Airbnb Case Study -2

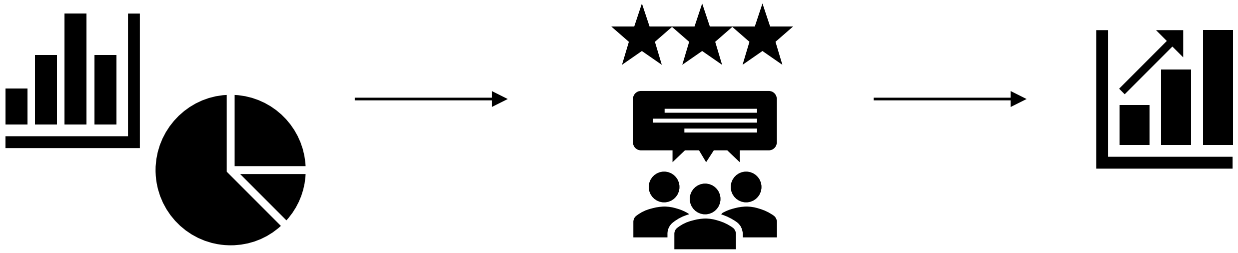
Ramya D & Jaydeep Nawagekar, DSC 27

September 2021

Contents

Objective

Identify key insights and customer preferences for Airbnb listings in NYC



If we can successfully unearth key insights, the revenue should go up as the sales team will now be focusing more on being customer centric and thus increase customer delight

Background

Identify key insights and customer preferences for Airbnb listings in NYC

For the past few months, Airbnb has seen a major decline in revenue.

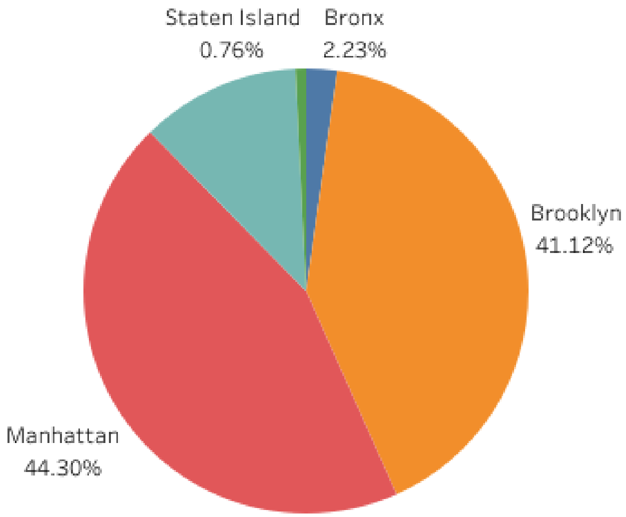
Now that the restrictions have started lifting and people have started to travel more, Airbnb wants to make sure that it is fully prepared for this change with enhanced customer experiences

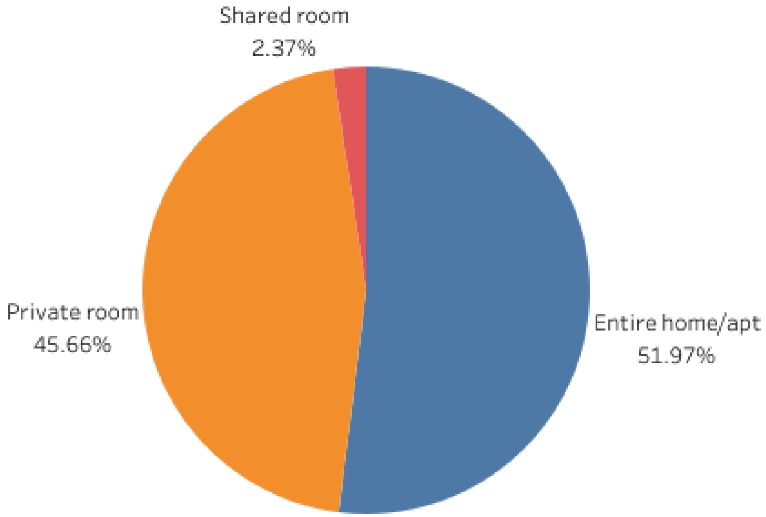
If we can successfully unearth key insights, the revenue should go up as the sales team will now be focusing more on being customer centric and thus increase customer delight

Key Findings (contd)

The segments Entire home/apt, Shared room at Brooklyn and Manhattan are the major drivers for revenue of Airbnb in NYC

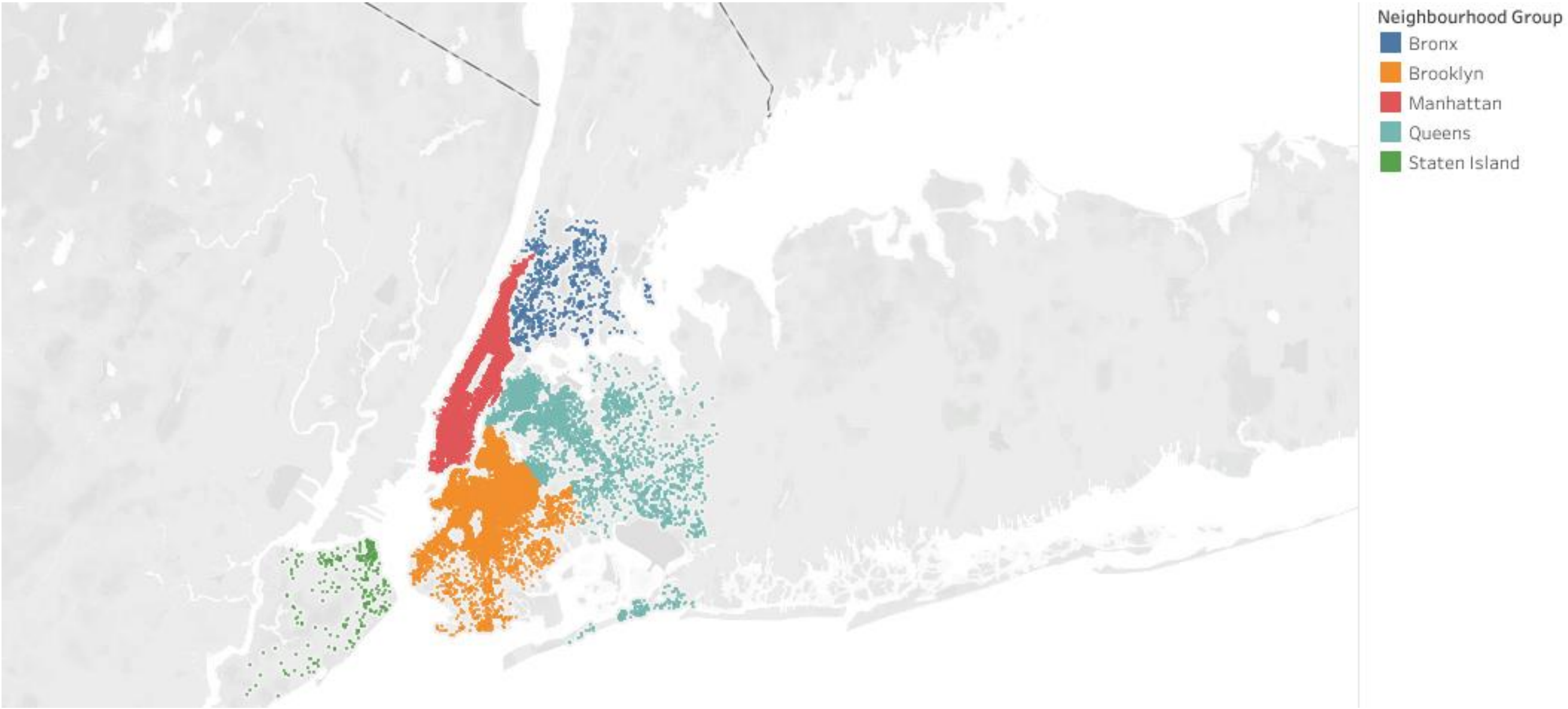
- There is very low interest for Shared rooms from Customers.
- Most of the customers prefer Entire home/apt indicating that customers tend to stay in groups





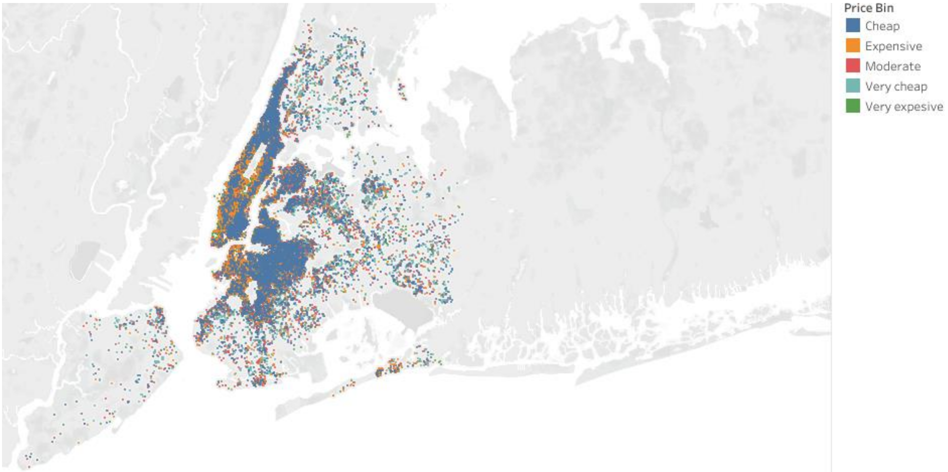
Key Findings (contd)

The listings are clustered very closely in Manhattan and Brooklyn. They tend to be highly sparse in the remaining neighborhood groups



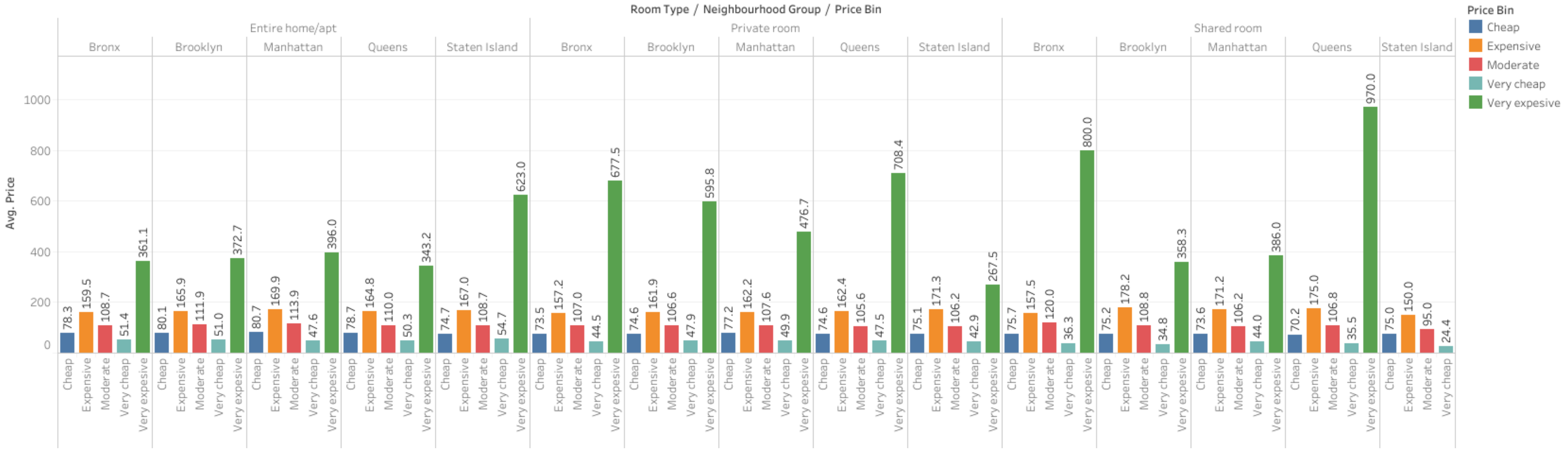
Key Findings (contd)

Majority of the listings have a very poor number of reviews associated with them (0 to 3 reviews) and very low availability (0 to 8 days in a year)



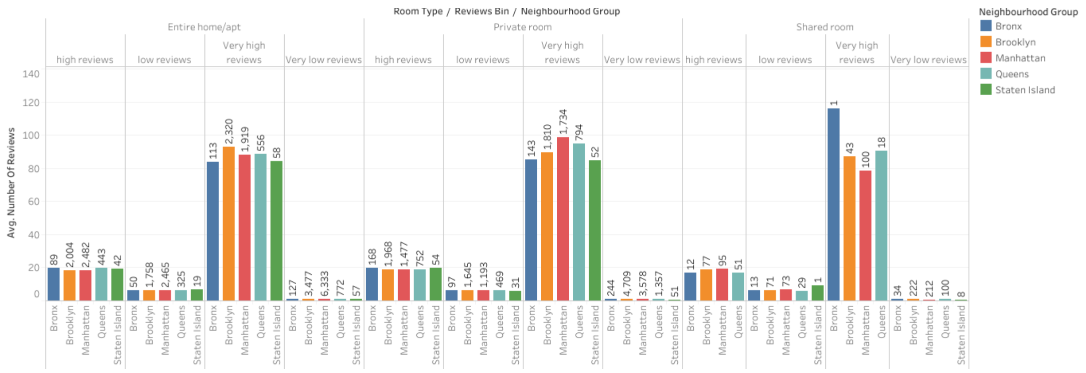
Key Findings (contd)

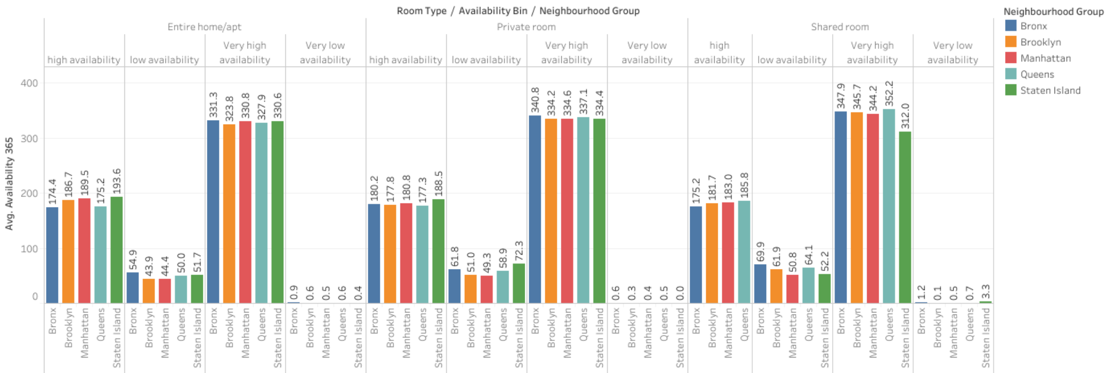
- Manhattan is the only exception in the neighborhood groups where it has higher entire home/apt listings more than that of Private room
- Majority of the listings in the Bronx range from Very cheap to cheap
- If someone is interested to stay in Brooklyn, it's better to stay in a Private room as they are very high in number and are very cheap compared to any other group or room type.



Key Findings (contd)

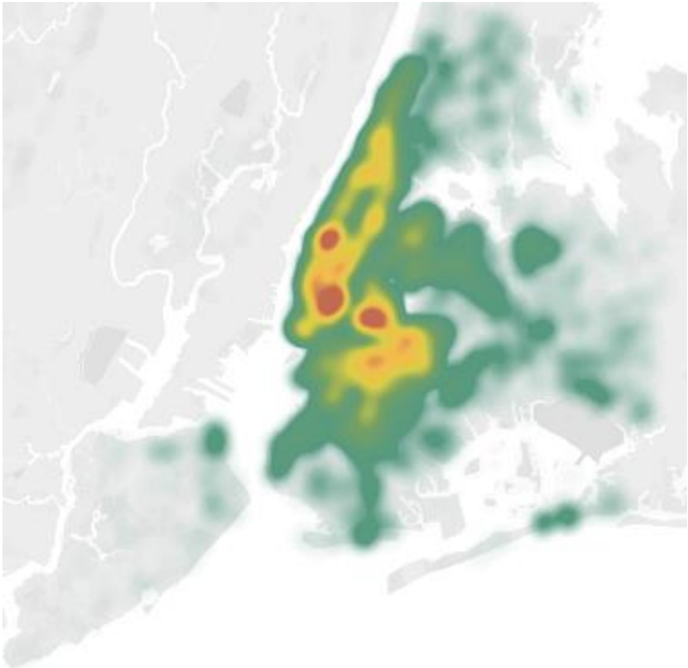
- Availability is usually very low in Brooklyn and Manhattan and very high at Staten Island. The median availability of all the listings is at most 4 across different groups of price and neighborhood
- The median minimum number of nights is the same (around 2) across neighborhood groups with a slightly high variation in shared rooms of Staten Island.
- On an average, the listings in Staten Island tend to receive higher number of reviews

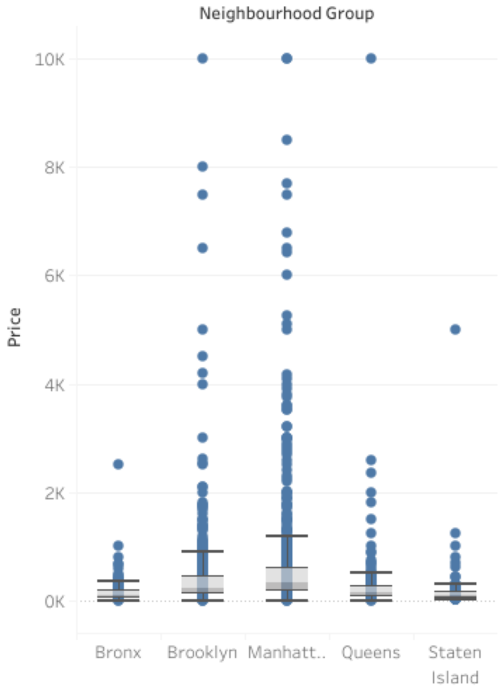




Key Findings (contd)

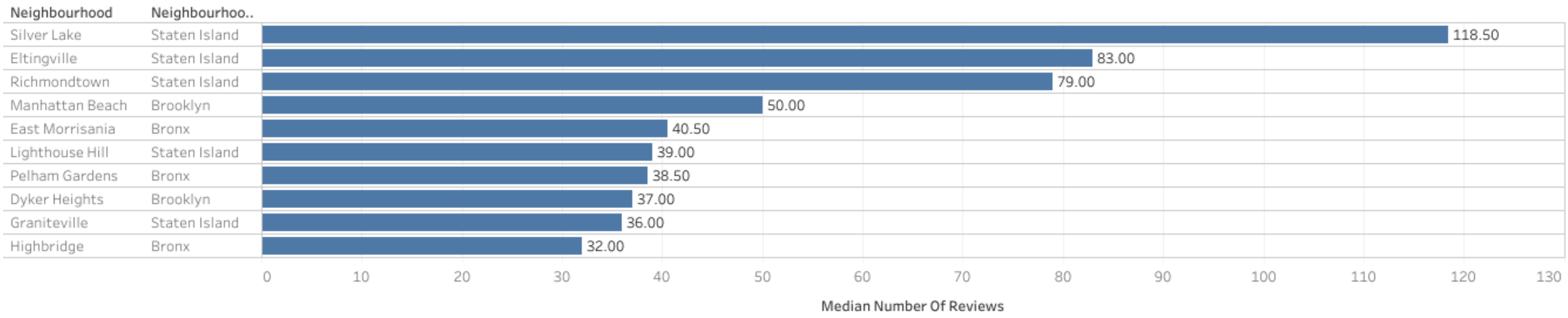
- Majority of the listings have a price of 30 to 130 usd pernight. There are very few listins with price as high as 5000 USD per night
- Expensive listings are mostly concentrated in the lower Manhattan and Upper Brooklyn area.





Key Findings (contd)

- Silver Lake, Eltingville, Richmondtown from Staten Island tend to receive higher number of reviews on an average. Top 10 neighborhoods of 221 contribute to almost 50% of listings in NYC
- All the top 10 neighborhoods with high listings are either in Manhattan or Brooklyn



Recommendations

- Concentrate on neighborhood groups like Staten Island, Bronx to improve the share of listings, reduce variation in prices
- Measures need to be taken to quantity improve the reviews associated with listings
- Target the top 10 neighborhoods with high listings as they tend to lag in generating sales and enough reviews even though they account for almost 50% of listings
- Explore alternative business models for the listings where the availability is very high but fail in attracting customers.
- A detailed analysis need to be conducted on the segment shared rooms as they tend to be very low in number and accounts for a very minute portion of sales
- Ensure that the listings in the neighborhoods like Bronx, Staten Island to not be sparse as opposed to the clusters that can be seen in Manhattan and Brooklyn

Appendix: Data Sources

- This dataset contains information about different Airbnb listings along with their hosts, locations, prices and other attributes.

Column	Description
id	listing ID
name	name of the listing
host_id	host ID
host_name	name of the host
neighbourhood_group	location
neighbourhood	area
latitude	latitude coordinates
longitude	longitude coordinates
room_type	listing space type
price	
minimum_nights	amount of nights minimum
number_of_reviews	number of reviews
last_review	latest review
reviews_per_month	number of reviews per month
calculated_host_listings_count	amount of listing per host
availability_365	number of days when listing is available for booking

Appendix: Data methodology

- We conducted a thorough analysis of the Airbnb datasets. The process included:
  - Cleaning the data set using fuzzy logic – Dropped columns such as name and last review which are not useful for analysis
  - Performed basic data checks to identify constant columns, missing values, duplicate rows etc
  - Imputed the missing values in the column “reviews per month” with a constant value of 0
  - Used scatter plots to identify abnormal data points and perform sanity checks
  - Although we have identified a few outliers in the price column, where the listings have a price of 0$ per night, we’ve decided to let them be part of analysis because of two reasons
    - They are very few (11 out of 40k+)
    - Important information in the remaining columns need to be retained for analysis
  - Later we created various pivot tables, univariate and bivariate graphs, statistical analysis on important columns to identify key relationships and customer preferences

Appendix: Data assumptions

- We assumed that the column price indicate the price per night a customer has to pay for a listing in USD
- A low value of Availability_365 indicates that the listing is not easily available to book and is usually occupied/busy
- Various bins are created to convert a continuous variable to a categorical variable so as to qualitatively analyze the insights based on quantile values.