

Capstone Project

Airbnb analysis

Team members

Mujahid Sayyed

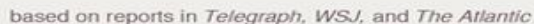
Ankit Tiwari

Lova Kumar Poluparti

Air Bed and Breakfast



- An online marketplace that involves the renting of property to travelers.
- Airbnb does not own any of the properties. It provides a platform on which people can rent out their properties or spare rooms to guests.
- Prices are set by the property owners and monies are collected via the Airbnb app.
- There are many different types of Airbnbs. You can rent a room in someone's house or a whole island and everything in between.
- Airbnb operates from >65000 cities and 191 countries.



City of interest: New York City

- Dataset provided for Analysis is of New York City
- The City is divided into 5 Boroughs namely **Manhattan, Bronx, Queens, Brooklyn, and Staten Island.**
- These Boroughs are further divided into distinctive neighborhoods.
- Today New York is USA's largest Short-term Rental market, with >30K hosts.



Discussion Topics :

1. How many bookings were made in each borough.
2. Number of Hosts in each Borough of NYC.
3. Types of rooms each borough has to offer.
4. Neighborhood with most bookings, which Hosts has most bookings.
5. Based on room type, an average number of nights the customer must stay.
6. Availability of room in each borough. Top reviewed Borough, Room type.
7. Price variation in each borough, based on the room types.
8. Most, a least expensive neighborhood in NYC, the average cost of a room in NYC.
9. Bird-eye view of the density of bookings, room type, price variation across NYC.
10. Words Host use most to attract the customers for bookings.

Data Analysis Steps:

Import Libraries:

In this part, we had imported the required libraries to perform Exploratory Data Analysis for the Airbnb dataset.

Descriptive Statistics:

In this part, we start by looking at descriptive statistic parameters for the dataset. We will use describe() for this

Missing Value Imputation:

We will now check for missing values in our dataset. In case there are any missing entries, we will impute them with appropriate values

Graphical Representation:

We will start with Univariate Analysis. We will be using a **bar graph** for this purpose.

Attributes of each variable:

- **id**: Unique for each listing available in the dataset. Total 48895 unique id's are available in data.
- **name**: Name of the listings available in the dataset. Total 47905 named hotels available in NYC.
- **host_id**: Unique for each host who has listed rooms for rent on Airbnb. Total 37457 hosts with their id's.
- **host_name**: Name of Host. Total unique 11452 host_names in data set.
- **neighbourhood_group**: Names of each borough in NYC city. In data all available hotel locations *Manhattan, Brooklyn, Queens, Bronx, State Island*.
- **neighborhood**: Locale area in each borough of NYC. Total 221 neighborhoods.
- **latitude, longitude**: Co-ordinated for each listing in the dataset. These are in between 40.72 and -73.95.
- **room_type**: Each type of room available for listing. In all areas *Private rooms, Entire home/apt, Shared room* types are available.
- **price**: Price of each listing in the dataset. Price between 0 to 10000.
- **minimum_night**: Mandatory number of nights room needs to be booked
- **number_of_review**: Number of reviews for each listed property. In 0 reviews and max 629 reviews.
- **last_review**: Last date the listing was reviewed.
- **review_per_month**: Number of reviews received per month.
- **Calculated_host_listing_counts**: Number of Listings each host owns.
- **Availability_365**: Number of days in the year the listing is available for booking. Availability zero to 365 days.

Sorting variable types:

Categorical Variables:

name, host_name, neighbourhood_group, neighbour, room_type.

Numeric Variables :

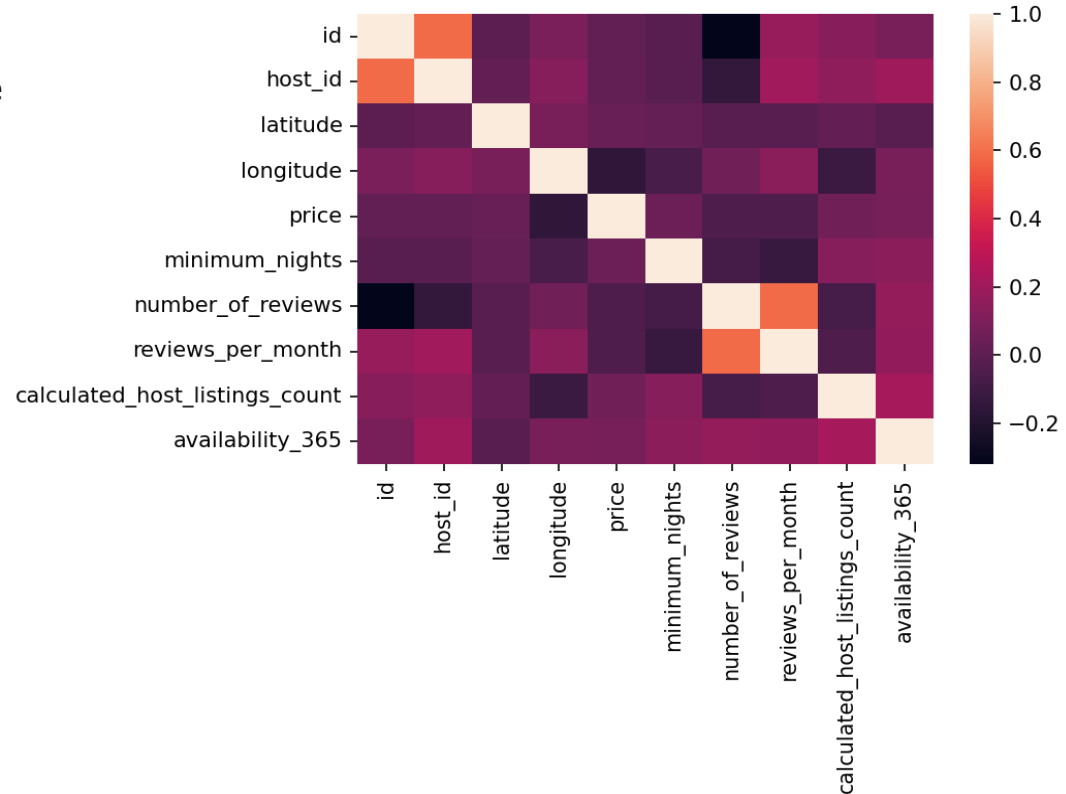
id, latitude, longitude, price, minimum_nights, number_of_reviews, last_review, review_per_month, calculated_host_listing_counts, availability_365.

Insights on variables:

- Each host_id could have multiple listings in the neighborhood, but each listing can have one and only one host_id.

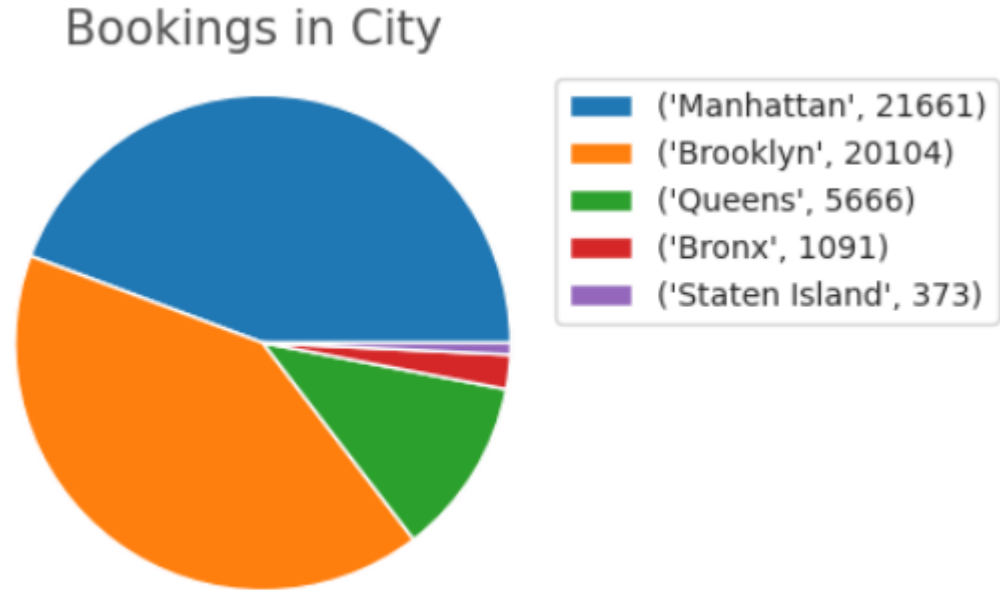
Correlation between Variables

- Attached heat map helps us understand the correlation between the variables of our dataset.
- (host_id ,review_per_month),(host_id, availability_365) are sets of positively correlated variables.
- Correlation for other sets is mostly close to zero.



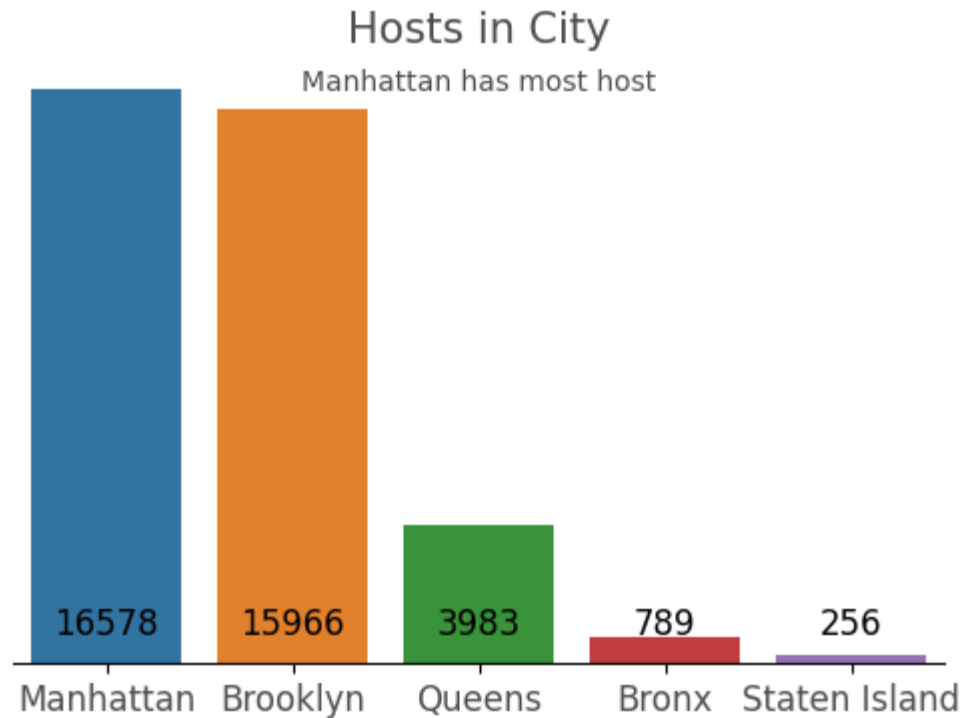
How many bookings were made in each borough.

- **Manhattan** has the most bookings with 21661.
- Brooklyn is the second-highest with 20104 bookings.
- The Bronx and Staten Island trail the pack with 1091 and 373 bookings.



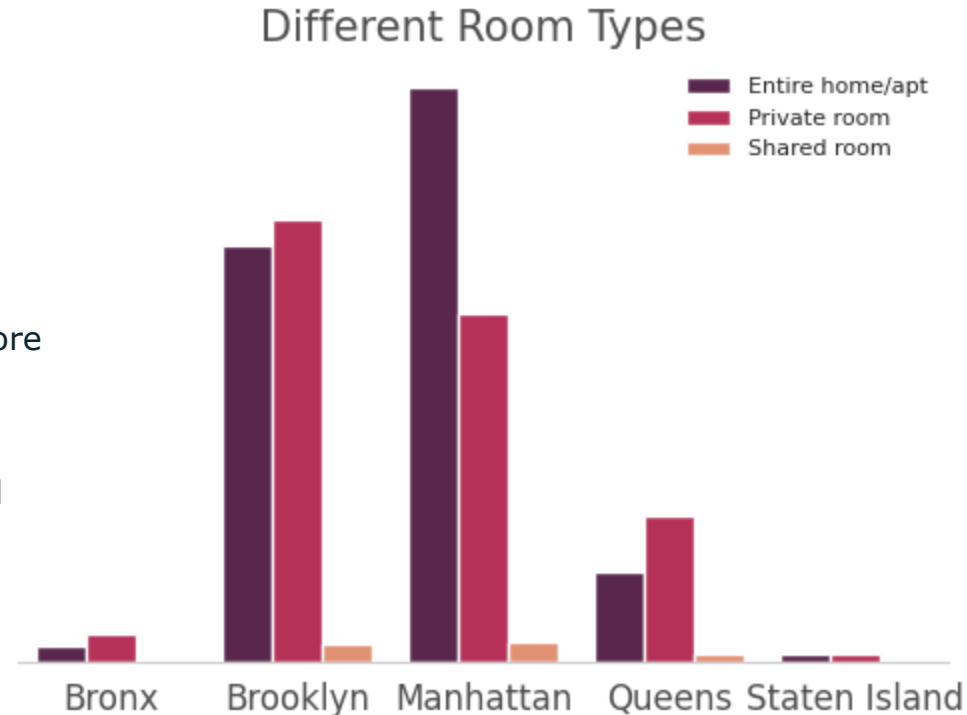
Number of Hosts in each Borough of NYC

- **Manhattan** has the most hosts, 16578.
- Brooklyn is the second-highest with 15996 bookings.
- Most customer stays explains the high number of hosts Manhattan has.
- Most customer stays are directly related to the number of hosts each borough has.



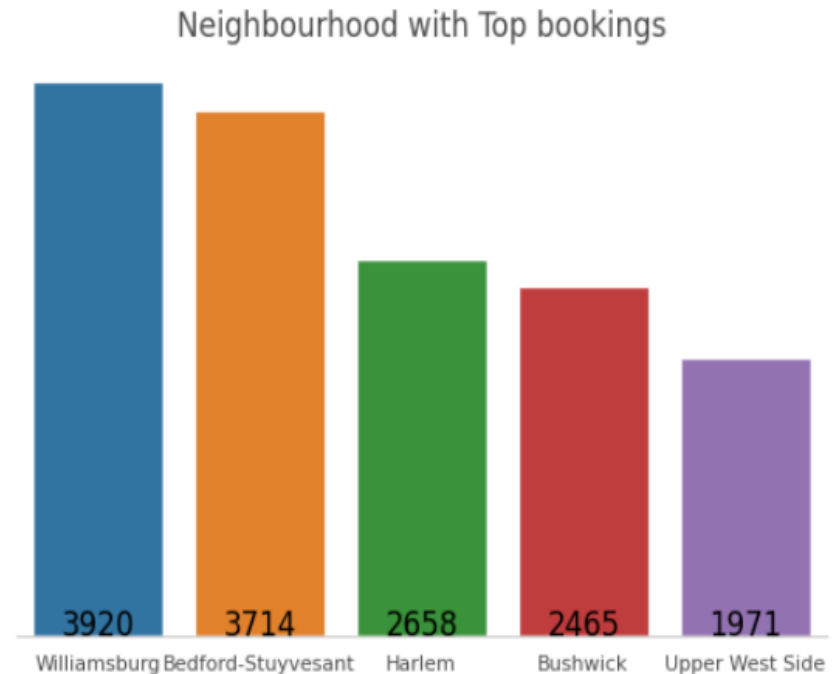
Types of rooms each borough has to offer

- **Manhattan** leads the group with the most rooms bookings, Customer prefer **Entire homes** more compared to Private rooms in Manhattan.
- Customer in Brooklyn prefer **Private rooms** more than **Entire rooms**.
- Except in Manhattan all the other boroughs have more bookings for a private room.
- Shared rooms are the least preferred type across all the boroughs of NYC.



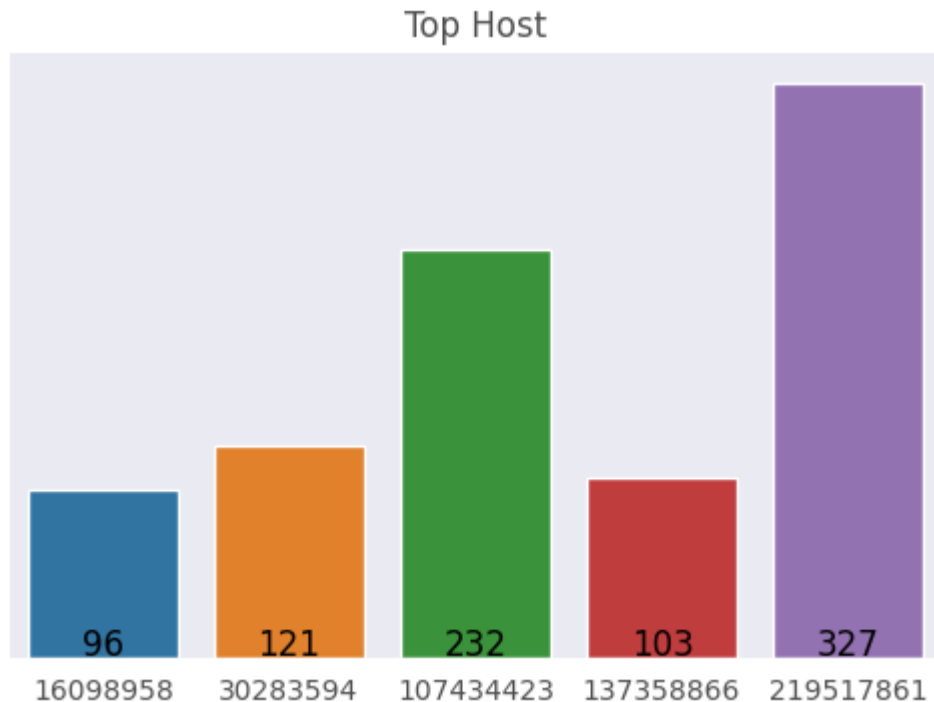
Top Neighborhoods with most bookings

- There are 221 unique neighborhoods in the given dataset.
- **Williamsburg** is the top neighborhood with 3920 bookings.
- Bedford-Stuyvesant has the second most bookings with 3714 bookings.
- 3 amongst the top 5 neighborhoods in bookings belong to Brooklyn and the rest of 2 belongs to Manhattan.



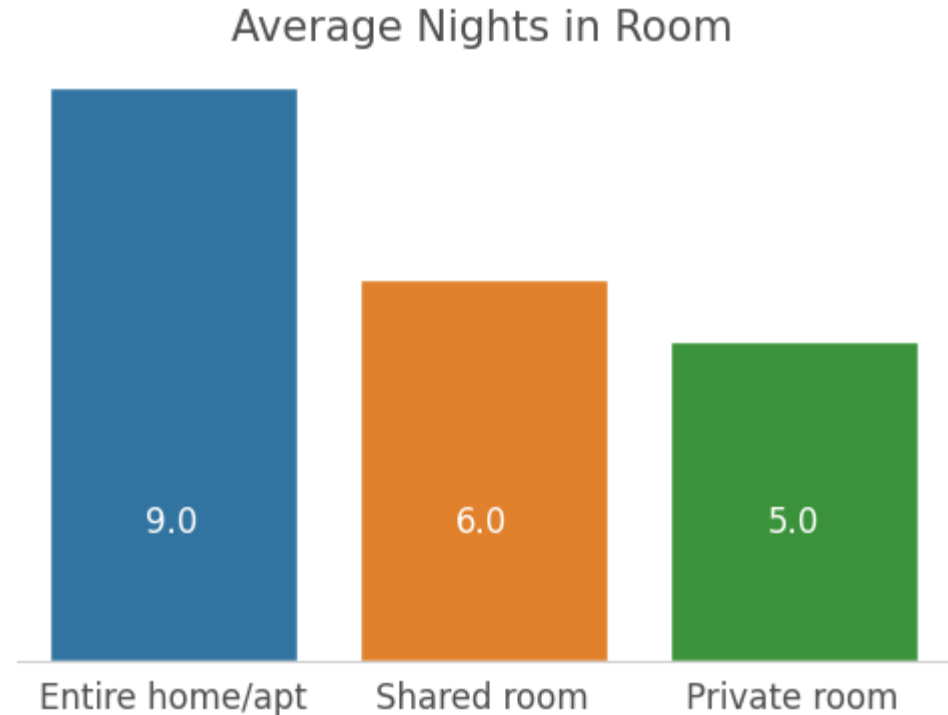
Top Hosts

- Host Sonder(NYC) with host_id 219517861 has hosted most customers **327**.
- Host, Blueground with host_id 107434423 is the second-highest with a total of 232 customers hosted.
- Host, Kara with host_id 30283594 has hosted 121 customers.



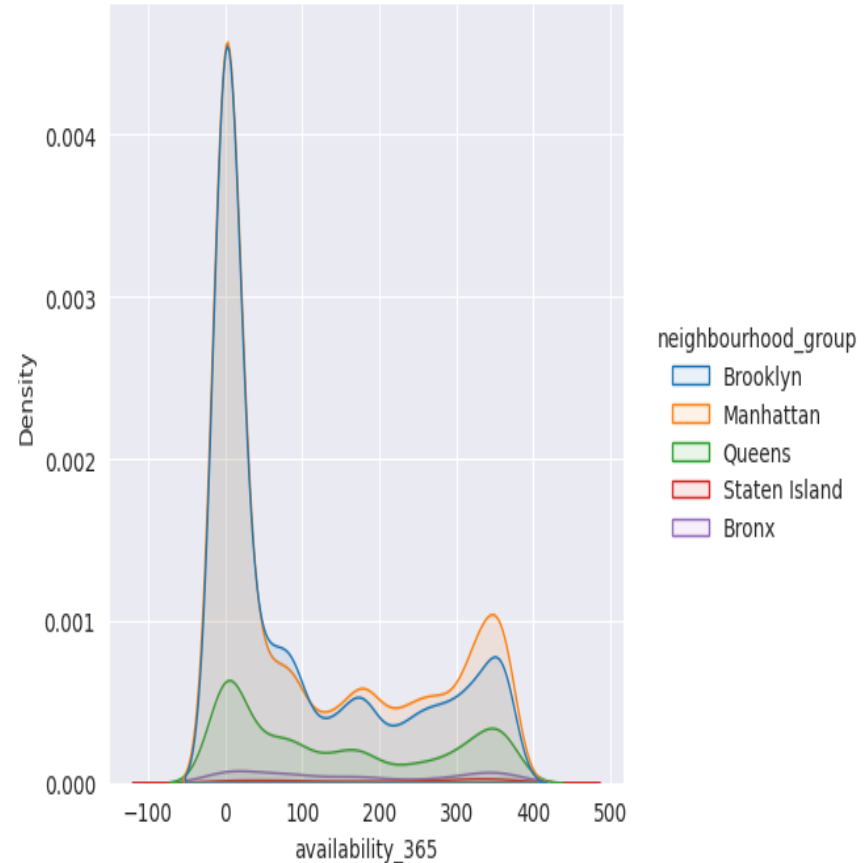
Average of Minimum number of Nights for each room type

- Average 9 nights are mandatory for stay for booking of Entire home
- Shared room is the second highest with average 6 nights stays mandatory
- Private room is third place with average minimum 5 nights.



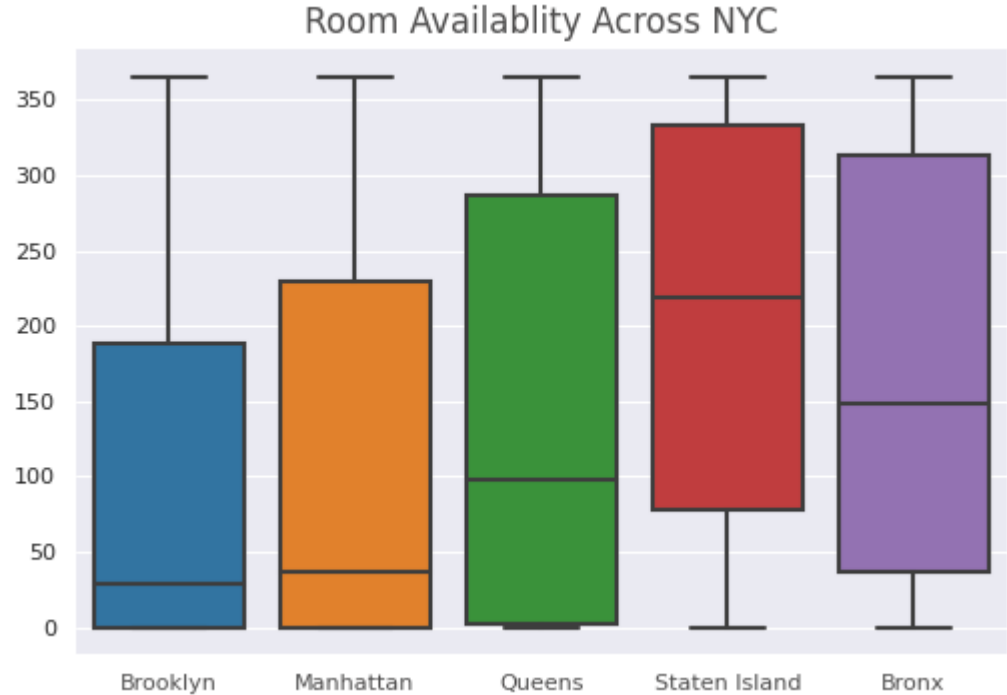
Availability distribution across New York City

- Manhattan and Brooklyn have very similar distribution with peaks close to zero.
- This shows most of the rooms are mostly booked in Manhattan and Brooklyn.
- Availability of rooms is more evenly distributed compared to Brooklyn and Manhattan.



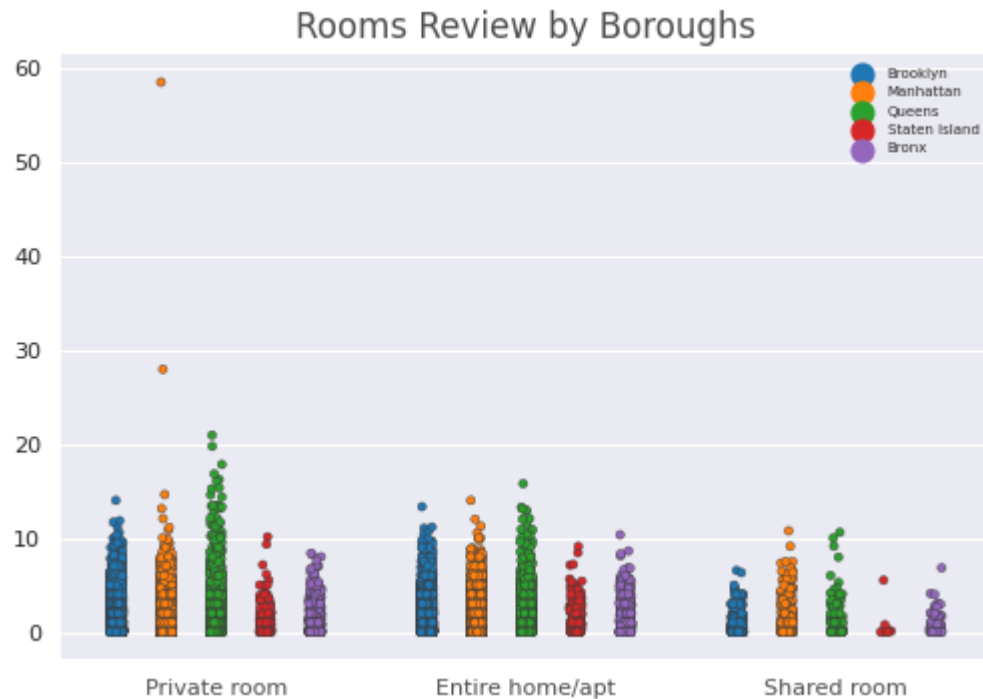
Availability of each room in Borough

- Brooklyn has the least availability followed by Manhattan-based on the distribution of room availability.
- Staten island has the highest availability of rooms.
- Availability of rooms in the Bronx is more evenly distributed compared to other boroughs.



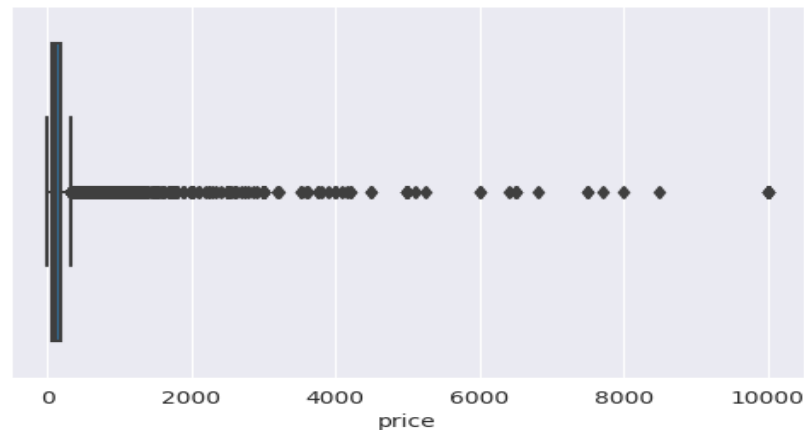
Reviews per month by borough

- **Private Rooms** are the most reviewed amongst the types of rooms
- **Manhattan** is the most reviewed followed by **Queens**.
- **Private room** and **Entire home room type** from **Staten island** are have similar kind Reviews.



Price observation and removal of outliers

- Boxplot of price column depicting the presence of high range outliers!



Using IQR process to remove outliers

Price column indicating negligible or 0 outlier values.

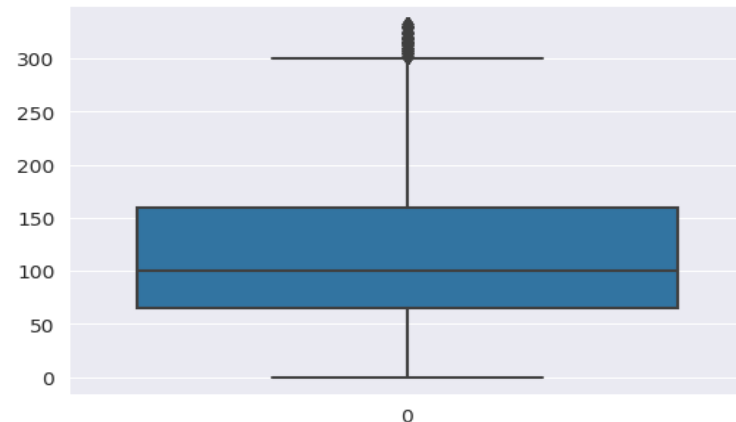
$Q1 = df.quantile(0.25)$

$Q3 = df.quantile(0.75)$

$IQR(Inter-quartile\ range) = Q3 - Q1$

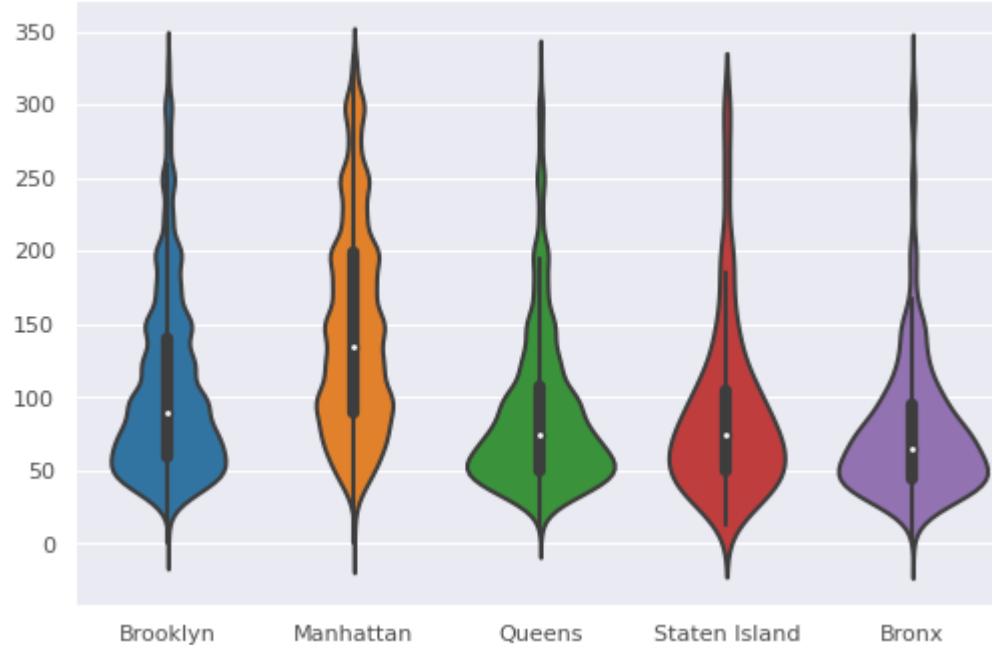
Lower range = $Q1 - 1.5 * IQR$

Higher range = $Q3 + 1.5 * IQR$



Price distribution across New York City

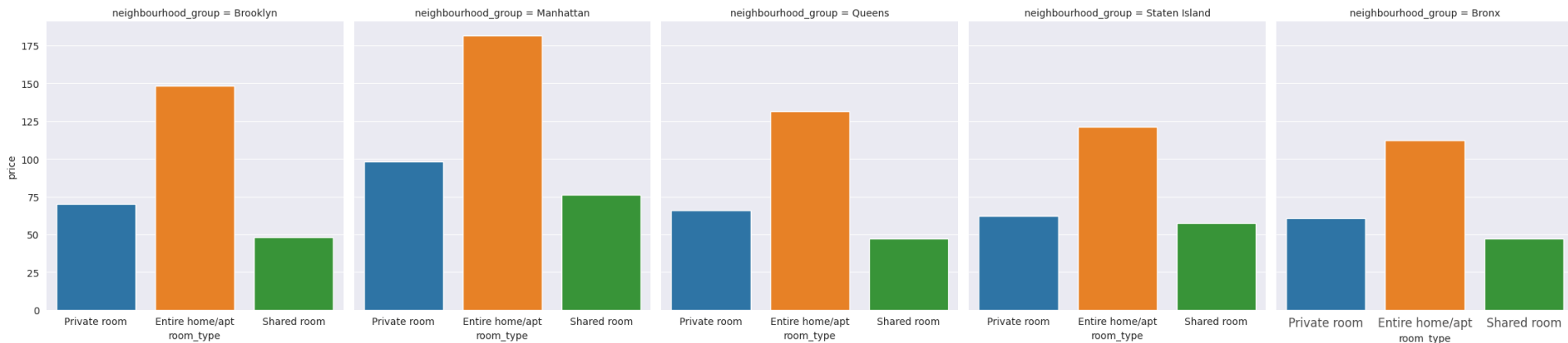
Price distribution for each borough



- **Manhattan** is the **most expensive** as the rental charges are more evenly distributed across all the **price ranges**.
- Median price in **Manhattan** is \$140 that's around double the Median price of the **Bronx** and Staten island.

Median price of each type of room in Boroughs

- Manhattan is the most expensive neighborhood for each type of room, followed by Brooklyn and Queens.
- Entire homes as expected have the highest median price across all the boroughs.
- Staten island have a similar median price for Private and shared room.



Most expensive Neighborhood

- Majority of most expensive listings are from **Manhattan**, that are located in **Midtown, Chelsea, Upper West Side, West Village, Lower East Side**.
- Mid-Century Museum Sleepover** from Williamsburg is the only listing from **Brooklyn** that comes under the most expensive Listing.

	name	host_id	host_name	neighbourhood_group	neighbourhood	price
15150	Midtown Sleep 6 Central Convenience	62031986	Adam	Manhattan	Midtown	333
38565	Parisian Palace in Heart of Manhattan	224001464	Aleszea	Manhattan	Chelsea	333
28062	27 FLR VIEWS!LINCOLN SQR-LUXURY 2BR MIDTOWN W ...	76104209	Rated	Manhattan	Upper West Side	333
25161	Flatiron Loft 3BR/1.5 Bath Best Location\n30 days	35635299	Mel	Manhattan	Midtown	333
48567	Your Private Duplex Mansion in the HEART of NYC!	269146714	Clinton	Manhattan	West Village	333
19641	SOMMwhere in NYC/ a unique, conscious artists ...	148108	Fatima	Manhattan	Lower East Side	333
45303	Five-star luxury Apt in Chelsea !	152747338	Paola	Manhattan	Chelsea	333
13798	Mid-Century Museum Sleepover	29065752	Cullen	Brooklyn	Williamsburg	333
32976	Luxurious 2bed/2.5bath with Central Park Views	4876826	Alex	Manhattan	Midtown	333
46439	CHELSEA CENTRAL LUXURY 2BATHS	184520918	Danilo & Larissa	Manhattan	Chelsea	332

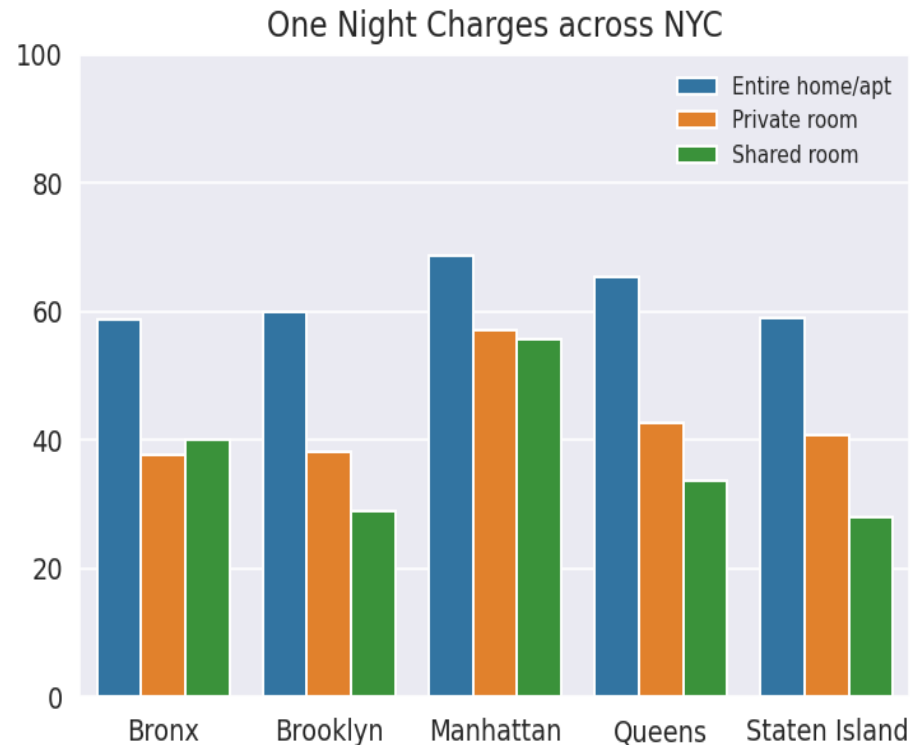
Least expensive Neighborhood

- Majority of the cheapest listings are from **Brooklyn**, that are located in **Bushwick, Greenpoint, Williamsburg, Bedford Stuyvesant**.
- **Hostel Style Room Ideal Traveling Buddies★** from the neighborhood of **East Morrisania** is the only cheapest listing from the **Bronx**.
- **The best you can find** is the only cheapest Listing from **Manhattan**.

	name	host_id	neighbourhood_group	neighbourhood	price
26841	Coliving in Brooklyn! Modern design / Shared room	101970559	Brooklyn	Bushwick	0
25433	★Hostel Style Room Ideal Traveling Buddies★	131697576	Bronx	East Morrisania	0
25634	MARTIAL LOFT 3: REDEMPTION (upstairs, 2nd room)	15787004	Brooklyn	Bushwick	0
25753	Sunny, Quiet Room in Greenpoint	1641537	Brooklyn	Greenpoint	0
25778	Modern apartment in the heart of Williamsburg	10132166	Brooklyn	Williamsburg	0
25794	Spacious comfortable master bedroom with nice ...	86327101	Brooklyn	Bedford-Stuyvesant	0
25795	Contemporary bedroom in brownstone with nice view	86327101	Brooklyn	Bedford-Stuyvesant	0
25796	Cozy yet spacious private brownstone bedroom	86327101	Brooklyn	Bedford-Stuyvesant	0
26866	Best Coliving space ever! Shared room.	101970559	Brooklyn	Bushwick	0
26259	the best you can find	13709292	Manhattan	Murray Hill	0
23161	Huge Brooklyn Brownstone Living, Close to it all.	8993084	Brooklyn	Bedford-Stuyvesant	0

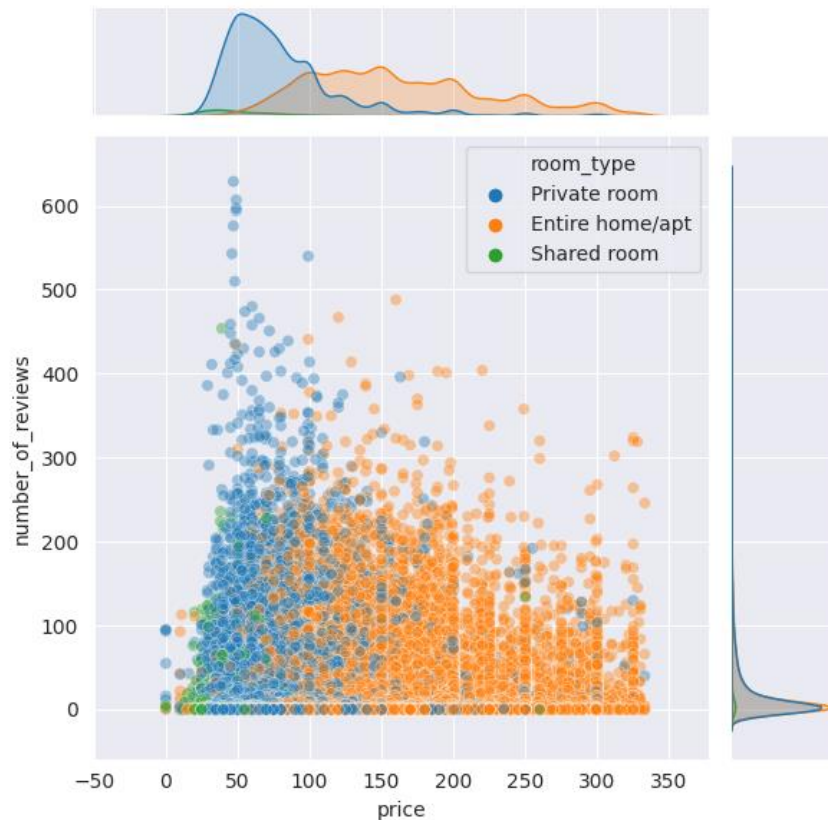
Per Night Cost in each Boroughs

- **Bronx** and **Brooklyn** have the lower night Cost compared to other boroughs across all room types.
- The **Bronx** is the only borough where the shared room charges per night is more than Private rooms.
- Difference between per night charges of Entire and private room is lowest for Manhattan, as compared to other boroughs.
- Per night charges of Private and Shared room are similar for Manhattan

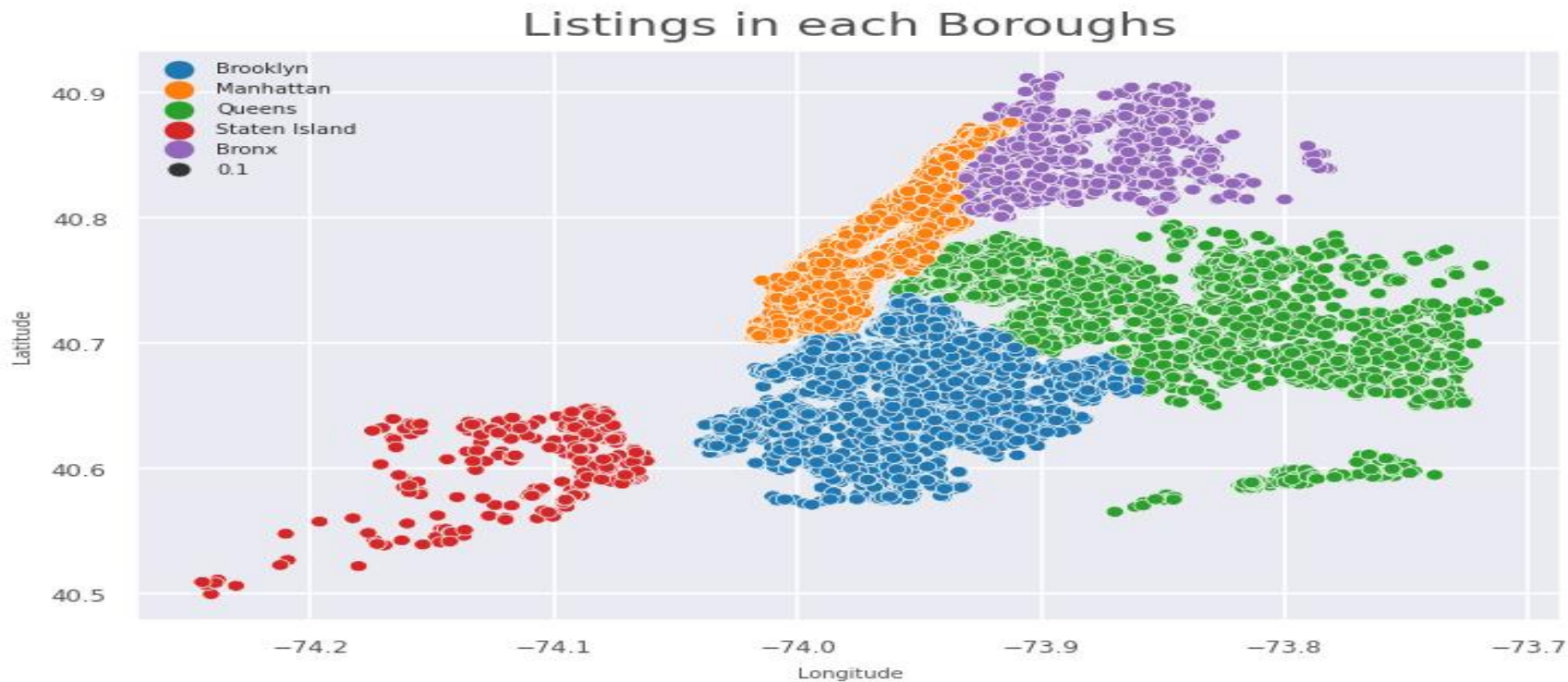


Pricing, Room reviews for room types

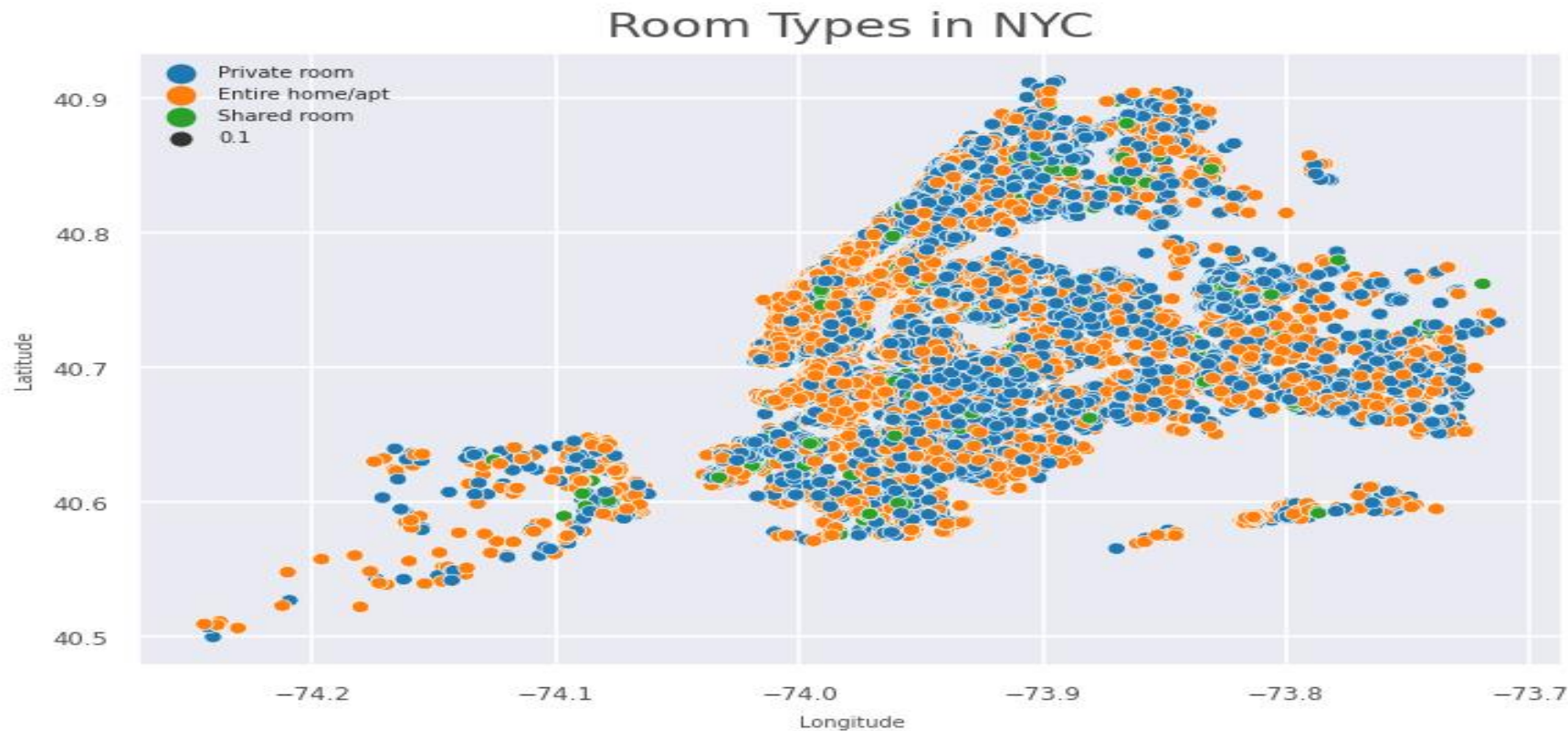
- Attached joint plot shows us the price variation for room types and the number of reviews.
- Private rooms** have positively skewed distribution with prices mostly in the range of **0-100 USD**.
- Entire home** has more evenly distributed pricing compared to Private, shared rooms
- Room review** is also **positively skewed** with most reviews condensed at zero.
- Shared rooms** are the least reviewed amongst the room types.



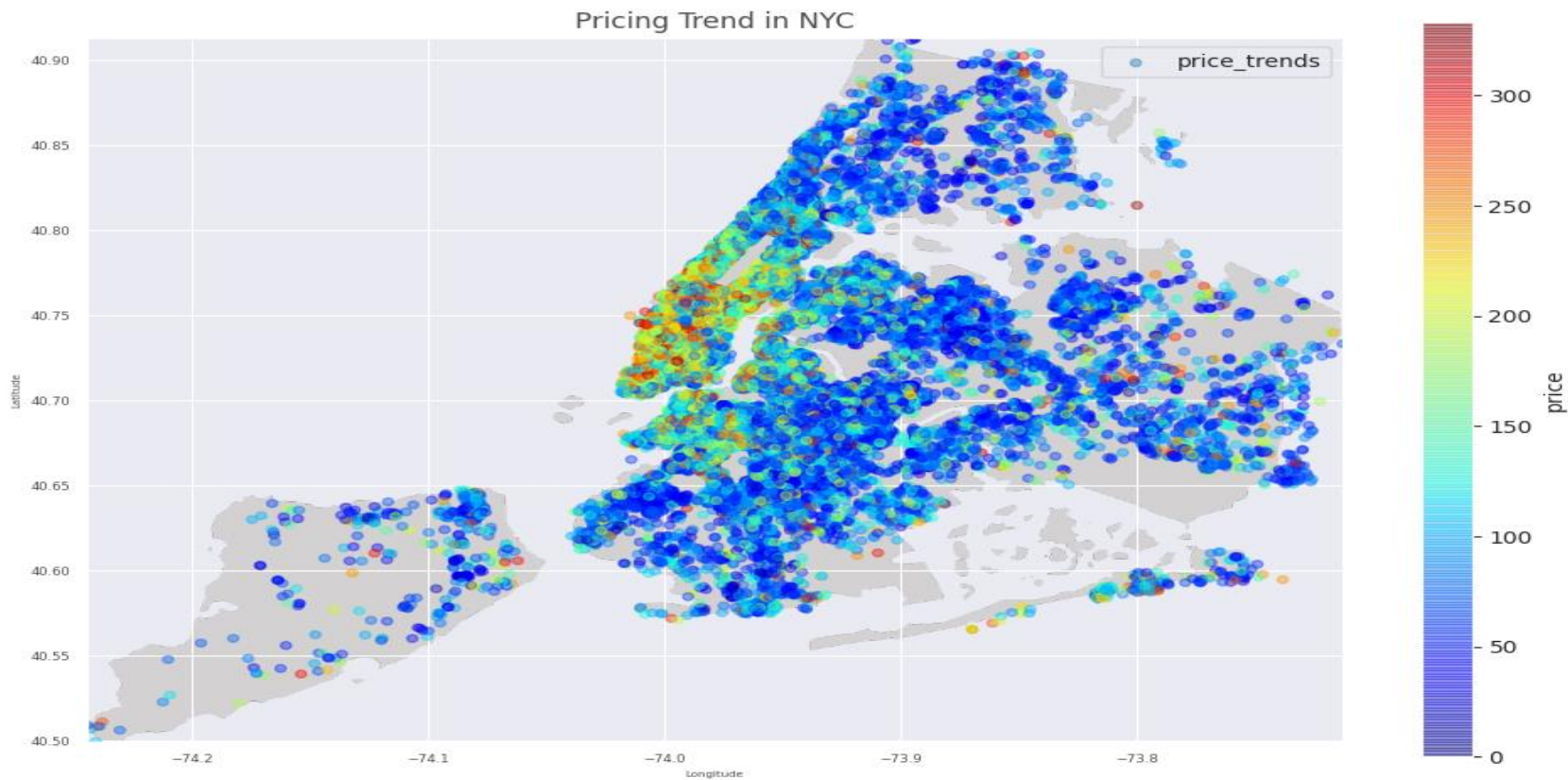
Visualizing variables with Longitude and Latitude



Visualizing the room types in the NYC city



Price distribution New York city



Observations from Airbnb analysis

Most booking, Most number of hosts: **Manhattan** followed by **Brooklyn**, Queens, Bronx, and Staten Island.

Most booked Room Type: **Private Room**, followed by Entire type, the Shared room is the lowest book.

- **Manhattan** has more bookings of Entire room compared to Private room. Other boroughs have more bookings of Private rooms compared to the Entire room.
- Most booked neighborhood is **Williamsburg**, followed by Bedford-Stuyvesant, Harlem, Bushwick, and Upper West Side. All these neighborhoods belong to **Manhattan** and **Brooklyn**.
- Average Mandatory number of nights for booking of Entire home is 9, the Shared room is 6 and **lowest** for **Private room** 5 nights.
- **Brooklyn** has the least availability of rooms available for booking followed by Manhattan.
- **Private rooms** are mostly reviewed, Entire homes get the second attention, Shared rooms are least reviewed. Manhattan and Queens are the most reviewed **Boroughs**.
- **Entire Homes** are most **expensive**, followed by Private, Shared rooms, **Manhattan** is the **most expensive** borough for all types of rooms.



Conclusion

Customer Needs Rental space with **Cheaper rent, good reviews, lower minimum nights, proximity to transportation services, good space.**

Based on the Word plot, the list of Words that are used to seek Customer attention are:

Private room, Brooklyn, Manhattan, Central Park, beautiful, spacious, subway, Williamsburg, Bushwick, times square

EDA Validation:

- Private room has lowest minimum nights avg, and cheaper compared to Entire room, and its prices are very similar to Share rooms when charges are compared on per night basis.
- Private rooms are the most reviewed across NYC
- Brooklyn has Cheaper per night charges for private rooms compared to other boroughs, 3 of the most booked neighborhoods are from Brooklyn, Williamsburg, Bushwick, Bedford-Stuyvesant.
- Manhattan is the hub of the center of activities like Central park, times square and has the busiest subway,
- Entire home provides more space, Entire home Charges in Manhattan are very close to Private room.

Q&A