

# Exploration of NYC Airbnb



## **Team Name**

Data Drinkers

## **Team Members**

Aayush Jain

Alexs Wijoyo

Madhura Jadhav

Mihir Zala

## Table of Contents

<b>1</b>	<b>Introduction .....</b>	<b>3</b>
<b>2</b>	<b>Data Preprocessing .....</b>	<b>3</b>
<b>3</b>	<b>Exploratory Data Analysis .....</b>	<b>3</b>
	<b>3.1 Listing and Price Comparison for all 5 boroughs .....</b>	<b>3</b>
	<b>3.2 Price Comparison for all private rooms in neighborhoods .....</b>	<b>6</b>
	<b>3.3 Do the criteria set by Airbnb for Superhost are strictly followed? .....</b>	<b>8</b>
	<b>3.4 Exploring Superhost Across Boroughs .....</b>	<b>10</b>
<b>4</b>	<b>Conclusion .....</b>	<b>13</b>

## Introduction

Airbnb is an online marketplace that connects people who want to rent out their homes with people looking for accommodation in that locale. It has massively disrupted the travel accommodation and rental market in the past decade. NYC being the most populous city in the United States, and one of the most popular tourism and business places globally, we decided to explore and visualize NYC Airbnb dataset using R and Tableau.

Airbnb has been providing unique opportunities for both entrepreneurs and travelers. For the first part of our analysis, we explored various aspects of Airbnb, such as room types, accommodations, location, amenities for entrepreneurs that are looking to host their property on Airbnb as well as factors that can help travelers make best choice for their stay.

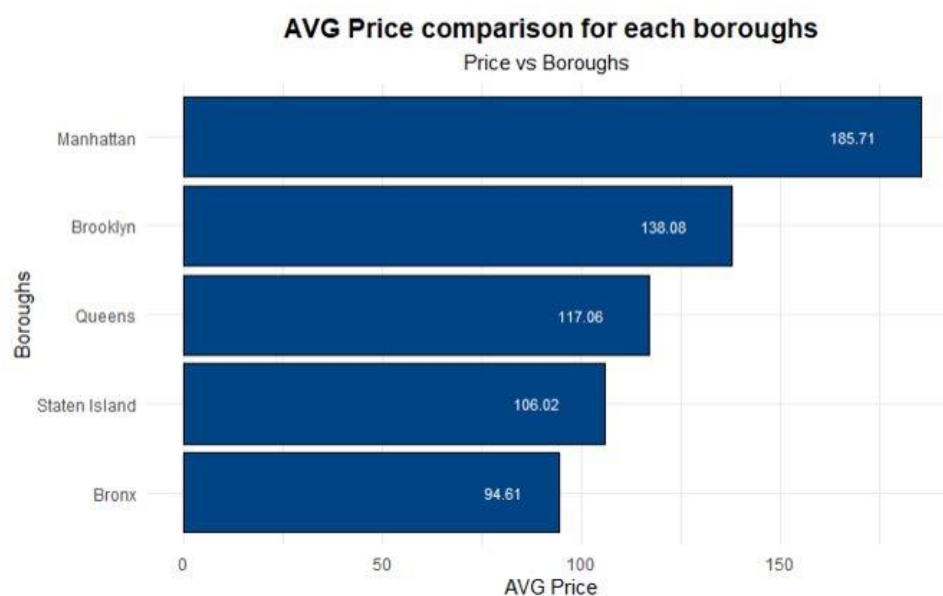
For the second part, we decided to explore the Super Host status of Airbnb hosts. We explored the data to see if the conditions set by Airbnb for Superhost status are strictly followed for all the hosts? We also analyzed the concentration of Superhosts in NYC map and looked deep to see how being a Super Host holds value as a consumer and entrepreneur.

## Data Preprocessing

The dataset we used is from 'Inside Airbnb', containing 37,713 records and 74 attributes. We preprocessed the data and removed 29 attributes that were either redundant or out of scope for our analysis. We transformed attributes like `host_response_time`, `host_response_rate`, `price` by removing special characters and converted their data type from character to numeric. As major part of our analysis focuses on Superhost status, we removed 31 NA records from 'host\_is\_superhost' attribute to get clean dataset for effective analysis.

## Exploratory Data Analysis

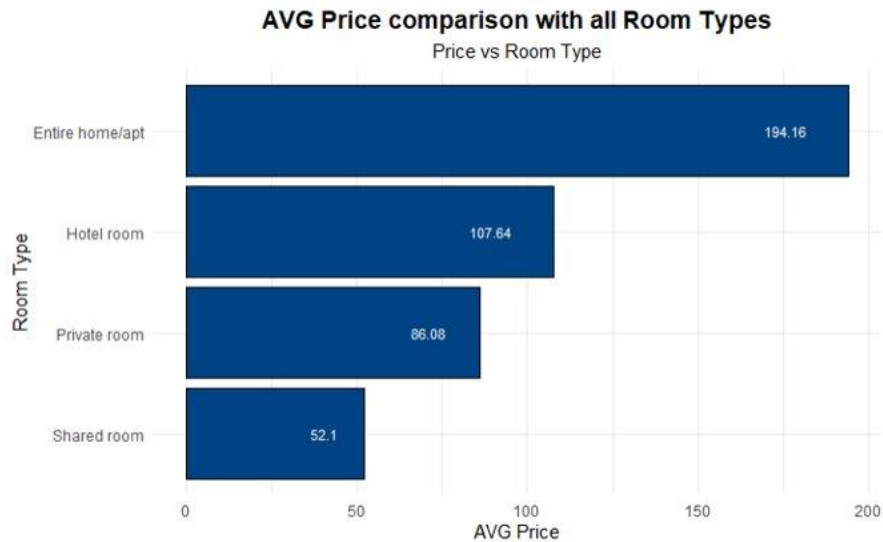
### Listing and Price Comparison for all 5 boroughs (Aayush)



What if a person wants to become a host by using your existing property or wants to rent out a property and then list it on Airbnb.

This analysis shows us AVG pricing for a listing throughout NYC. We can make assumptions from this analysis that

1. Average listings price is the highest for Manhattan (211.72 USD) followed by Brooklyn (136.43). One possible reason for high average price in Manhattan could be that whole apartments/home are the most common type of listings there (we will see an analysis on this in the upcoming slides).
2. Bronx has the cheapest listings with an average price of 105.75 USD.



This analysis shows us the price comparison of different types of rooms that are offered throughout NYC.

Analysis: Average price is the highest for hotel rooms followed by Entire home/apartment which is quite expected because of the services offered by a hotel and security when compared it to an entire home/ apartment. Therefore, according to this analysis, shared rooms have a higher average price because the number of rooms available throughout NYC is less and the frequency at which they are booked is lower.



This graphical analysis shows us the type of listing available in each borough ranging from entire home, private room, shared room, or hotel room.

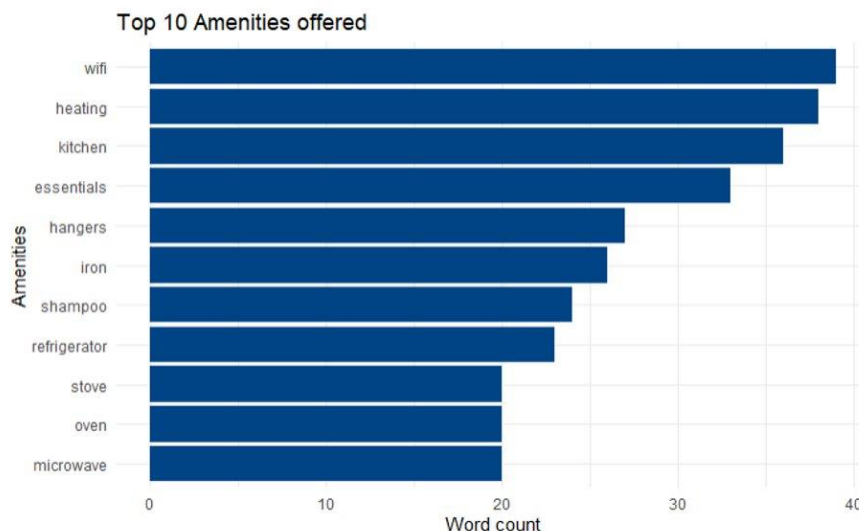
Assumptions:

1. Private rooms are the most common listing type in all neighborhoods except Manhattan where entire Home/apartment is the most common type.
2. Hotel rooms are the least common in all neighborhoods (reason: hotel rooms have a quota to certain number of rooms on Airbnb but the prices are usually higher because of the services offered by the hotel).



This analysis shows us what are the top 10 most expensive neighborhoods for an Airbnb and if a person is planning to be a host, this is how they should price the property.

In reference to the previous analysis where we compared prices of all the 5 boroughs, avg price might be higher in Manhattan but when we look at this analysis the most expensive neighborhood is not in Manhattan. This can be because of the size of property and the area, for example properties in Fort Wadsworth, has views of Manhattan skyline.



This analysis shows us what are the 10 most common amenities a host should offer. For example, Wi-Fi, heating, kitchen, essentials and these are some of the important factors which matter to the customers as some of these are advertised in listing descriptions

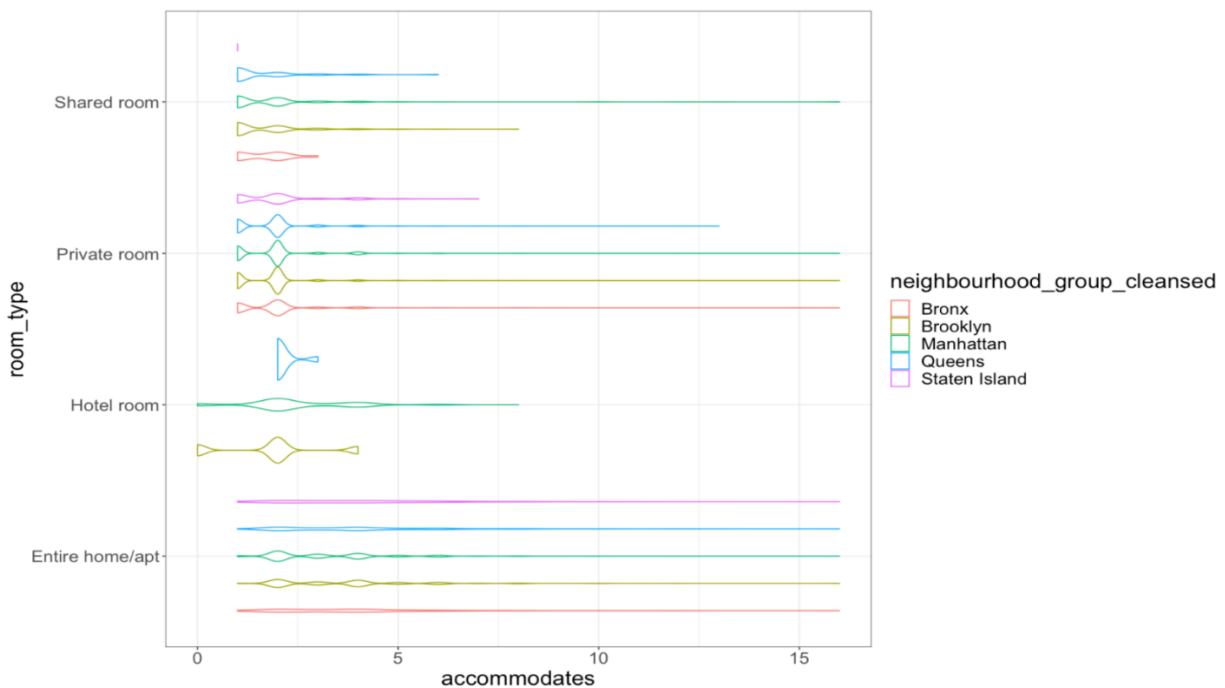
### Comparing the price range of private rooms in neighborhoods (Mihir)

This graph shows the comparison of private rooms. By seeing the boxplot graph, we can state that Manhattan and Brooklyn are the two neighborhoods who have the highest number of private rooms with the price range from \$50 to \$1000. In Manhattan, Brooklyn, and Queens we can see the major number of private rooms and In Bronx compared to other 3 have less rooms and Staten Island have a smaller number of private rooms with price range near \$50 to somewhere \$600. Where we can have the lowest range as well, but the number of private rooms is less in that price range.



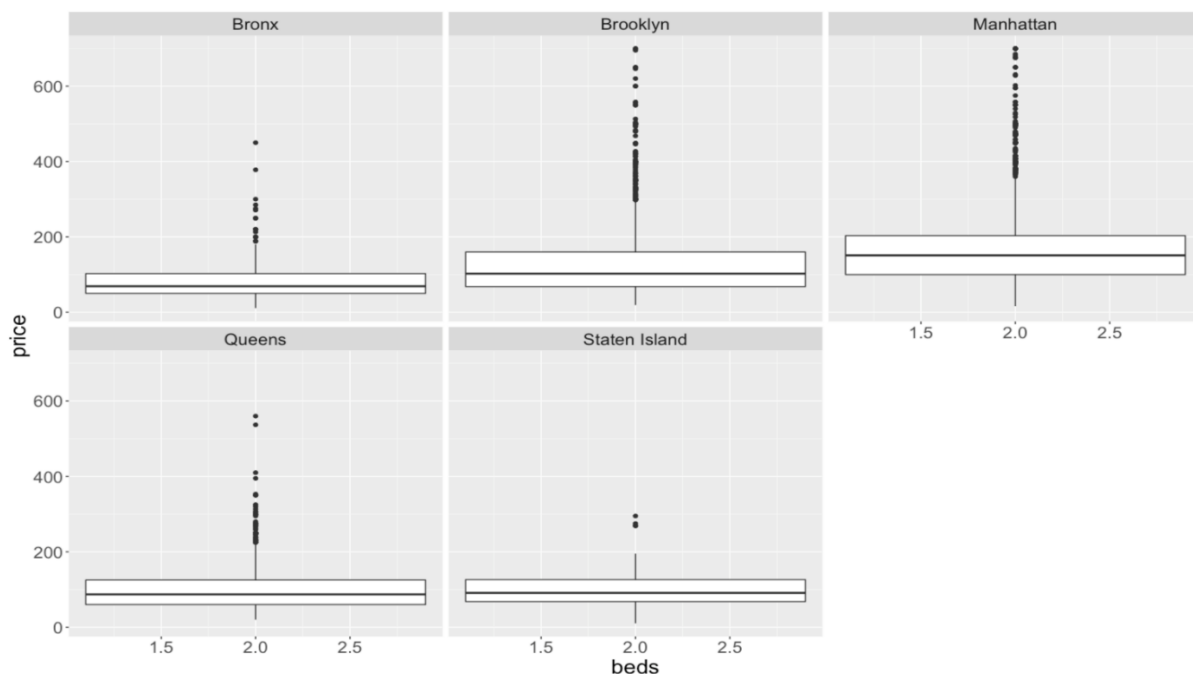
### Types of rooms can accommodate most people

In this visualization, we see that which type of rooms can accommodate more people? So, a shared room in Manhattan can accommodate more than 15 people after that Brooklyn can accommodate not more than 7 people. Shared rooms in Queens can allow 5 people. But, in Staten Island it only accommodates 1 person and in Bronx which can accommodate only 2 people, whereas private rooms in Queens and Staten Island can accommodate a smaller number of people but other Three neighborhoods can accommodate more people. Hotel rooms in Manhattan, Queens, and Brooklyn can hardly accommodate more than 3 people as they range from 0. In entire home/apt we can see in every neighborhood more than 15 people can accommodate.



Someone comes to New York City and searching for beds with their appropriate budget.

For example, if someone plans to visit NYC for vacation and they are searching for rooms with 2, 2.5 or 3 beds where they can live for months or for few days with their appropriate budget, so here is a boxplot on beds and price which states that in Bronx prices range start from \$200 to near \$500 but in Brooklyn and Manhattan we can see the difference in price because of the views and the more tourist's sites these both neighborhood has the highest price range and highest rooms. In Queens there are rooms which cost from \$200 to \$600 and Staten Island only rooms cost around \$250 to \$300.



## Do the criteria set by Airbnb for Superhost are strictly followed? (Madhura)

While looking for an Airbnb, we are often hit with a few different selections that often have the same amenities and descriptions, however the differentiating factor is the fact that the listing has superhost status or not. This is where we come in to see what it means to be a superhost.

The creation of the Airbnb Superhost program was a response to the lack of consistent experiences on the platform. Airbnb's earlier rating system proved an inadequate way of deciphering between professional quality operators and novice hosts. With 80% of Airbnb properties achieving at least a 4.5-star rating, it was difficult for a guest to confidently select a premium property. The Superhost program fills this void by creating a more comprehensive quality metric that considers a host's entire portfolio of properties.

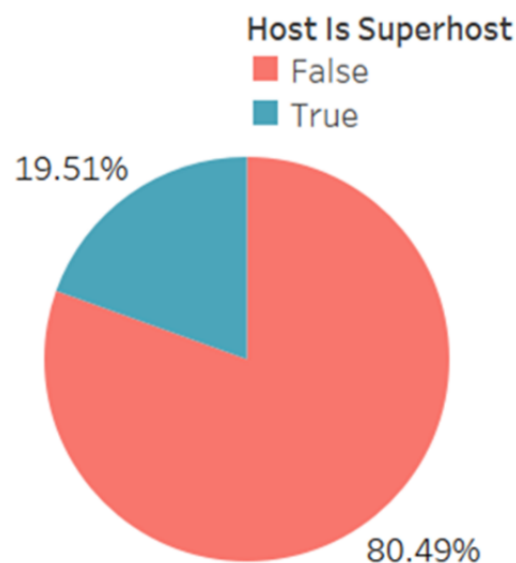
Based on Airbnb website, there are 4 criteria that a host must meet to become a Super Host:

1. **4.8+ overall rating:** Superhosts have a 4.8 or higher average overall rating based on reviews from their Airbnb guests in the past year.
2. **90% response rate:** Superhosts respond to 90% of new messages within 24 hours
3. **<1% cancellation rate:** Superhosts cancel less than 1% of the time. This means 0 cancellations for hosts with fewer than 100 reservations in a year.
4. **10+ stays:** Superhosts have completed at least 10 stays in the past year or 100 nights over at least 3 completed stays.

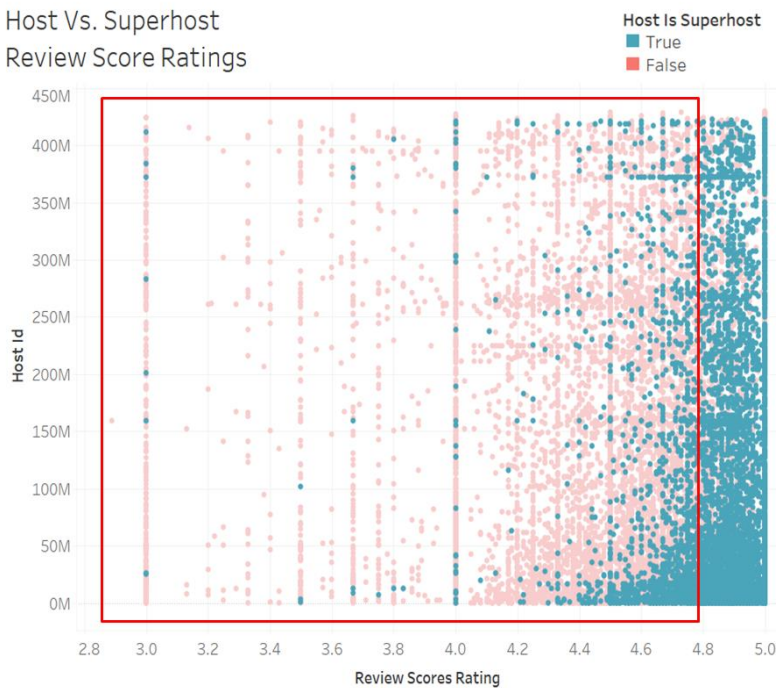
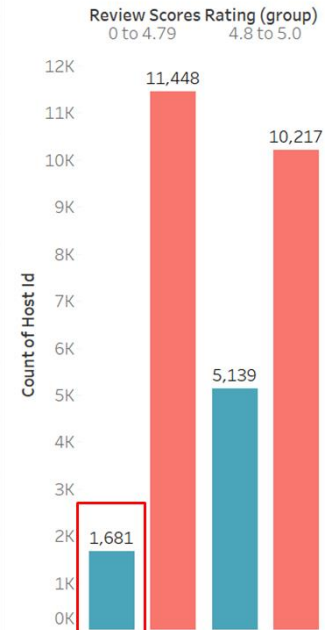
Out of these, we were able to get data for just two criteria focused on ratings and response from the host.

For NYC dataset, we found about Host vs. Superhost distribution to be 80%-20%. Out of total 37682 hosts, just 7353 qualifies to be a Superhost.

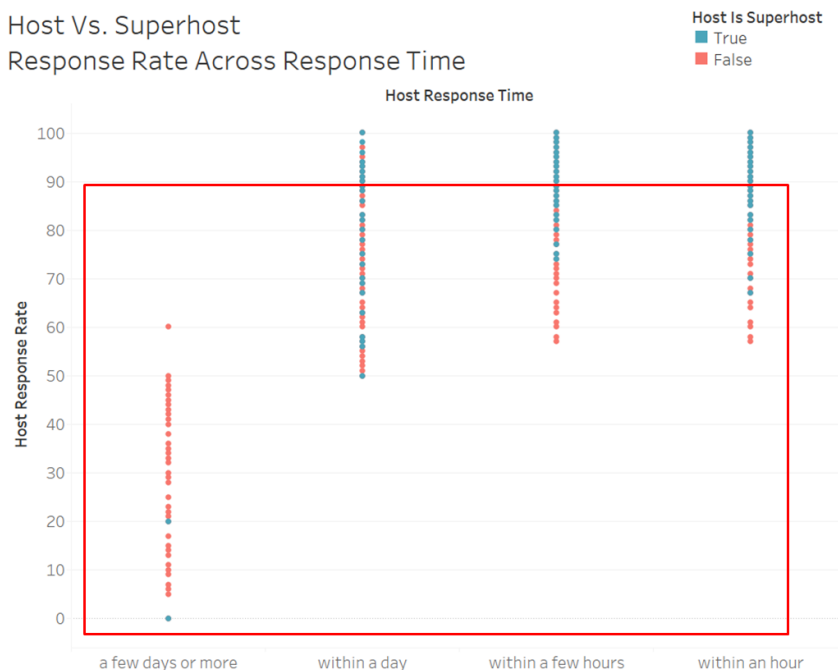
### Host Vs. Superhost



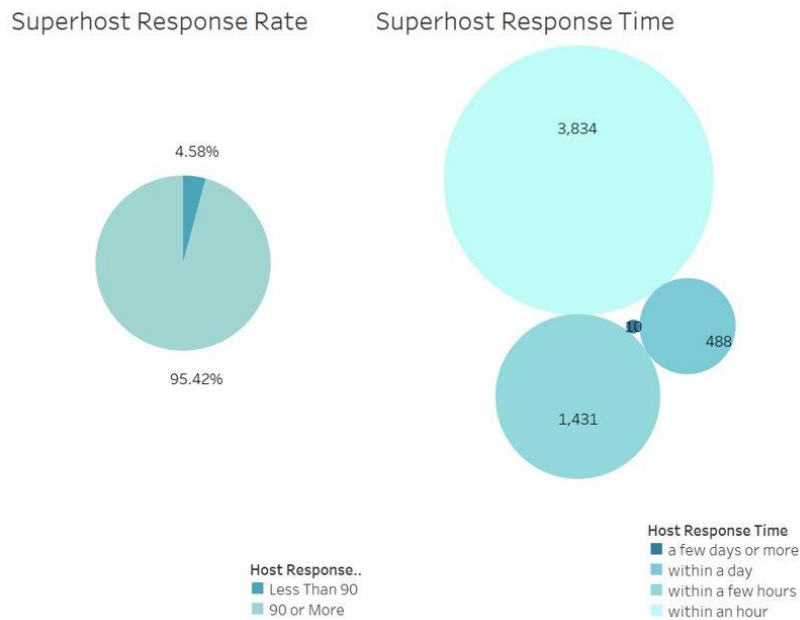


**4.8 or higher overall ratings**Host Vs. Superhost  
Review Score RatingsHost Vs. Superhost  
Review Score Ratings

The scatter plot has hosts placed across ratings on the X-axis. The teal color dots in the box here represents superhost that have ratings less than 4.8. To explore deeper, the bar chart shows that there are around 1681 superhosts i.e., about 23% of the superhost don't qualify for this criteria.

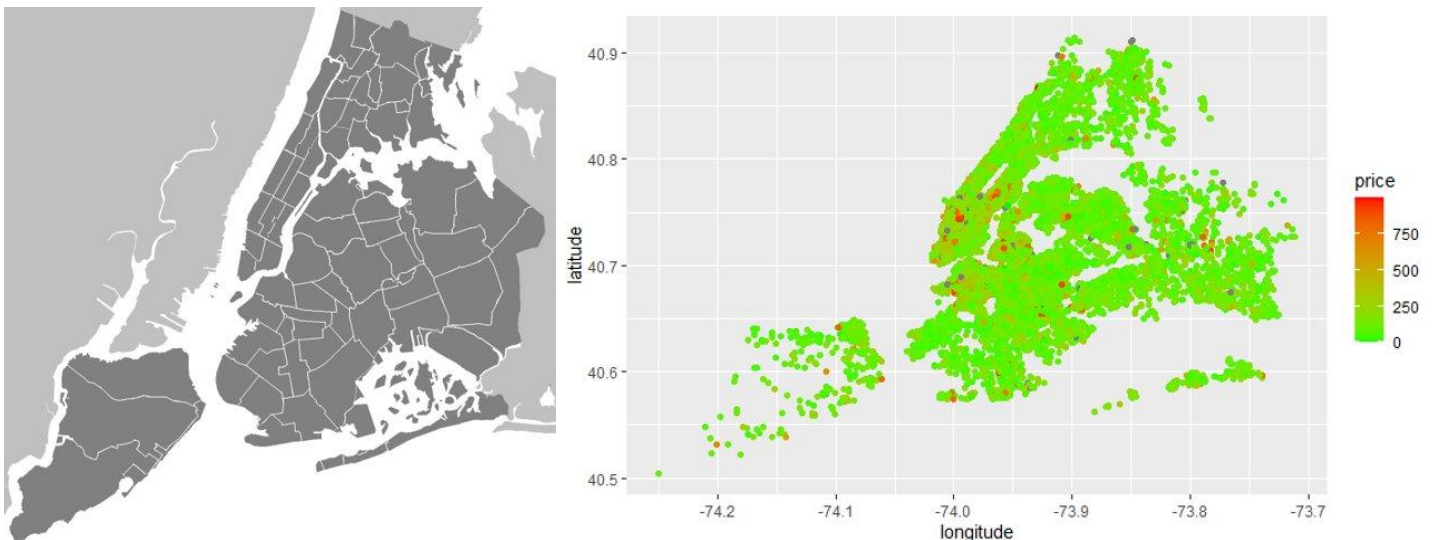
**Respond to guests within 24 hours and maintain a response rate 90% or higher**Host Vs. Superhost  
Response Rate Across Response Time

Based on this plot, it is clear that some Superhosts highlighted in box do not exactly follow this criteria. They are either not 'responding within 24 hours' or have 'lesser than 90% response rate'.

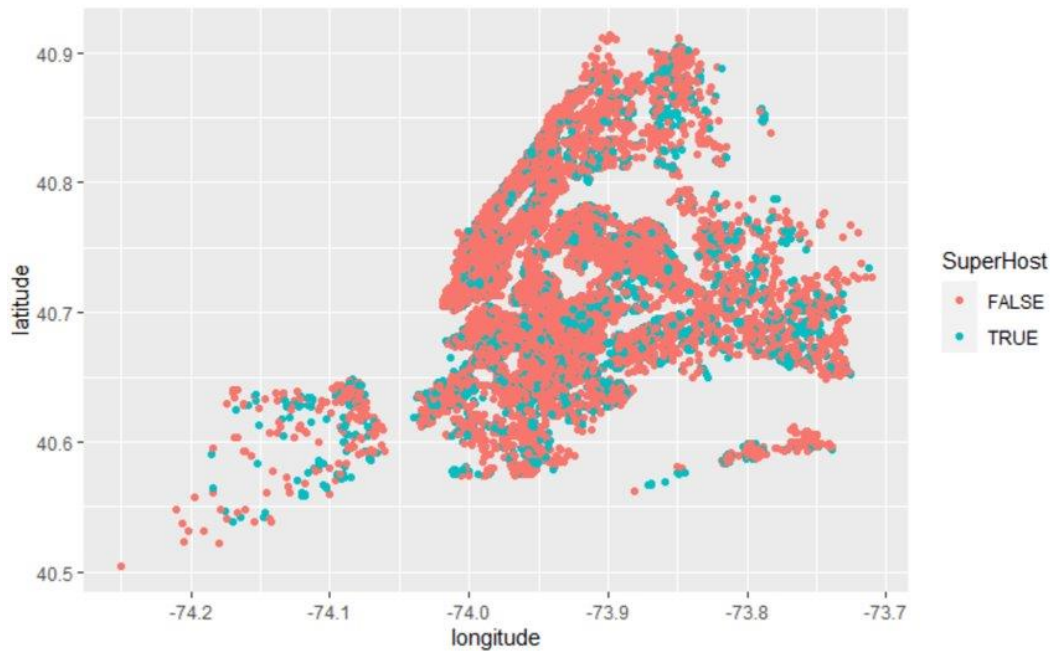


Upon taking a deeper look, a small fraction of superhosts i.e., 10 superhosts we are taking "a few days or more" to respond. Also, about 5% of superhosts are having lesser than 90% response rate.

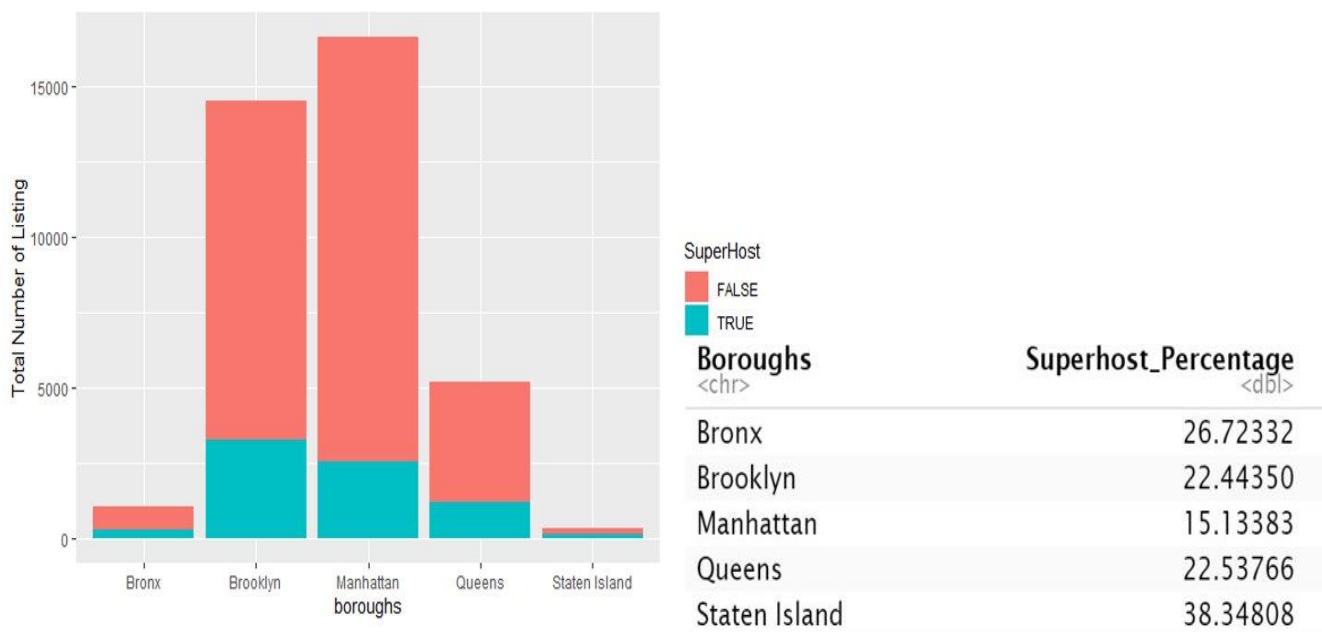
## Exploring Superhost Across Boroughs (Alexs)



As a resident within the New York City area, we all know that Manhattan is the heart of the city. It is filled with all the exciting adventure, the historical buildings, train stations and many different tourist attractions that bring them to New York City repeatedly. With our initial interest of looking at the Airbnb data through the spectacle of being a super host or not, we must first understand the dispersion of other data Metrix in comparison. This is intended to fight or prove some common assumption we have about Airbnb in our city. First, we look at the pricing data

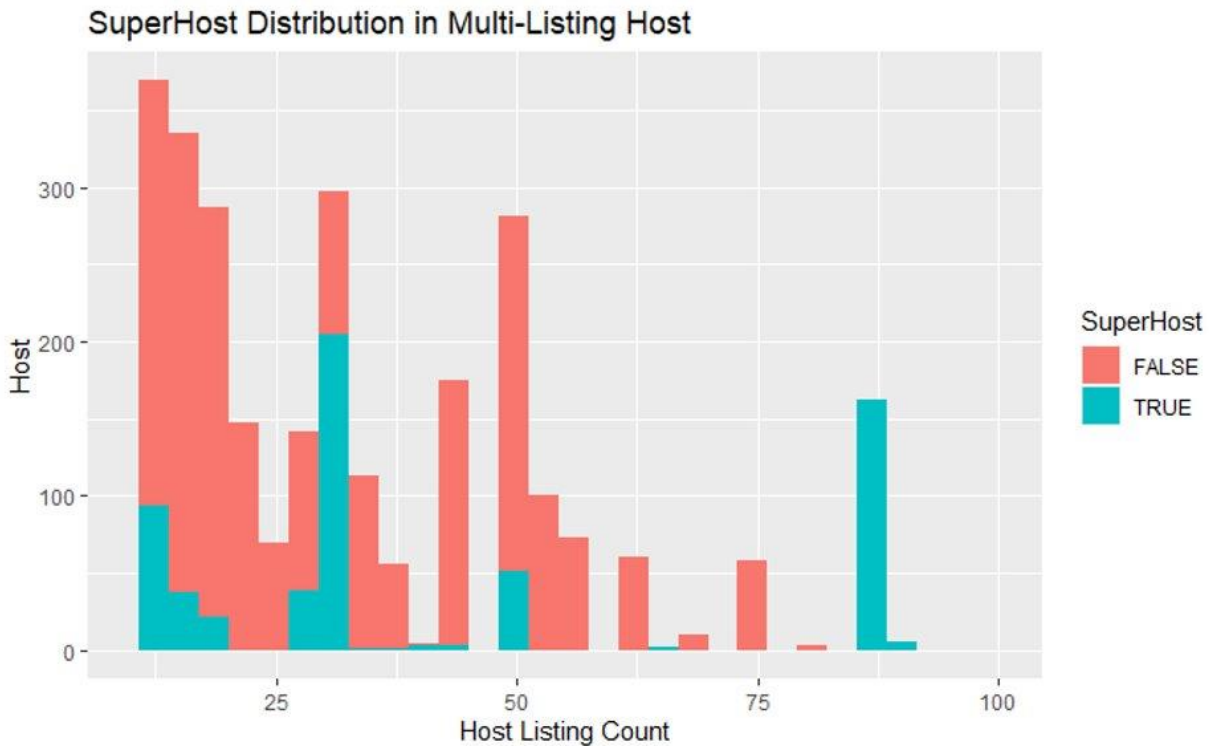


As we explore the pricing data in comparison to many different boroughs, while we do see a density forming within the lower tips of the Manhattan borough, we see also a wider scatter of high-priced Airbnb in the other residential area which is very interesting. We see places in Jamaica as well the tip Staten Island, which is interesting since these areas are the further from other location as well has a very long commute to the city. Then this made me wonder about the super host spread in the city, hoping that the super host would care about their listing more we would assume that the majority would be located where the money maker is, the city.



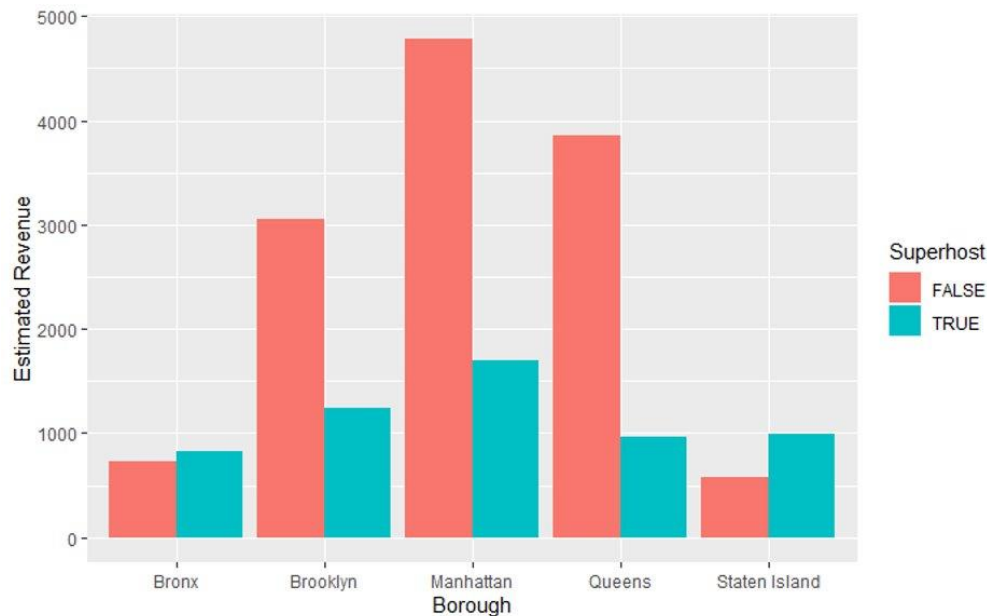
But then, we were proven wrong, with the number of super hosts spreading around the city. This is a very interesting phenomenon which puts the questions of what the host's game plan are owning these Airbnb around the city and becoming super host, while the big proportion of the Manhattan borough is not super hosted. If we check the proportion of the data, with Staten Island and the Bronx having the highest proportion of super hosted listings, this validates the idea

we had about the Airbnb host having different kind of strategy that they are applying to their listings. Further but cheaper or they accumulate more people that are more loyal and just stay with them due to the host. For that reason, we were more curious, the host that has more listings place within the city would care more about their locations and listings.



But although this number shows a visible characteristic of that behavior, we still see a very large host which owns more than 5 different listings not being a super host. And that begs the question why do people who have more listings not want to be super host? We associate with taking a proactive approach such as getting verified on social media and getting a certain badge on our eBay shop as a sign of slightly higher importance, but this is not displaying 100% on the dataset. Can there possibly be other motives by the host?

neighbourhood_group_cleansed <chr>	host_is_superhost <chr>	TotalListing <int>	AvgReviewsPerMonth <dbl>	avgrating <dbl>	medianminnights <dbl>	avgprice <dbl>	EstRevenueMonth <dbl>
Bronx	FALSE	776	1.930729	4.473352	5	75.0	724.0232
Bronx	TRUE	283	3.961599	4.852305	3	70.0	831.9357
Brooklyn	FALSE	11255	1.095990	4.537526	30	93.0	3057.8123
Brooklyn	TRUE	3257	2.826470	4.852032	4	110.0	1243.6468
Manhattan	FALSE	14109	1.106646	4.474959	30	144.0	4780.7087
Manhattan	TRUE	2516	2.753322	4.841631	4	154.0	1696.0463
Queens	FALSE	4011	1.838718	4.482317	30	70.0	3861.3086
Queens	TRUE	1167	3.487788	4.859680	3	92.0	962.6295
Staten Island	FALSE	209	1.643500	4.623625	4	87.5	575.2250
Staten Island	TRUE	130	3.434048	4.859762	3	96.5	994.1568



We were able to find some clues that could have something to do with a flat monthly profit strategy. By using a 30-day minimum stay, hosts do not need to care about being super host or not as they need one person to book their Airbnb for the entire month. Do not get me wrong, this move has its advantages and disadvantages. With the host applying this revenue cap, regardless of the time of year, they cannot potentially earn more, their only move is to increase the prices per night before they book the listing. This revenue cap is remarkably interesting as it is displayed within a group that is not a super host. For the super host, their limit is endless, the blue bar shows their minimum estimated revenue and as more people come during the holidays, they can become very much more profitable than the other strategy.

## Conclusion

As we go through many of the dataset provided by Airbnb, we see much different convincing evidence that were able to prove how popularity of certain areas would affect pricing as well as other aspects of listings. However, it also disproved many kinds of speculation that one would have when they hear about Airbnb in New York City. Superhost was a concept within the system that we were interested in, and it played a role. Even though we found a small number of violations, but they exist. Maybe there are some other exception rules to superhost criteria which are not known. But based on data findings, it is probably safe to say both the requirements laid out by Airbnb themselves might not be strictly followed. Also, Superhost is a factor that allows you to determine the true intention of the host, whether they are trying to provide a better experience to a person or just pursuing Airbnb as a business venture. None the less, it was a remarkably interesting find.