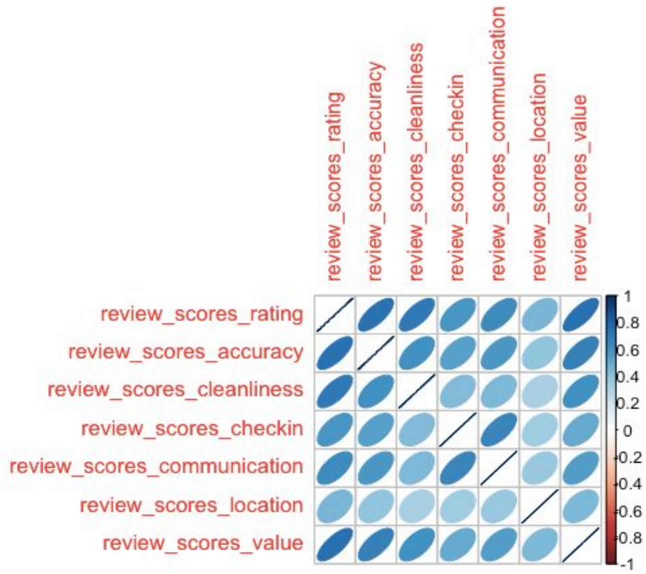
To see the number of superhosts per room type, a cross-tabulation was done. In the dataset, the value “t” is assigned if the host is a superhost and “f” if the host is not a superhost. From the cross tabulation, it can be seen that the number of superhosts for listings that were entire apartment category were the highest and were least for listings that fall under the shared room category. It can be assumed that this is because of the number of listings under the entire room category are highest which have skewed the results for this analysis. The result can be seen in Table 3.

Table 3:

Count (t/f)	Entire Apartment	Private Room	Shared Room
	5	4	0
t	4063	3505	175
f	21982	19313	994

To find if individual ratings for accuracy, cleanliness, check-in, and communication had a statistically significant effect on the overall ratings that the host received, we did a correlation matrix (See Figure 12) and found that none of the individual ratings had an impact at 0.90. However, the review score ratings and the ratings for accuracy were the most correlated at 0.73 which could be considered a good correlation for factors in the hospitality industry.

Figure 12:



Empirical Methodology:

Here, we will be discussing models we used for each of our hypothesis. All tests were conducted in R Studio, here we will be discussing the mathematical equations for each of those tests.

Hypothesis 1(1):

H1(1): There is a statistically significant difference between listing prices in different boroughs

Model:

one-way fixed effects ANOVA model. This is where we have a single 'treatment' factor (= borough) with several levels (5 boroughs = 5 levels) and replicated observations at each level. We are interested in comparing the means of the observations between the different levels. The levels being compared are fixed by the researcher, rather than being chosen at random. This is because we are sampling from all 5 boroughs instead of randomly selecting a few of them.

One-way fixed effects ANOVA is based on a Mathematical Model: $Y_{ij} = \mu + \alpha_i + \varepsilon$, where:

- Y_{ij} is the observed value of the jth listing of borough i,
- μ is the combined population mean of all five boroughs
- α_i is the fixed deviation of the mean of borough i from the grand mean μ

- Under the null hypothesis all αi equal zero; Under the alternate hypothesis some or all αi are nonzero, and their value does not vary.
- ε is a random error effect.

Results:

Results show a statistically significant difference between groups (boroughs), which is in line with our hypothesis, that a listing's borough location will affect that listing's price significantly.

Limitations:

- External Validity: Random sampling: we only sampled listings from June 2018 and applied the results to today's situation. There might be changes in the listings, some listings might be off the market and some new listings might have emerged.
- Internal Validity: Our sample sizes of each borough differ and their variances vary considerably, our ANOVA P-value may be misleading.

Hypothesis 1(2)

H1(2): There is a statistically significant difference in price between Private Rooms and Shared Rooms

Model:

Welch t-test (unequal sample sizes, unequal variance)

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

Where,

$\bar{X}_1$= Mean of the price of a private room

$\bar{X}_2$= Mean of the price of a shared room

S_1^2 = Standard deviation for the price of a private room

S_2^2 = Standard deviation for the price of the shared room

N_1= Total number of private room listings

N_2= total number of shared room listings

Results:

Results show a statistically significant difference in means in the two samples, which is in line with our hypothesis, that prices of the shared room are much lower than the prices of private rooms.

Limitations:

- External Validity: We are only taking the June 2018 listings and applying it to the current market.
- Power: We did not control other variables that might affect price, like Location (borough), which may reduce the power of the test.

Hypothesis 2:

H2: An increase in certain aspects of review score ratings will lead to an increase in the overall review score rating

Model:

Multiple Linear Regression Model

lm(overall ratings) = β0 + β1*review_score_ratings_accuracy + β2* review_score_cleanliness+...+ βn*review_score_ratings_value+βm...x*as.factor(control_variables), βj ≠ 0 for at least one j, j = 1, . . . , n

Here, we have kept other factors controlled.

Results:

Results show that all six specific review score ratings have statistically positive effects on the overall review score ratings, which is also in line with our hypothesis. Although tests revealed something we didn't know: accuracy, cleanliness, and value has the strongest positive impact on overall review score ratings.

Limitations:

- External Validity: Random sampling: we only sampled listings from June 2018 and applied the results to today's situation. There might be changes in the listings, some listings might be off the market and some new listings might have emerged.

- Internal Validity:
    - As seen from EDA, the dependent variable is not normally distributed (it's negatively skewed).
    - For each variable, the conditional distributions have equal variances.
    - The positive relationships do not imply causal relationships, eg we know review score ratings and overall review scoring ratings are positively related, but it doesn't mean the increase of accuracy ratings would cause the increase in overall review score ratings.

Hypothesis 3:

H3: An increase in listing counts, listing price, minimum nights of stay, availability, number of reviews, review score ratings and reviews per month will increase the likelihood of a host becoming a superhost.

Model:

Logistic Regression Model

logit(superhost or not) = $\beta0$ + $\beta1$*price + $\beta2$*number of reviews in total +...+ βn* number of reviews per month $\neq 0$ for at least one j, j = 1, . . . , n

Results:

Results are consistent with our hypothesis, that these variables would affect the superhost status; however, listing counts and minimum nights are not statistically significant.

Limitations:

- External Validity: Random sampling: we only sampled listings from June 2018 and applied the results to today's situation. There might be changes in the listings, some listings might be off the market and some new listings might have emerged.
- Internal Validity: Interaction Effect - some of the variables might have interaction effect between each other which might skew the results: for example, review per month and number of reviews, if review per month (in this case June), it might very much lead to a high total number of reviews.

Results:

Table of outputs:

After running tests for each of the hypothesis in R, these were the following results:

Hypothesis 1:

H1(1): There is a statistically significant difference between listing prices in different boroughs

The results of running the ANOVA test can be seen in table 4:

```
> oneway.test(price ~ neighbourhood_group_cleansed , data=list_1, var.equal=FALSE)

	One-way analysis of means (not assuming equal variances)

data:  price and neighbourhood_group_cleansed
F = 1279.6, num df = 4.0, denom df = 1990.9, p-value < 2.2e-16
```

Results show a statistically significant difference between groups (boroughs), which is in line with our hypothesis, that a listing's borough location will affect that listing's price significantly. Hence, we accept the hypothesis and reject the null hypothesis.

H1(2): There is a statistically significant difference in price between Private Rooms and Shared Rooms

The t-test run for this hypothesis in R Studio gives the following output:

```
> t.test(Private_rooms$price, Shared_rooms$price, var.equal = FALSE)

	Welch Two Sample t-test

data:  Private_rooms$price and Shared_rooms$price
t = 6.5828, df = 1263.7, p-value = 6.75e-11
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  9.630319 17.807498
sample estimates:
mean of x mean of y
 82.75635  69.03744
```

Results show a statistically significant difference in means in the two samples, which is in line with our hypothesis, that prices of the shared room are much lower than prices of private rooms. Hence, we accept the hypothesis and reject the null.

We didn’t test the entire home/apt because from EDA we already know that the average price of an entire home is much higher than the other two.

Hypothesis 2:

An increase in certain aspects of review score ratings will lead to an increase in the overall review score rating.

The output for the Multiple Linear Regression tests performed in R gave us the following output:

```
                            Estimate Std. Error t value Pr(>|t|)    
(Intercept)                 -4.90013    0.32026  -15.30   <2e-16 ***
review_scores_accuracy       2.29724    0.03532   65.05   <2e-16 ***
review_scores_cleanliness    2.22573    0.02405   92.54   <2e-16 ***
review_scores_checkin        0.91942    0.03764   24.43   <2e-16 ***
review_scores_communication  1.64336    0.03935   41.76   <2e-16 ***
review_scores_location       0.72591    0.02802   25.91   <2e-16 ***
review_scores_value          2.57455    0.03270   78.73   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.814 on 50034 degrees of freedom
Multiple R-squared:  0.7347,    Adjusted R-squared:  0.7346 
F-statistic: 2.309e+04 on 6 and 50034 DF,  p-value: < 2.2e-16
```

Results show that all six specific review score ratings have statistically positive effects on the overall review score ratings, which is also in line with our hypothesis. Hence, we accept the hypothesis and reject the null hypothesis.
Although tests revealed something we didn’t know: accuracy, cleanliness, and value has the strongest positive impact on overall review score ratings.

Hypothesis 3:

An increase in listing counts, listing price, minimum nights of stay, availability, number of reviews, review score ratings and reviews per month will increase the likelihood of a host becoming a superhost.

When the Logistical Linear Regression Model was run to prove the validity of the hypothesis, the following output was received:

```
Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)         -2.383e+01  4.326e-01 -55.090  < 2e-16 ***
host_listings_count -4.146e-04  2.582e-04  -1.606 0.108335
price                4.178e-04  1.213e-04   3.444 0.000572 ***
minimum_nights      -1.710e-03  1.034e-03  -1.654 0.098113 .
availability_30      9.667e-03  1.773e-03   5.454 4.93e-08 ***
number_of_reviews    1.751e-02  3.748e-04  46.718  < 2e-16 ***
review_scores_rating 2.210e-01  4.405e-03  50.161  < 2e-16 ***
reviews_per_month    2.446e-01  9.494e-03  25.765  < 2e-16 ***
---
```

Results are consistent with our hypothesis, that these variables would affect the superhost status; however, listing counts and minimum nights are not statistically significant. We accept the hypothesis since some of these factors do have a significant impact on the overall status of the host being a superhost.

Conclusion:

Main Findings:

After conducting statistical tests for hypothesis validity as well as EDA on specific variables to find characteristics for the variables, our findings can be summarized as:
(a) Majority rentals were for entire home/apt.
(b) Most listings are priced $200 or less per night.
(c) Manhattan has the highest number of listings while Staten Island has the least.
(d) Most host ratings are between 80-100.
(e) Manhattan listings are most expensive while the Bronx is least expensive
(f) Statistically significant difference in price for listings in different boroughs
(g) Room type affects price significantly.
(h) All review score ratings affect overall review score rating
(i) Accuracy, cleanliness and value have the maximum effects on the overall score rating.
(j) The number of review score ratings and listing price effect if the host is a super host, positively related.

When the test for Hypothesis 2 was conducted, our results showed that accuracy, cleanliness, and value has the strongest positive impact on overall review score ratings. While this is partly in line with our hypothesis, In this case, Luo asserts that:

> “...the results suggest that there are five lodging aspects that customers most care about: communication, experience, location, product/service, and value. Customers have different sentiments toward different lodging aspects, which lead to different aspect evaluation weights and aspect ratings when they judge their overall stays with the

accommodations offered on sharing accommodation platforms. Specifically, customers value the aspects of location and experience more than any other lodging aspects." (Luo, 2018).

Here, while we do see that value was one of the most important (statistically significant) influencers on the overall review score ratings, location was not as important to the guests in our results. In our dataset, we were also missing the field for "Review Score Experience" that Luo has mentioned as a key driver of the overall review score ratings in the presence of which, our results could have been different.

Business Implications

Based on our results we could see that while Manhattan had the highest number of listings and Brooklyn in comparison had fewer listings, the number of superhosts in Brooklyn was higher as compared to Manhattan which could have been higher due to higher listings. The review scores of the Brooklyn hosts can be studied in dept to understand which score categories they scored highest in which helped them get higher overall review scores and helped them become superhosts. These practices could then be deemed "best practices" which could be implemented by the Manhattan hosts to help drive up their status. This would also help Airbnb to ensure a greater customer experience. To improve a listing's overall rating, the easiest way is to improve the cleanliness of the room and the accuracy of the description. Although having a listing in Manhattan might profit more than in other boroughs, but becoming a superhost is another way to stand out and to become a superhost, it is important to ask guests to leave reviews. Another consideration for Airbnb to this effect would be to limit the listings in Manhattan since the market seems to be overcrowded and allow for and promote more listings in the other boroughs like Brooklyn and Staten Island which have higher customer satisfaction ratings.

According to Stephen Shephard and Andrew Udell, Airbnb can also raise real estate prices for properties that are within 300 meters of an Airbnb listing by 6% to 9% which is based on the model the listing and the property follow. Sheppard, S., & Udell, A. (2018). Do Airbnb properties affect house prices? Working Paper.

This is an interesting way of looking at Airbnb's business model and objective and implementing ways to insulate itself against inflating prices which could drive away potential guests back to traditional hotel chains that have a relatively stable price model.

References:

(1) Airbnb Data Site: http://insideairbnb.com/get-the-data.html

(2) Luo, Y. (2018) *What Airbnb Reviews can Tell us? An Advanced Latent Aspect Rating Analysis Approach.* Iowa State University. Graduate Theses and Dissertations.

(3) Wachsmuth, D. (2018). The high cost of short-term rentals in New York City. McGill University.

(4) Sheppard, S., & Udell, A. (2018). Do Airbnb properties affect house prices? Working Paper.