

# AirBnB New York Case Study Methodology



- By Aamir Farhan Sheikh

# Introduction

Airbnb is a global online marketplace that connects travelers with hosts offering unique accommodations, ranging from cozy apartments and luxury villas to treehouses and boutique stays. Founded in 2008, Airbnb revolutionized the travel industry by enabling individuals to rent out their properties or spare rooms to guests, providing an alternative to traditional hotels.

The platform operates in over 220 countries, offering flexible booking options, personalized experiences, and cost-effective stays. With a user-friendly interface, secure payment system, and a robust review system, Airbnb ensures trust and transparency between hosts and guests.

## Business Objective

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.

Airbnb wants to focus on the listings in various cities in the state of New York.

# Data Source

The analysis has been done on a dataset provided by Airbnb which contains details of New York city listings in certain neighbourhoods for the years through 2011 to 2019. However, The data mostly represents the records from the year 2019.

The data has some null values which did not affect the analysis and hence had not been treated.

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

# Assumptions

As we are not aware about the nature of reviews, we have assumed them as positive and that the properties which received higher number of reviews have a better customer liking and preference.

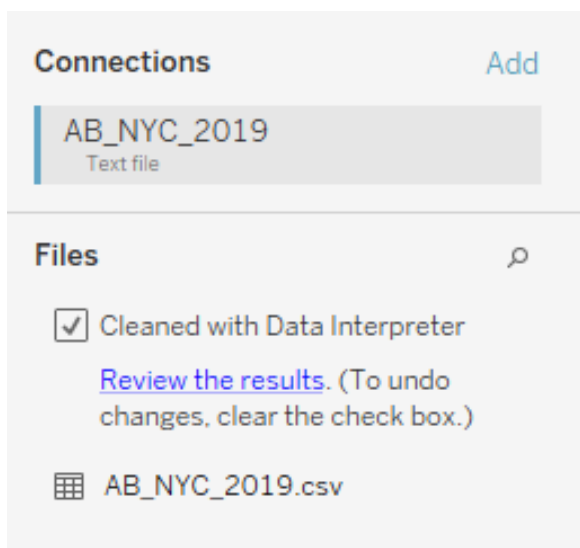
Minimum nights could be interpreted as

- Minimum nights a listing offers the booking to be
- Minimum nights the customers prefer to book


We are considering the second situation in order to understand customer preference.

## Data Wrangling

- Imported the CSV file in Tableau
- Used Data interpreter to fix the headers and data type

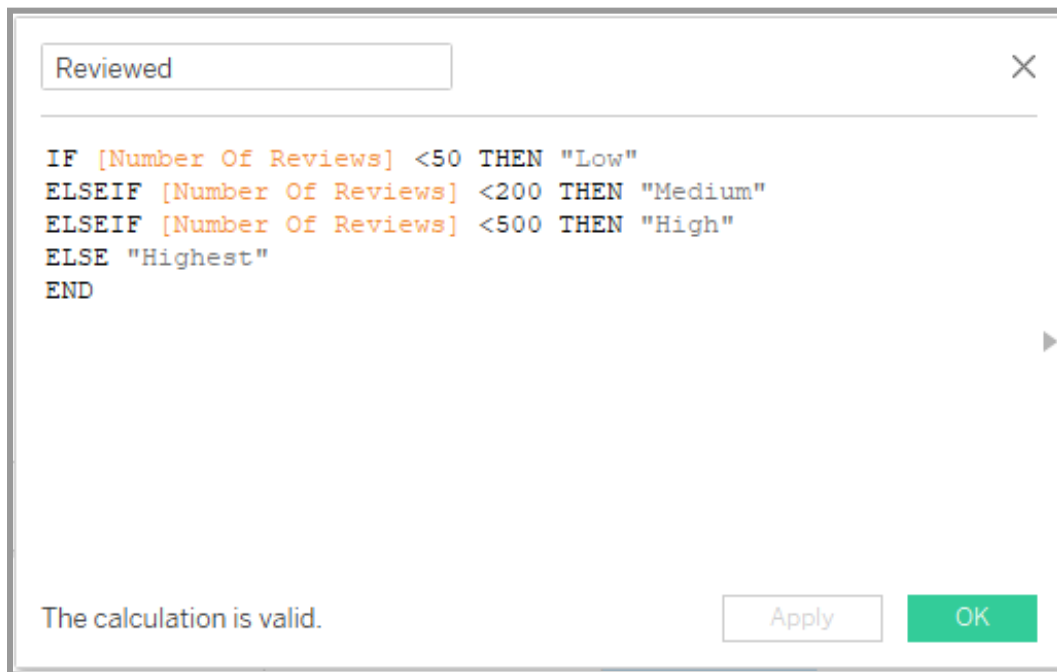


- Data contains 48895 rows and 16 fields(columns)
- Did univariate analysis using Tableau on the fields to see their distributions, the unique values in a field, the missing values and to check for outliers if any
- The data had some null values which did not affect the analysis and hence had not been treated


AB_NYC_2019.csv
<b>Last Review</b>
12-05-2019
21-06-2019
<i>null</i>
22-06-2019
<i>null</i>
05-07-2019
21-06-2019
27-05-2019
20-06-2019

- Changed data type of ID and Host ID to string

- Created a calculated field named "Reviewed" using the Number of reviews and bucketed the results in low, medium, high and highest.



A screenshot of a software dialog box titled "Reviewed". The dialog box has a title bar with a close button (X). Below the title bar is a text input field containing the word "Reviewed". The main area of the dialog box contains a multi-line text field with the following SQL-like code:

```
IF [Number Of Reviews] <50 THEN "Low"  
ELSEIF [Number Of Reviews] <200 THEN "Medium"  
ELSEIF [Number Of Reviews] <500 THEN "High"  
ELSE "Highest"  
END
```

At the bottom of the dialog box, there is a status message "The calculation is valid." and two buttons: "Apply" and "OK".

- Extracted month from the last review date to understand the pattern



A screenshot of a software dialog box titled "Month of review". The dialog box has a title bar with a close button (X). Below the title bar is a text input field containing the words "Month of review". The main area of the dialog box contains a multi-line text field with the following SQL-like code:

```
LEFT(DATENAME('month', [Last Review]), 3)
```

At the bottom of the dialog box, there is a status message "The calculation is valid." and two buttons: "Apply" and "OK".

- Excluded null values from reviews per month axis
- Created buckets for minimum nights using 'Minimum nights' columns and created 'Minimum nights grouped' column



# Methodology

The important factors taken into consideration while doing analysis were

## **Presentation - 1**

- Geography based bookings
- Bookings based on room type
- Number of reviews
- Minimum number of nights

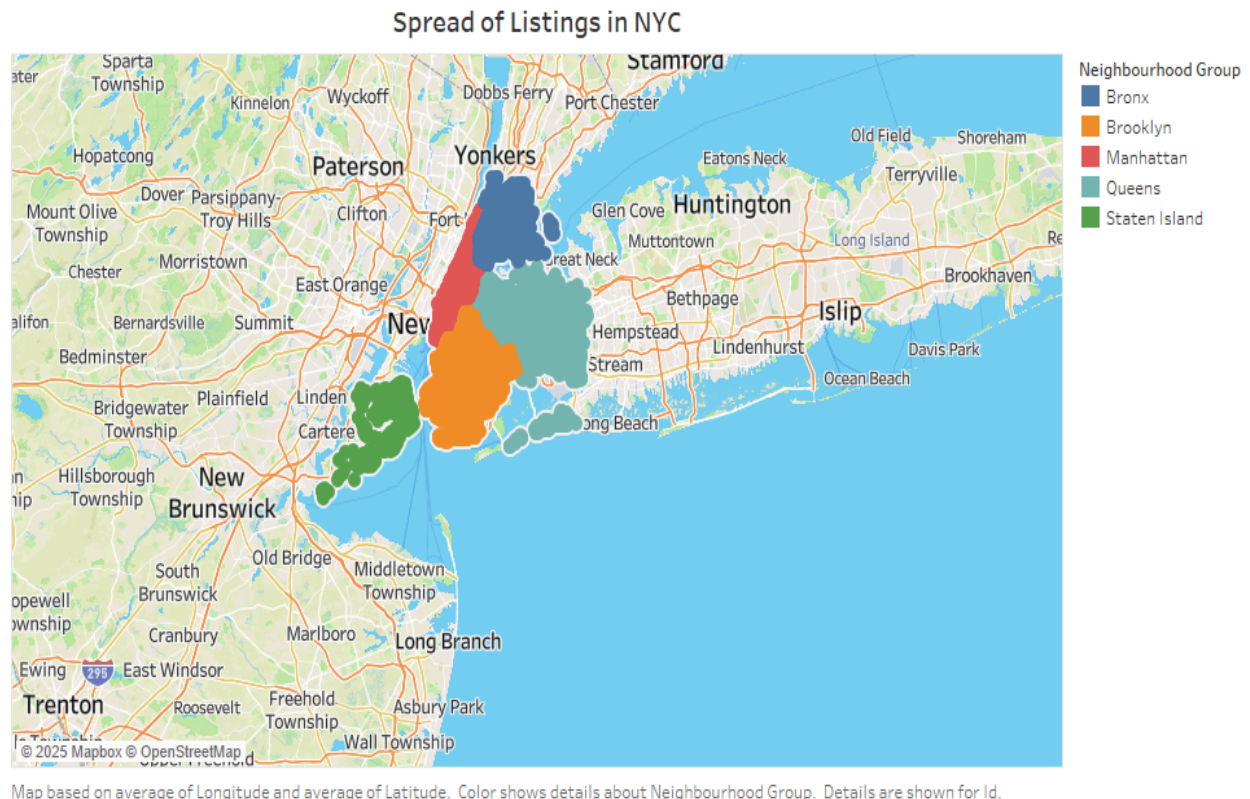
## **Presentation - 2**

- Customer bookings and preference
- Customer experience w.r.t Neighbourhood, Room type & minimum nights
- Price variations
- Customer Reviews
- Listing Availability

Recommendations have been made with respect to these parameters to impact the business positively.



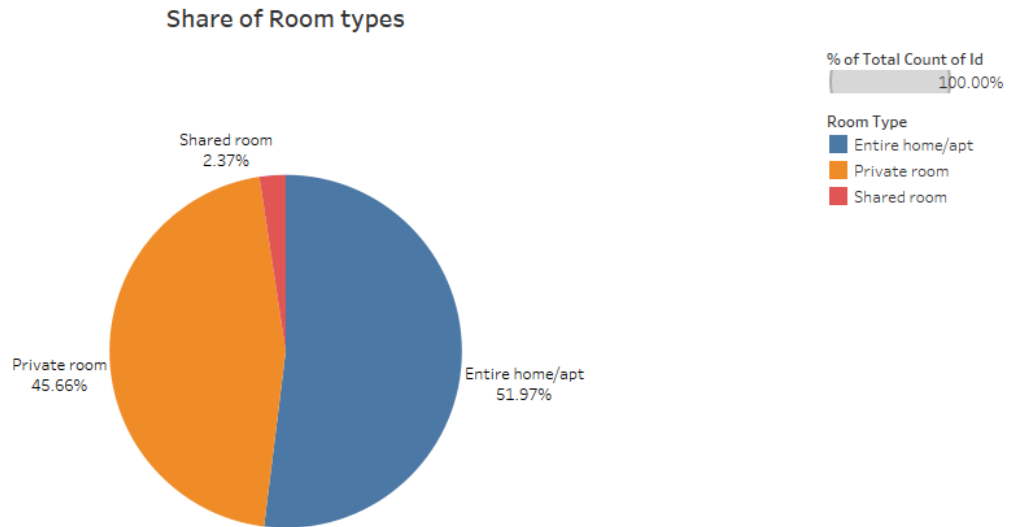
# Presentation 1



From the above visualisation, we studied how Airbnb listings spread out in NYC.

To create the visualisation, we used

- Latitude and Longitude : To create the Map
- ID in Details mark
- Neighbourhood Group in Colour Mark



Room Type and % of Total Count of Id. Color shows details about Room Type. Size shows % of Total Count of Id. The marks are labeled by Room Type and % of Total Count of Id.

## From the above visualisation, we studied the share of different Listing space type

To create the visualisation, we used the PIE chart option using

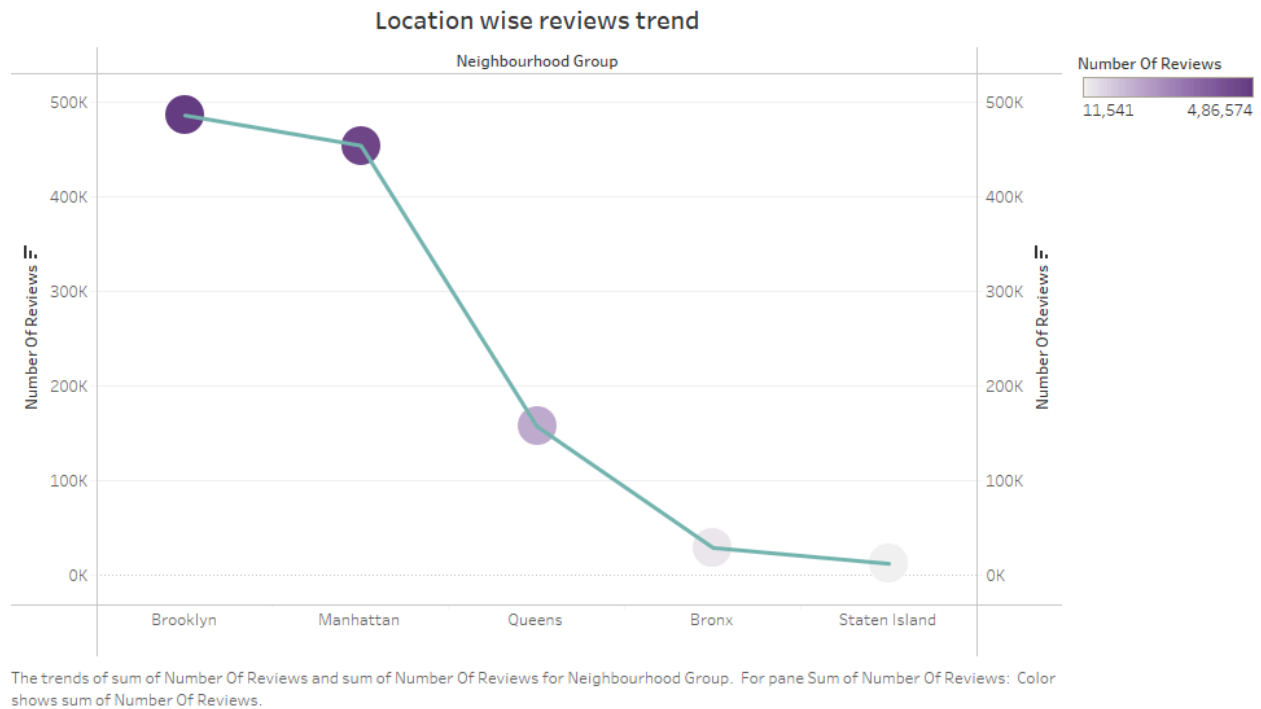
- Count of ID and converted the measure into a %Total across the table using 'Quick table calculation'
- %total count of ID across table is used twice, i.e. in Label mark and in Size Mark
- Room Type field is also used twice, i.e.in Label Mark and in Colour Mark



**From the above visualisation, we studied customer booking preference w.r.t minimum nights of booking**

To create this Bar Chart visualisation we used

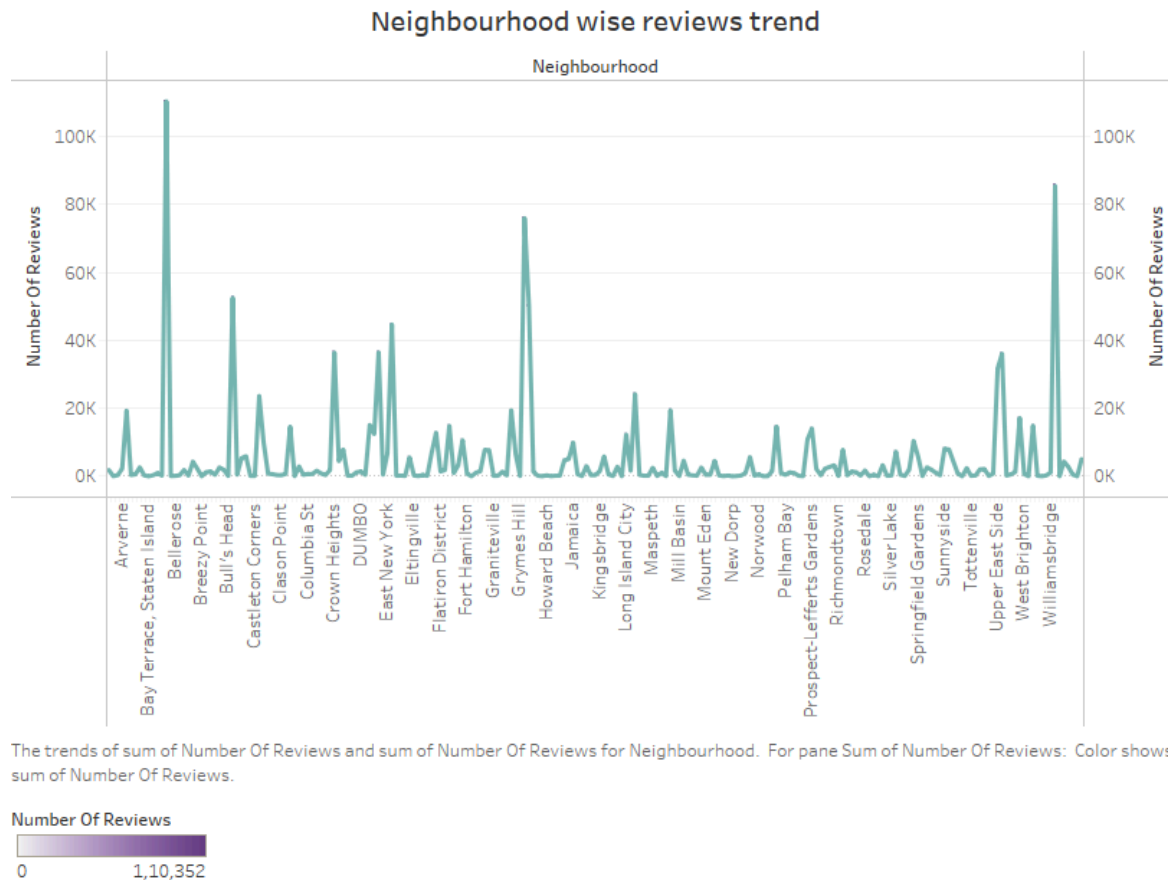
- Count of ID as a measure(rows) and colour mark
- 'Minimum nights grouped' as a field



**From the above visualisation, we studied customer reviews trend per location**

To create this trend line chart visualisation we used

- Neighbourhood Group as a field
- Sum of number of reviews as a measure
- We created two charts with this data and merged them creating a dual axis chart
- We synchronized the axes of both these charts to visualize the trend
- We also added sum or number of reviews to colour mark to show the variation

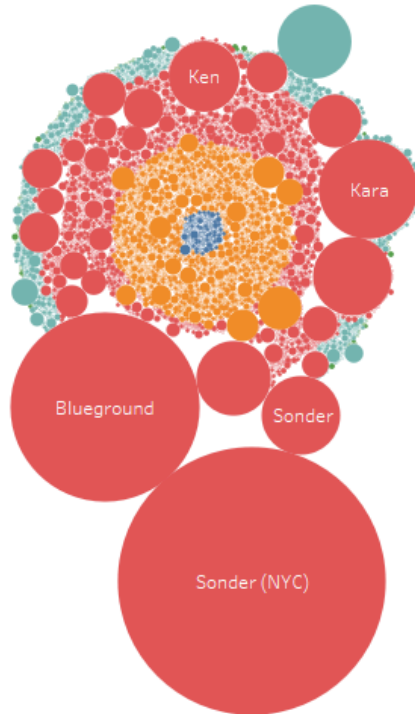


**From the above visualisation, we studied customer reviews trend per Neighbourhood**

To create this trend line chart visualisation we used

- Neighbourhood Group as a field
- Sum of number of reviews as a measure
- We created two charts with this data and merged them creating a dual axis chart
- We synchronized the axes of both these charts to visualize the trend

### Tops Hosts with the most Listings



Host Name. Color shows details about Neighbourhood Group. Size shows sum of Calculated Host Listings Count. The marks are labeled by Host Name.

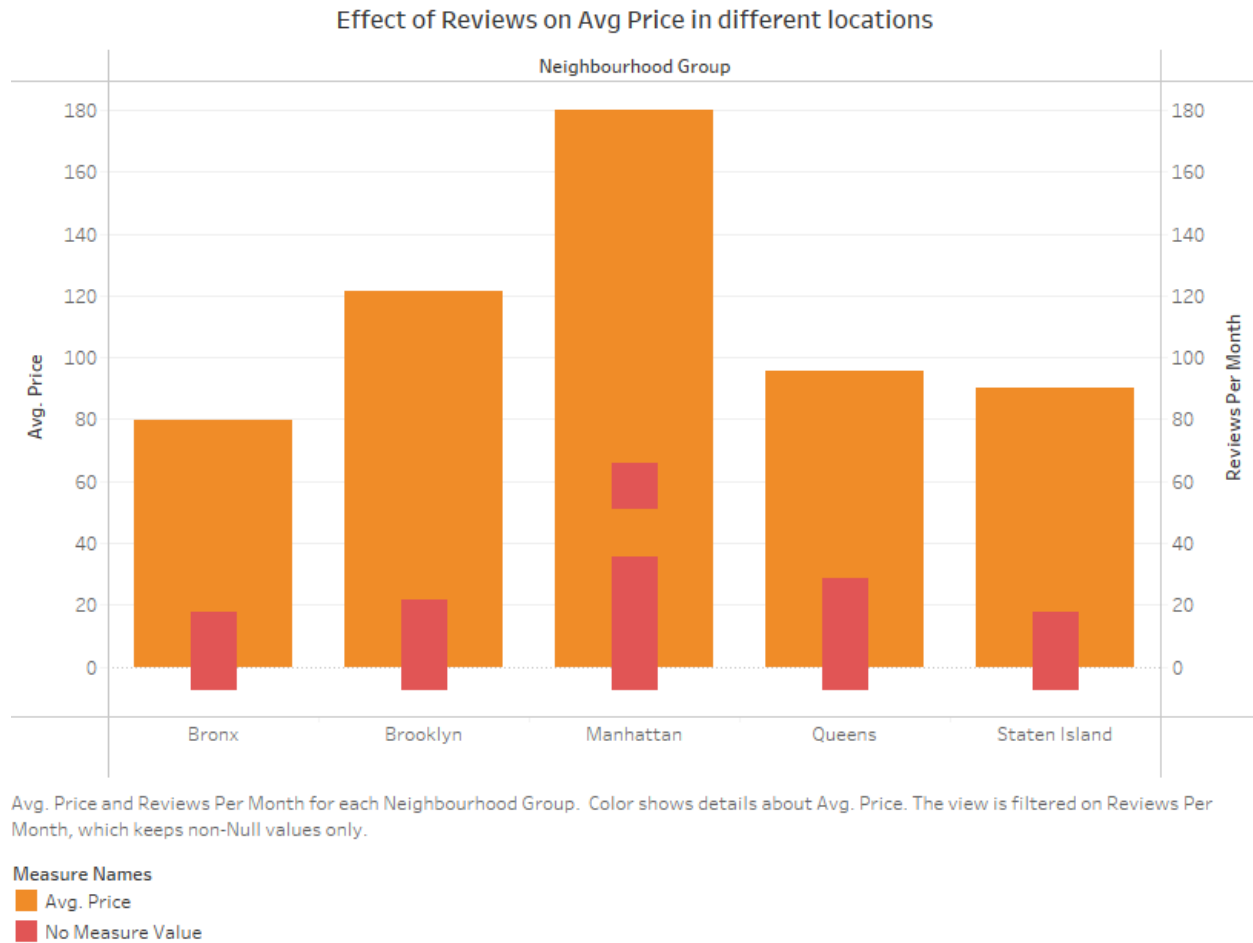
#### Neighbourhood Group

- Bronx
- Brooklyn
- Manhattan
- Queens
- Staten Island

**From the above visualisation, we studied Hosts having the most listings divided into neighbourhood groups**

To create this Packed Bubbles visualisation we used

- Neighbourhood Group: Colour Mark
- Sum of 'calculated host listing count': Size Mark
- Host Name: Label Mark



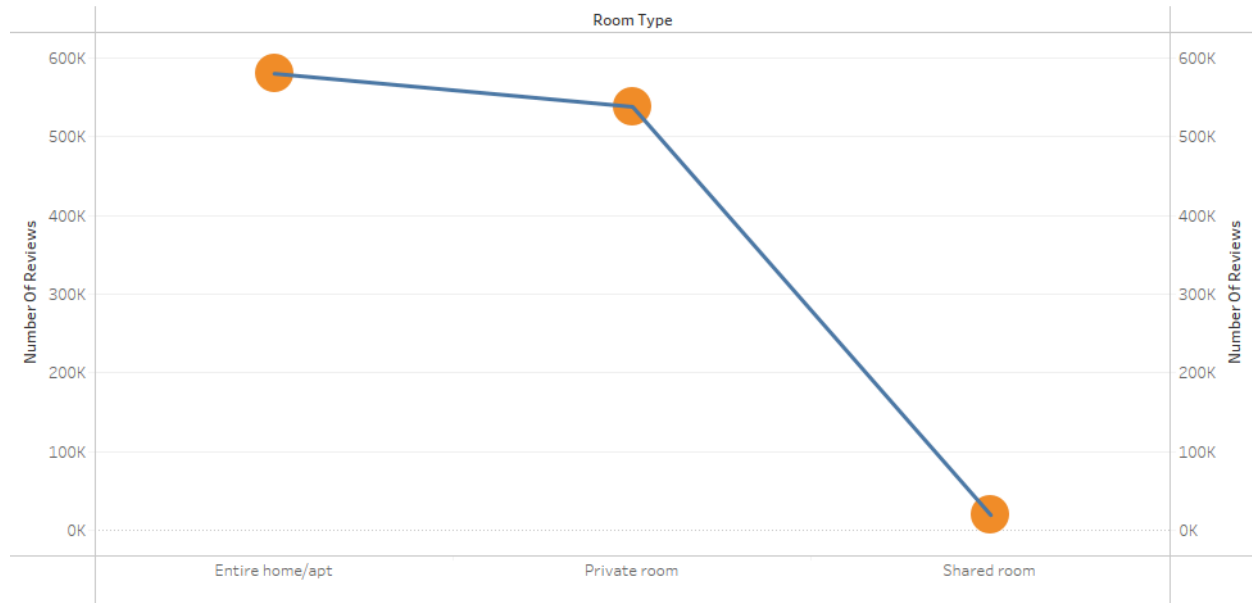
From the above visualisation, we studied the relation between the **Average Price** of listings per location and **Reviews per month** for each location

To create this Gantt view visualisation we used

- Neighbourhood Group as a field
- Average Price and Reviews per month as measures
- We created two charts with this data and merged them creating a dual axis chart
- We synchronized the axes of both these charts to visualize the relation

# Presentation 2

Sheet 16



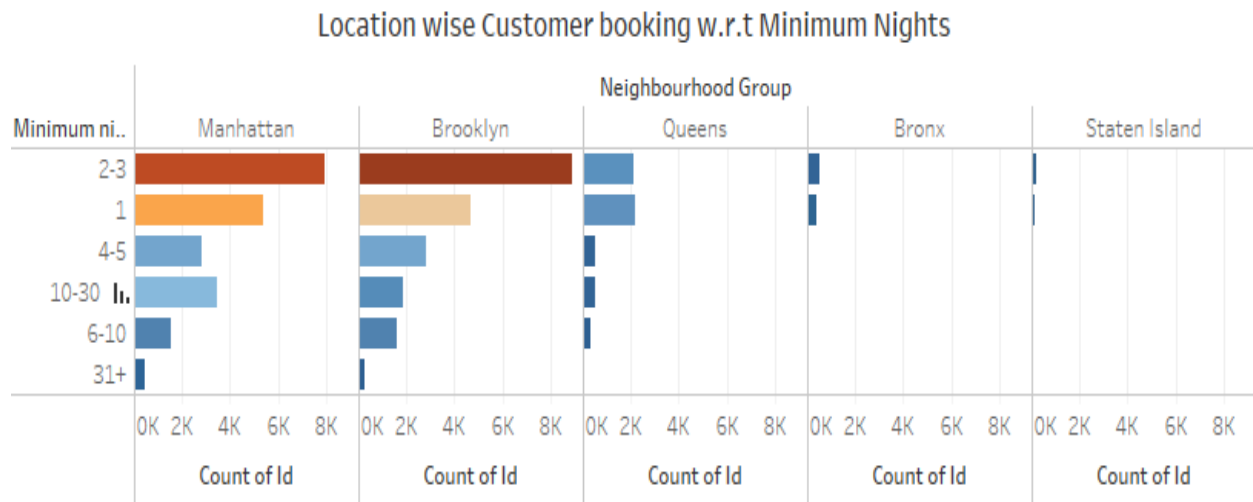
The trends of sum of Number Of Reviews and sum of Number Of Reviews for Room Type.

**From the above visualisation, we studied customer reviews trend per Room/listing space type**

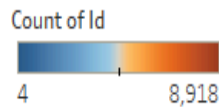
To create this trend line chart visualisation we used

- Room Type as a field (Column)
- Sum of number of reviews as a measure
- We created two charts with this data and merged them creating a dual axis chart
- We synchronized the axes of both these charts to visualize the trend





Count of Id for each Minimum nights Grouped broken down by Neighbourhood Group. Color shows count of Id.

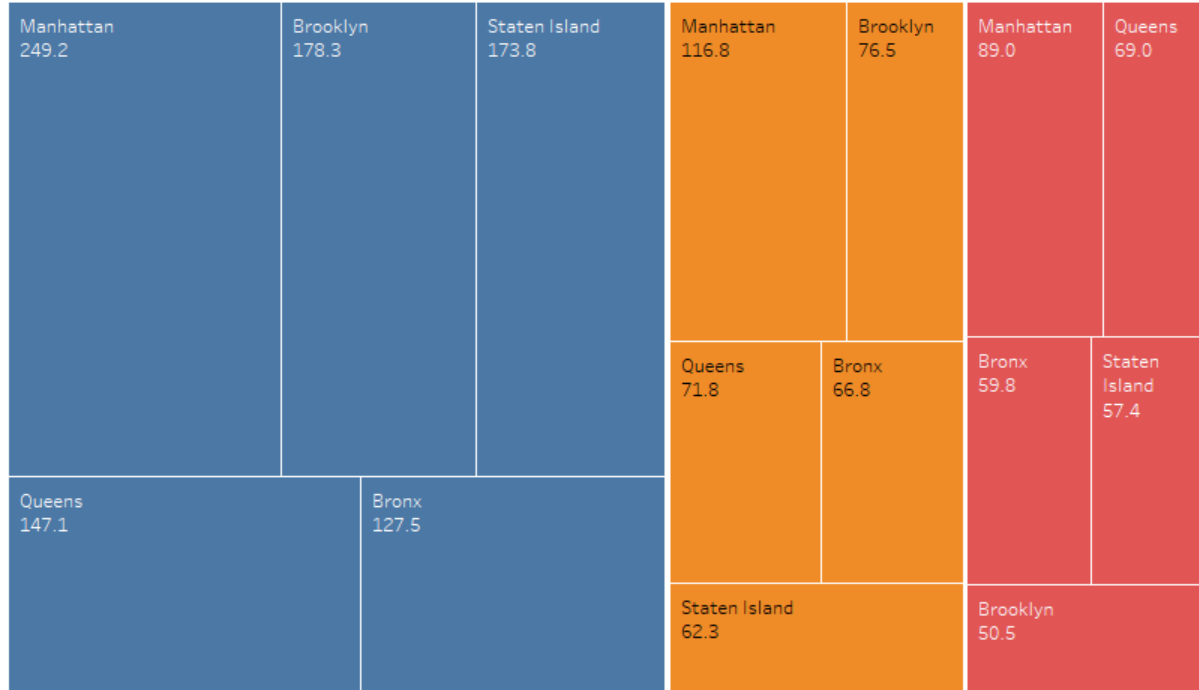


**From the above visualisation, we studied customer bookings per minimum nights group across the 5 locations**

To create this Distributed bar chart visualisation we used

- Neighbourhood Group as a field
- Count of ID as field
- Minimum nights grouped as measure
- We also added count of ID to colour mark to show the variation

Distribution of Avg Price per Location



Neighbourhood Group and average of Price. Color shows details about Room Type. Size shows average of Price. The marks are labeled by Neighbourhood Group and average of Price.

Room Type

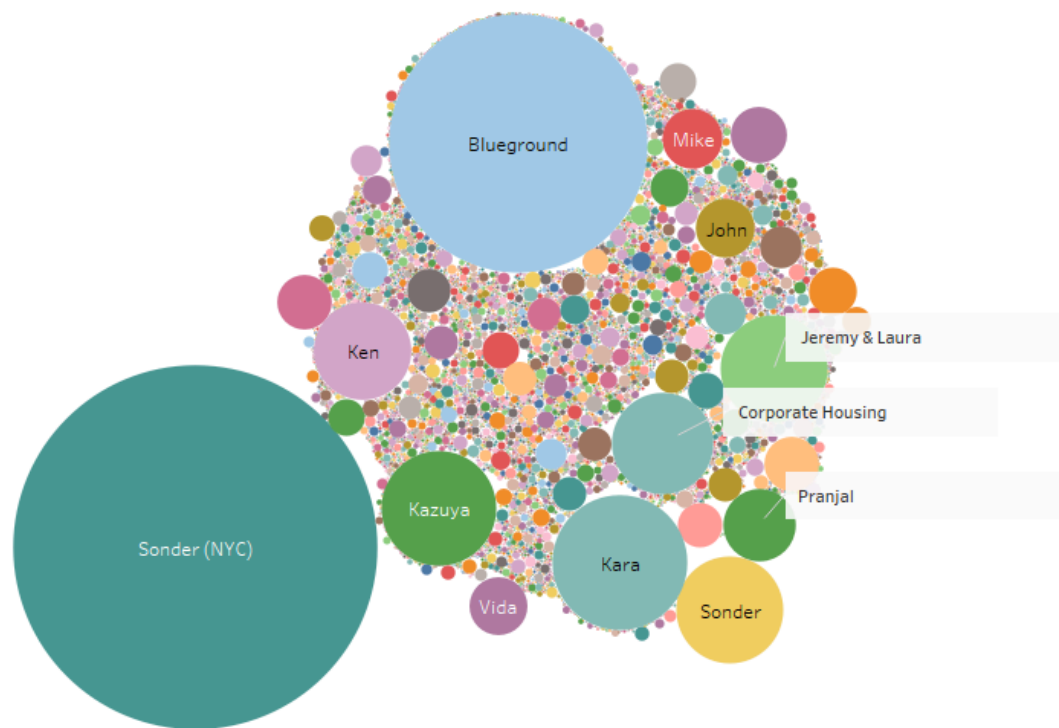
- Entire home/apt
- Private room
- Shared room

**From the above visualisation, we studied Average price distribution per neighbourhood with respect to Room type**

To create this Tree map visualisation we used Average price, Neighbourhood group & Room Type

- The size of each section corresponds to average price
- Room type is indicated by Colour
- The labels are Avg price of each section and Neighborhood group

### Tops Hosts with the most Listings



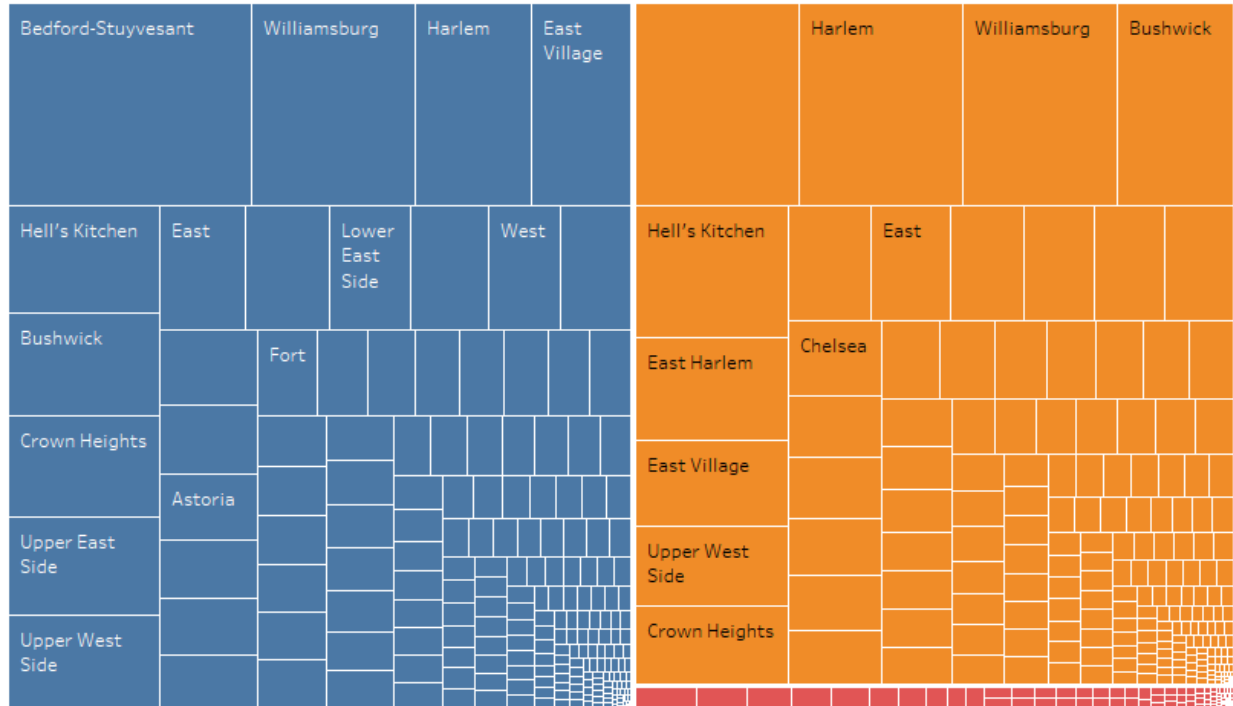
Host Name. Color shows details about Host Name. Size shows sum of Calculated Host Listings Count. The marks are labeled by Host Name.

**From the above visualisation, we studied Top Hosts with the most number of listings**

To create this Packed Bubbles visualisation we used

- Host Name as field and Sum of calculated listing count as a measure
- Sum of calculated listings count: Size Mark
- Host Name: Colour Mark
- And the labels also show Host Name

### Neighbourhood with Highest Reviewed Listings



#### Room Type

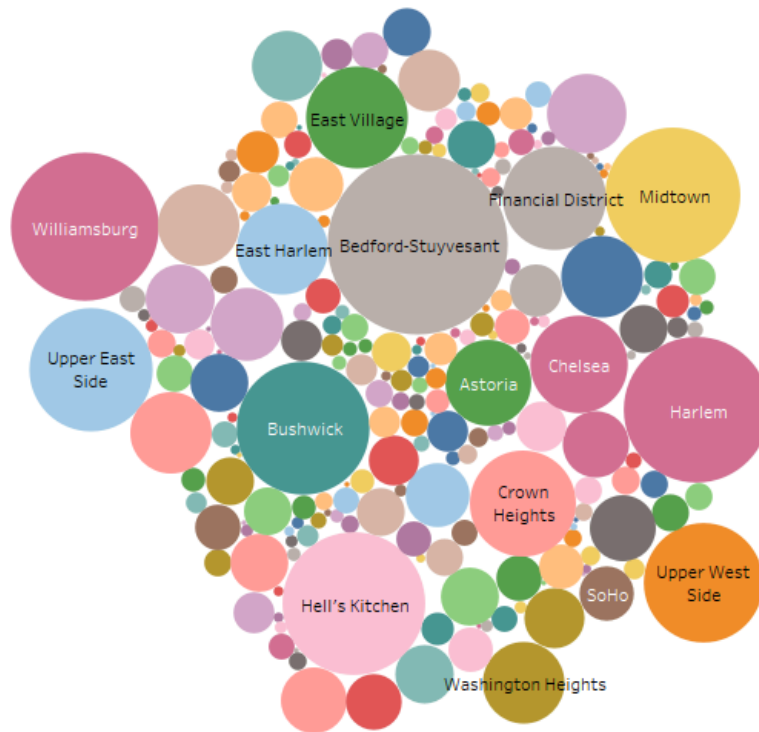
- Entire home/apt
- Private room
- Shared room

From the above visualisation, we studied Customer preference of neighbourhood with respect to Room type

To create this Tree map visualisation we used Sum of Number of reviews, Neighbourhood & Room Type

- The size of each section corresponds to Sum of Number of reviews
- Room type is indicated by Colour
- The labels are names of the Neighborhood

### Neighbourhoods with most listing availability



Neighbourhood. Color shows details about Neighbourhood. Size shows sum of Availability 365. The marks are labeled by Neighbourhood.

**From the above visualisation, we studied availability of various listings throughout the year in different neighbourhood**

To create this Packed Bubbles visualisation we used

- Host Name as field and Sum of Availability 365 as a measure
- Sum of Availability 365 is represented by the size of each bubble
- Neighbourhood is represented by colours
- And the labels also show name of the Neighbourhoods

# Recommendations

- Promotion of shared rooms with targeted discounts or eradication of the type entirely to save cost
- Discounts for bookings of more than 20 nights can increase business as customers can rent out apartments on monthly basis
- Price range can be reduced for the areas in Manhattan as it generates the most business and is preferred by most customers
- Approach hosts with maximum listings to acquire listings in other areas like Queens as they are more experienced in the market
- Focus on increasing number of listings in for private rooms
- Focus on improving customer satisfaction in the listings in areas with low reviews
- Target the areas with high availability to attract more customer bookings
- Acquire more properties in areas with low availability to increase customer base

**\*\*END\*\***