The University of North Carolina at Charlotte

****

**The Effect Of COVID-19 Pandemic**

**On NYC’s Airbnb Market**

**Group 1:**

Brendan Dignan, Kevin Ovendorf, Togzhan Tlegenova,

Juan Ricardo, Rohan Mutalik, Olga Osipov

Dr. Kexin Zhao

DSBA 6211: Advanced Business Analytics

August 5th, 2020

**TABLE OF CONTENTS**

[**BUSINESS OVERVIEW**](#_heading=h.1fob9te) **2**

[**OBJECTIVE**](#_heading=h.3znysh7) **4**

[**DATA AND PREPROCESSING**](#_heading=h.tyjcwt) **5**

[**HOW DOES THE COVID-19 PANDEMIC AFFECT THE NYC’S AIRBNB MARKET?**](#_heading=h.3dy6vkm) **8**

[**INITIAL ANALYSIS AND BACKGROUND**](#_heading=h.1t3h5sf) **8**

[**MARKET SUPPLY**](#_heading=h.4d34og8) **13**

[**MARKET DEMAND**](#_heading=h.2s8eyo1) **21**

[**PRE-COVID REVIEWS**](#_heading=h.17dp8vu) **23**

[**POST-COVID REVIEWS**](#_heading=h.3rdcrjn) **24**

[**OVERALL IMPACT ON MARKET**](#_heading=h.26in1rg) **26**

[**WHAT FACTORS HAVE AFFECTED AIRBNB HOSTS’ MARKET EXIT DECISIONS**](#_heading=h.lnxbz9) **26**

[**WHAT TYPES OF PROPERTIES HAVE BEEN AFFECTED THE MOST?**](#_heading=h.35nkun2) **30**

[**WHAT INDICATOR(S) QUANTIFY INDIVIDUAL LISTINGS’ PERFORMANCE?**](#_heading=h.1ksv4uv) **43**

[**DATA COLLECTION SUGGESTIONS FOR FUTURE ANALYSES**](#_heading=h.2jxsxqh) **51**

[**SUMMARY/CONCLUSION**](#_heading=h.44sinio) **53**

[**APPENDIX**](#_heading=h.1y810tw) **54**

[**SOURCES**](#_heading=h.4i7ojhp) **54**

# 

# 

# BUSINESS OVERVIEW

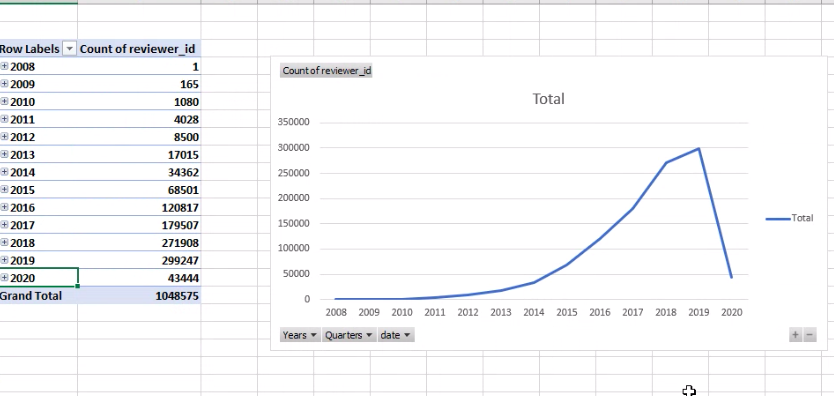
Airbnb, originally “Air Bed And Breakfast”, was launched in 2008 as a few young designers decided to rent space in their apartment on air mattresses while there was a convention in town and they realized there were going to be hotel shortages. This was their plan to make a few extra dollars to cover their high San Francisco living expenses. Fast forward 12 years and Airbnb is seen as one of the great disruptors and unicorn companies of its time that has announced its plans to soon go public. Although initially investors had doubts about this “air mattress rental” concept, Airbnb really took off after 2009, when Sequoia Capital did a $600,000 seed investment in Airbnb which really started to project them forward. By the end of 2010, Airbnb had raised $7.2 million in venture capital and expanded to countries all over the globe. By 2011, Airbnb had over a million bookings through their platform.

Although it doesn’t go without say, Airbnb has had its own fallbacks throughout the years which they had to settle with permanent policy and rental changes, lawsuits, and added fees. Overall, the company has been on a steep growth incline, significantly cutting into the typical hotel, resort, stay, and rental market share. Their concept has proved to be very successful and has only been launching forward.

Over Airbnb’s business history, multiple events have led them to have some fallbacks. For example in 2012, hosts complained of their properties being ruined by guests; this took them back a few steps and led them to the implementation of a coverage policy. Other hosts were also having some problems with renting out their places and some were even being evicted for renting out. This led to implementation of more taxes and fees.

Massive numbers of Airbnb rentals have also been seen to lead to shortages in the housing availability in different cities, where politicians and cities started to take notice and started implementing limits and regulations. Although events and disturbances like these may have made minor dents in the growth of Airbnb as a company, Airbnb continued to prosper. That is, until 2020 where a world-wide and previously unforeseen event took place - the 2020 Covid-19 pandemic. After following along the rapid spread across the world, on March 11th, 2020, The World Health Organization officially announced that “COVID-19 can be characterized as a pandemic.” By March 13th, 2020, massive closings started to take place, people were warned to stay home, and nonessential jobs were transferred to being online. Travel restrictions were put in place and it was recommended to stay at home. This was a hit not only to all of the world, but also to platforms such as Airbnb. People were staying home, everyone cut down on travel significantly if not completely, and thus demand for rentals went down.

*Figure 1:Measuring the growth of AirBnB growth from 2008 to 2020 by the metric: reviews*



In the graphic above, we see an exponential growth in the company according to reviews posted from 2008 until June 2020 in New York City’s Airbnb market. We will consider reviews during a specific time period as an indicator of the number of stays during the time as not only it is very common to leave reviews following a stay, but it is a way to guarantee that the data we use is accurate and there are no false positives of no one staying at a rental while it appears rented in the dataset. We see a slower start during the company launch in 2008 until about 2011, and then a rapid growth since their multiple venture investments took place, and this growth continues until the beginning of 2020. Then, it is evident that there is a dramatic drop in 2020, which we can infer to be caused by the spread of COVID-19 and the resulting shutdown worldwide.

# OBJECTIVE

In this project, we will focus specifically on New York City’s Airbnb Market and look at how the Covid-19 Pandemic affected it. Although we can see the rapid drop from the simple graph above, we want to explore what exactly happened, analyze listings, hosts, reviews, renters, draw conclusions about who and what was affected the most and try to quantify listing performance during the pandemic and beyond. Our goal is to analyze the Airbnb data and draw conclusions that would be significant for the Airbnb business.

There are a lot of factors from the covid pandemic that have affected NYC’s Airbnb market, including the number of active listings in the market (supply), and the number of reviews per listing (demand). We want to focus our attention on doing the following: analyzing how covid has affected NYC’s Airbnb market, taking into consideration market supply and demand, customer comments; exploring which factors have affected Airbnb hosts’ exit decisions from the market; figuring out which types of properties have been affected the most; and quantify the performances of individual listings. The questions are listed below in summary.

Questions:

How does the COVID 19 pandemic affect NYC's Airbnb market?

Comprehensive list of matrices, such as: Market supply: the number of active listings

Market demand: occupancy rate, average monthly reviews, etc.

Customer comments: topics, sentiments, etc.

What factors have affected Airbnb hosts’ market exit decisions?

What types of properties have been affected the most?

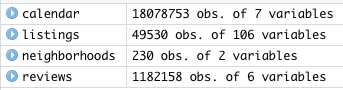
Please propose a reasonable indicator(s) to quantify individual listings’ performance.

# 

# DATA AND PREPROCESSING

The data for this project was derived from the “Inside Airbnb” Database. The initial downloaded data consisted of listings, reviews, calendar, and neighborhoods files from 2008 when Airbnb was launched until June of 2020. Below is the overview of the original data we received.

*Figure 2: Overview of original data downloaded from OpenAirBnB*



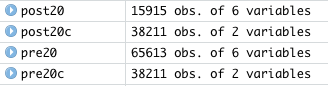
We used these original files to make general observations, get a first look into the data, as well as to decide what was important to use for the project and what needed to be cleaned, discarded, or altered. The original four data files consisted of numerous variables not all of which we thought were useful for our project. For example, there were many variables consisting of urls, various id’s, repetitive location names, location summaries, verifications, and those that had identical values for all its rows. These along with other similar variables were discarded. Our variables were then altered as needed and transformed to the appropriate types.

Although we have used the raw and complete files for the purposes of initial visualization, for our main analysis and modeling we decided to only use the data from the year 2020. We used Friday, March 13, 2020 as the cutoff date for when COVID’19 started to make a major impact on the United States economy and people’s lives. March 13th was one of the first days major closures started occurring throughout the USA. Therefore, we used this date to divide our data as that which would be considered prior to COVID, meaning 1/1/2020 to 3/13/2020, and that which was post the start of COVID-19, meaning 3/13/2020 to 6/10/2020.

The main question that we are trying to answer in this project is the effect of COVID-19 on the Airbnb market, therefore we also felt it was unnecessary to include listings that were part of the dataset but were not active both prior to covid or since it had started. We then decided to make a number of different data partitions in order to make it easier to analyze data for exactly what we were looking for although it was often combined for other analysis. Some of the partitions we did were isolating those units that were listed but haven't received a single review since COVID, those which were listed for rent and have been rented, properties with at least 2 reviews pre-Covid 2020 but now not listed at all. We thought these were interesting and could drive meaningful insights.

For the purposes of this project, we used reviews as our target variable. The number of reviews received at a specific period in time is seen to be a generally good indicator of the number of rentals which were occupied at the time. Therefore the presence or absence of a review at a given time is our Target variable.

*Figure 3: pre and post covid, with and without comments*



*Figure 4: Units listed that have not received a single review since the start of COVID,*

*With and without reviews*



*Figure 5: Units listed for rent and have been rented,*

*With and without reviews*



*Figure 6: properties with at least 2 reviews pre-COVID, but now not listed at all*



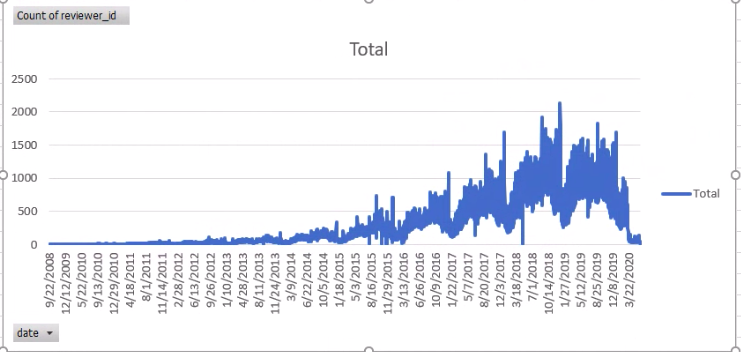
Throughout the project and for analysis and modeling we used various tools including Excel, Tableau, RStudio, and Python

## HOW DOES THE COVID-19 PANDEMIC AFFECT THE NYC’S AIRBNB MARKET?

# INITIAL ANALYSIS AND BACKGROUND

Looking at the data ranging back to 2008 from the launch of the company, we see that Airbnb has been on a steep and exponential growth incline, that is, until COVID-19 unexpectedly hit the world in 2020. However, prior to COVID-19, we found that although the mean of the rentals has been increasing, the growth was not perfectly smooth. We see that the Airbnb market has similar seasonality trends that can be seen year to year.

*Figure 7: Growth of Airbnb from 2008 - 2020 by measuring the amount of reviews left.*

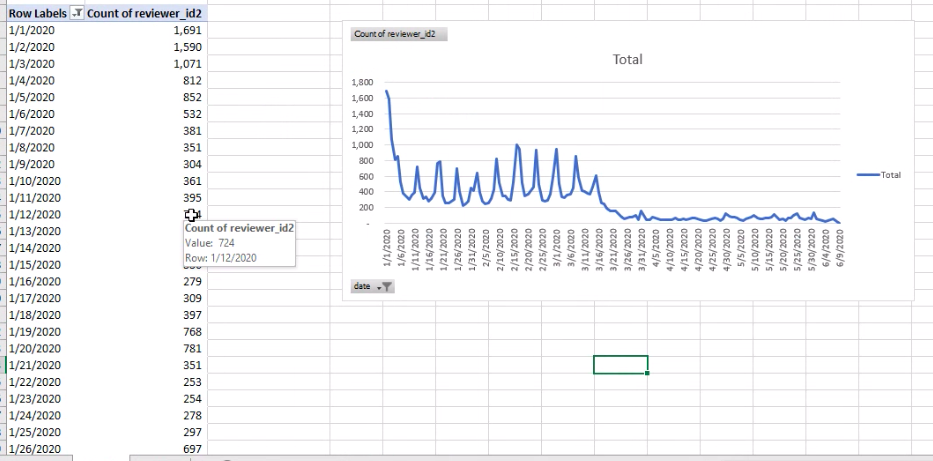


In the graph above we see a timeline of reviews received throughout the years. We see that there is a steep decline in reviews every January and that from that point on they start increasing until the beginning of summer in May/June. At this time, there is usually a spike in reviews for the summer months and again a decline once fall starts. We then see a significant increase in reviews during the holidays until we again have the dramatic drop in January.

These seasonalities in the Airbnb Market make sense. People typically go back to work in January after the holidays, and may travel less during the winter months. Closer to spring and during spring time, it is typical for people to take a spring vacation and book an airbnb. Once summer starts, people take time off from work, students have break, and this is typically the time people reserve the time to travel hence the spike. Starting in the fall, all school starts and schedules become less flexible for travel until fall break and the December holidays when we see the highest numbers of reviews.

It is important to keep in mind these seasonalities because if we only look at the January through June data from 2020 that we will use for analyzing the effect of COVID-19 on the Airbnb Market, we would not know that the sudden drop in reviews in January is not actually from COVID-19 but a common event that has been happening every January.

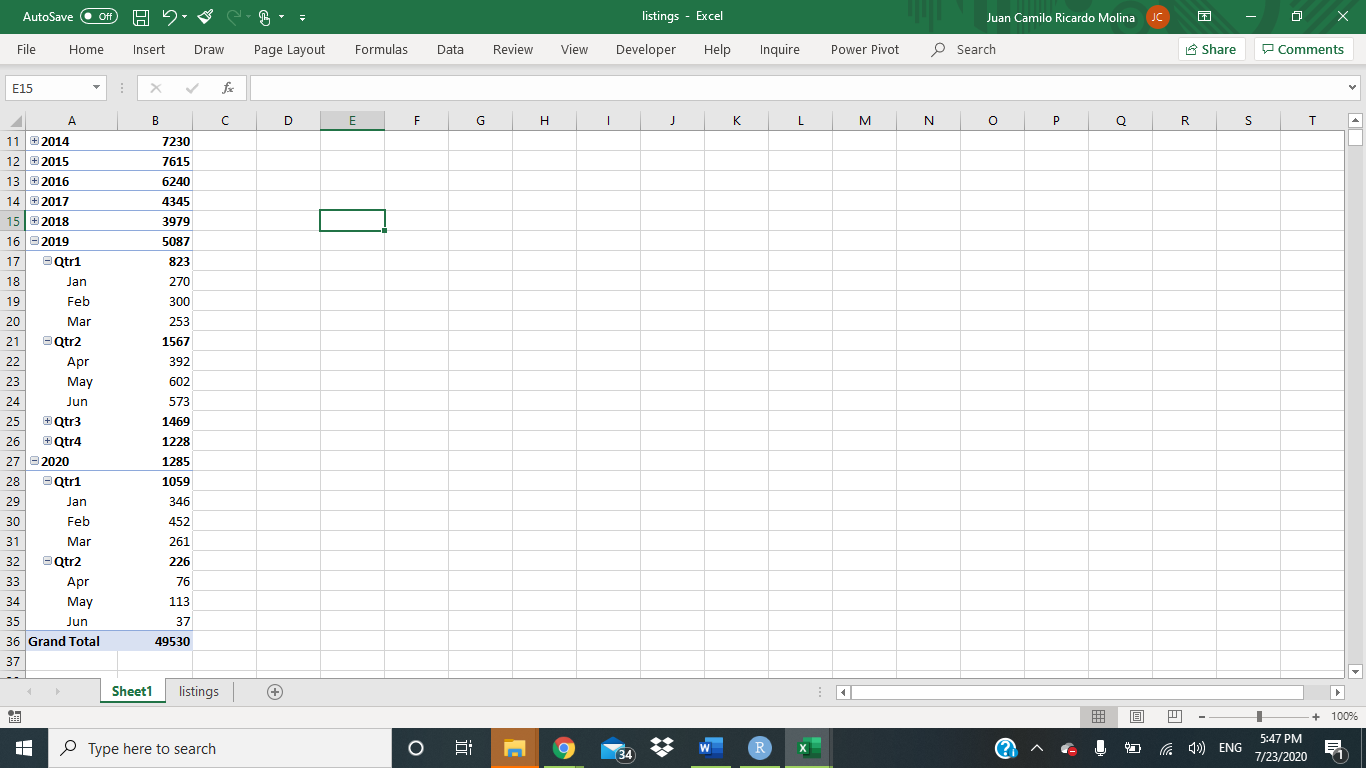
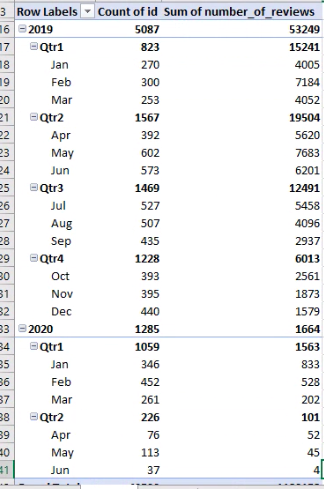
*Figure 8: January - June 2020 data scaled to show trends on a weekly basis*



Now, taking a closer look at the January through June 2020 data, we also see that there are some trends that happen on a weekly basis. For example, we see from the graph above that the number of reviews increases during the weekends, showing us that it is typical to have more bookings during the weekends as compared to the weekdays.

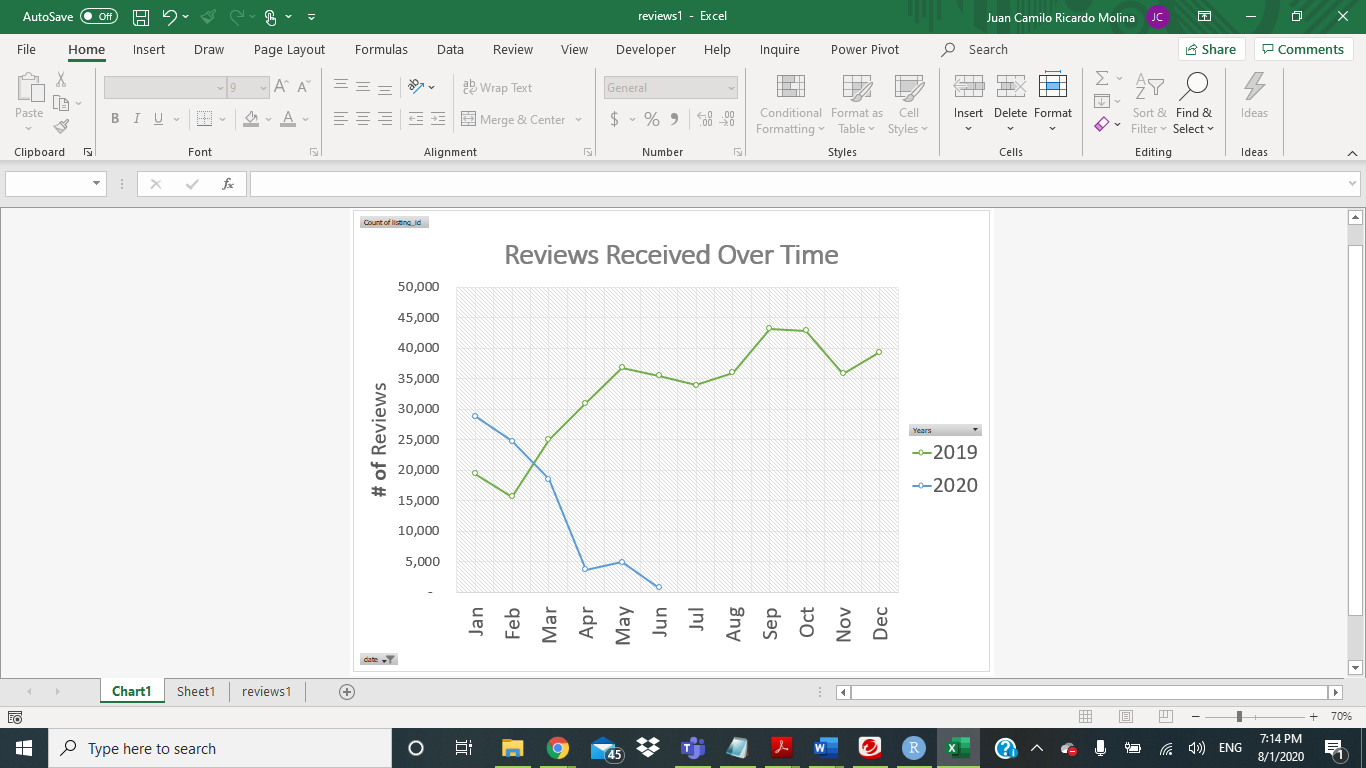
It is also important to note and not disregard the fact that increased numbers of bookings occur on special occasions and holidays. Above, we also see a small spike starting in mid-February. This is likely due to Valentine’s Day getaways and February break for students in some schools. Following this we see a significant decline in reviews due to COVID-19 which we will explore in further detail throughout this project.

*Figure 9: New Listings by Quarter and Reviews By Quarter*

Interestingly, if we look at the numbers of new listings by month and quarter, we see that January, February, and March of 2020 overall had higher numbers of new listings as compared to the same time during 2019. This is likely due to typical yearly company growth and popularity. We then see a significant decline in quarter 2 of 2020.

*Figure 10: Comparing reviews in 2019 to 2020 to visualize the negative effects of COVID*



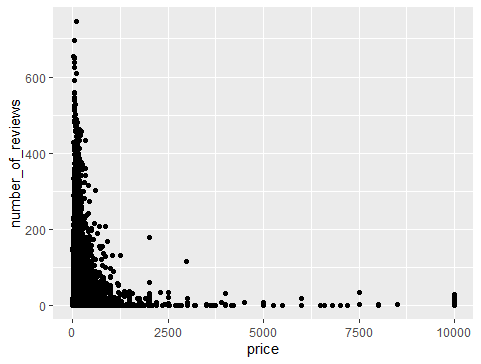
The reviews, on the other hand, tell a different story. In the charts above, we compare January to June received review totals in 2019 as compared to 2020. Although we see that in January and February of 2020 we had more reviews as the previous year at the same time, the overall trend changed. Instead of increasing from February and on throughout the year with common yearly seasonality trends, we see a rapid and ongoing decline that does not bounce back.

Although for the purpose of this project we are using March 13th, 2020 as the date after which we are considering as post the start of COVID since this is when the economy dramatically changed with stay-at-home orders, closings,etc., we want to mention that we also predict that the effect of covid began much earlier and possibly as soon as the end of the previous year.

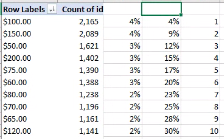
According to the New England Journal of Medicine, the first case of COVID-19 in the USA was confirmed on January 20, 2020 after the initial outbreak in Wuhan, China late 2019 and the spread throughout the world. The first confirmed case in the State of New York was confirmed on March 1st, 2020, and according to the New York Times by April 4th, 2020 there were 12,000 official cases in New York. By April 10th, the USA had a record number of cases and more confirmed cases than any other country. By August 1st, 2020, over 6 million tests have been taken and over 400,000 cases have been confirmed. Therefore, the effect of COVID on Airbnb might have started much earlier even before the spread began in the United States and before the worldwide closures, travel bans, and stay-at-home orders.

# MARKET SUPPLY

To learn more about the Airbnb market, we started our research by looking into various variables to see what insights could be drawn from them and what we needed to analyze deeper.

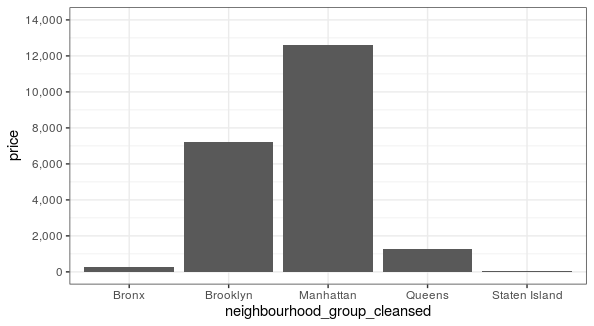
*Figure 11: Scatterplot showing the relationship between the number of reviews of a listing and its price.*

*Figure 12: Table showing the most frequent listing prices*



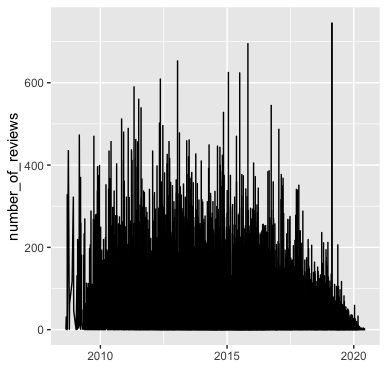
We first looked at how the pricing of properties was reflected in the data. Although there were some outliers and the prices overall ranged from $0.00 to $10,000.00, the average of the listing prices was $684.10 and the median was $700.00. The top 30% of the listings were between the price of $50.00 to $200.00, with the most common price being at $100.00 per night for an Airbnb stay. As shown in the graph above, even popular listings were priced low.

*Figure 13: Price margins of the boroughs*



When looking at price by location and specific neighborhood, we see that the highest priced listings came from Manhattan, the second most expensive came from Brooklyn, and then later Queens, Staten Island and Bronx. The price margins between Manhattan, Queens and the other three boroughs were very large, signifying that there were probably more listings in Manhattan and Queens. These two boroughs were also the neighborhoods hardest hit by the pandemic

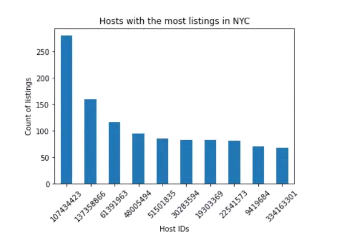
*Figure14: Histogram showing the Host Since Date vs. Number Of Reviews*



In the graph above, we looked at the relationship between the number of reviews received as compared to how long the hosts have been on Airbnb. We do see that in general newer hosts have less reviews and those who have been hosts for a few years now have more.

However, we also see some extreme outliers that do not make sense. For example, the noticeable spike in reviews for a host in 2019 seems suspicious. The number of reviews received in the time period is larger than there are even days in the year that the property could be rented for. Even if this host could have listed many properties, it would still be unlikely to get more reviews than seasoned hosts who are renting out larger numbers of properties. We think that this specific host may be falsifying their reviews.

*Figure 15: Histogram showing the anonymized Hosts’ who own the most individual listings*

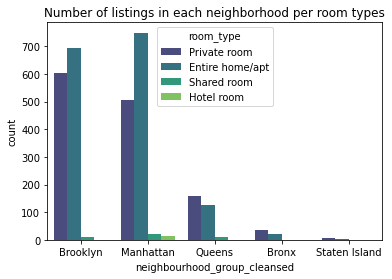


The above table shows host IDs with the most listings in NYC. The superhost with the highest number of listings in NYC, upwards of 250 listings, already has a 100-listing advantage over the other superhosts and has a larger monopoly over the supply market. As seen in this data, the superhosts have a larger market share than regular hosts. They are the people who meet requirements of having a large number of reviews, great ratings and almost no cancelations, and we can note that most of them probably took mortgages on several properties, expecting to get profit from Airbnb guests to cover those payments. Covid reduced cash flow a lot. So most probably this pandemic affected some superhost to exit the market.

However, referring towww.airbnb.com, we can state that Airbnb’s employees created a Fund to support Airbnb’s superhosts considering the hard pandemic situation and donated USD 1 million which will be granted until September 2020. The owners and investors of Airbnb have increased this Fund up to USD 17 million by contributing USD 9 million and 7 million correspondingly. This allowed many superhosts to stay alive at the market.

Something we can also infer from this and the nature of the listings is that Airbnb and its separate hosts are not the only business model, but there are also separate businesses operating within Airbnb’s infrastructure. These businesses are renting out large numbers of properties and doing business close to how a hotel would. Those operating under such a business model contribute to the data being skewed, however, because of Covid and because everyone has been affected in some way or another, we decided to keep this within the data and see how all were affected.

*Figure 16: The number of listings of each room type for each Borough*

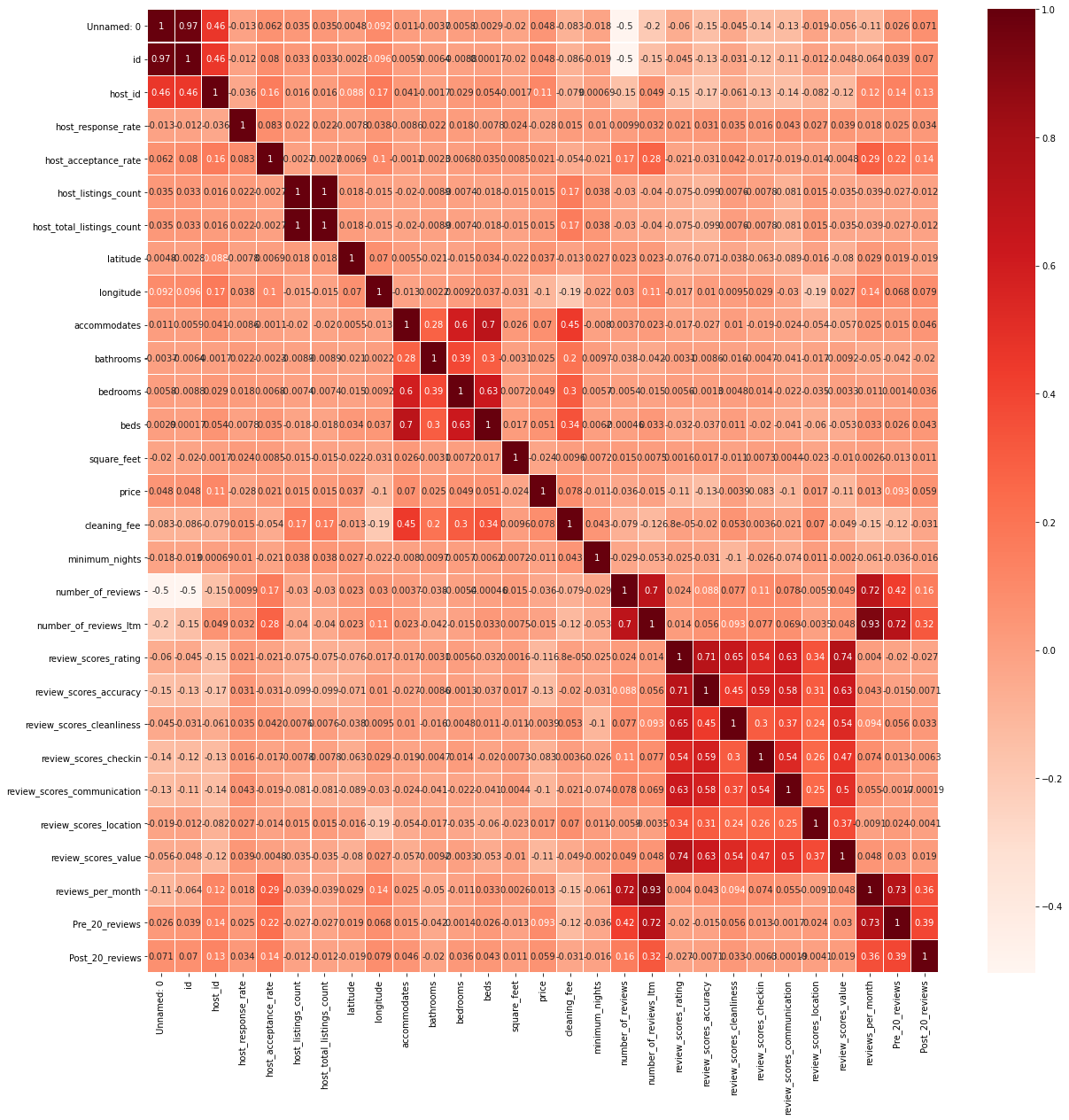


We did some analysis on the number of listings per neighborhood group categorized by room type. It is observed above that people preferred to stay apart from others before the COVID period, namely to rent private rooms or entire homes. In both Brooklyn and Manhattan, people preferred to have an entire home/apt to themselves versus just a private room.

Further, we did some analysis on checking relationships between variables in each dataset “Pulled”, “Cantrent” and “Liked” separately.

The “pulled” dataset included those listings that existed prior to Covid, but were taken off the market after March 13th, post the start of Covid. The “can’t rent” dataset included those that were listed before and throughout the pandemic but were not rented out after the pandemic, and the “liked” dataset included those that were listed and rented before and after the start of Covid.

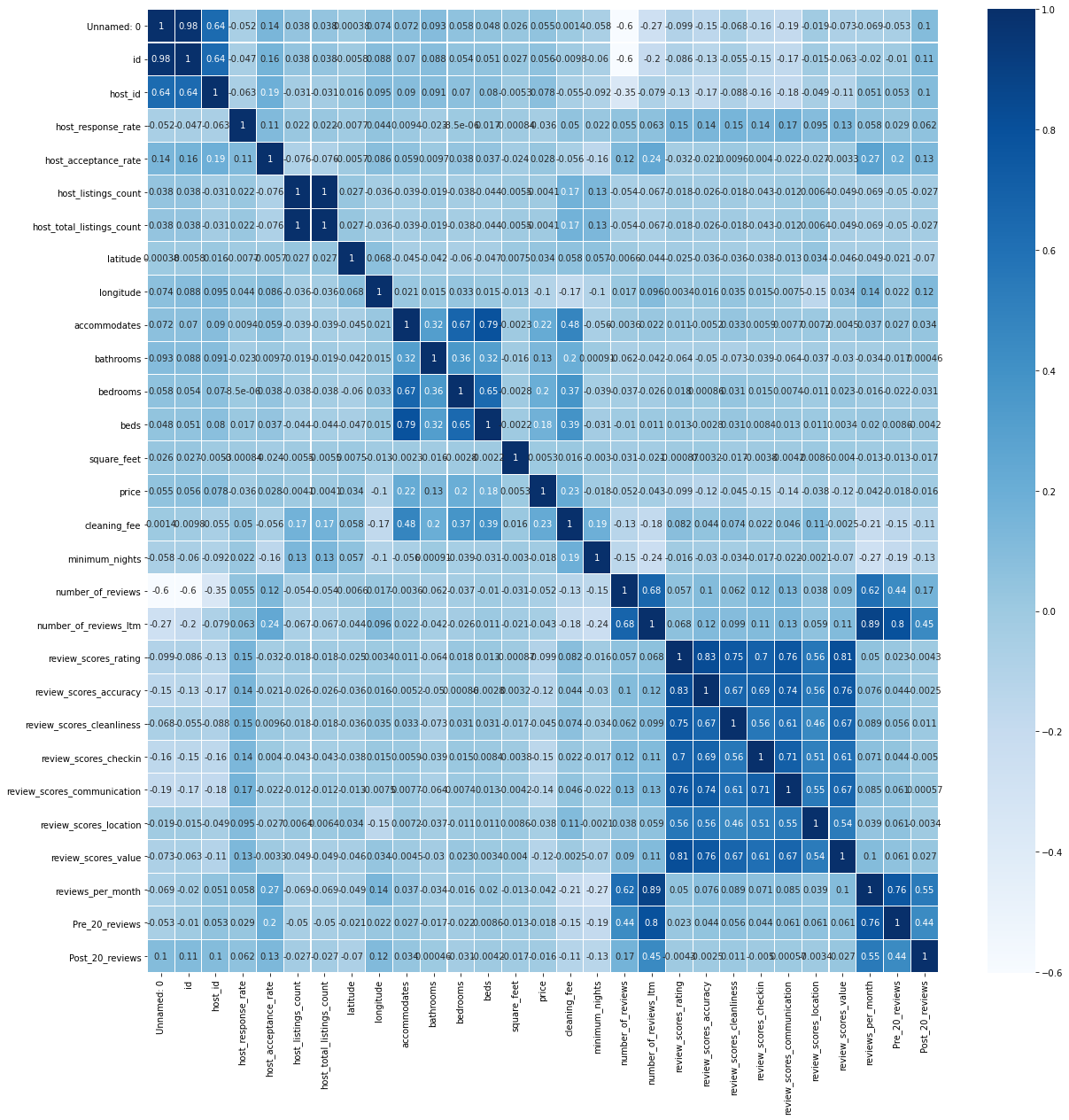
Below matrices show the results followed by their explanations.

*Figure 17: Correlation matrix of the “pulled” dataset*

Checking correlation of the variables from the “Pulled” dataset gave us the result that “Review\_scores\_value” is highly affected by “Review\_scores\_rating” and “Review\_scores\_accuracy” (74% and 63% correspondingly). The highest positive correlation (93%) is shown between “Review\_per\_month” and “Number\_of\_reviews\_ltm”. Variable “Pre\_20\_reviews” has 73% positive correlation with “Reviews\_per\_month”, while “Post\_20\_reviews” has only 36% positive correlation with the same variable.

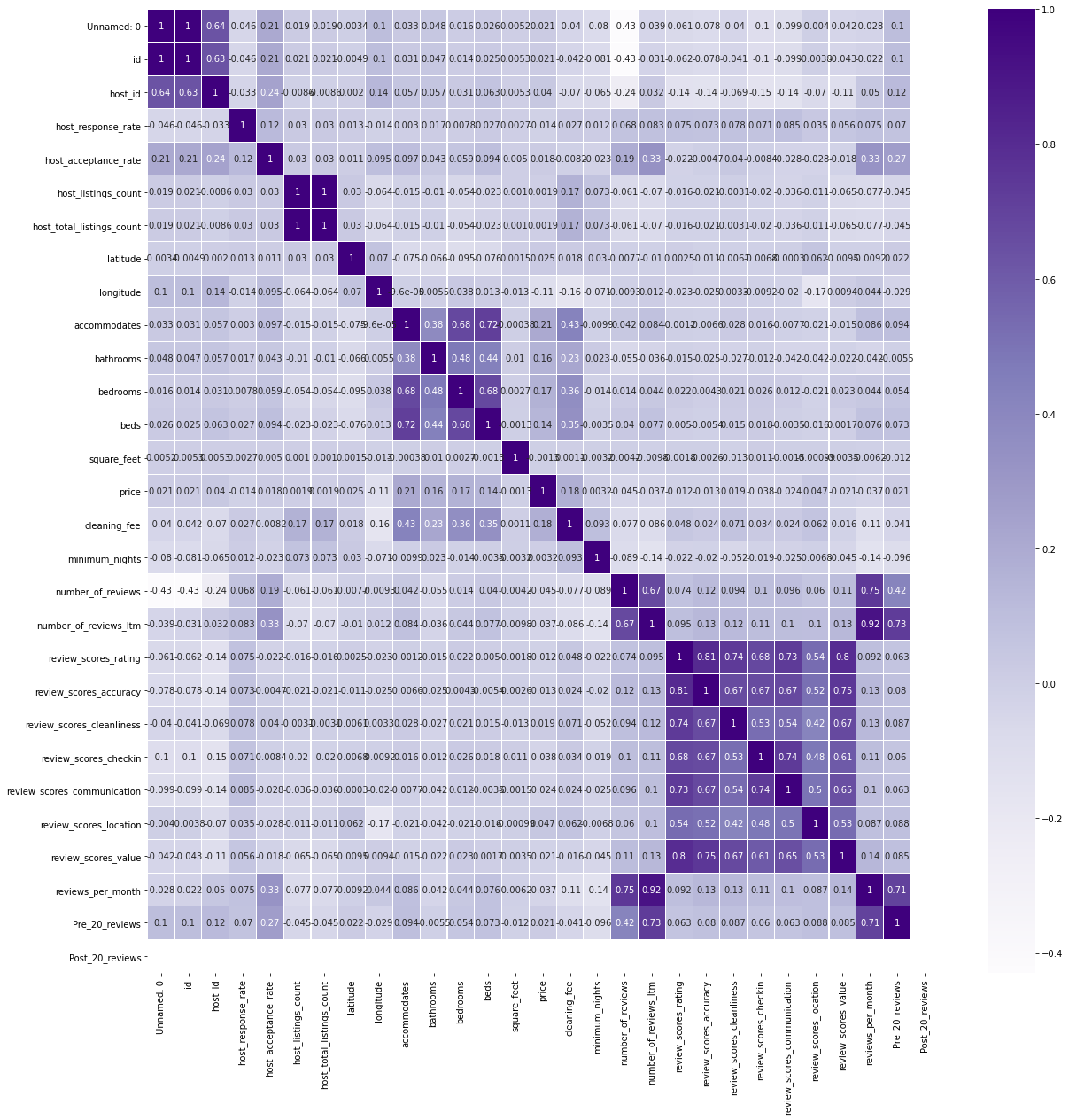
There are no variables with big % of negative correlation. Variables “Review\_scores\_accuracy”, “review\_scores\_cleanliness”, “review\_scores\_checkin”, “review\_scores\_communication” “review\_scores\_location”, “review\_scores\_value”, “review\_scores\_rating” are positively correlated among each other with 24% to 74%.

*Figure 18: Correlation matrix of the “liked” dataset*



Almost the same applies regarding variable correlation from the “Liked” dataset. All types of Reviews have relatively high correlation among each of them (46%-83%). For these datasets we can observe stronger correlations between “Pre\_20\_reviews” and “Reviews\_per\_month” (77%), and between “Post\_20\_reviews” and “Reviews\_per\_month” (55%).

*Figure 19: Correlation matrix of the “cant rent” dataset*

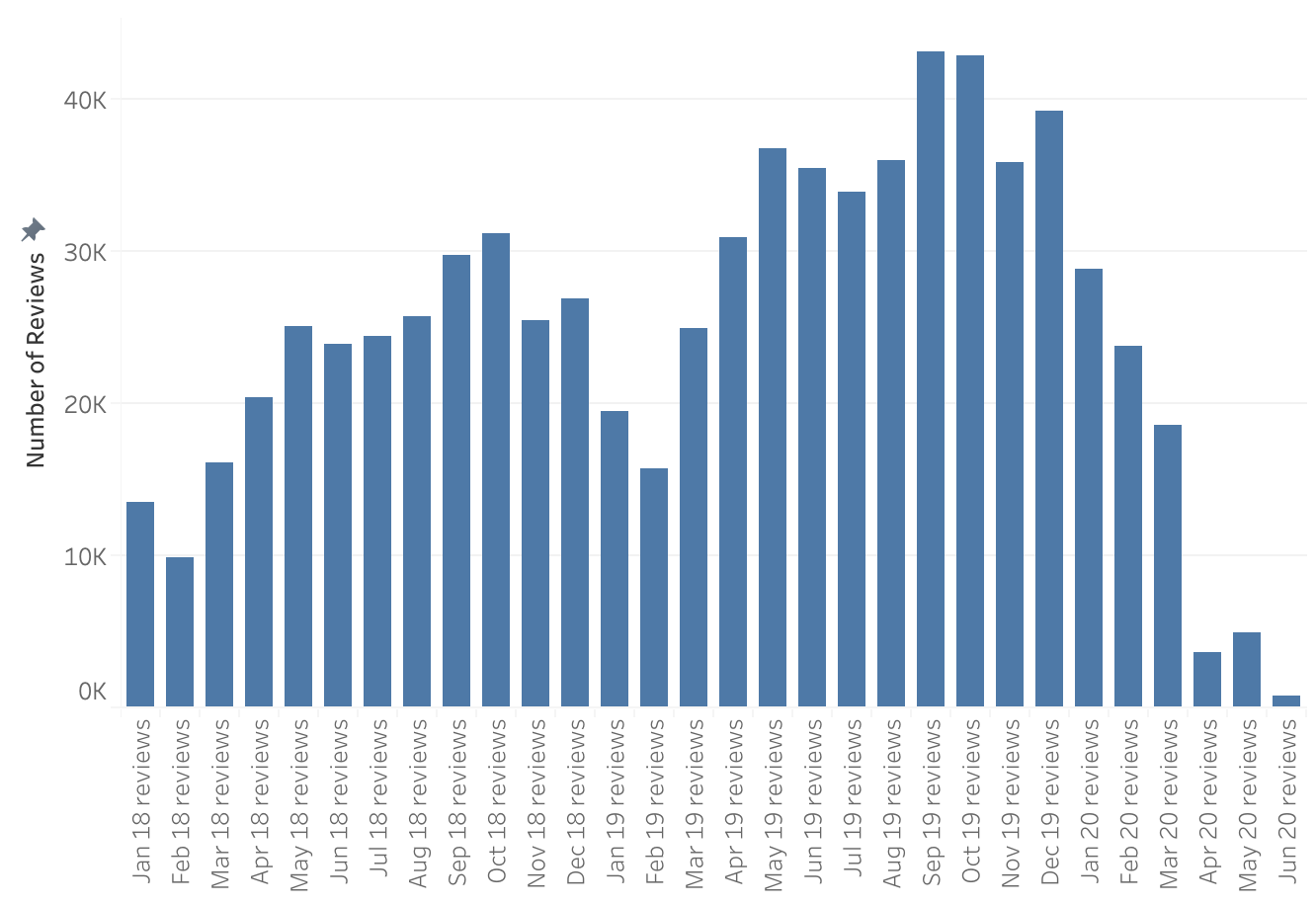


As per “Cantrent” dataset there are no reviews for “Post\_20\_reviews” since Covid, therefore the correlation matrix gives a relative picture between “Pre\_20\_reviews” and “Reviews\_per\_month” as 71%. The highest positive correlation (92%) is shown between “Reviews\_per\_month” and “Number\_of\_reviews\_ltm”. “Accommodates” has about 70% correlation with “Bedrooms” and “Beds”.

# MARKET DEMAND

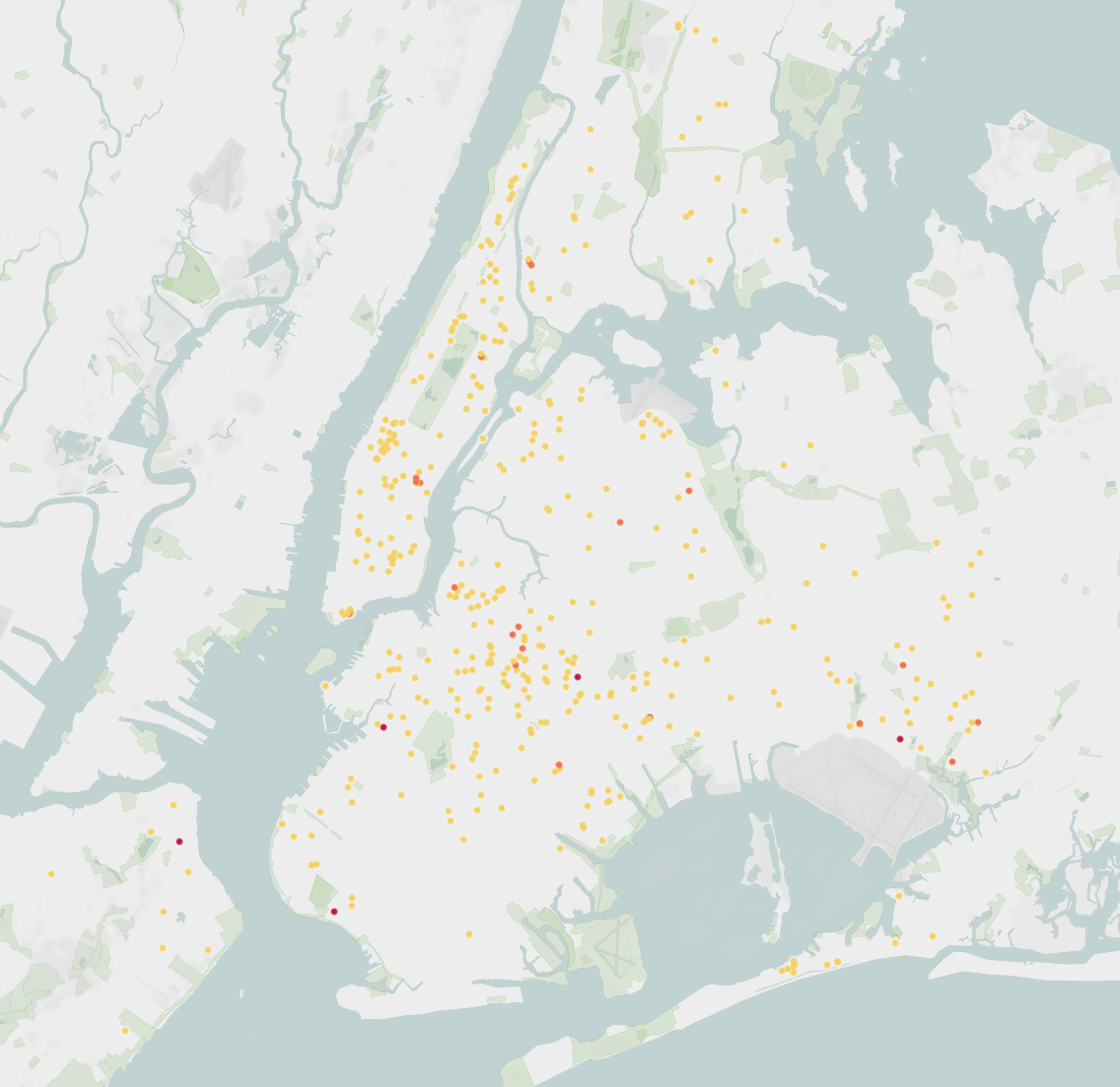
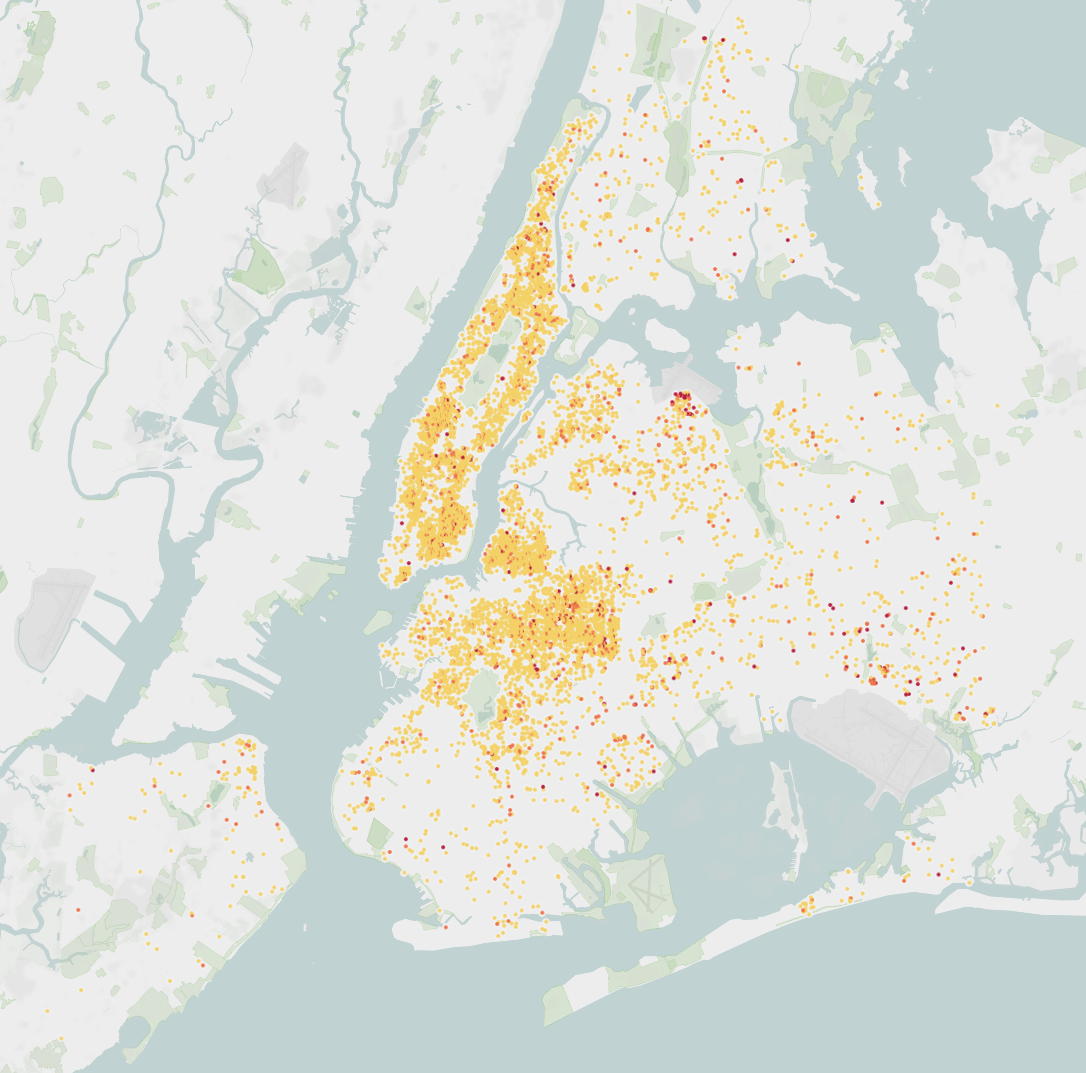
AirBnB’s New York City market has grown year by year with a quadratic seasonal trend. Every year starting with January, they start with the lowest amount of rentals for that specific year, and it begins to rise as the season changes from winter to spring.

*Figure 20: Histogram showing the total number of reviews per month from January 2018 through June 2020*



This trend doesn't slow down until after the holidays are over and the new year is ready to begin again. Now, by observing the bar graph above you would think differently, but that is because of an unprecedented pandemic. As you can see, when COVID-19 was declared a pandemic and New York was declared the epicenter of the United States, the amount of rentals dropped immediately. The maps below correspond to the graph seen above and visualize exactly how much the available rentals dried up.

*Figure 21: Two maps showing the listings that hosted customers in pre-COVID March 2020 (left), and post-COVID June 2020( right)*

****

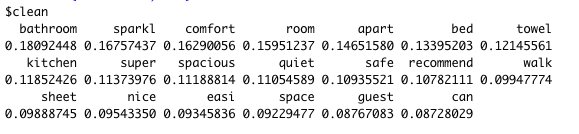
The two maps above show the number of reviews for the month of March 2020 on the left compared with June 2020 on the right. The yellow dot signifies that the rental was booked at least once but as the color saturates to red it shows the rentals that were booked the most and the most popular or successful. When looking at the June 2020 map, the remaining properties are spread out and because of this the yellow is a lot harder to see but it clearly shows the dramatic drop in AirBnB availability due to the pandemic. Overall, the drop in June 2020 compared to June 2019 is a 98% decrease in bookings. The drop from spring (April, May, and June) 2019 to spring 2020 was a drop in 91% of total bookings for the NYC area. These two figures are astronomical in terms of how a company can operate, as so far COVID-19 has almost completely wiped out the New York AirBnB market which has the potential to have lasting and even permanent effects, such as reducing the amount of investors in property in New York. These effects though are not entirely negative as the pandemic has the potential to weaken the competitive housing market in New York which has in turn led to steep increases in property in the already expensive and upscale city.

# PRE-COVID REVIEWS

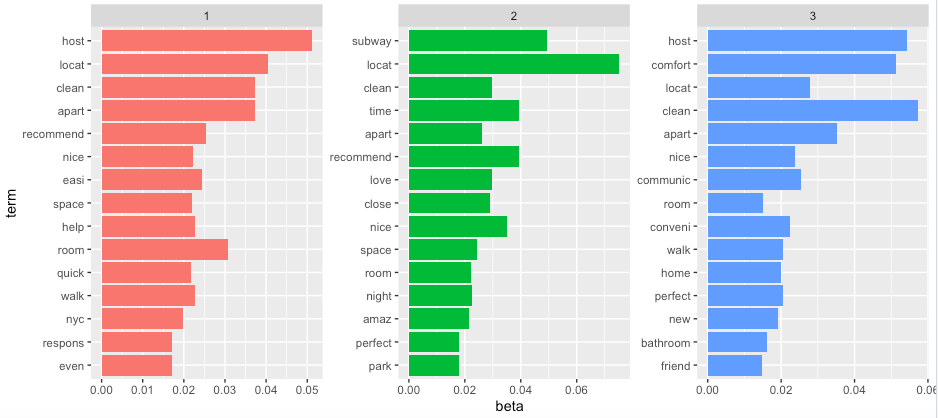
*Figure 22: Word cloud of pre-COVID reviews*



*Figure 23: Top 20 words related to “Clean”, pre-COVID*



*Figure 24: Term/beta plots for the top three topics derived from pre-COVID reviews with the most relevant 15 terms for each topic.*



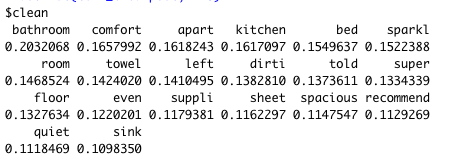
For pre-covid, it looks like being in a good location, close to local destinations, and comfort were people’s main priorities, from the three topics in the figure above. It was more about the host and the living situation. Interestingly, cleanliness was just as important pre-covid. The words most related to ‘clean’ were about the cleanliness of the rooms.

# POST-COVID REVIEWS

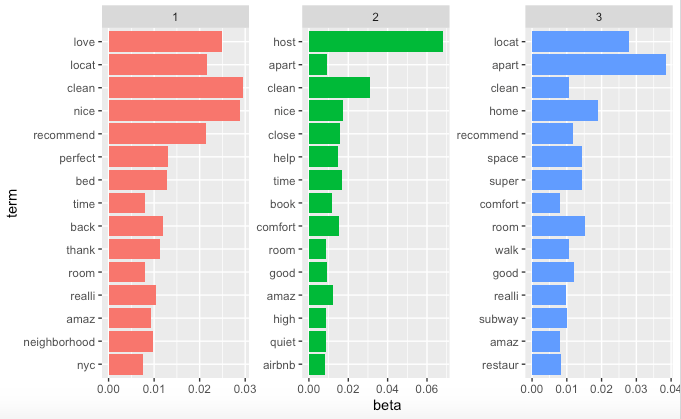
*Figure 25: Word cloud of post-COVID reviews*

****

*Figure 26: Top 20 words related to “Clean”, post-COVID*

****

*Figure 27: Term/beta plots for the top three topics derived from post-COVID reviews with the most relevant 15 terms for each topic.*



We predicted that as compared to before COVID, we would see more reviews talking about sanitation and cleanliness of the rental properties as compared to those before COVID. However, after doing text mining analysis, we actually found few variations between the comments that were posted on listings before covid and those during/after the start of COVID. It seems that the post-covid reviews had little to do with covid itself. As long as the host and living situation was fine, that was all that mattered, as noted in the topic figure above. The topics discussed were about having a clean space, about location, convenience, and close distance to surrounding restaurants and locations, as well as straightforward communication with the host. This was quite similar for both pre and post COVID topics.

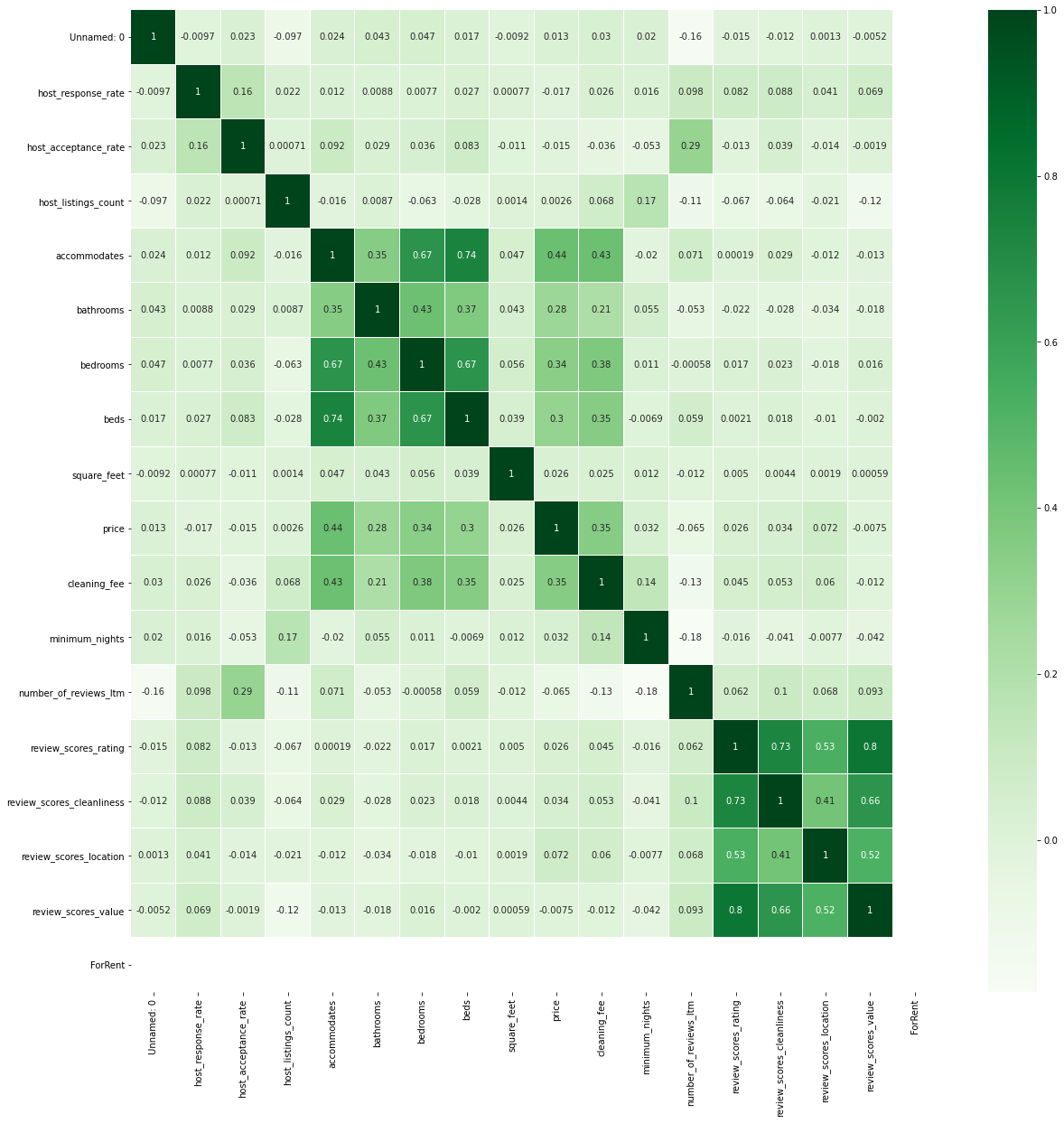
# OVERALL IMPACT ON MARKET

The overall impacts on the market can easily be mistaken as short term because once a vaccine for COVID-19 is created and widespread, the market could return to its normal growing trend. This is too reliant on an unknown date and time so that is entirely speculative, so this leaves the current NYC AirBnB with short term and potential long term side effects from the impact that the pandemic has laid on the city. First, the short term effects have been devastating as the decrease in bookings from last year's data is a staggering 90% since the pandemic was first declared. The vast and saturated AirBnB market in New York has been deflated and AirBnB itself has been given a $2 billion loan from the US government to stay afloat, so while AirBnB, the company is protected long term, its Hosts are not covered and this may have various dangerous long term effects.

# WHAT FACTORS HAVE AFFECTED AIRBNB HOSTS’ MARKET EXIT DECISIONS

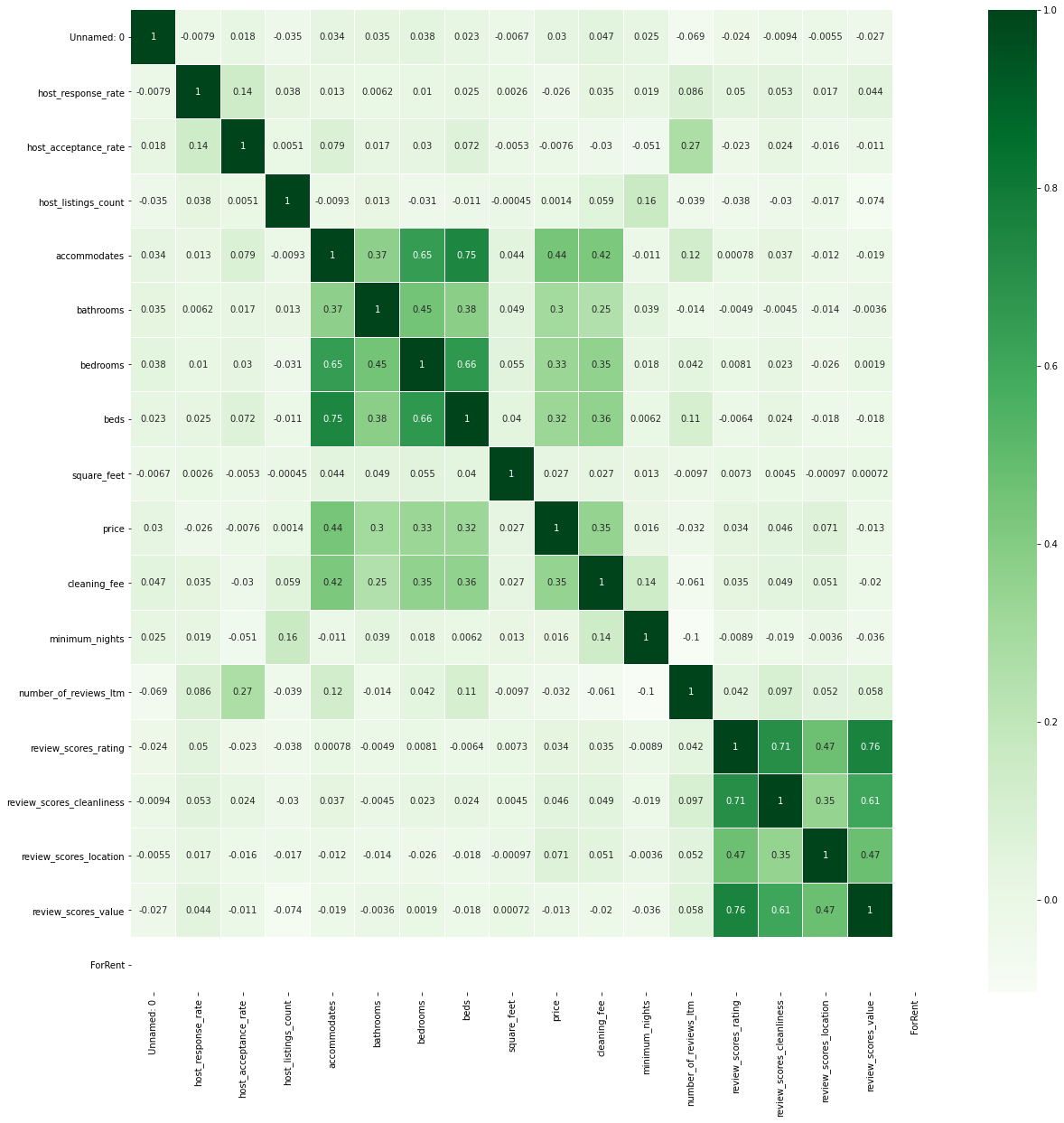
Using a correlation matrix, we can observe any relations between variables. The matrix below provides information about several variables when the “ForRent” column is equal to 1 in our dataset “ModelData4Supply”. “Accommodates” is strongly positively correlated with “Beds” and “Bedrooms”, 67% and 74% correspondingly. There are more than 40% positive relations of “Accommodates” with “Price” and “Cleaning fee”. We can also find out that “Review\_scores\_rating” has sufficiently high correlation with “Review\_scores\_cleanliness”, “Review\_scores\_value”, “Review\_score\_location”. “Host\_acceptance\_rate” has a positive but weak (29%) effect to “Number\_of\_reviews\_ltm”. Of course, the matrix confirms that “Price” is positively correlated with “bathrooms”, “bedrooms” and “beds”; more bedrooms, and price goes up. We can also see some negative correlations between our variables. For instance, “Review\_scores\_value” is negatively and very weakly interacted with “Host\_listings\_count”, “Host\_acceptance\_rate”, ‘Accommodates”, “Price”, “Cleaning\_fee” and so on. However, very small numbers indicate that they are basically not correlated; there is very little association between them.

*Figure 28: Correlation matrix for the “ForRent” dataset where variables are equal to 1*



The Correlation matrix below shows where the “ForRent” column = 0. There is not much difference between these two tables, just a small difference in correlation percentage.

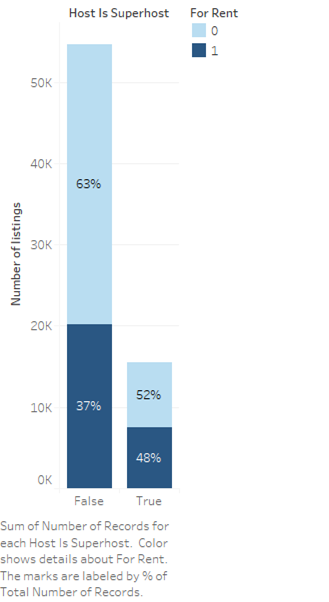
*Figure 29: Correlation matrix for the “ForRent” dataset where variables are equal to 0*



# WHAT TYPES OF PROPERTIES HAVE BEEN AFFECTED THE MOST?

In summary, we believe our analysis shows that Airbnb ‘hobbyist’ hosts tended to withdraw from the market while ‘super’ hosts remained in the profession while apartments and entire houses in Brooklyn, Manhattan, and Queens withdrew from the Airbnb market to seek longer term tenants on alternative sites. Our discussion of these conclusions is below.

*Figure 30: Comparison of listings that are available and unavailable for rent, dependent on whether or not the host is a Superhost or not.*



In order to explore the question of what factors have affected the Airbnb hosts’ market exit decision, we turned to a linear regression model to help inform our understanding. We curated the dimensionality of the listing characteristics made available by Airbnb to focus on a few areas we believed were key for our independent variables:

1) Characteristics of the Host

2) Location

3) Property Characteristics

4) Cost of Rental

5) Reviews

We transformed the raw data into a subset of listings by joining the ‘Listings’ file with the ‘Calendar’ file to create a subset of only the properties which were made available for rent during the active COVID-19 period. The calendar of availability data began in early June. We limited the rental period through the end of September as we believed that this represented a good proxy for the period for which COVID was anticipated to be active in New York City. This also allowed us to work with a data frame of much more manageable size, but still large enough to be represented by the units which were made available for supply in the market.

For our target (dependent) variable, we utilized the true and false column in the ‘Calendar’ raw data file. We transformed the true/false column into a binary 0/1 column and utilized it as our dependent variable as to whether a host made a property available (T), or said otherwise, exited the market (F).

For the characteristics of the host, we looked at their response rate, the acceptance rate of renter inquiries, whether they were a superhost (owned several Airbnb properties), the actual number of rentals offered by the host, and whether the host had verified their identity. Outside of the last variable, which was a factor of 2 levels (T,F) , these variables were all numeric.

For location, we looked at the borough in which the rental existed. While within the Airbnb data there is greater granularity in terms of individual neighborhoods within the boroughs, we felt that this led to excessive dimensionality (with over 300 neighborhoods). Focusing on the boroughs allowed us to work with a 5-factor variable which offered us insight, without overwhelming the regression.

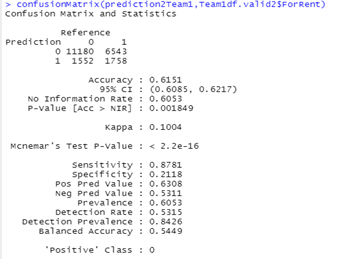
For Property Characteristics, we looked at quantitate characteristics of the rentals, such as how many people it accommodates, the number of bathrooms, number of bedrooms, total number of beds, and the square footage of the unit. In addition, we pulled in a factor variable, ‘Room Type’ which offered four levels, ‘Entire home/apt’, ‘Hotel Room’, ‘Private Room’, and ‘Shared Room’.

For cost of rental, we focused on two numeric variables, the price of the unit and the cleaning fee.

Lastly, the Airbnb data set featured robust data around the reviews received by each property. As mentioned earlier, we utilized whether reviews were available in the Pre- and Post- COVID-19 outbreak period as a proxy for whether a unit was rented. This was an important consideration in our demand model, but not so in our supply model, due to the availability of calendar listings from the firm. However, we still believed that review data would be a rich source of predictive information, so we decided to pull in ratings around the overall score, cleanliness, location, and value.

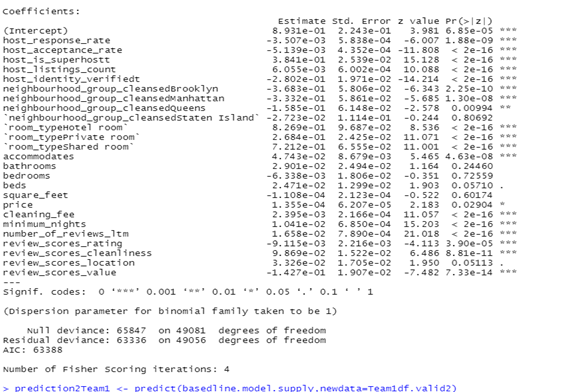
Upon running our model, we reviewed the following Confusion Matrix and Statistics:

*Figure 31: Console results of the Confusion Matrix of the demand model*

****

The factors we focused upon were able to produce a reasonable amount of predictiveness as it relates to whether a Airbnb host would leave their property on the market, or, said another way, it offered a reasonable amount of predictiveness as to whether a host would exit the market. With an accuracy of 0.6151, we were pleased with the overall performance of the model, but would have liked for a bit greater predictiveness. We had a strong sensitivity value of .8781 but were disappointed in our specificity of 0.2118. However, we knew that the model was not the only input into our desire to understand what factors have affected the Airbnb hosts’ market exit decision, but we also knew a review of the model output could help inform our inquiry. The regression output is as follows:

*Figure 32: Console output of the regression model’s coefficients and their significance*

****

In reviewing the significance of the variables, we framed their interpretation in the way described earlier:

1) Characteristics of the Host

2) Location

3) Property Characteristics

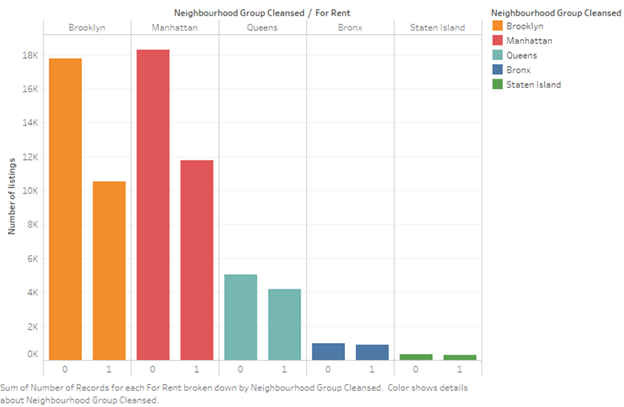
4) Cost of Rental

5) Reviews

The variables associated with the host pointed us to the conclusion that it was mainly superhosts with multiple properties who continued to list their properties, and likely non-superhosts with only one or perhaps few properties that had exited the market. This is to say that career hosts had decided to stick through the COVID-19 period, but those who perhaps listed properties as a side-job or supplemental income, looked to have exited. The positive coefficient associated with the ‘superhost’ variable and the ‘listing count’ variable alongside its very low p-value point to this conclusion. In addition, the negative coefficients and low p-values on the ‘response rate’ and ‘ acceptance rate’ point to a host who is more sophisticated in filtering inquiries on their properties. They do not have the need to respond to, nor accept, every indication of interest in their properties. This would appear to point toward more inexperienced hosts making an exit decision.

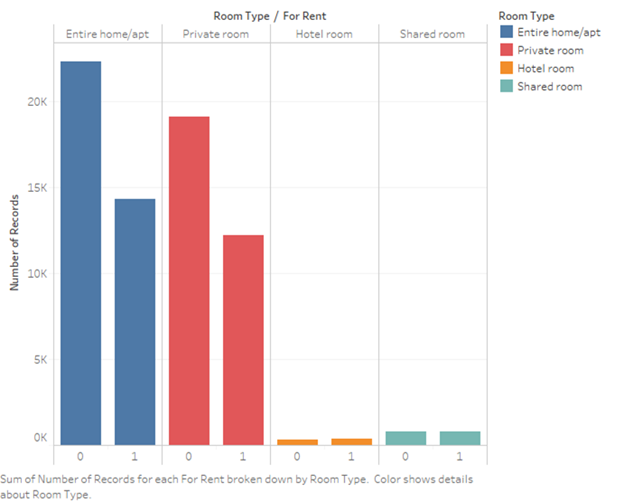
In regard to location, we saw properties in Brooklyn, Manhattan, and Queens (to a lesser extent based on the size of the coefficient) being removed at a greater pace than in the Bronx. Staten Island was not statistically significant. At first this puzzled us, but in researching the dynamics of the Airbnb market during the COVID period, we learned that many hosts pulled properties off of Airbnb and went to a more traditional long-term lease model of 1,3,6, or perhaps 12 months at a time to make it through the period. A CNBC article1 discussed this development, indicating that “more than 10,000 apartments were listed on the market in June.” In addition, an article in Forbes2 , “hosts have opted to list their properties on the long-term rental market such as Zillow, Craigslist or Apartments.com for 3-, 6- or 12-month leases.” This helped us understand our location variable, as units in Brooklyn, Manhattan, and Queens would be more marketable for long-term leases, as opposed to the Bronx or Staten Island.

*Figure 33: Bar plot comparisons of listings available for rent divided by neighborhood*



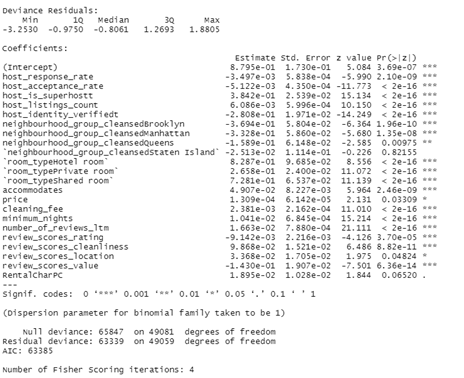
This information also played into our interpretation of our Property Characteristics, hosts with Hotel Rooms, Private Rooms, and Shared Rooms showed a statistically significant positive coefficient in regards to having the unit lists on Airbnb. As would be expected, the ability to rent a ‘Entire home/apt’ on a long-term basis, as described in the CNBC and Forbes articles, would lead to these properties being the bulk of those pulled off of Airbnb as hosts tried to rent them longer-term to cover their expenses. The characteristics of the units’ size, beds, etc. was not statistically significant to any one variable. These characteristics would be naturally thought of by a model user, so we left them in but in our final model, we decided to group these into a single PC to reduce dimensionality.

*Figure 34: Bar plot showing the different types of room available for rent*



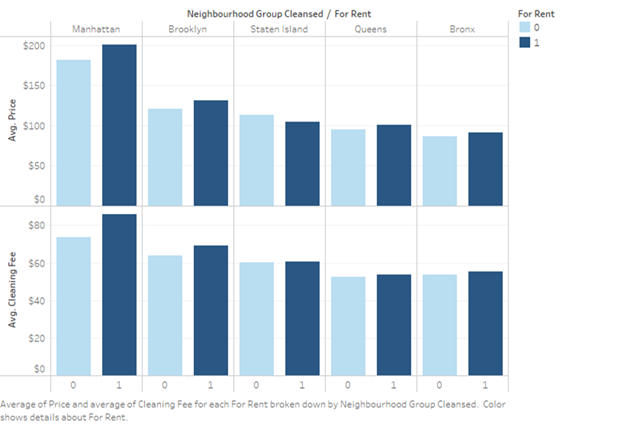
We labeled this primary component as ‘RentalCharPC’ as shown below:

*Figure 35: Console output of the regression model’s coefficients and their significance*

****

As it relates to the cost of the rental, the price of the unit was of marginal statistical significance at the 5% level to our model. This would seem to indicate, given the positive coefficient, that more expensive units were typically left on the market. This aligns with the thought that single rooms and hotel rooms tended to be left on the market, while entire apartments and houses were pulled down. Also, the cleaning fee had a positive coefficient and statistical significance. This aligns with the hypothesis that super or professional hosts tended to not withdraw from the market. Higher cleaning fees are typically shown by hosts who are more sophisticated in maximizing their net profit from their listing.

*Figure 36: Bar plot exploring the significance of the variables cleanliness and average price and how it relates to availability*



As it relates to the ‘Review’ variables, they showed various degrees of significance, with cleanliness showing a notably positive coefficient. We suspect this is related to the discussion of higher cleaning fees above. However, our VIF analysis did not show signs of multicollinearity. A negative coefficient with value is also notable, as it would seem to indicate that these are properties where the host can maximize revenue. A negative correlation would indicate the lower ratings for value still remained on the market.

# WHAT INDICATOR(S) QUANTIFY INDIVIDUAL LISTINGS’ PERFORMANCE?

Having discussed the effects of COVID-19 on hosts’ decisions to keep their properties listed (Supply), we must also explore whether the pandemic has affected the decisions that renters are making as to which properties to utilize (Demand). In order to do so, we examine data concerning the reviews that a unit received pre- and post- the COVID-19 outbreak in New York City. In our approach, we joined data from all three files made available from Airbnb, ‘Listings’, ‘Reviews’, and ‘Calendar’.

We utilized the ‘Reviews’ file offered by Airbnb and limited the data to the 2020 period. We then split the data into pre- and post- outbreak time periods, using mid-March as the dividing point for the data. After joining the ‘Review file to the ‘Listings’ file, we transformed the ‘Reviews’ column into a binary. Now pulling in the calendar file, we joined the ‘True/False’ column and filtered only for values that were ‘True’. We believe that this allows us to understand the demand from renters for certain properties.

Essentially, we created a list which shows us all properties that were for rent (‘Available’ = T from the calendar file) and then looked at whether renters chose to rent the properties (‘Post-Covid Review > 0). With this approach, we were able to build a model which helped us understand which property characteristics were important to renters. Said otherwise, we were examining demand.

We utilize the same independent variables as in our ‘Supply model, as discussed above. To refresh the reader, we looked at the following major characteristic groups:

1) Characteristics of the Host

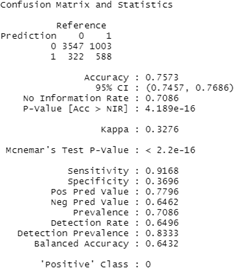
2) Location

3) Property Characteristics

4) Cost of Rental

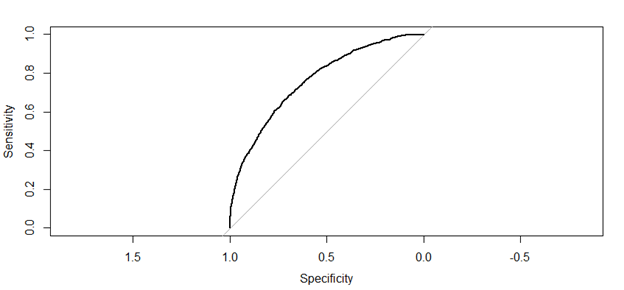
5) Reviews

We were pleased with the predictive ability of our model, as shown below:

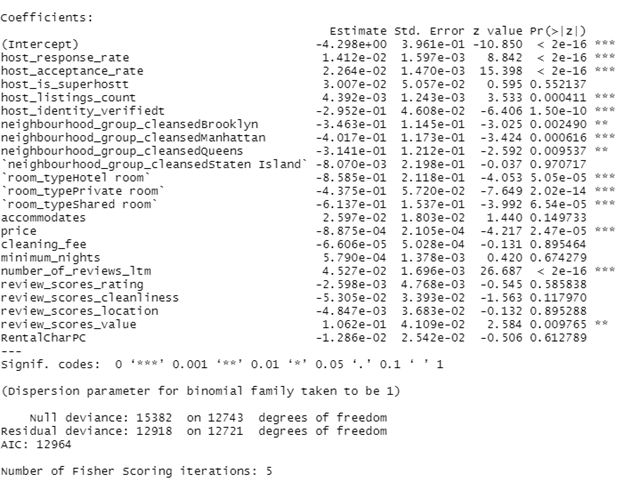
*Figure 37: Console output showing the confusion matrix and statistics of the final model*

Area under the curve: 0.7609

*Figure 38: ROC curve of the final model*

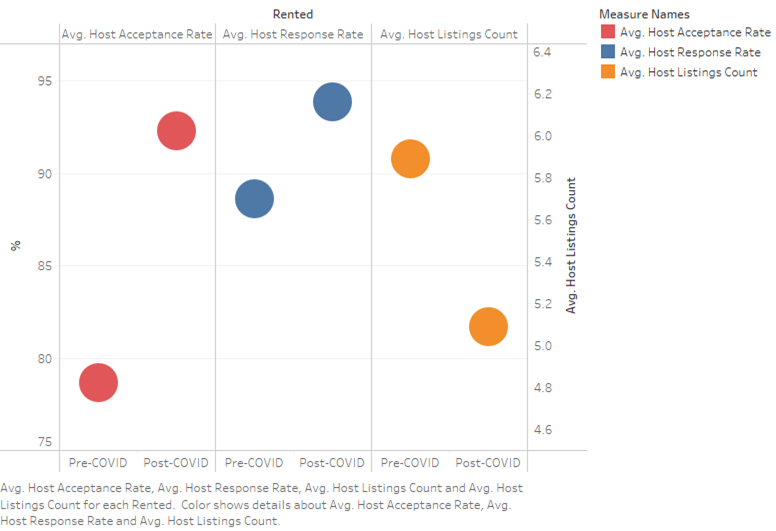
 In considering what types of properties were affected the most, we now can understand both the supply hosts were making available and the characteristics of properties that renters were demanding.

*Figure 39: Console output of the final regression model’s coefficients and their significance*



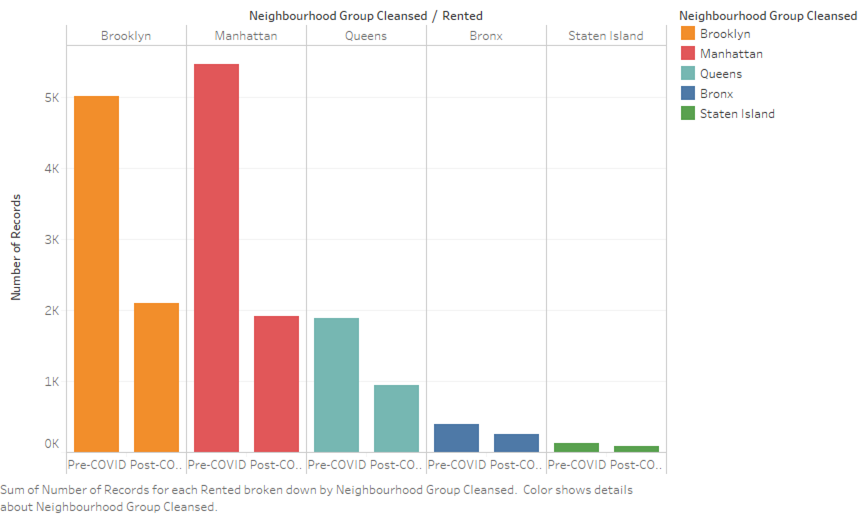
As it relates to the characteristics of the host, renters showed a clear preference for those hosts who had a high response rate and a high acceptance rate, as evidenced by the positive coefficients for these variables, coupled with a statistically significant p-value. Renters also favored hosts who had a higher number of listings, though the designation as a ‘superhost’ was not of statistical significance to renters. In looking at these variables holistically, it appears that service matters to renters in the post-COVID outbreak period even more so than in the pre-COVID period.

*Figure 40: Circle plot exploring the preference of renters for high acceptance and response rate variables.*



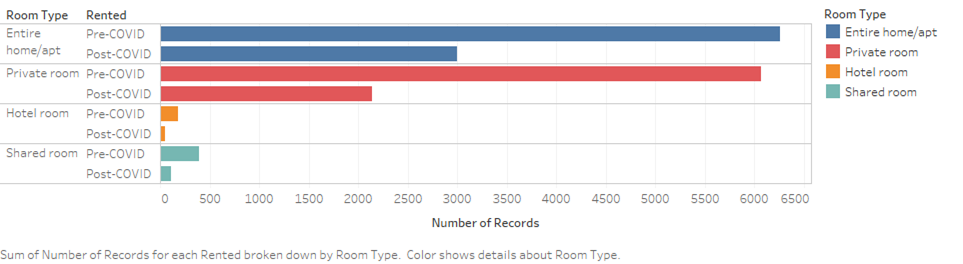
Not surprisingly, we saw a drop-off in the listings concentrated boroughs of Brooklyn, Manhattan, and Queens, particularly in Manhattan. As discussed in our ‘Supply’ model description, these boroughs saw the greatest amount of listing retraction relative to the Bronx and Staten Island during the outbreak. Additionally, these boroughs would be most suitable for tourists visiting New York City attractions, so it is somewhat predictable that they would see the biggest amounts of demand drop-off relative to the Pre-COVID-19 outbreak period.

*Figure 41: Bar plot showing decrease in rentings divided by neighborhood*



As evidenced in our ‘Supply’ discussion, hosts continued to make ‘Hotel Rooms’, ‘Private Rooms’, and ‘Shared Rooms’ available, but we see in our ‘Demand’ model that renters were not interested in these property types. All three types show a negative coefficient with a strong level of statistical significance. This outcome makes sense to us. Given the high level of transmission potential of this virus, renters didn’t trust to be in shared accommodations and favored units which didn’t expose them to others.

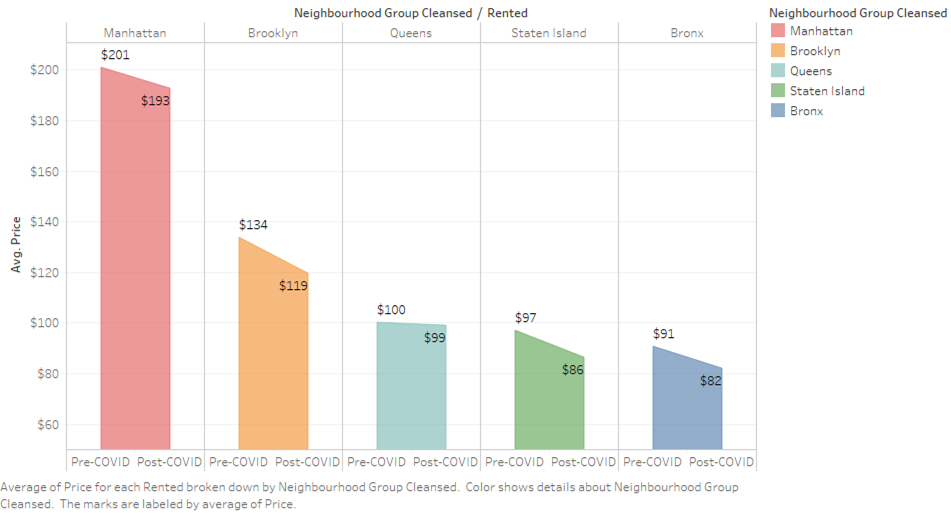
*Figure 42: Number of Records in Pre-Covid and Post-Covid by Room Type*



In considering both the cost of the rental, it is clear that renters showed a strong preference for lower priced units, as evidenced by the large negative coefficient associated with the ‘price’ variable and its statistically significant status. There was no significance as it relates to the cleaning fee. Given the uncertainties around the virus and the tremendous drop-off in demand for units, it is of little surprise that renters demanded lower prices during the outbreak. This phenomenon is also evidenced in the CNBC and Forbes articles referenced previously.

This preference for lower prices is also reflected in our ‘Review’ variables. The only ‘Review’ variable which showed significance was the ‘review scores – value’ variable. This further reinforces the notion that renters continued to demand good value during outbreak.

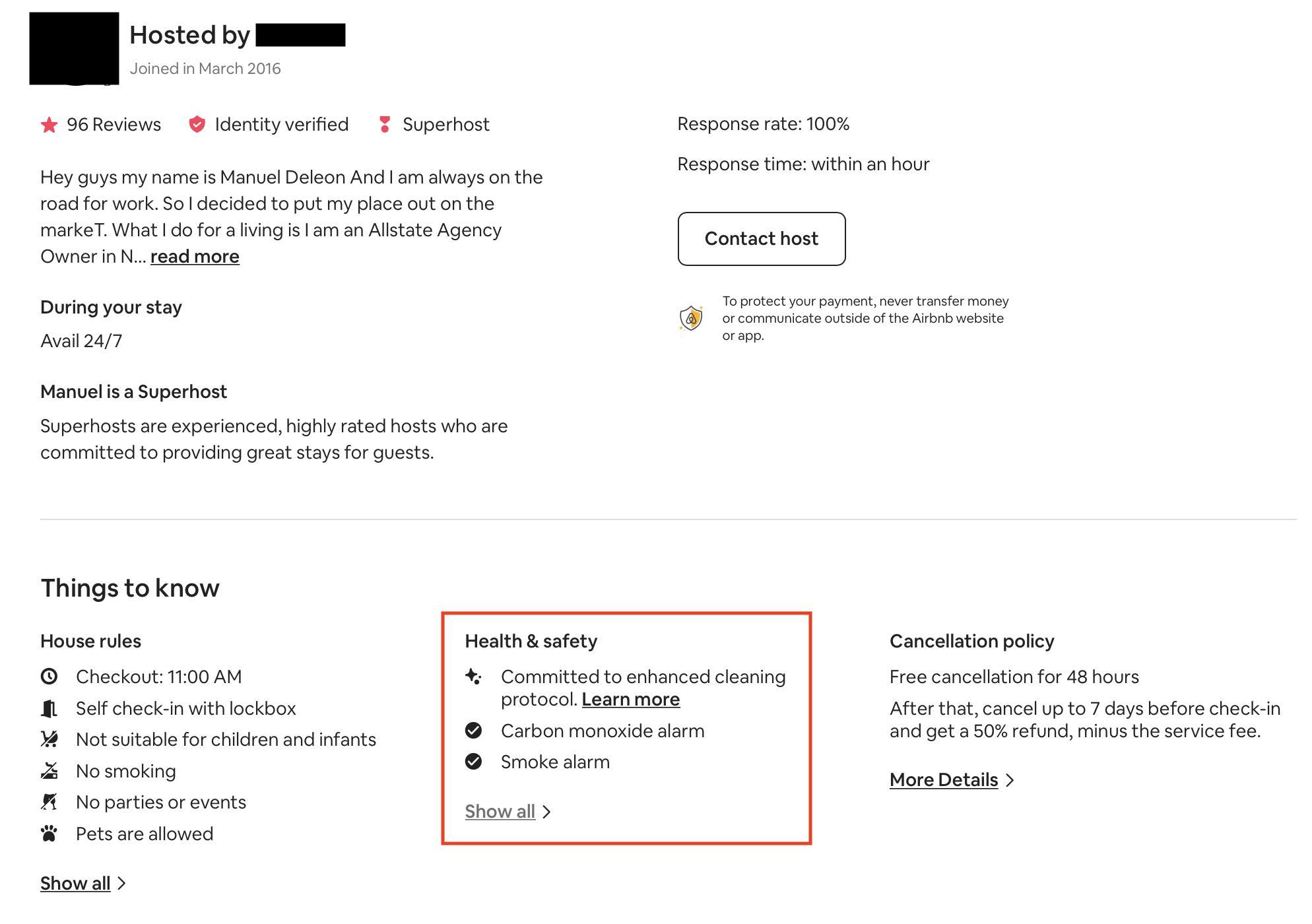
*Figure 43: Average price drop of Pre-Covid and Post-Covid by neighborhood*



# 

# DATA COLLECTION SUGGESTIONS FOR FUTURE ANALYSES

*Figure 43: Screengrab of AirBnb’s new health and safety feature to show which Superhosts correspond with new COVID-19 cleaning protocols*



# In mid-June 2020, Airbnb introduced a covid-related Health & Safety protocol that “includes a five-step cleaning process, room-by-room checklists, and more. The guidelines include how to use personal protective equipment, how to clean every room in a home, what to sanitize, and what supplies to have on hand. If hosts reviewed the guidelines, completed a quiz, and adhered to the guidelines, they will receive a special call-out on their listing so guests know that they are committed to an enhanced cleaning and sanitization routine.” (Airbnb News) For future analyses, it would be interesting to see how the cleaning protocol affected the market demand and see if it played a big role in a listing getting rented. It would also give more insight to the customers’ comments, because although cleanliness is important, it would help show how valuable sanitization was to a renter.

# SUMMARY/CONCLUSION

In forming our conclusions as to which listings were most affected by the outbreak, we considered the previous ‘Supply’ discussion, coupled with the results of our ‘Demand’ model listed above. There were a few clear takeaways from our work.

The properties most affected were those which required to share a space with a stranger, ‘Hotel Rooms’, ‘Private Rooms’, and ‘Shared Rooms’. While hosts pulled their ‘Entire home/apt’ type listings to seek longer term rental agreements, Airbnb renters had strong demand for the greater isolation of these accommodations. We also saw a clear drop-off in demand from tourists and pleasure-seekers which are more associated with the boroughs of Brooklyn and Manhattan, while demand in the outer boroughs of the Bronx and Staten Island remained relatively steady. These outer boroughs are more likely demanded by out of town workers and those with some local ties, given their distance from Manhattan.

Our data also helps us understand that ‘superhosts’, those more likely to utilize Airbnb as their primary source of income, continued to put their units up for rent. Meanwhile, hosts that likely utilize Airbnb as a supplemental source of income tended to pull their listings as a result of the COVID-19 uncertainty. While these professional hosts, with their high ratings across multiple categories, continued to supply their units, the smaller contingent of renters sought out properties that offered greater value as reflected in those reviews.

# APPENDIX

**CODE**



All code can also be found at this link: <https://drive.google.com/file/d/11k0RzI55WPN_JM-Enrg4w9D93JRLMpV4/view?usp=sharing>

# SOURCES

Business Insider (Founding of Airbnb)

<https://www.businessinsider.com/how-airbnb-was-founded-a-visual-history-2016-2#in-march-2009-the-company-finally-scrapped-the-air-bed-and-breakfast-name-and-simplified-it-to-airbnb-no-more-confusing-associations-with-air-mattresses-16>

Wikipedia (Airbnb Timeline)

<https://en.wikipedia.org/wiki/Timeline_of_Airbnb>

World Health Organization (Covid-19 Timeline)

<https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19>

(COVID-19 Timeline)

<https://en.wikipedia.org/wiki/COVID-19_pandemic_in_the_United_States#March_2020>

Data Derived From “Inside Airbnb”

<http://insideairbnb.com/get-the-data.html>

Cases|News:

<https://www.nytimes.com/2020/07/12/us/florida-coronavirus-covid-cases.html>

<https://covid19tracker.health.ny.gov/views/NYS-COVID19-Tracker/NYSDOHCOVID-19Tracker-Map?%3Aembed=yes&%3Atoolbar=no&%3Atabs=n>

<https://www.wsj.com/articles/first-case-of-coronavirus-confirmed-in-new-york-state-11583111692>

<https://www.nejm.org/doi/full/10.1056/NEJMoa2001191>

<https://ourworldindata.org/coronavirus>

# Empty Manhattan apartments reach record levels, landlords slash rent <https://www.cnbc.com/2020/07/09/empty-manhattan-apartments-reach-record-levels-landlords-slash-rent.html>

# How Bad Are Covid-19 Pandemic Effects On Airbnb Guests, Hosts?

# <https://www.forbes.com/sites/lealane/2020/06/09/how-bad-are-covid-19-pandemic-effects-on-airbnb-guests-hosts/#34d9b7be7432>

Airbnb Cleaning Protocol  
<https://news.airbnb.com/introducing-airbnb-enhanced-clean/>