**Storytelling case study**

**Airbnb, NYC**

**Akshay Jain**

# Problem background

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.

The different leaders at Airbnb want to understand some important insights based on various attributes in the dataset to increase the revenue.

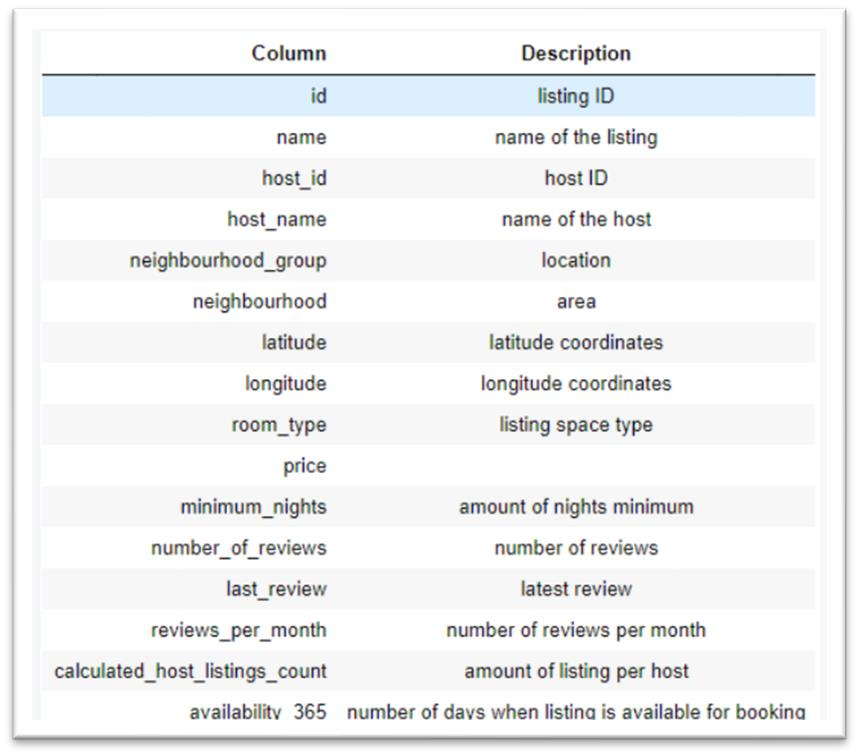
# Objective

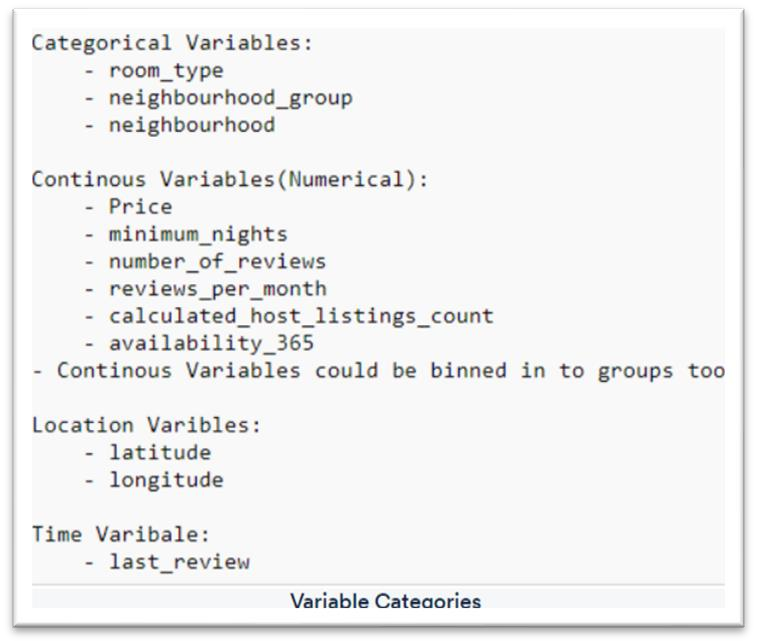
To prepare for the next best steps that Airbnb needs to take as a business, we have been asked to analyze a dataset consisting of various Airbnb listings in New York as a data analyst.



# Dataset

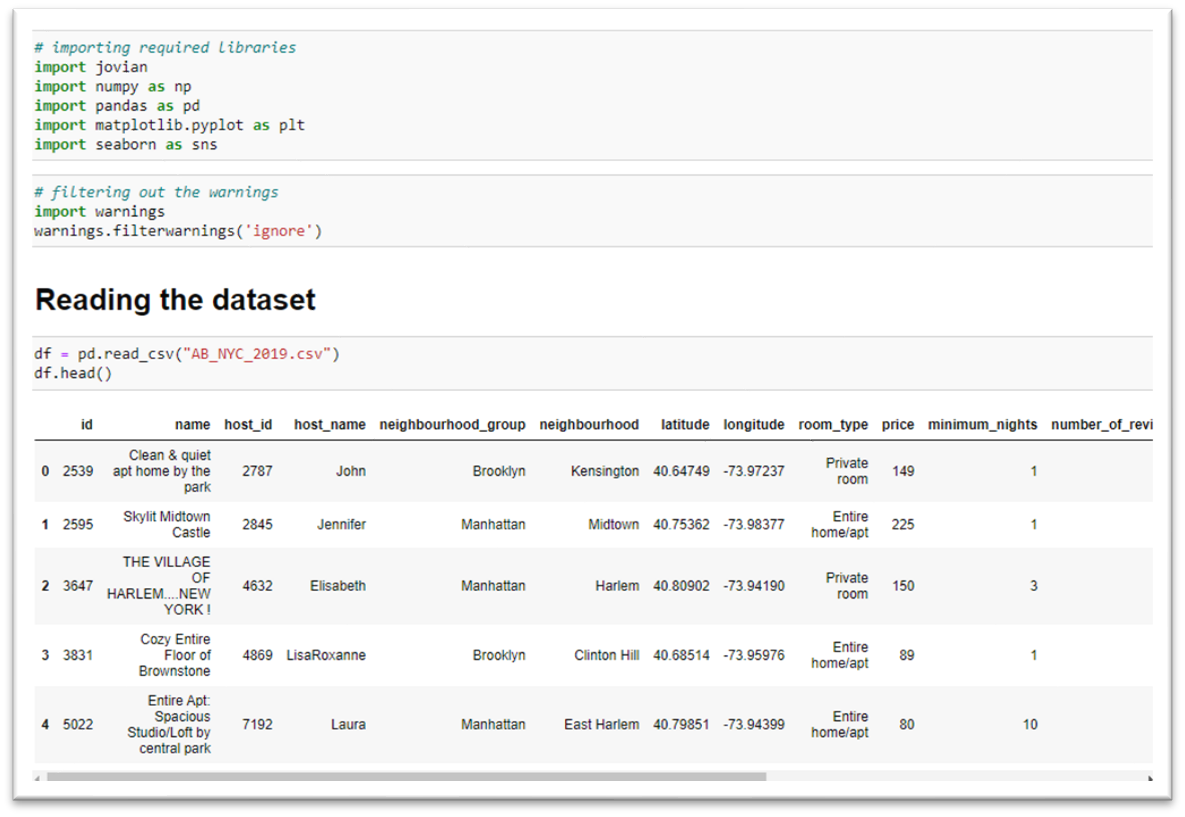
Contains information about different Airbnb listings along with their hosts, locations, prices and other attributes. Below is the column description –



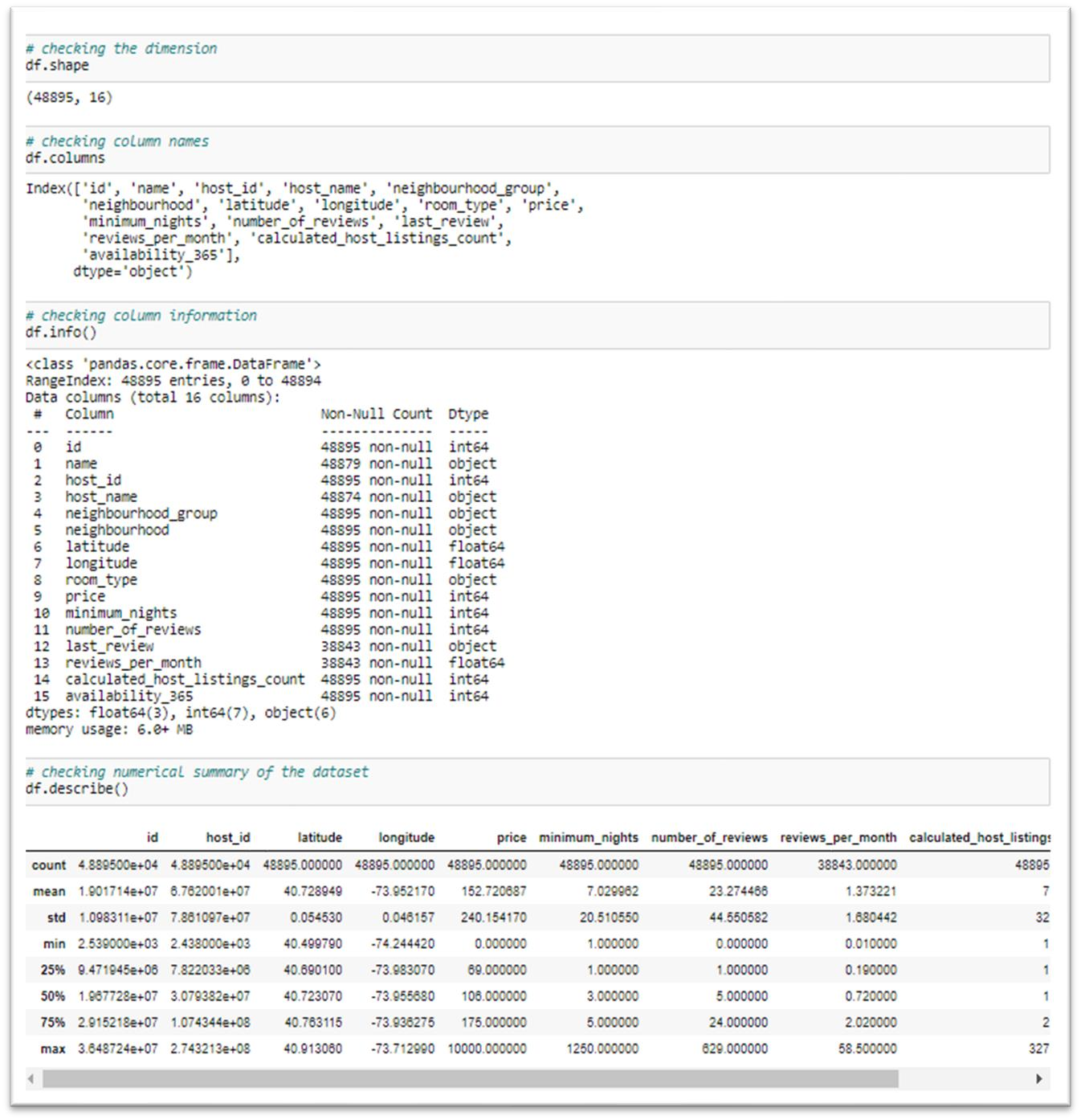


# Data Analysis

**Step 1: Imported appropriate libraries and read the data into a data frame.**



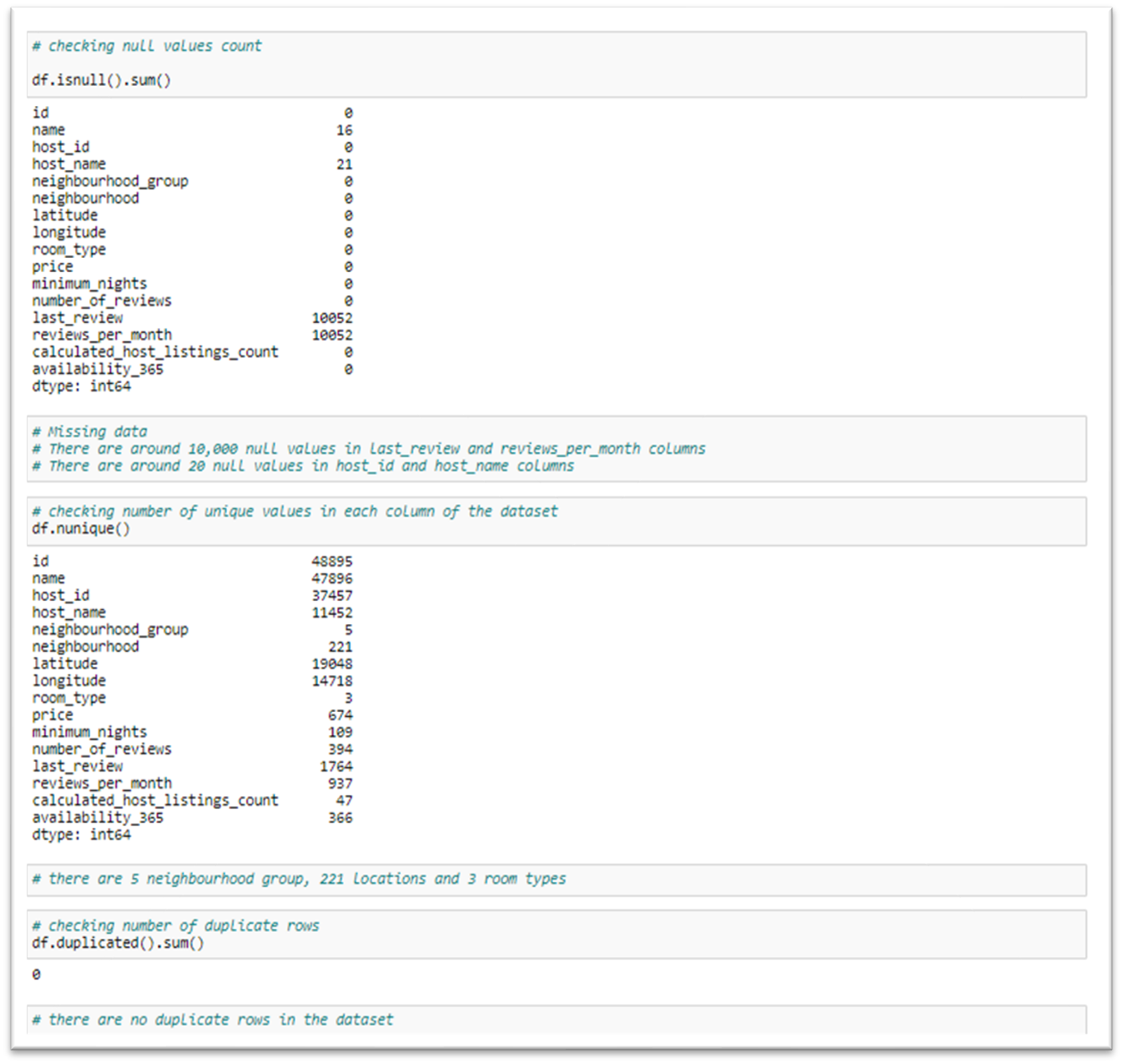
## Step 2: Data understanding/ Normal routine check



**Observations:**

* Dataset contains **48895** rows and **16** columns.
* 3 columns are of float type, 7 columns of int type and 6 columns of object type.
* Average price of rentals in NY is **$153**, average no. of nights spent is **7 days**, and average availability of an Airbnb is **112 days.**
* There are some listings which appear to be free.
* Few customer(s) are also exists who have rented for almost **3 years**.
* There are rentals which are available for all times of the year too i.e. **365 days**.
* Outliers are also present in columns: minimum\_nights, number\_of\_reviews, reviews\_per\_month and calculated\_host\_listings\_count columns.

## Step 3: Data Quality inspection

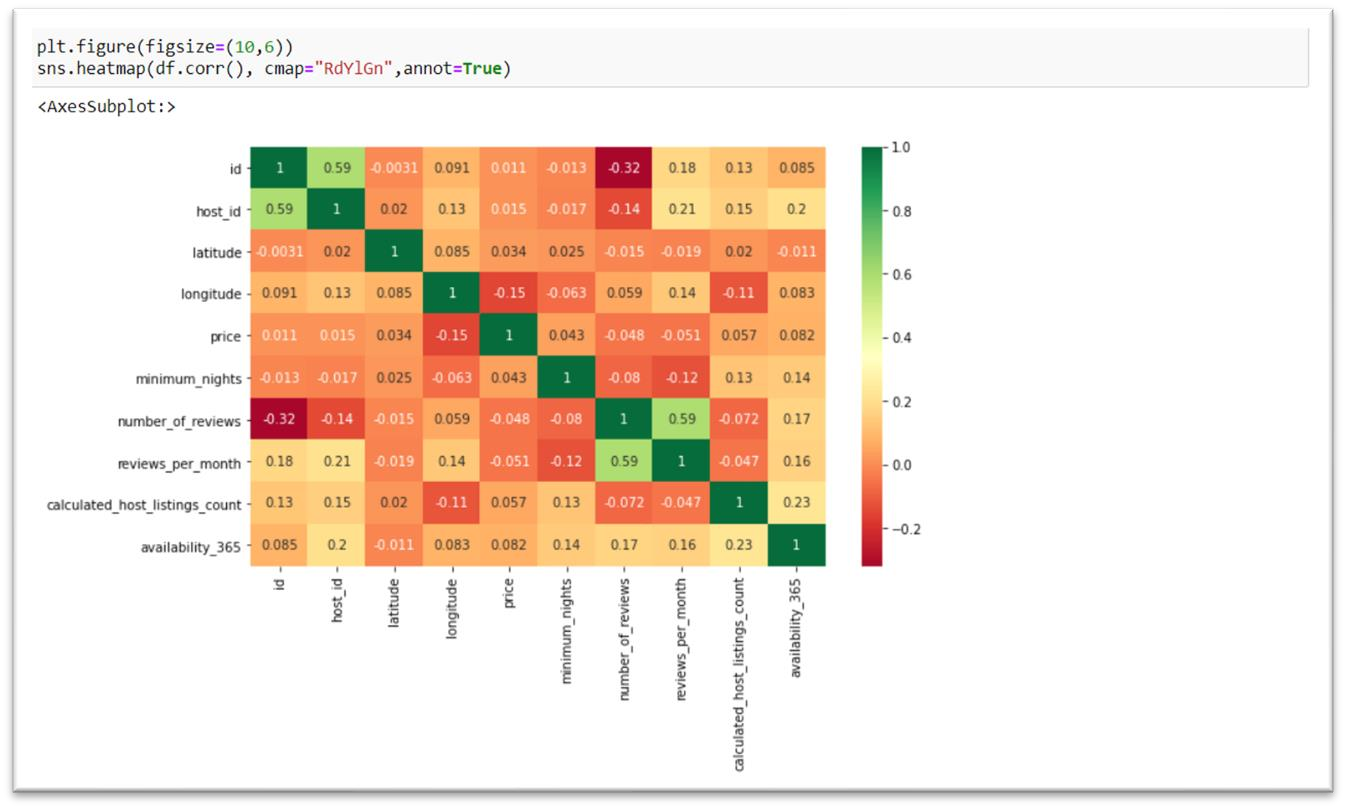


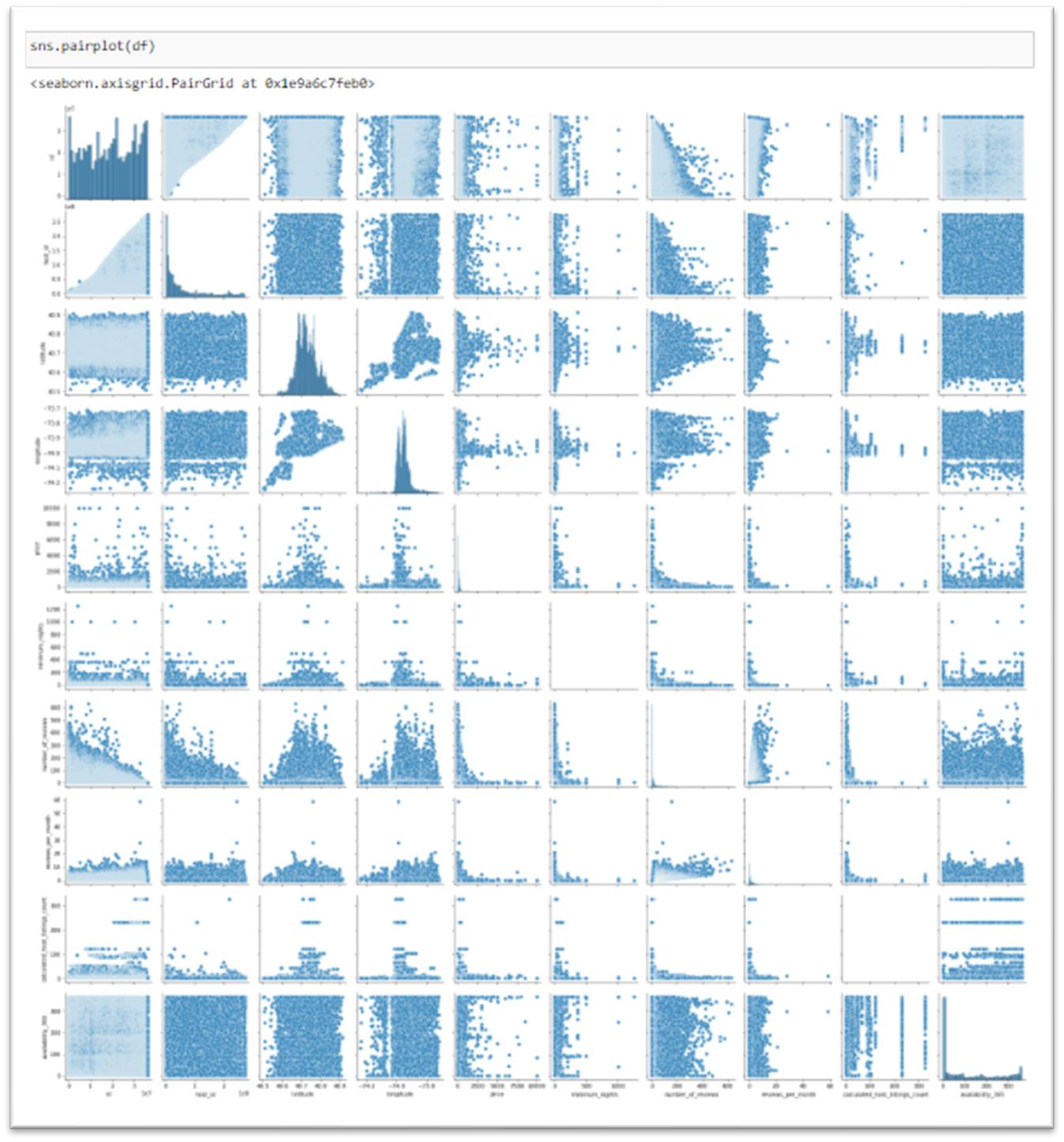
## Step 4: Data Cleaning

Four columns have NaN values, and the missing values must be dealt with.



## Step 5: Bivariate analysis





• There are no significant correlations between columns in the dataset.

## Step 6: Exported cleaned data frame to a CSV file



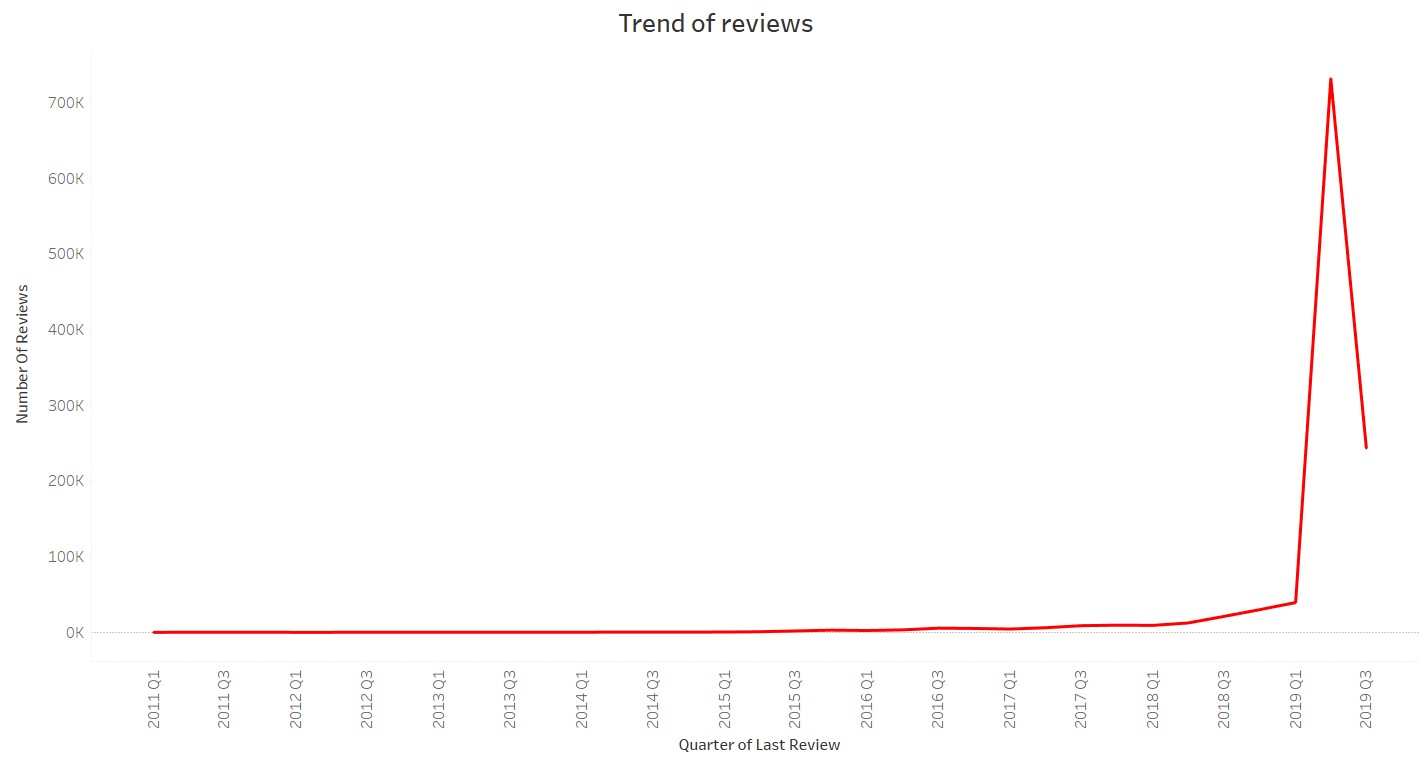
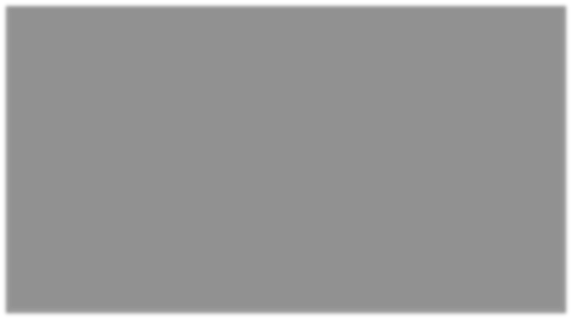
## Step 7: Data visualization

**COVID**

**-**

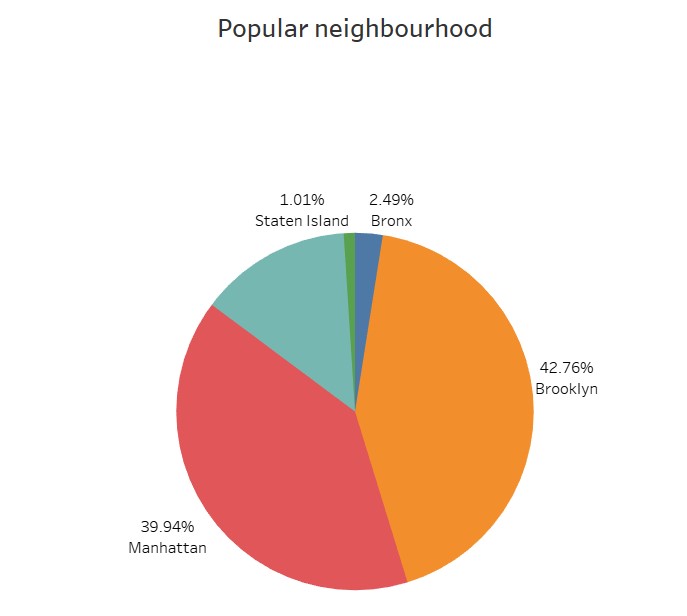
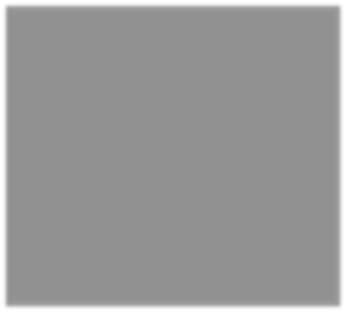
**19**

**pandemic effect on Airbnb business**



* COVID-19 pandemic affected Airbnb business due to travel restrictions.
* Revenue took the largest hit in NYC in the **Q2** and **Q3** of **2019**.

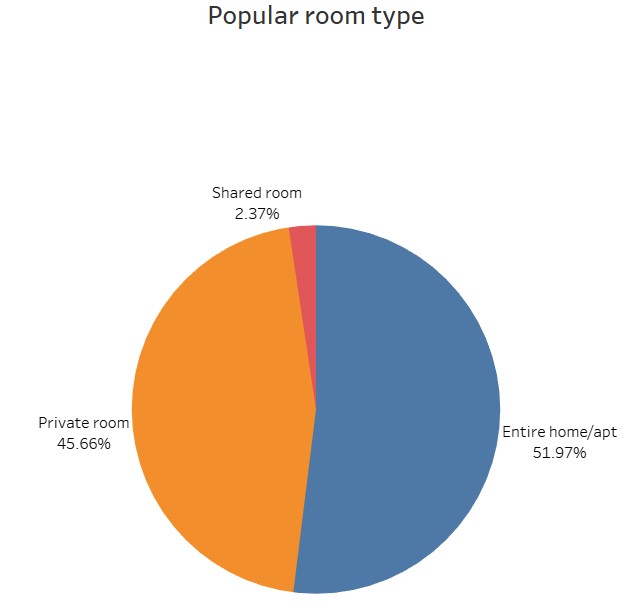
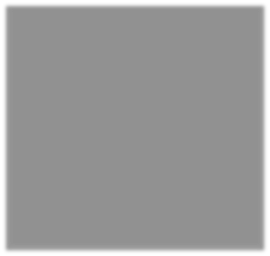
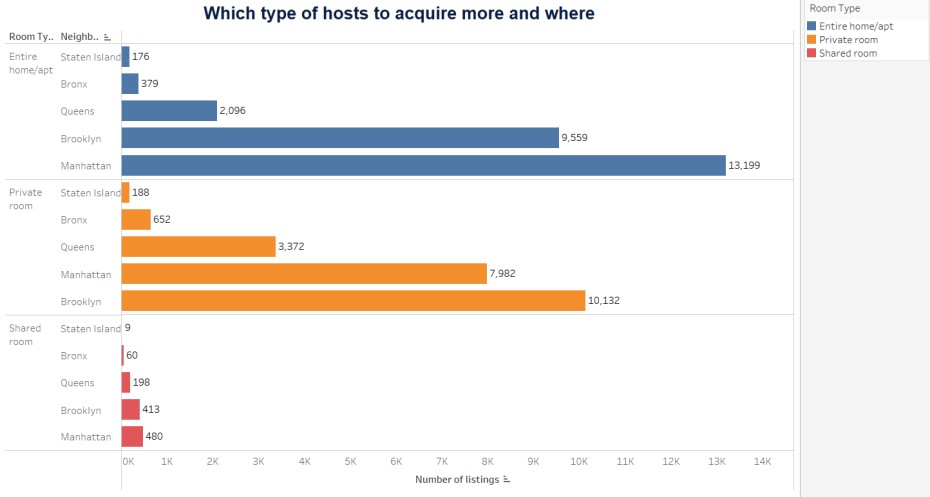
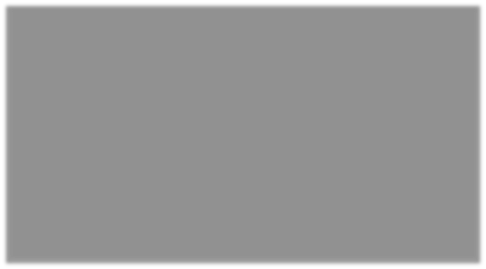
**Neighborhood with maximum listing**



* **Manhattan** and **Brooklyn** are most popular neighborhoods with regards to total number of listings. These two neighborhoods account for **82.7%** of the listings in NYC.

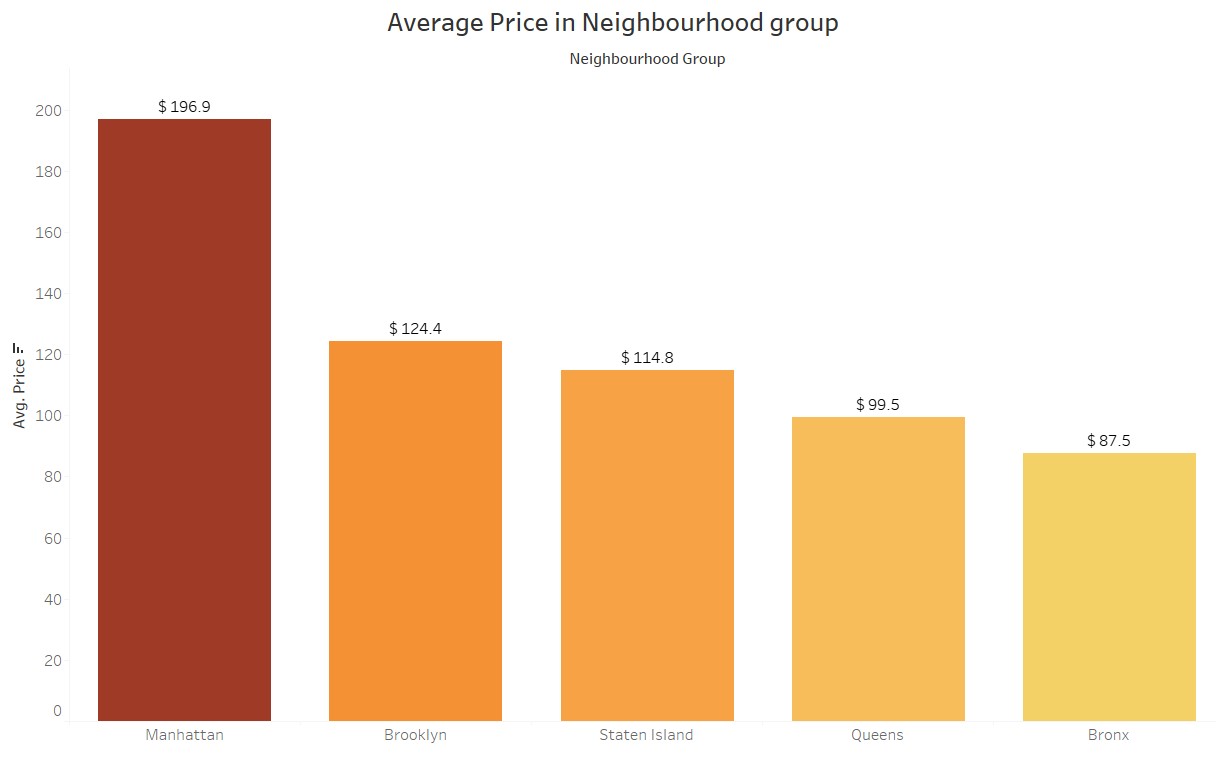
# Insights

**1. Which type of hosts to acquire more and where?**



• Although, shared room accounts for only **2.37%** of all listings in NYC, Airbnb must acquire more no. of Entire home/apt and private room listings across all neighborhood groups as these will be preferred by customers during the pandemic to avoid coming in contact/cohabiting with strangers.

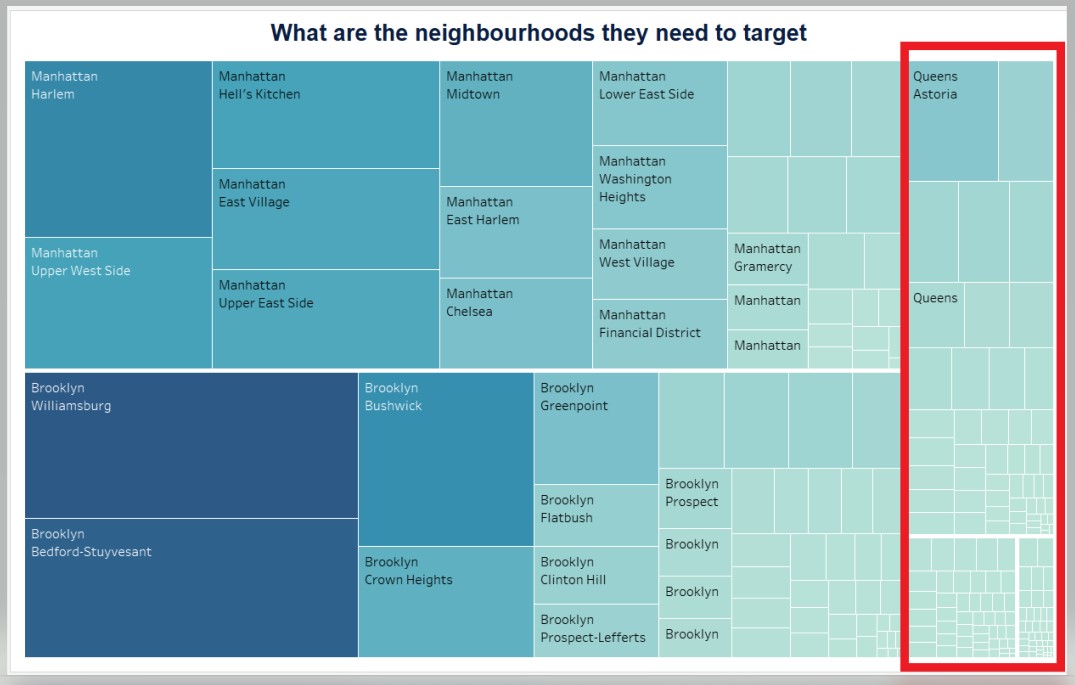
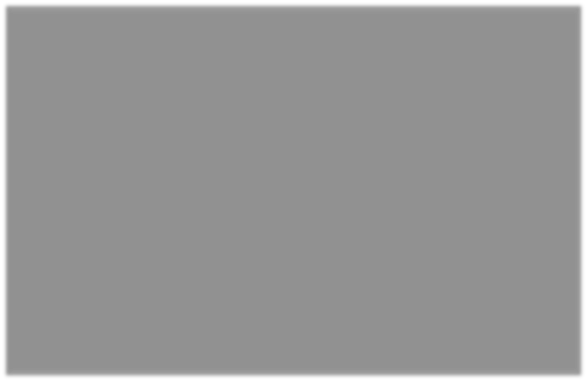
## Average price of listing per neighborhood



• Average price of listing per neighborhood is **$197** for **Manhattan** which is much higher than the other neighborhoods.

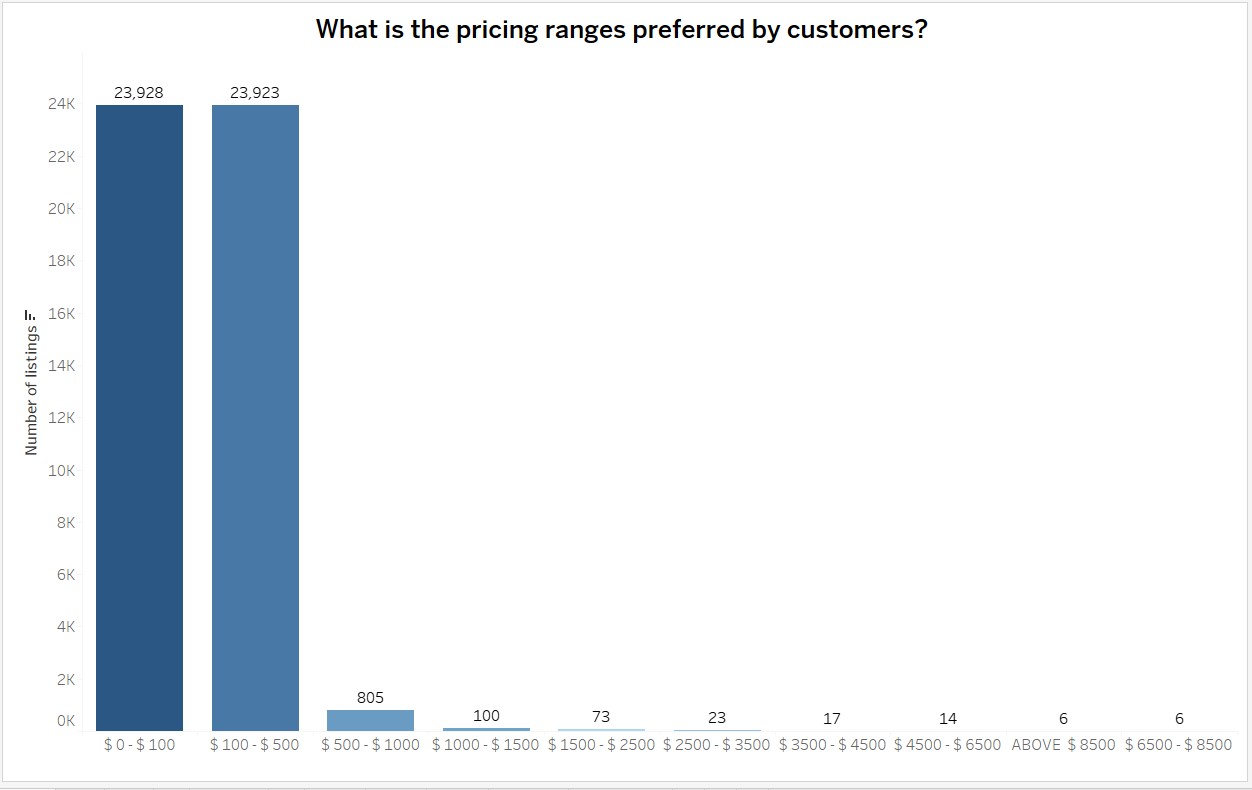
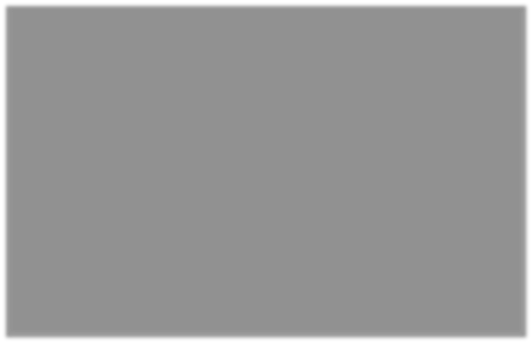
**2. The categorization of customers based on their preferences:**

**i) What are the neighborhoods they need to target?**

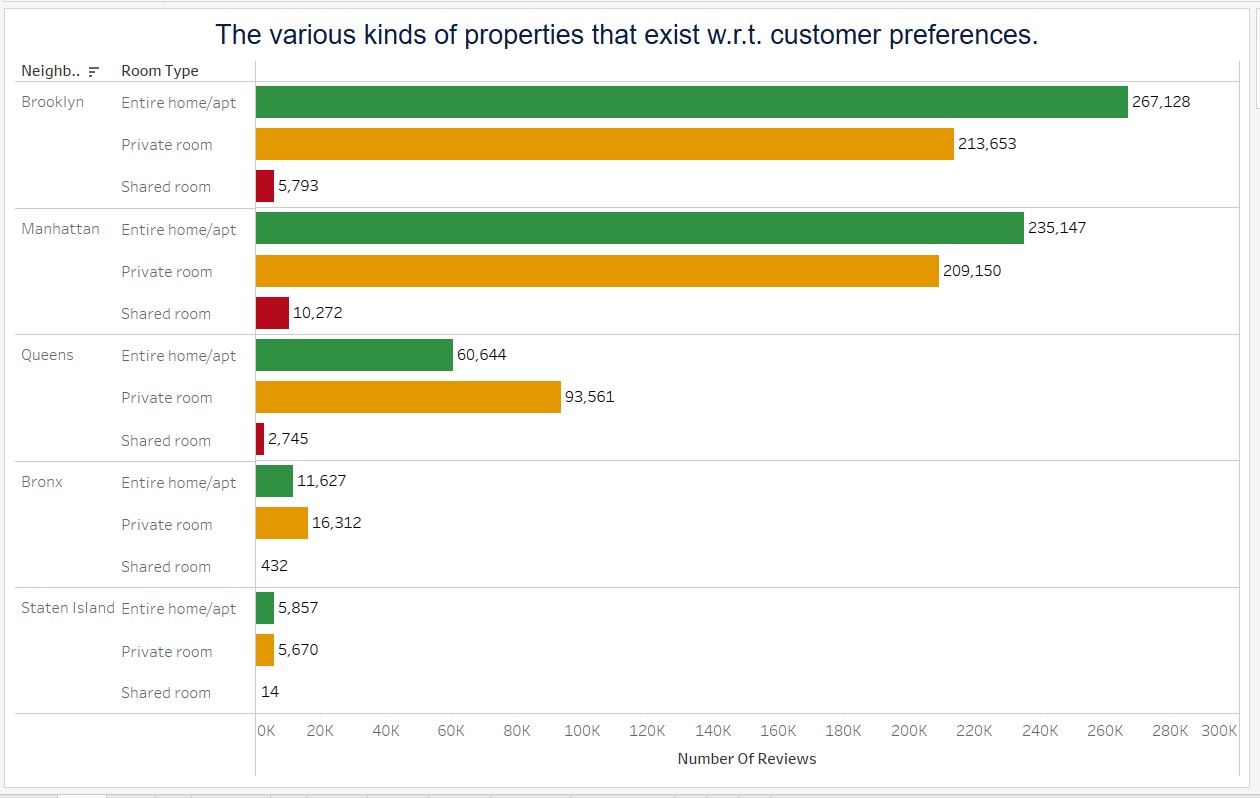
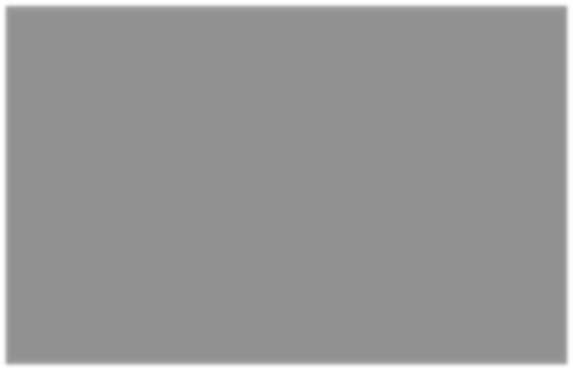


* Airbnb must acquire more customers and listings at **Bronx**, **Staten Island** and **Queens**.

* 1. **What are the pricing ranges preferred by customers?**



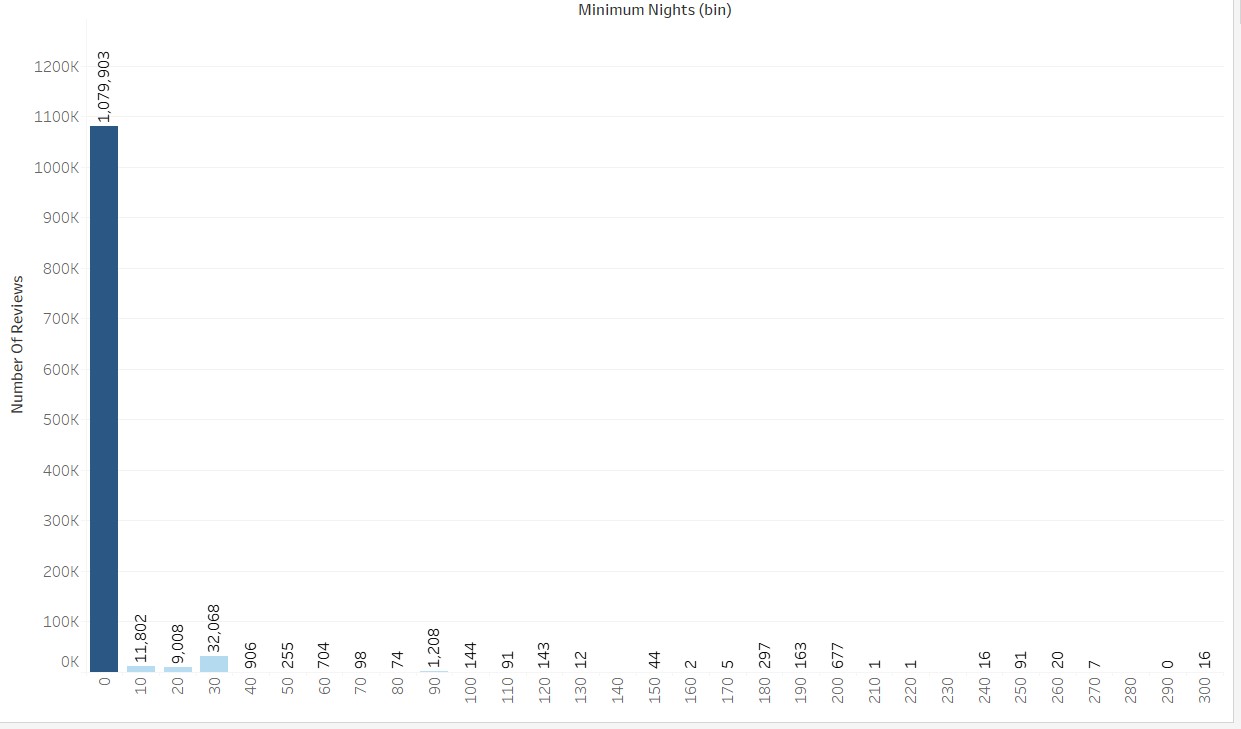
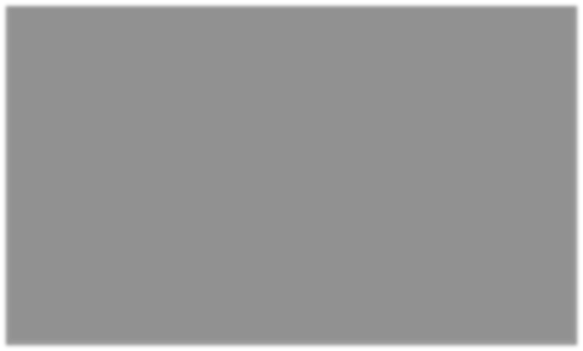
* **$0 to $500** price range per night is preferred by customers while visiting NYC.
  1. **The various kinds of properties that exists w.r.t. customer preferences.**



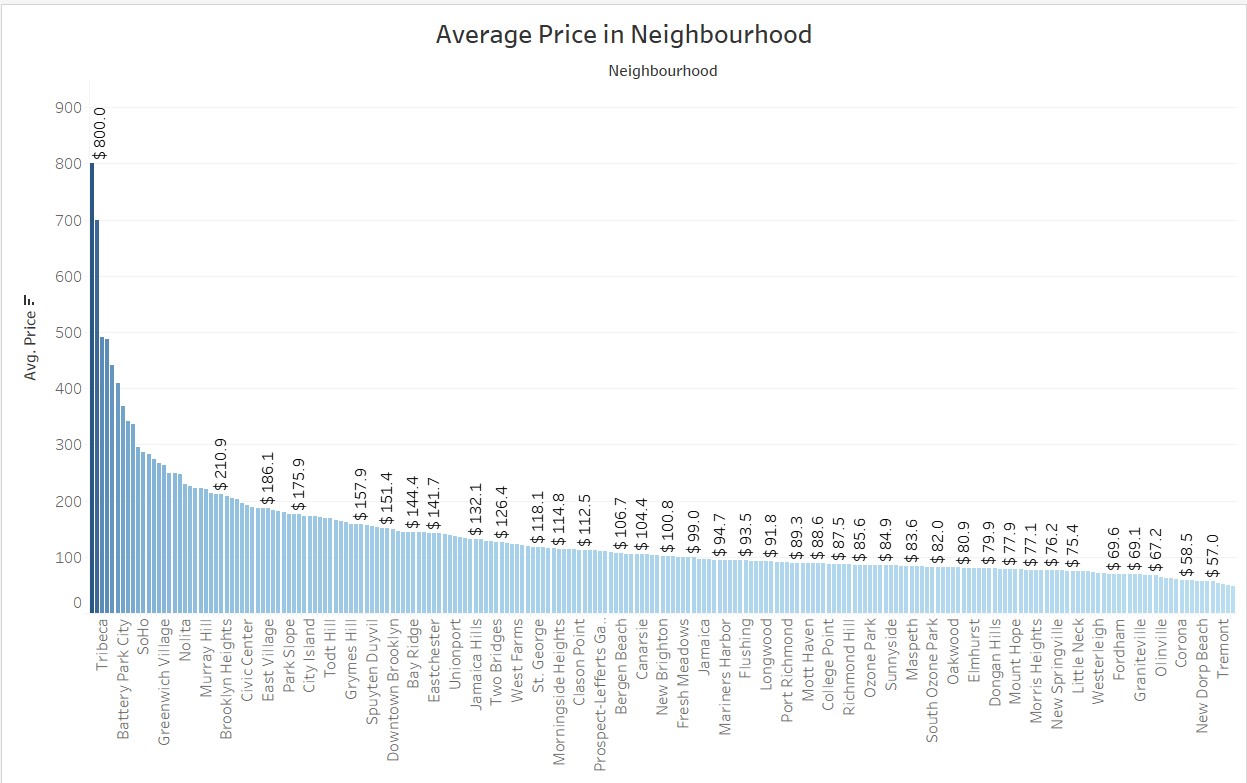
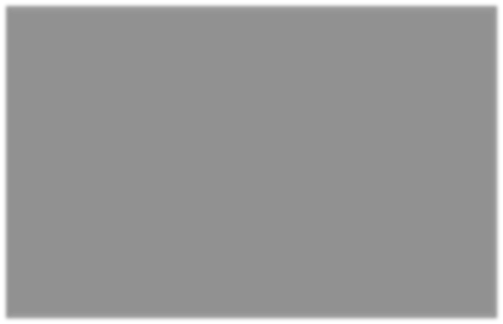
* **Manhattan** and **Brooklyn** prefer Entire Home/Apartment and Private Rooms.
* **Bronx** and **Queens** neighborhood prefer Private rooms over Entire Home/Apartment.
* At **Staten** **Island**, customer doesn’t have any specific preference over Entire Home/Apartment and Private Rooms.
* **Shared rooms** are least preferred among all the neighborhoods might be because of pandemic.

* 1. **Adjustments in the existing properties to make it more customers oriented.**

* Since, customers prefer **entire home/apt or private rooms** more. So, shared rooms can be converted to private rooms as customers would not prefer sharing rooms with strangers due to ongoing pandemic.
* Customers prefer **$0 to $500** per night price points; hence, decrease the listings’ price to attract more customers towards existing properties.
* Customers usually prefer **1-10 days** of minimum nights per stay. But due to ongoing pandemic, customers may prefer to book properties for 30, 60 or 90 days to avoid frequent travelling. Hence, the listings must have high availability.

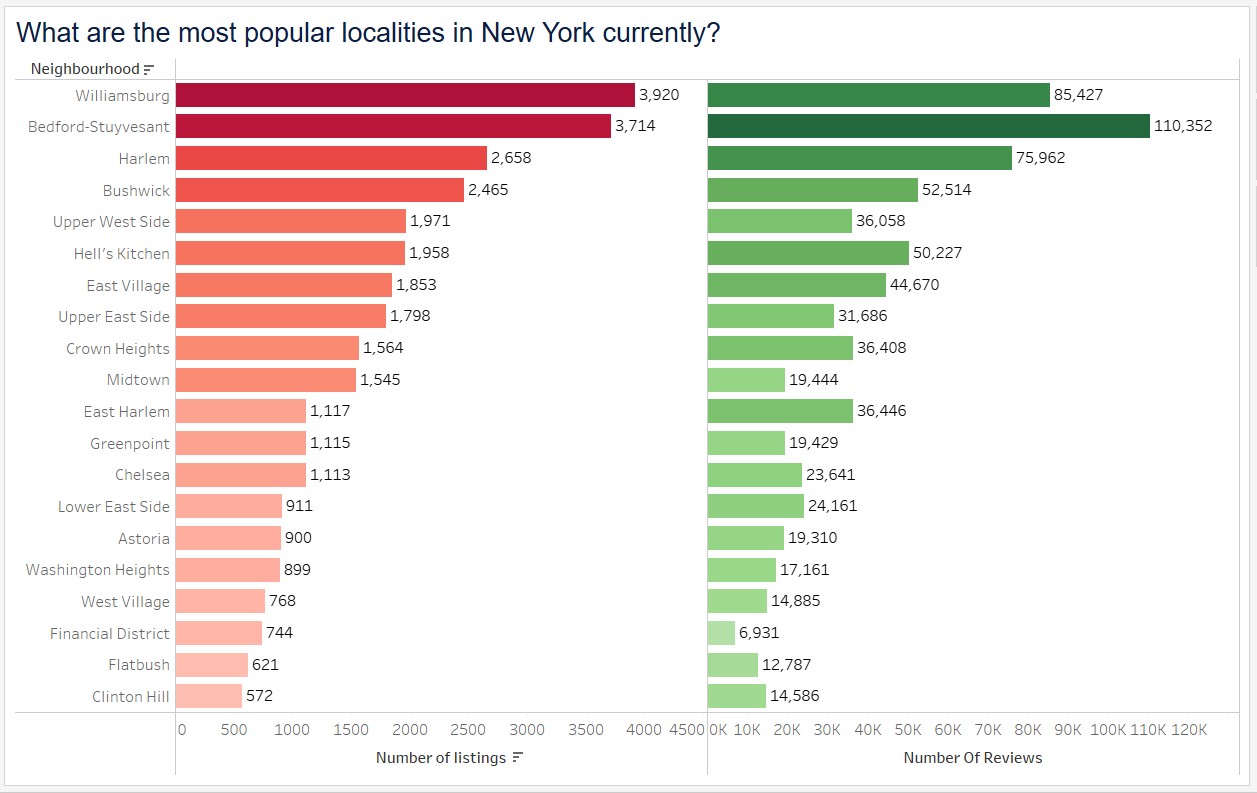
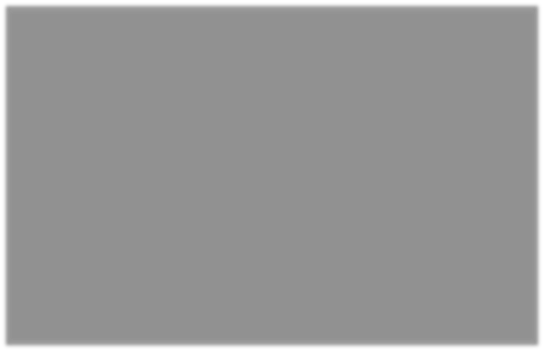


## Average price per neighborhood

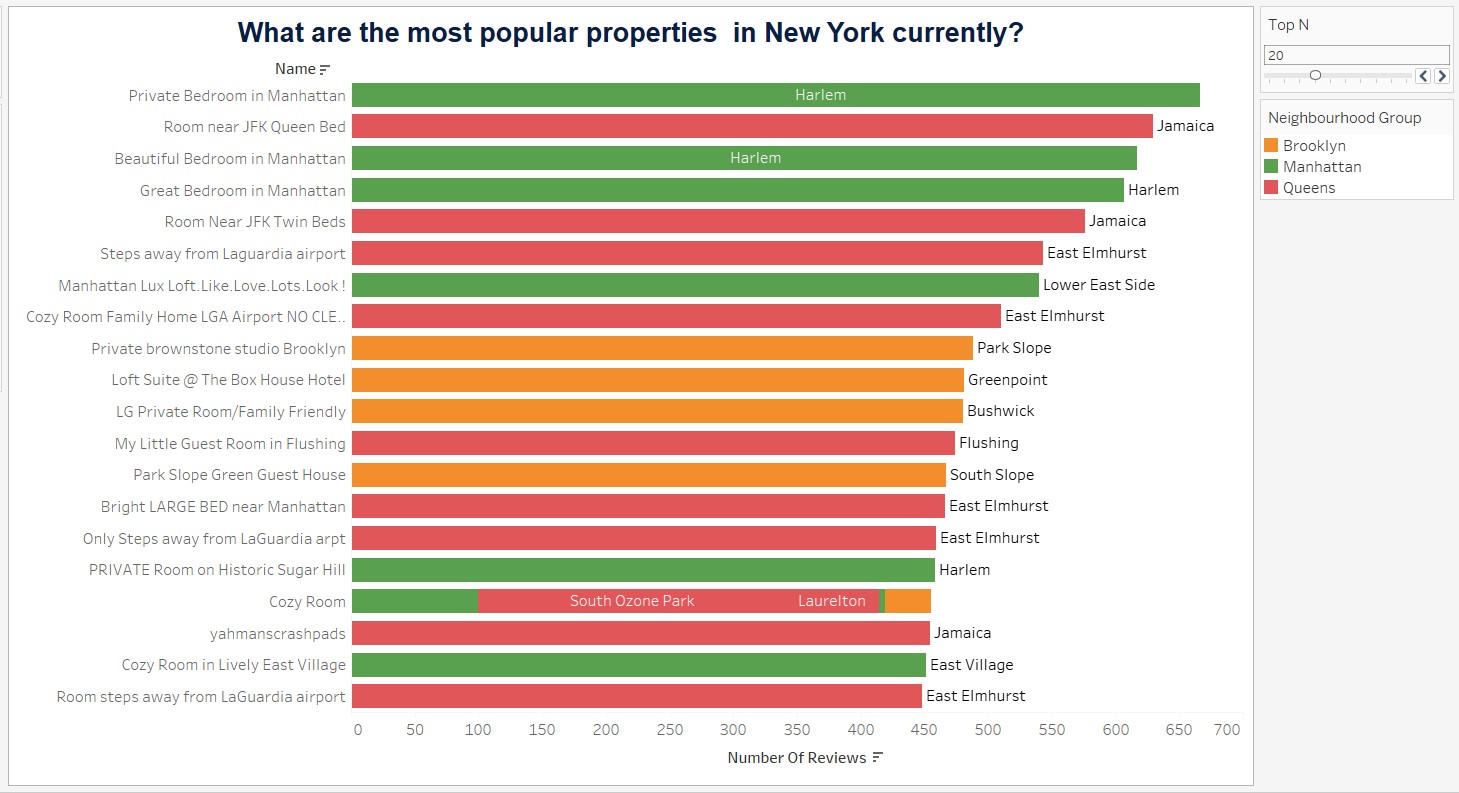
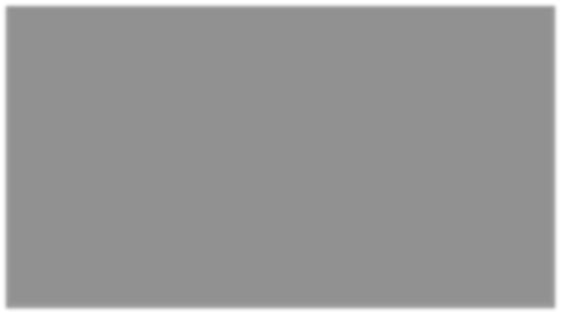


* Two neighborhoods in **Staten** **Island** have high average price according to above graph. This is due to availability of only one property in these neighborhoods.
* The cheapest locality at **Manhattan** is **Inwood** (**$88.9**) and that in **Brooklyn** is **Borough** **Park** (**$63.1**)

**3. What are the most popular localities and properties in New York currently?**



* **Williamsburg**, **Bedford-Stuyvesant**, and **Harlem** have the highest number of reviews and listings in NY neighborhoods.
* Although, **Williamsburg** has the highest number of listings, but its number of reviews are comparatively less.

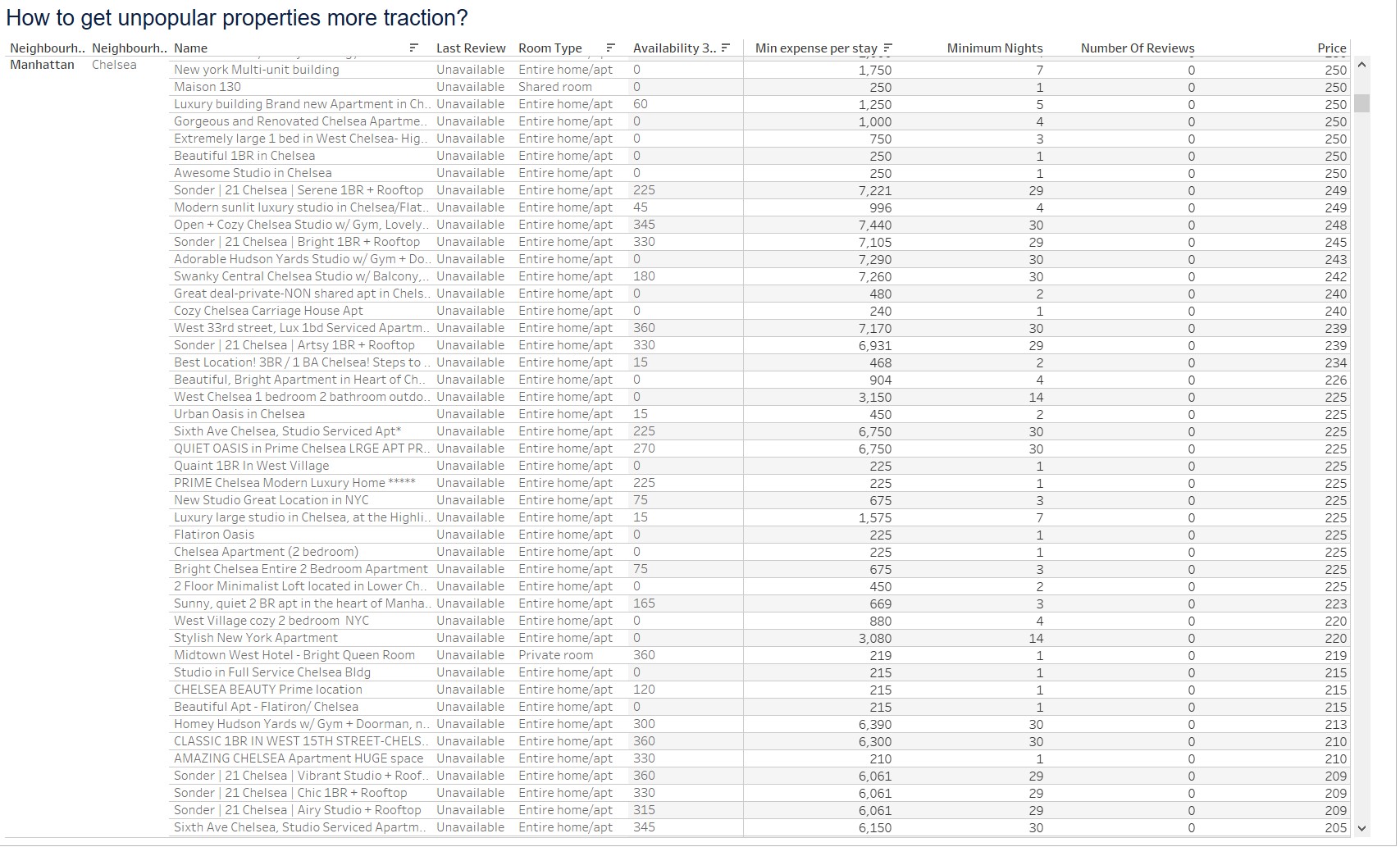
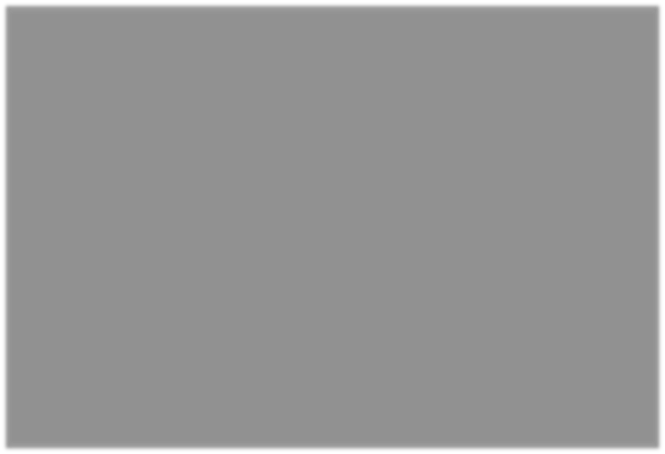


* There are few properties with the same name across neighborhoods.
* Although, the number of listings in **Queens** is less, properties feature a lot in the top N results w.r.t the highest number of ratings. This may be attributed also to the fact that fewer listings are available in Queens and hence, listings have higher number of ratings available.

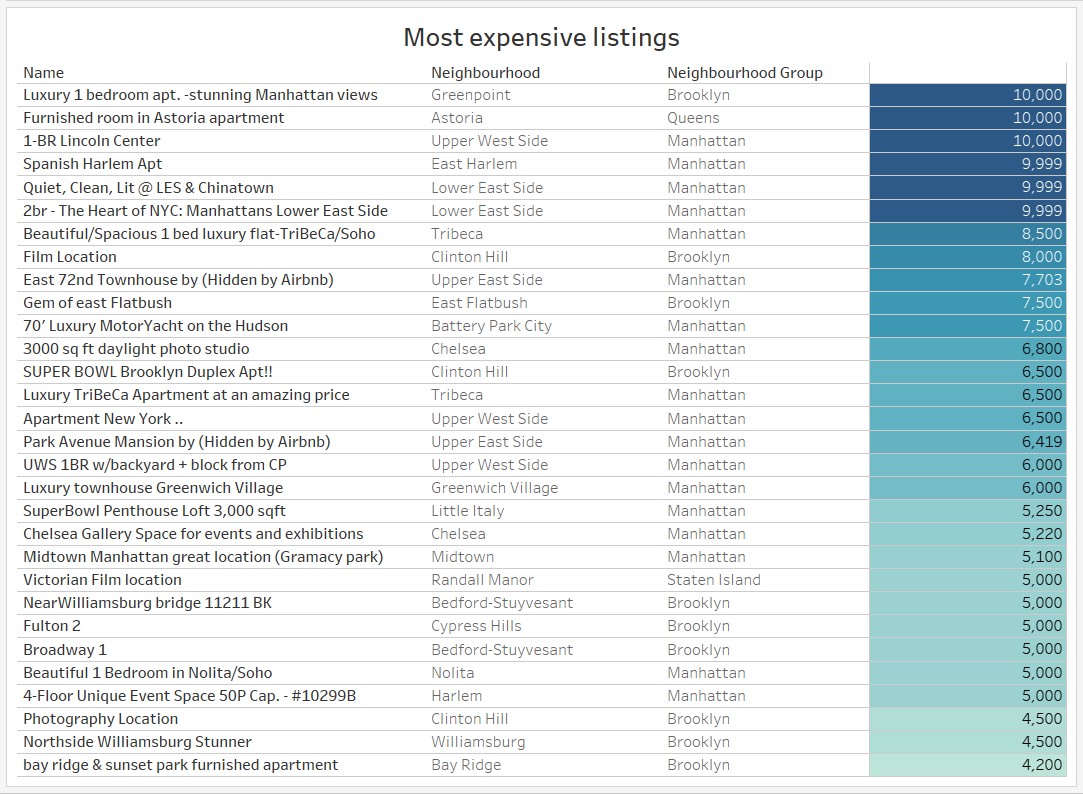
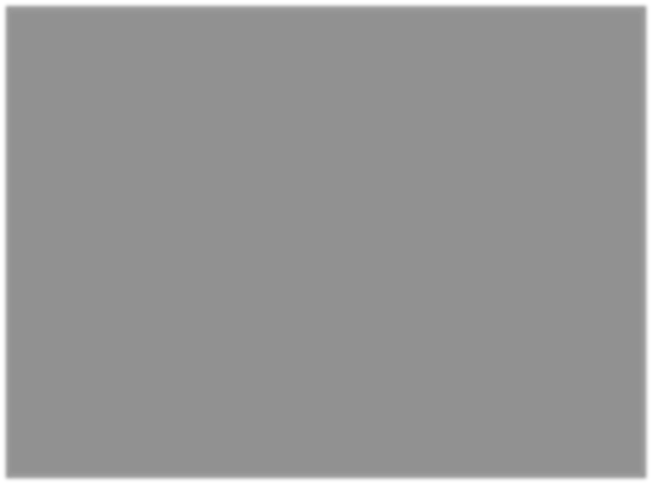
**4. How to get unpopular properties more traction?**

**Assumption:** Unpopular properties are the ones that have not been visited even once.

* Customers prefer **$0 to $500** price points per day. Hence, decrease the listings’ price to attract more customers towards existing properties.
* Customers prefer **1-10** days of minimum nights per stay. Due to ongoing pandemic, customers may prefer to book properties for 30, 60 or 90 days to avoid frequent travelling. Properties with higher days of minimum nights per stay should also be acquired as they could have market traction.
* Listings must be highly available to target higher number of bookings.

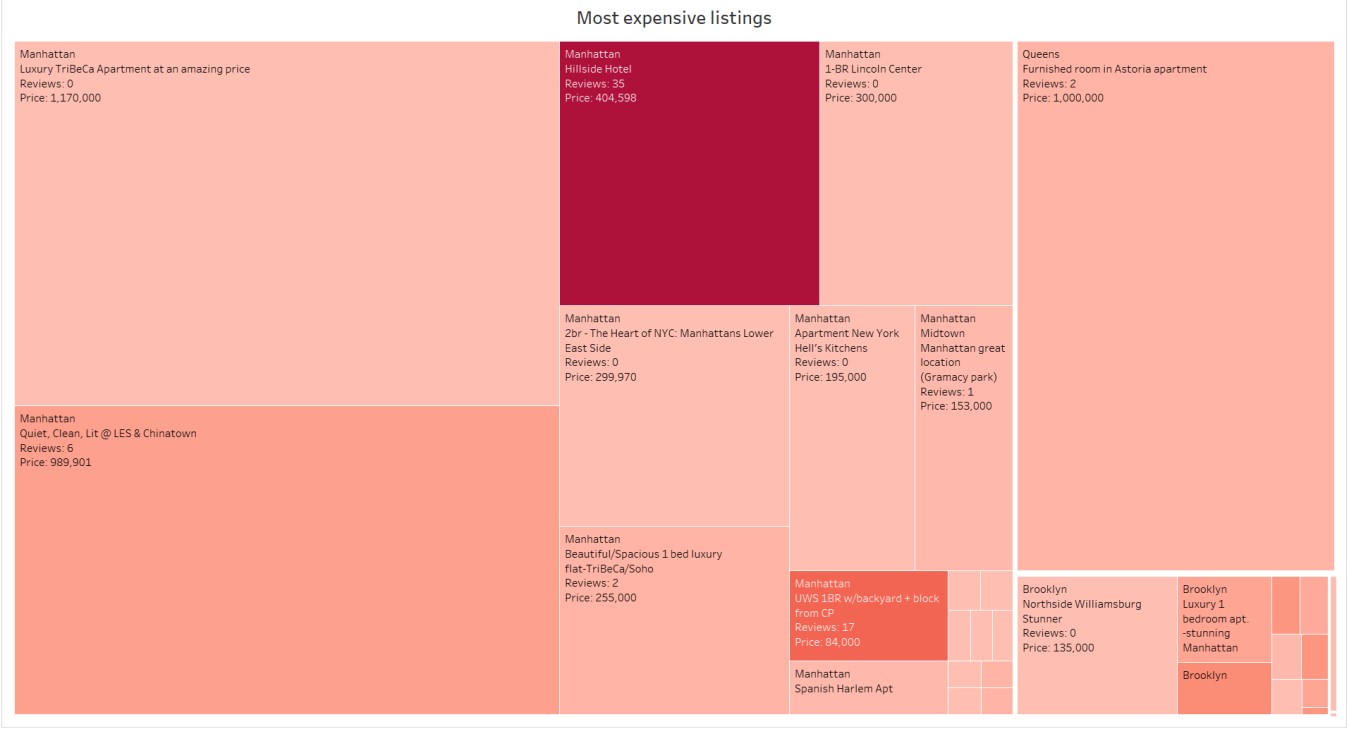
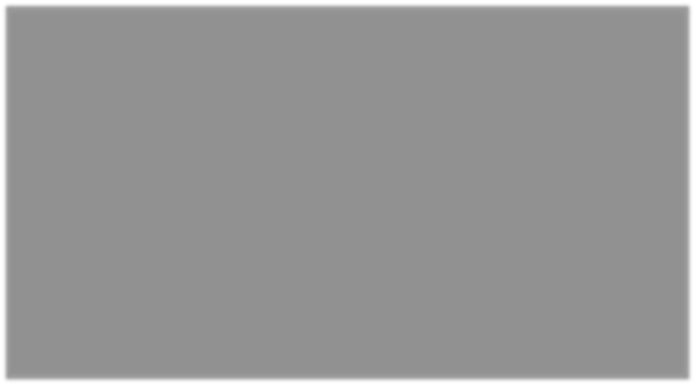


**5. Most expensive listings by price per night**



* Listing at **Queens** is available for **$10,000** per night and that in **Staten Island** is available for **$5,000**. These are **premium** listings for Queens and Staten Island respectively.
* There are multiple **premium** listings for **Brookyln** and **Manhattan**.

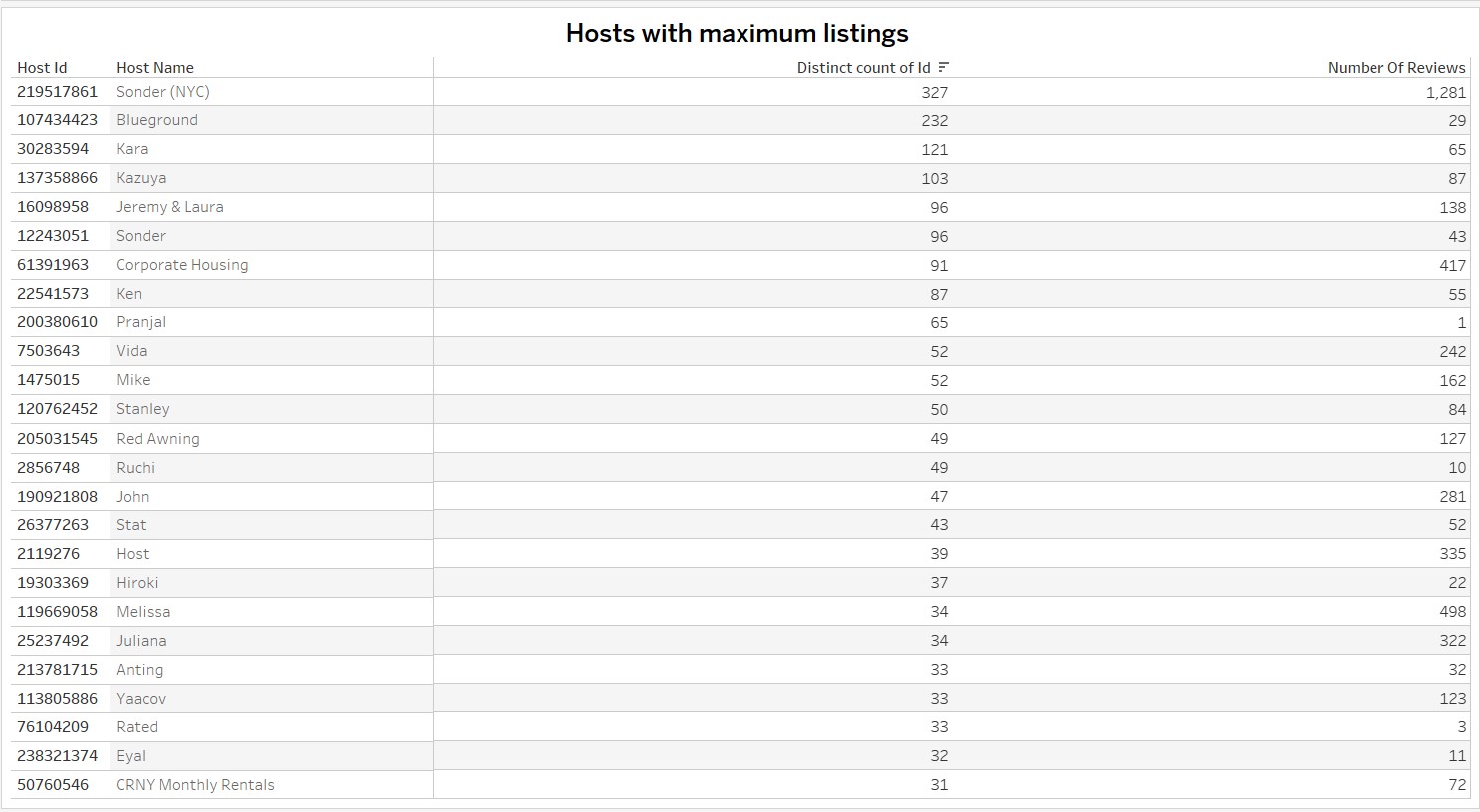
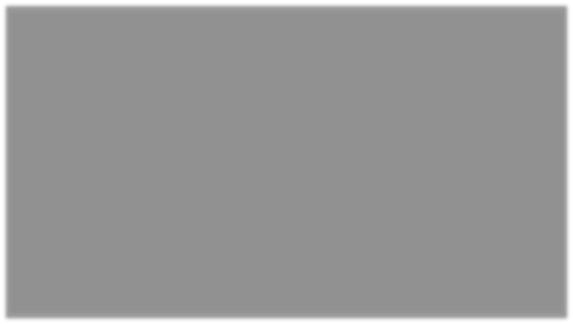
**6. Most expensive listings by mininum expense per stay**



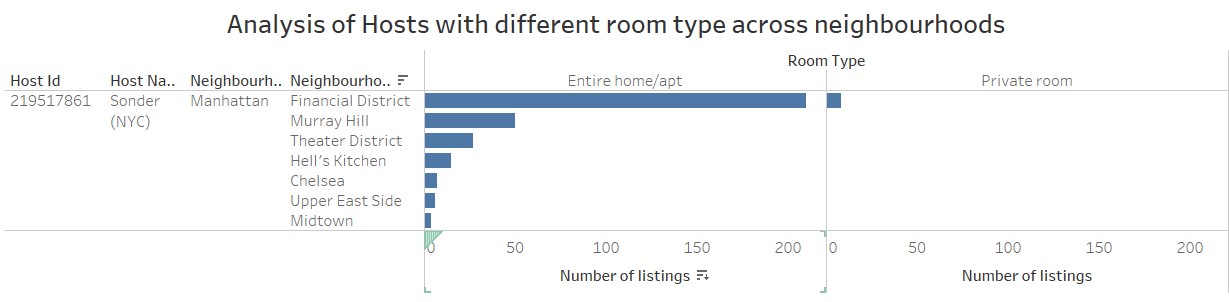
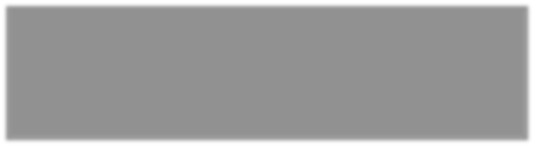
* Most expensive listing costs **$1,170,00** per stay but no reviews available for this listing.
* **Hillside Hotel** and **UMS 1BR w/backyard+block** from **CP** at **Manhattan** are most visited **premium** listings.

**7.**

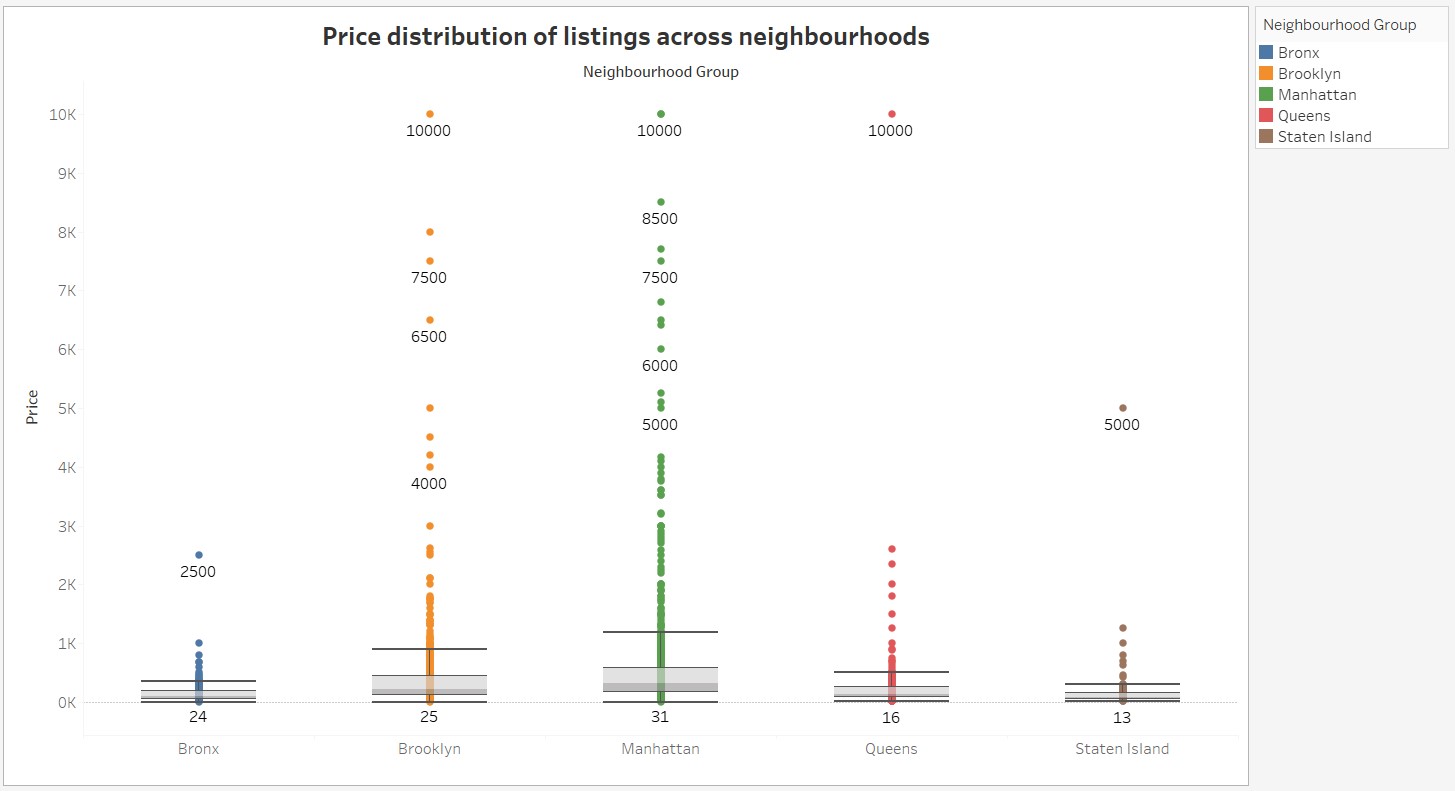
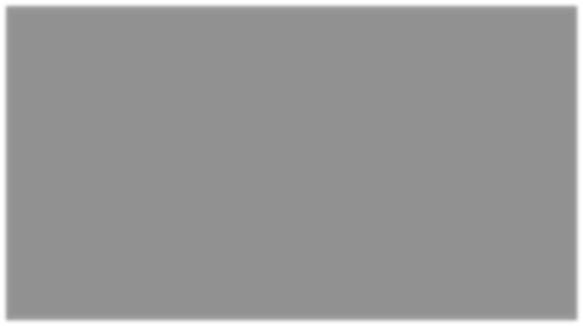
**Hosts with maximum listings**



* There are many hosts with more than 10 listings at different neighbourhoods.
* Only **Sonder(NYC)** has a large number of reviews, hence many customers have stayed at his listings.
* All other listings have very less reviews, either customers haven’t reviewed the listings as much or due to poor reviews, these listings have not been visited frequently. The quality of these listings must be reviewed by Airbnb.
* **Sonder(NYC)** has highest number of listings across NY and he primarily owns Entire room/apt.



**8. Price distribution of listings across neighborhoods**



* **Premium properties** in **Bronx** and **Queens** must be targeted as they cost an upwards of **$2500**.
* **Non-premium** **properties** in **Manhattan** and **Brooklyn** must be targeted. Premium properties are exorbitant and may not attract a lot of bookings. Cheaper listings are usually preferred in these localities.