

Exploratory Data Analysis



Airbnb bookings in New York City

Team Gladiator

Ankit Kumar Gupta

Ganta Praneeth Kumar

Rohit Bhat

CONTENTS

1. Introduction

2. Problem Statement

3. Dataset Analysis

4. Plot Analysis

5. Conclusion



Abstract

Since 2011, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. Today, Airbnb became one of a kind service that is used and recognized by the whole world. Data analysis on millions of listings provided through Airbnb is a crucial factor for the company.

These millions of listings generate a lot of data and in our experiment we have analyzed this data which can help us in taking various business decisions, understanding of customers' and Host's behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

Introduction

Airbnb (ABNB) is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in specific locales.

The company has come a long way since 2011, when its co-founders first came up with the idea to invite paying guests to sleep on an air mattress in their living room. According to Airbnb's latest data, it has in excess of 6 million listings, covering more than 100,000 cities and towns and 220 plus countries worldwide.

How Airbnb Works

Airbnb has revolutionized the hospitality industry. Prior to 2011, travelers would have likely booked a hotel or hostel for their trip to another town. Nowadays, many of these same people are opting for Airbnb.

The idea behind Airbnb is simple: Find a way for local people to make some extra money renting out their spare home or room to people visiting the area. Hosts using this platform get to advertise their rentals to millions of people worldwide, with the reassurance that a big company will handle payments and offer support when needed. And for guests, Airbnb can offer a homey place to stay that has more character, perhaps even with a kitchen to avoid dining out, often at a lower price than what hotels charge.

How Does Airbnb Make Money?

Airbnb's business model is quite profitable. essentially making money renting out property that it doesn't own.

Every time a reservation is made, Airbnb takes a cut. When you click on a property, you'll find to the right of the page a breakdown of the fees you'll be charged if you go ahead and book. One of these fees is a service fee, which covers the cost of running the platform and providing support; this basically makes up the bulk of Airbnb's revenue.

Problem Statement

For this project we are analyzing Airbnb's New York City data of 2019 . As we know New York City is the world's 11th biggest city (Population wise) and Airbnb is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in specific locales. Dataset is provided by millions of listings that happened on the Airbnb website.

- **Our main objective is to find the key metrics that influence the listings of properties. And our major focus will be on below mentioned points.**
 - ❖ Top 10 host id with most number of listings ?
 - ❖ Which neighbourhood_group has the most number of listings?
 - ❖ Which room types people prefer in each neighbourhood_group?
 - ❖ Find the average price in every neighbourhood_group with respect to room type?
 - ❖ What is the maximum , minimum for each neighbourhood_group ?
 - ❖ Most Expensive host with less availability ?
 - ❖ Top 5 hosts with high profit ?
 - ❖ Total no. of nights spent per room types ?
 - ❖ Most Expensive hosts in neighbourhood_group which are suitable for business meetings?
 - ❖ Which quarter has the most number of listings ?
 - ❖ Which quarter has the most avg number of reviews?
 - ❖ What is the average availability in each Quarter with the respective neighbourhood_group ?
 - ❖ Which quarter has the highest profits earned?
 - ❖ List out the most expensive host for each quarter?
 - ❖ Find out the correlation between the features?
 - ❖ Find which part of New York City has more listings and their geographical advantages?

Dataset Analysis

This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values.

- ❖ **Id**- It provides a unique id to each listing in the dataset.
- ❖ **Name**-It describes the property in a concise manner.
- ❖ **Host_id**-It is used to ascertain the ownership of a certain individual
- ❖ **Host_name**-It is the name of the property owner
- ❖ **Neighbourhood_group**-It basically categorizes neighborhood into five groups Bronx,Brooklyn,Manhattan,Staten Islands and Queens
- ❖ **Neighborhood**-It designates the location of a property.
- ❖ **Latitude**-It Specifies the Latitude of the property.
- ❖ **Longitude**-It denotes the Longitude of the property.
- ❖ **Room_type**-It classifies the type of property into Private_Room,Shared_Room and Entire_Apartment
- ❖ **Price**- It provides the price of each listed property.
- ❖ **Minimum_Nights**-It gives the minimum night one needs to pay for in the property.
- ❖ **Number_of_reviews**-It is used to ascertain the number of reviews received by a property.
- ❖ **Reviews_per_month**-It denotes the number of reviews per month.
- ❖ **Calculated_host_listings_count**-It corresponds to the number of properties hosted by the unique host_id
- ❖ **Availability_365**-The number of days the property is available in a year.

New York City Map



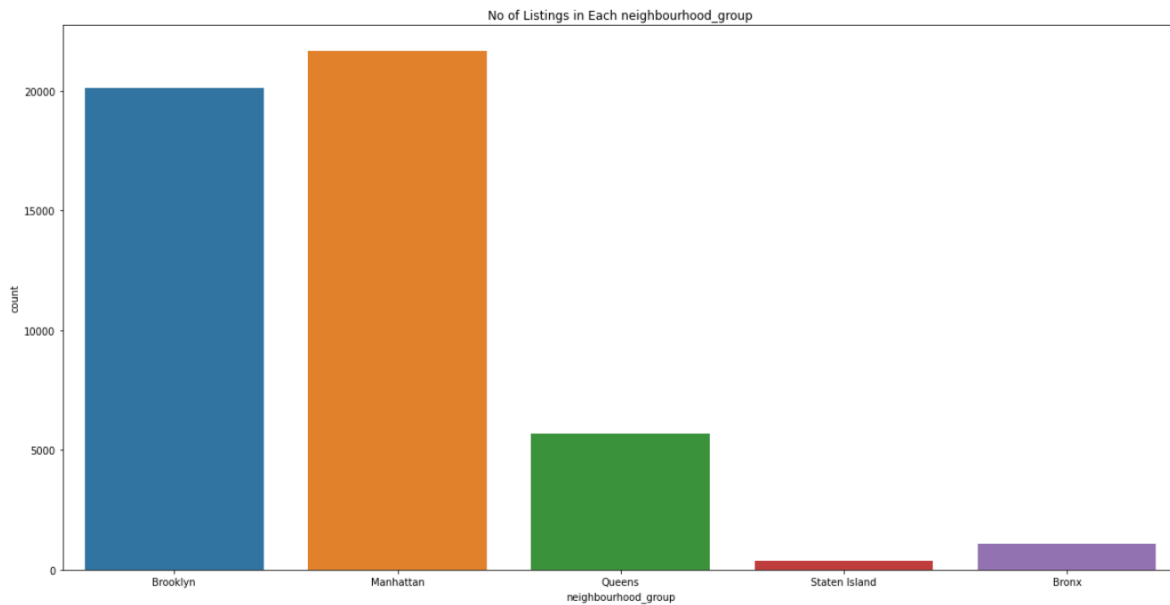
Plot Analysis

1. Top 10 host id with most number of listings.

	host_id	number_of_listings
0	219517861	327
1	107434423	232
2	30283594	121
3	137358866	103
4	16098958	96
5	12243051	96
6	61391963	91
7	22541573	87
8	200380610	65
9	7503643	52

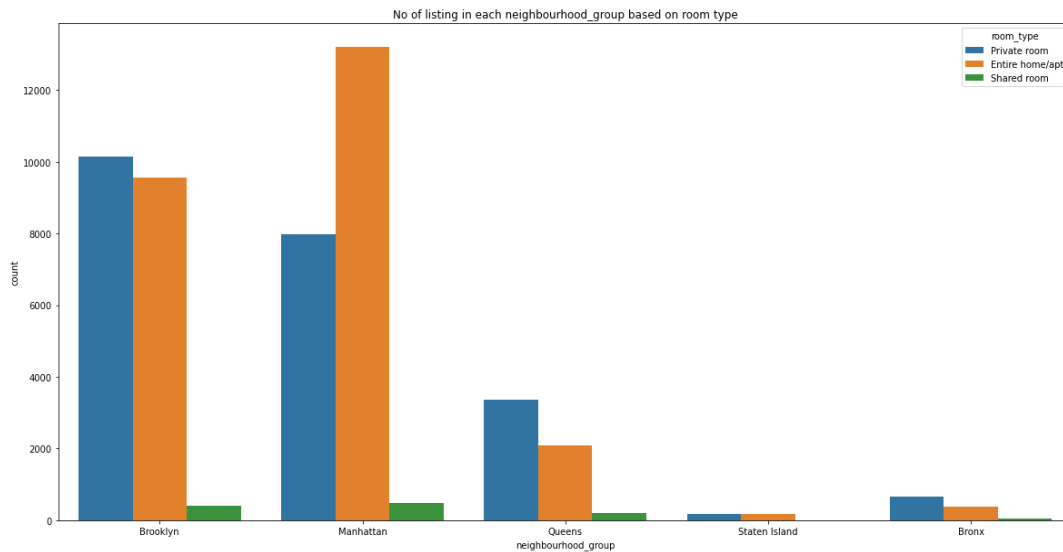
- The country with the most Airbnb listings is the USA, which recorded 877,694 active listings in 2019 alone.
- In May 2019, an average of 2 million bookings were made every day
- Made a data frame of top 10 most no of listing host id in this visualization and the host id 219517861 has the most no of listing (327).

2. Which neighbourhood_group has the most number of listings?



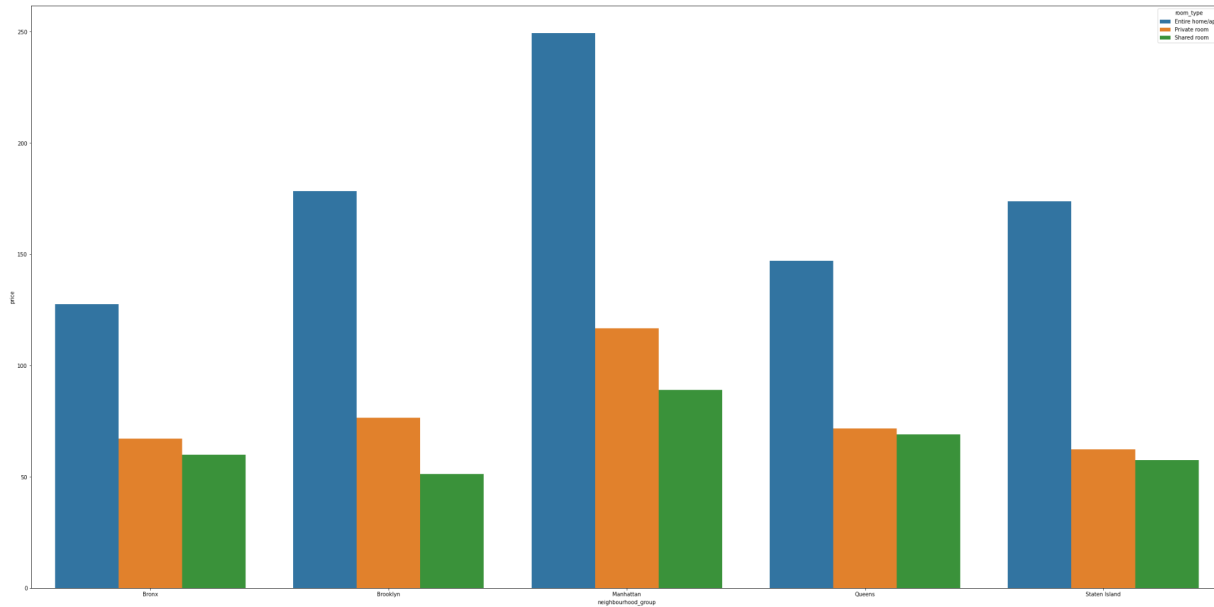
- According to the survey of New York City USA it is found that the number of neighborhood listings is most in Manhattan followed by Brooklyn, Queens, Bronx and Staten Island.
- In neighborhood listing Manhattan has more than 20000 hotel listings on Airbnb platform .

3. Which room types people prefer in each neighbourhood_group



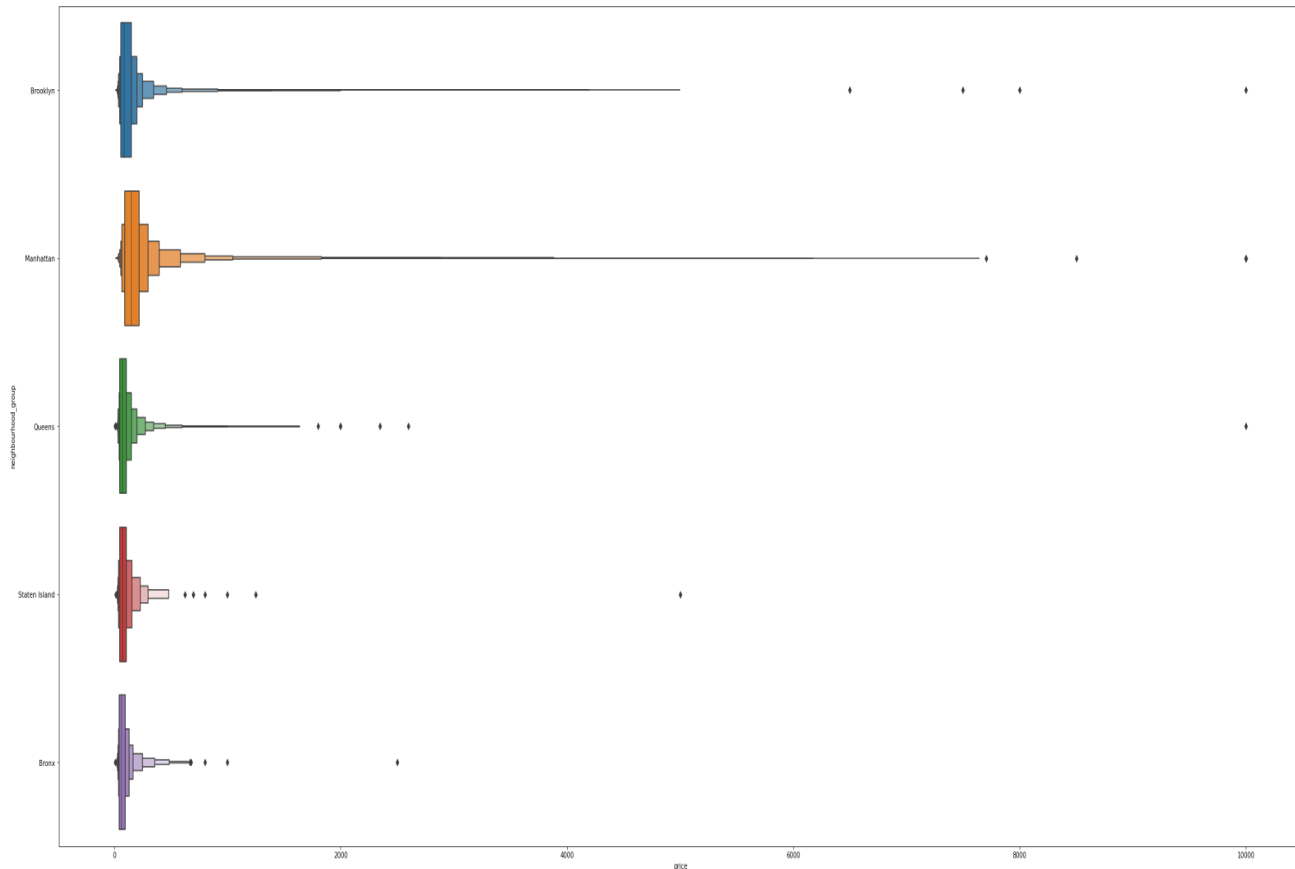
- According to the survey I had found that each room type the Entire home apt has the most no of preferred by people was in Manhattan and followed by Private room was in Brooklyn and so on in this graph visualization.
- Shared room is the least preferred by people in the Bronx.

4. Find the average price in every neighbourhood_group with respect to room type



- According to the survey I had found the different average price model of each room type and I had made a data frame with each room type with each neighborhood and I found that the entire home apt has the highest average price (249.25) was in Manhattan and followed by Entire home apt was in Brooklyn.
- Shared room has the lowest average price (51.26) in Brooklyn.

5. What is the maximum , minimum for each neighbourhood_group ?



- According to my survey I had found the max and min room price of each neighborhood and as I had seen these 3 places (Manhattan, Brooklyn, Queens) have the maximum room price (10000) and followed by Staten Island and Bronx.
- And Manhattan, Brooklyn, Queens, Bronx have the minimum room price (10) followed by Staten Island.

6. Most Expensive host with less availability

	name	host_id	host_name	neighbourhood_group	room_type	price
17692	Luxury 1 bedroom apt. -stunning Manhattan views	5143901	Erin	Brooklyn	Entire home/apt	10000.0
9151	Furnished room in Astoria apartment	20582832	Kathrine	Queens	Private room	10000.0
6530	Spanish Harlem Apt	1235070	Olson	Manhattan	Entire home/apt	9999.0
3049	Spacious center hall colonial	9430973	Donna	Staten Island	Entire home/apt	700.0
8521	1 Room in a 2 Bedroom Available	7074531	Jonathan	Bronx	Private room	680.0

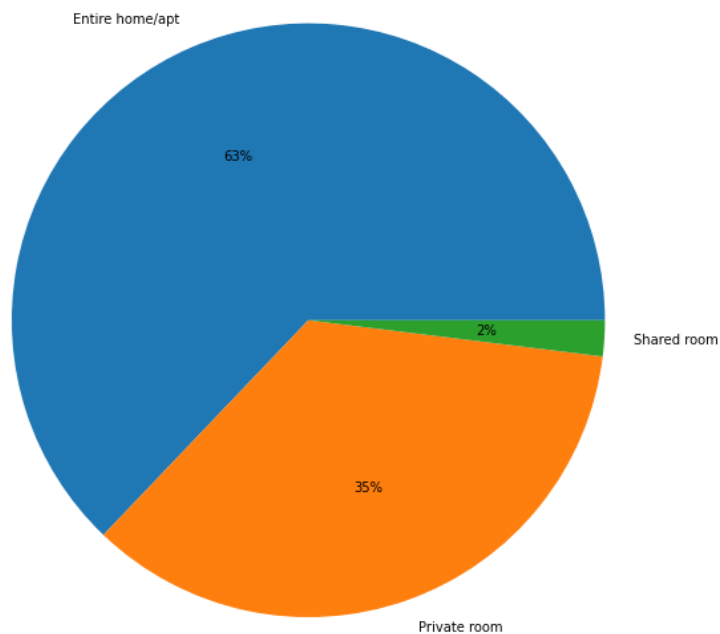
- In this Survey we have tried to find the most expensive as well as busiest host of all
- Erin from brooklyn comes out to be the winner here. She has a luxurious one bedroom apartment with stunning Manhattan views. She is the busiest host of all due to which the rent is on the higher side (Demand-Supply Concept).

7. Top 10 host with high profit

		price
host_name	host_id	
Sonder (NYC)	219517861	82795.0
Blueground	107434423	70331.0
Sally	156158778	37097.0
Red Awning	205031545	35294.0
Kara	30283594	33581.0
Sonder	12243051	20451.0
Jeremy & Laura	16098958	20060.0
Henry	836168	19500.0
Pranjal	200380610	18865.0
Kevin	3750764	18780.0

- Here we have tried to find the top 10 most profitable hosts of all.
- This survey will help the company to do the behavioral analysis on customer preference while choosing the host. It will allow the company to adopt the practices followed by these top 10 hosts and will apply the same on other hosts who are not so profitable.

8. Total no. of nights spend per room types



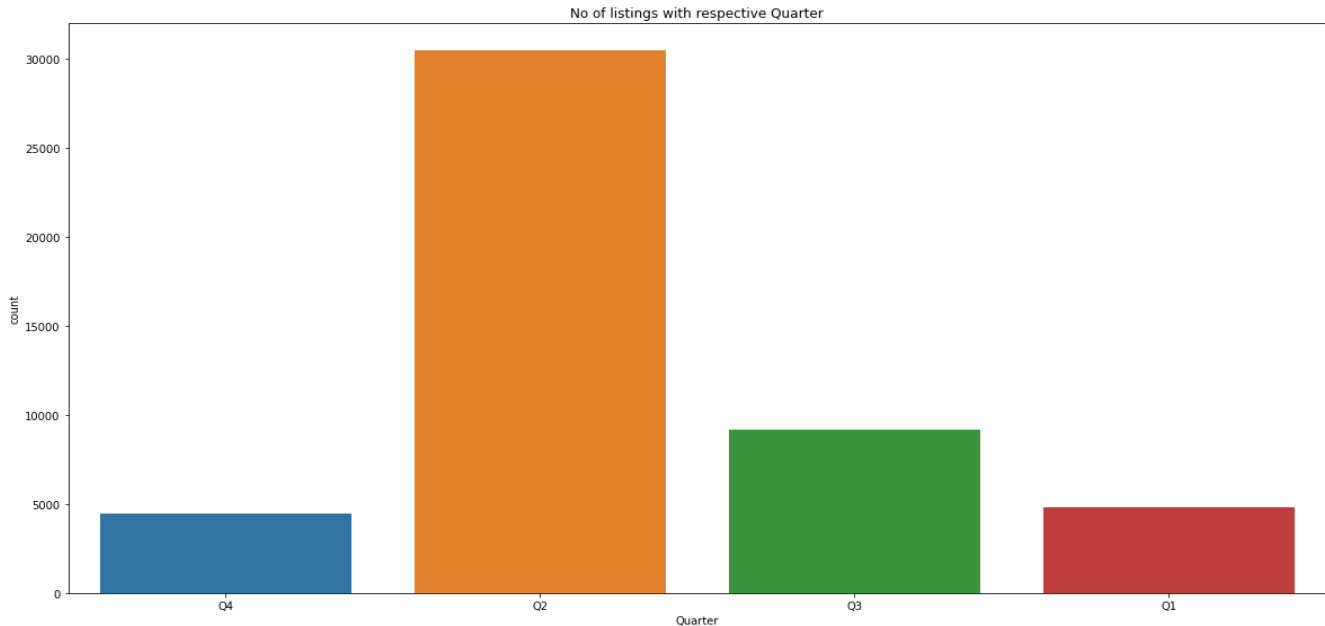
- In this survey we have done the analysis to find the preference of the customers with regards to the choice of type of rooms for their stay.
- Surprisingly even after the high cost customers have preferred the entire home (63%) for their stay followed by private room (35%) and least preferred is shared rooms (2%)
- This behavior of the customer indicates the luxury lifestyle of New York City and also that the customer takes his privacy seriously and is ready to pay a luxury amount for the same.

9. Most Expensive hosts in neighbourhood_group which are suitable for business meeting

	name	host_id	host_name	neighbourhood_group	room_type	price
29238	1-BR Lincoln Center	72390391	Jelena	Manhattan	Entire home/apt	10000.0
17692	Luxury 1 bedroom apt. -stunning Manhattan views	5143901	Erin	Brooklyn	Entire home/apt	10000.0
9151	Furnished room in Astoria apartment	20582832	Kathrine	Queens	Private room	10000.0
22353	Victorian Film location	2675644	Alissa	Staten Island	Entire home/apt	5000.0
24477	"The luxury of Comfort"	131826530	Kathy	Bronx	Private room	2500.0

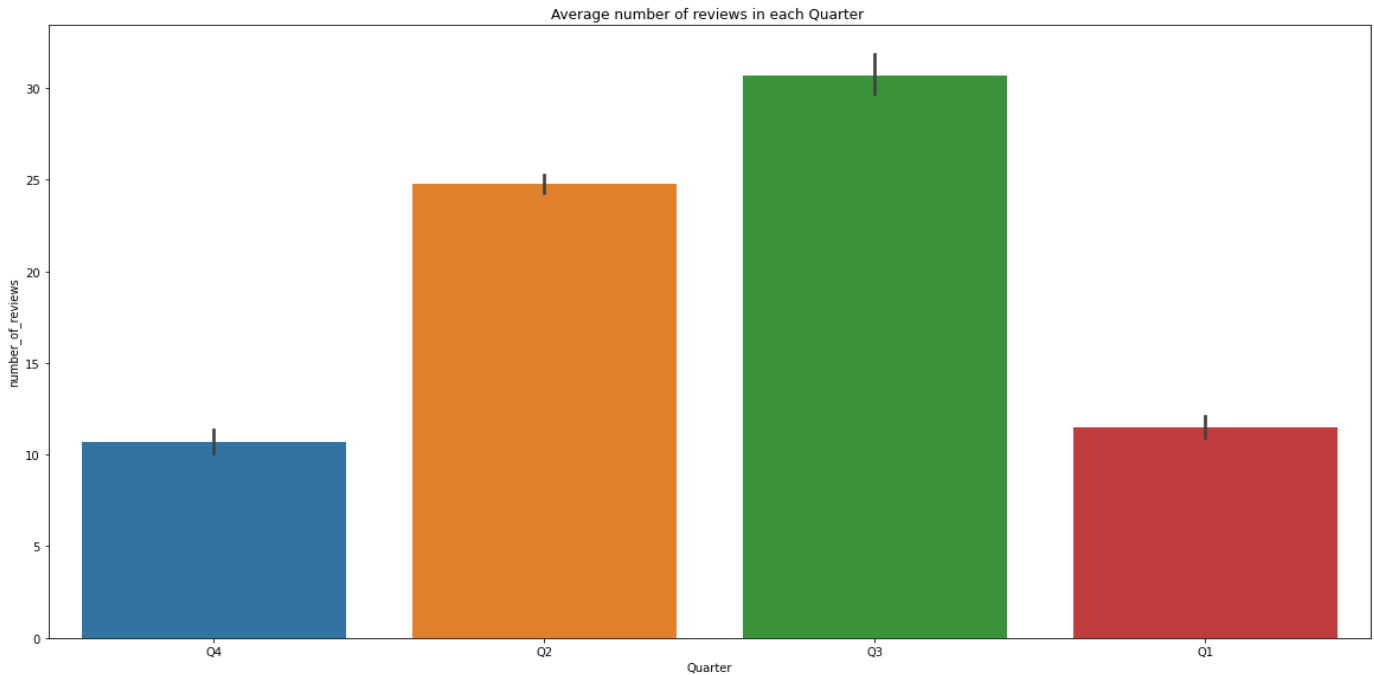
- In this survey we are finding the most expensive host available in the neighborhood group that is available for business meetings.
- Here, not to our surprise Manhattan appears on top of our results as Manhattan is home to the world's two largest stock exchanges by total market capitalization, the New York Stock Exchange and Nasdaq. Many multinational media conglomerates are based in Manhattan, and the borough has been the setting for numerous books, films, and television shows. Manhattan real estate is among the most expensive in the world.

10. Which quarter has most number of listings



- In this survey we are analyzing the data to find the busiest quarter in terms of listings in New York City
- Q2 has been the busiest quarter of all in terms of listings because in New York City peak season months (May to September) offer high inventory, equaling more choices.
- The best time of the year to rent in New York City is between late spring and early summer. This is true for most people as more apartments hit the market for rent from May-July than any other time of the year.

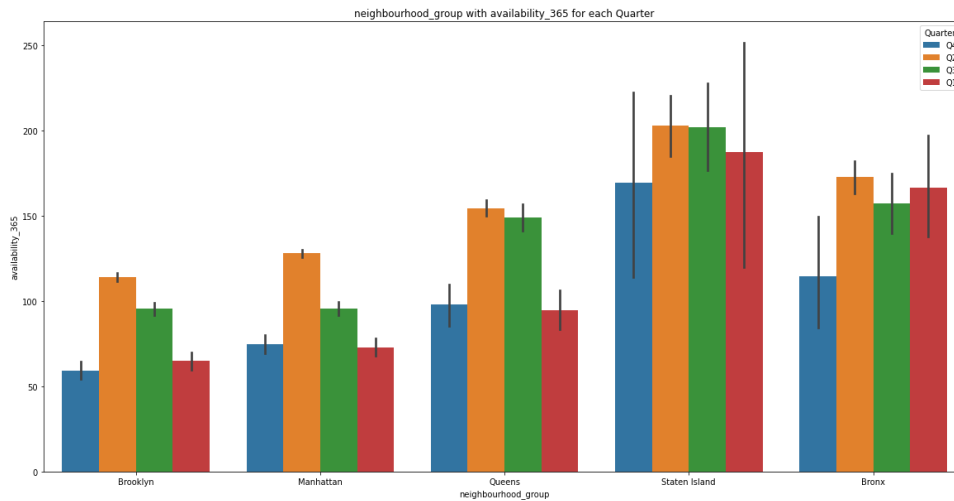
11. Which quarter has most avg number of reviews



Quarter 3 has the highest average in number of reviews then followed by Quarter 2. Because most number of Host listings are done during Quarter 2 and Quarter 3

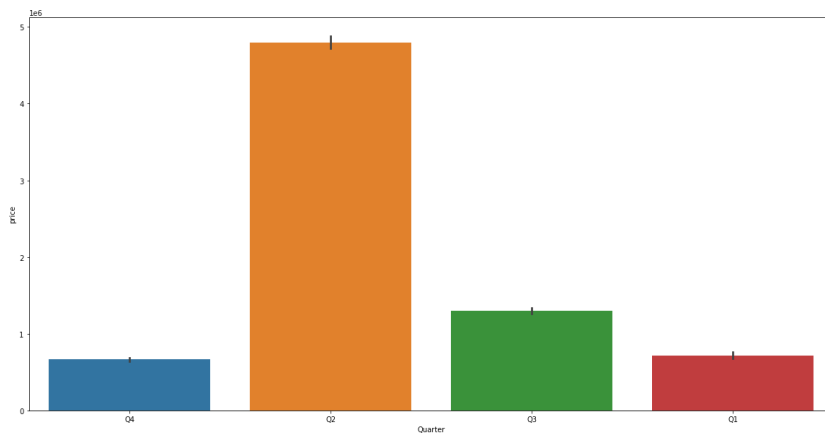
We can see the huge difference in the number of reviews increased after Quarter 2 .The Average was 12 in Quarter 1 then in Quarter 2 it's near to 25 then in Quarter it's near to 30 then in Quarter it's near to 13. Because the Reviews for Quarter 2 might be given in Quarter 3. As we know Quarter 2 has more number of listings

12. Avg availability of each quarter wrt ng



- Brooklyn and Manhattan has the least average availability among all neighbourhood_group
- The Host is Staten island have highest average availability among all neighbourhood_group
- In all neighbourhood_group Q1 and Q4 have least average availability
- In all neighbourhood_group Q2 has the highest average availability

13. Which quarter have the highest profits earned



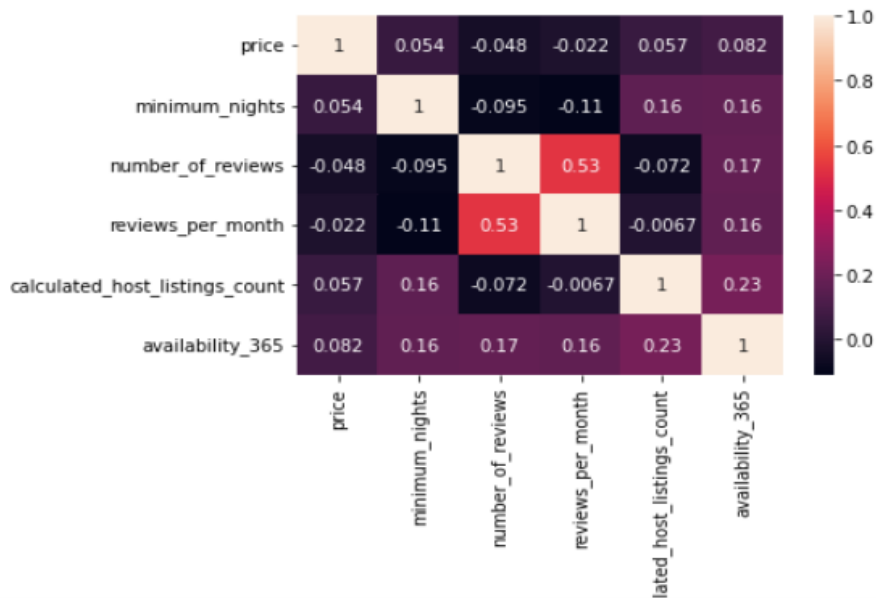
- According to the Graph the profits in Quarter 2 are more than any other which is in the months of April, May, June. Which is generally a summer vacation time for the people so they came to enjoy themselves .
- The start of the financial Year. So many new business deals and stays will be done during this period in New York City .
- During this period the sports events like NFL, NBA, and WrestleMania will be held in New York City.

14. Expensive host in each quarter

	name	host_id	host_name	neighbourhood_group	room_type	price	Quarter
29238	1-BR Lincoln Center	72390391	Jelena	Manhattan	Entire home/apt	10000.0	Q2
17692	Luxury 1 bedroom apt. -stunning Manhattan views	5143901	Erin	Brooklyn	Entire home/apt	10000.0	Q3
9151	Furnished room in Astoria apartment	20582832	Kathrine	Queens	Private room	10000.0	Q1
26739	Broadway 1	153497815	Sarah-B	Brooklyn	Entire home/apt	5000.0	Q4

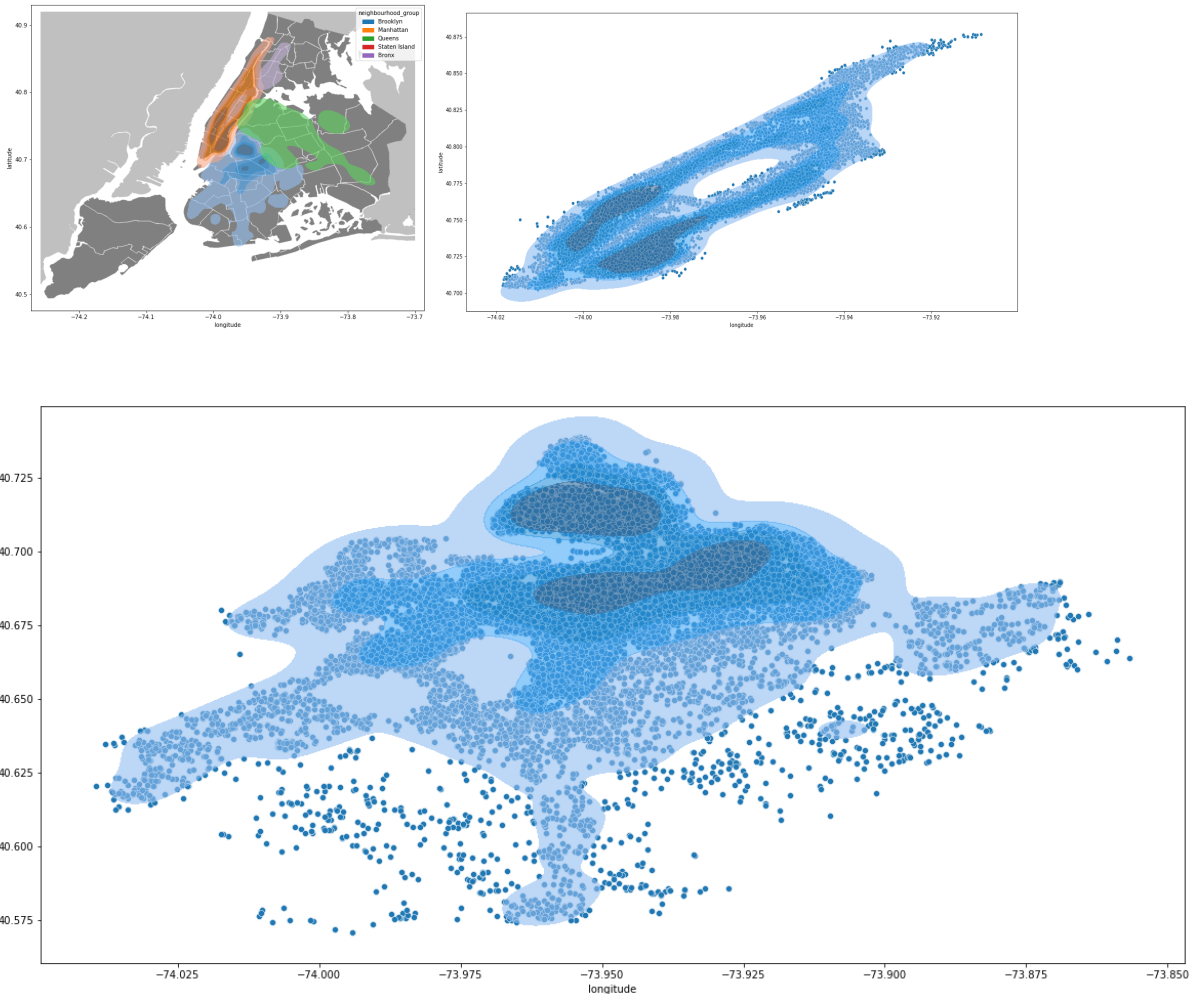
- The most expensive host for each Quarter are and the reasons why those are expensive
- 1-BR at Lincoln Center because it has stunning views central park in Quarter 2 because Spring in Manhattan is so good
- Luxury 1 bedroom apt. -stunning Manhattan views in Quarter 3 which is summer and autumn
- Furnished room in Astoria apartment in Quarter 1 , it's winter time and new year time . So seeing Firework on New Year over the East river will be so good
- Broadway 1 which is in brooklyn and near to church in the time of Christmas and Halloween

15. This is the correlation matrix using heatmap



- we can see there is only 53% correlation between the features review_per_month and number_of_reviews
- all the continuous features are very less correlated and some are negatively correlated

16. Find which part of New York City has more listings and their geographical advantages?



- We can see two hotspots in Manhattan and two in Brooklyn . These hotspots say about the area with the most number of listings in these areas .
- The hotspots in **Brooklyn** and their Geographical advantages
- The first hotspot is Greenpoint which is on the bank of the East river and the host in this area will get the amazing view of Manhattan's skyscraper and the East river.
- The next hotspot in Brooklyn is Bedford-Stuyvesant. This area comes in central Brooklyn . From here we can go to any part of New York City within a 1hr time.
- The Hotspot in **Manhattan** and Their Geographical advantages These hotspots in Manhattan are mostly around Central Park and mostly these are apartments in Skyscrapers with the view of Central Park and the Hudson River.

Conclusion

- From the entire analysis, it can be concluded that,
- Most visitors don't prefer shared rooms, they tend to visit private rooms or the entire home.
- Manhattan and Brooklyn are the two distinguished, expensive & posh areas of New York City.
- Most People Prefer Quarter 2 to visit New York City among all Quarters.
- The average availability of hosts in Manhattan and Brooklyn is very low compared to Other parts of city in all the Quarter
- Some properties have Minimum Nights to stay more than 365 Days which can be favorable among Students, Low-Income Employees & Immigrants.
- The hosts with Manhattan views and Central Park surrounding hosts are in high Demand
- Though location of property has high relation on deciding its price, a property in a popular location doesn't mean it will stay occupied most of the time.
- It'd be better if we had avg guest ratings of a property, that would be beneficial in understanding the property more and could also be a factor in deciding price. A low rated property tends to lower their price.
- Manhattan has the highest average price for all three room types compared with others . So we can say Manhattan is the most expensive area in New York City.

