

**upGrad**



**upGrad & IITB**

**Data Science Program – April 2022**

# **METHODOLOGY DOCUMENT**



## **Storytelling Case Study: Airbnb, NYC**

**Submitted By:**

**Roshani Srivastava | Avni Srivastava | Muna Sahu**

# METHODOLOGY DOCUMENT

## Index

1. Research Problem.....	2
2. Objectives.....	2
3. Data Assumptions .....	2
4. Data Methodology .....	2
a. Data Understanding and Preparation:.....	3
b. Handling Missing Values and Outliers: .....	3
c. Data Analysis:.....	3
i. Exploring and Visualizing Data .....	3
ii. Bi-Multivariate Analysis done in Tableau .....	5
1. Customer Preferences of the Three Property Types.....	5
2. Prices vs Room Type in Neighbourhood Group .....	6
3. Customer Preferences of Minimum Nights Per Stay .....	6
4. Average Price Range of the Neighbourhood Groups.....	7
5. Count of Top 10 Host Listing in the Neighbourhood Group .....	8
6. Neighbourhood vs Avg Price – Neighbour Group .....	9
7. Host listings Vs Neighbourhood Group.....	10

## 1. Research Problem

- For the past few months, Airbnb business has seen a significant decline in revenue due to travel restrictions because of the Covid-19 pandemic.
- The revenue took the largest hit in NYC in the Q2 of 2020.
- Now that the restrictions have started lifting and people have started traveling, Airbnb wants to make sure it is fully prepared for the change.

## 2. Objectives

- Improve our strategies to revive the impact of Covid-19 on the economic and market conditions of Airbnb, NYC.
- Understand the customer preference and user experience trends for Airbnb, NYC.
- Provide recommendations for new acquisitions and improve customer experience.

## 3. Data Assumptions

- Assumed that the data prior to the Covid-19 period was achieving the desired goals.
- Airbnb wants to continue its business in NYC and has no plans of expanding to other territories.
- The strategies decided were considered keeping in mind that there will be no further travel restrictions.

## 4. Data Methodology

### -Tools used -

Python and Python libraries, For visualization Tableau is used.

## a. Data Understanding and Preparation:

- The following relevant libraries were imported.  
import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
import matplotlib.image as mpimg  
%matplotlib inline  
import seaborn as sns
- The dataset was loaded, datatypes of variables were checked and along with that the dimensions and size of the data frame was checked

## b. Handling Missing Values and Outliers:

- The missing values and outliers were checked in the data frame.
- The following columns had missing values - last\_review, reviews\_per\_month, host\_name, and name.
- These columns had NaN values - last\_review and reviews\_per\_month indicating some listed properties didn't receive reviews.
- The following columns had outliers - price, minimum\_nights, number\_of\_reviews, reviews\_per\_month, and calculated\_host\_listings\_count.
- The missing values and outliers were not treated for.

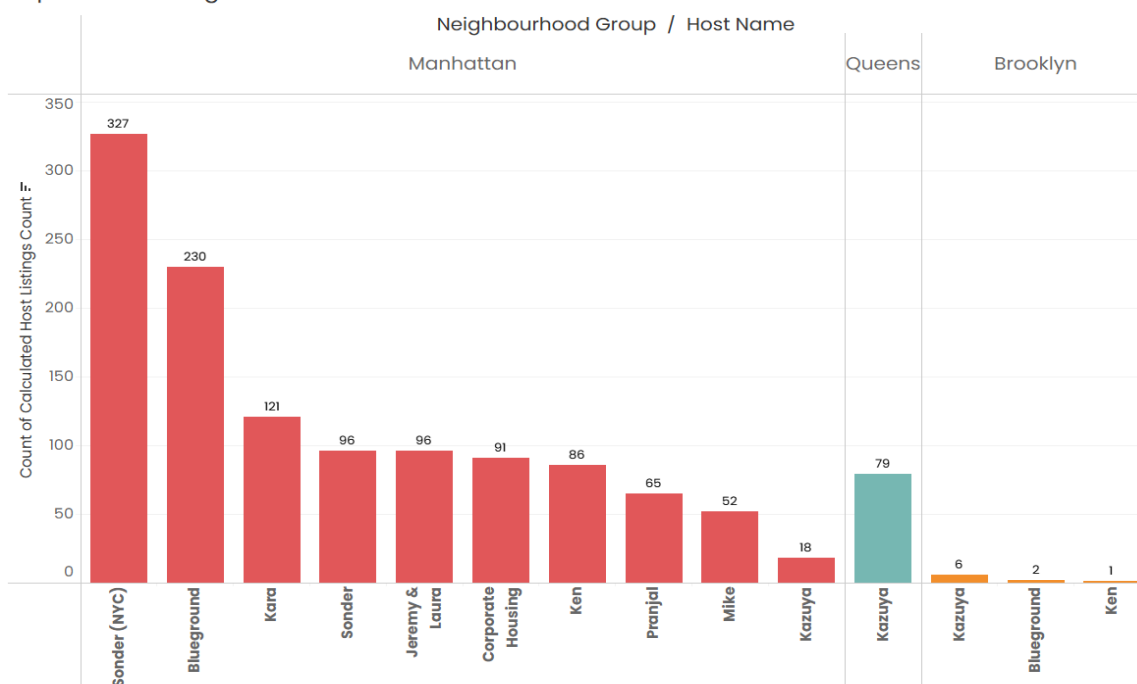
## c. Data Analysis:

### i. Exploring and Visualizing Data

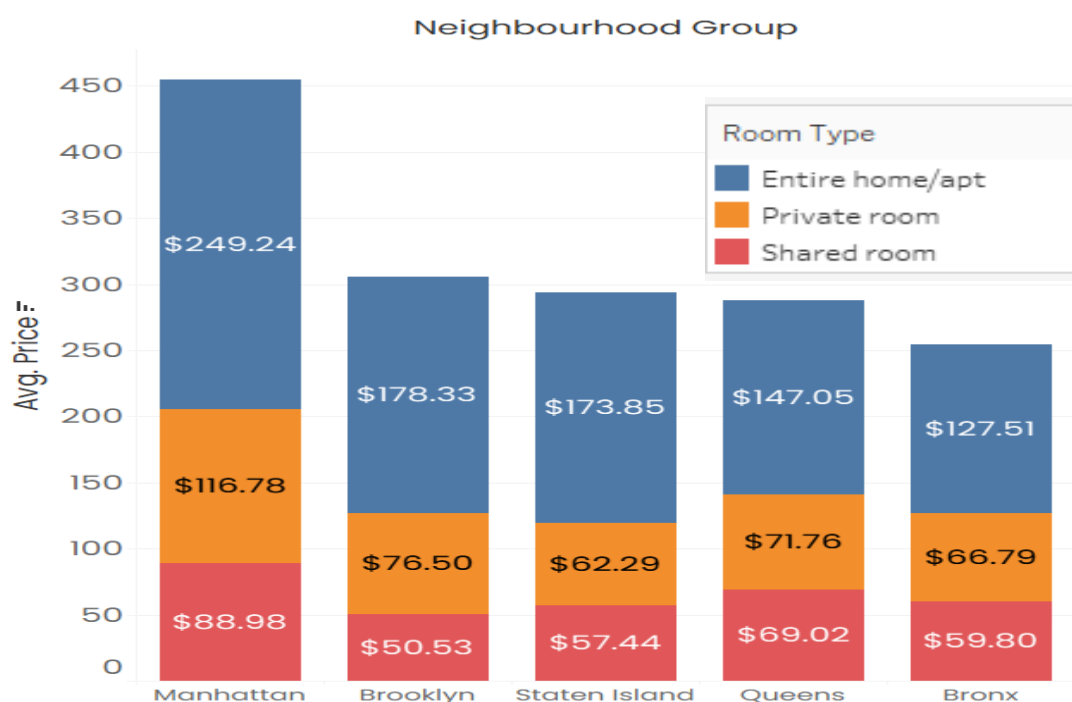
- Exploring the data by analysing its statistics and visualizing the values of features and correlations between different features. Explaining the process and the results
- Now that we are ready for an exploration of our data, we can make a rule that we are going to be working from left to right. The reason some may prefer to do this is due to its set approach - some datasets have a big number of attributes, plus this way we will remember to explore each column individually to make sure we learn as much as we can about our dataset.

- We have created a barplot to see that there is a good distribution between top 10 hosts with the most listings. First host has more than 300+ listings.

Top 10 Host Listing



Average Price Range

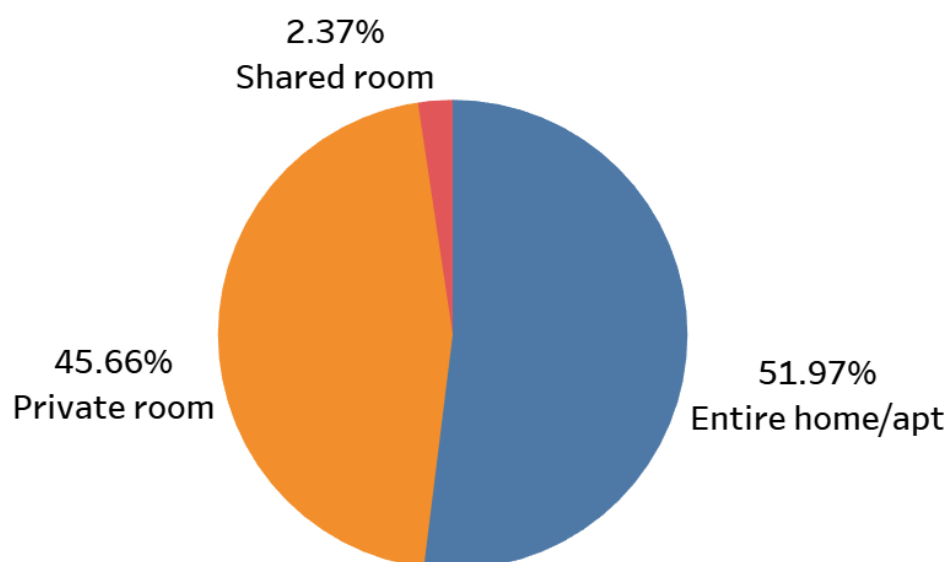




With a statistical table and a stacked bar chart we can definitely observe a couple of things about distribution of prices for df in NYC boroughs. First, we can state that Manhattan has the highest range of prices for the listings with \$150 price as average observation, followed by Brooklyn with \$90 per night. Queens and Staten Island appear to have very similar distributions, Bronx is the cheapest of them all. This distribution and density of prices were completely expected; for example, as it is no secret that Manhattan is one of the most expensive places in the world to live in, where Bronx on other hand appears to have lower standards of living.

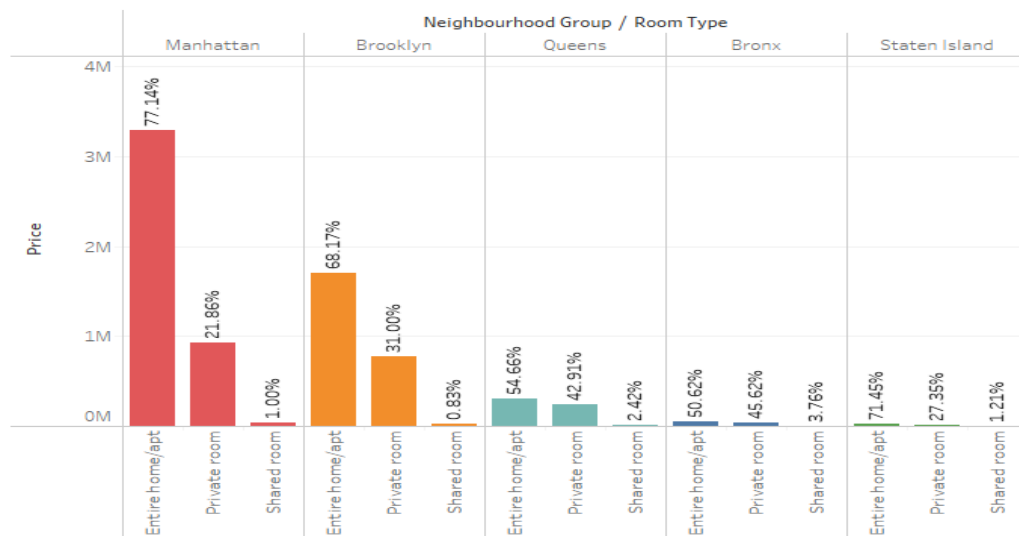
## ii. Bi-Multivariate Analysis done in Tableau

### 1. Customer Preferences of the Three Property Types



- The properties of Entire home/apt and Private room are preferred over shared rooms by Airbnb hosts offering rentals in NYC.
- The room types 'Entire home/apt' and 'Private room' accounts for around 97.6% of the listed properties in NYC.
- Shared rooms account for only 2.4% of the total listed properties.

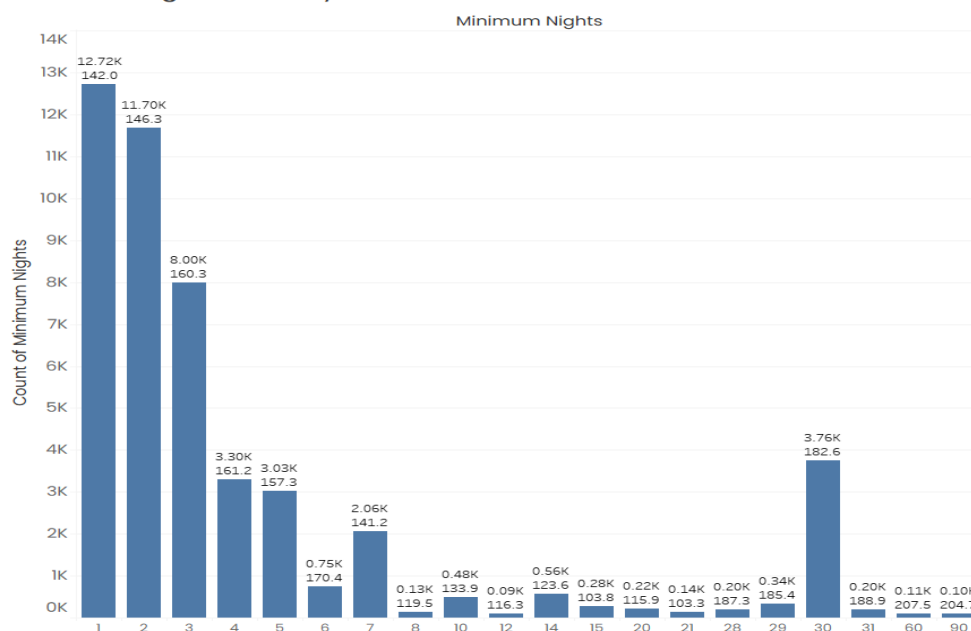
## 2. Prices vs Room Type in Neighbourhood Group



- Manhattan has the most expensive properties followed by Brooklyn while the Bronx and Staten Island are the cheapest.
- Entire homes/apt in both Manhattan, as well as Brooklyn, are the most expensive room type followed by Private rooms.
- In Queens and Bronx, the prices of room types of entire homes/apt and private rooms have a close difference of 5-10% as compared to the other neighbourhood groups.

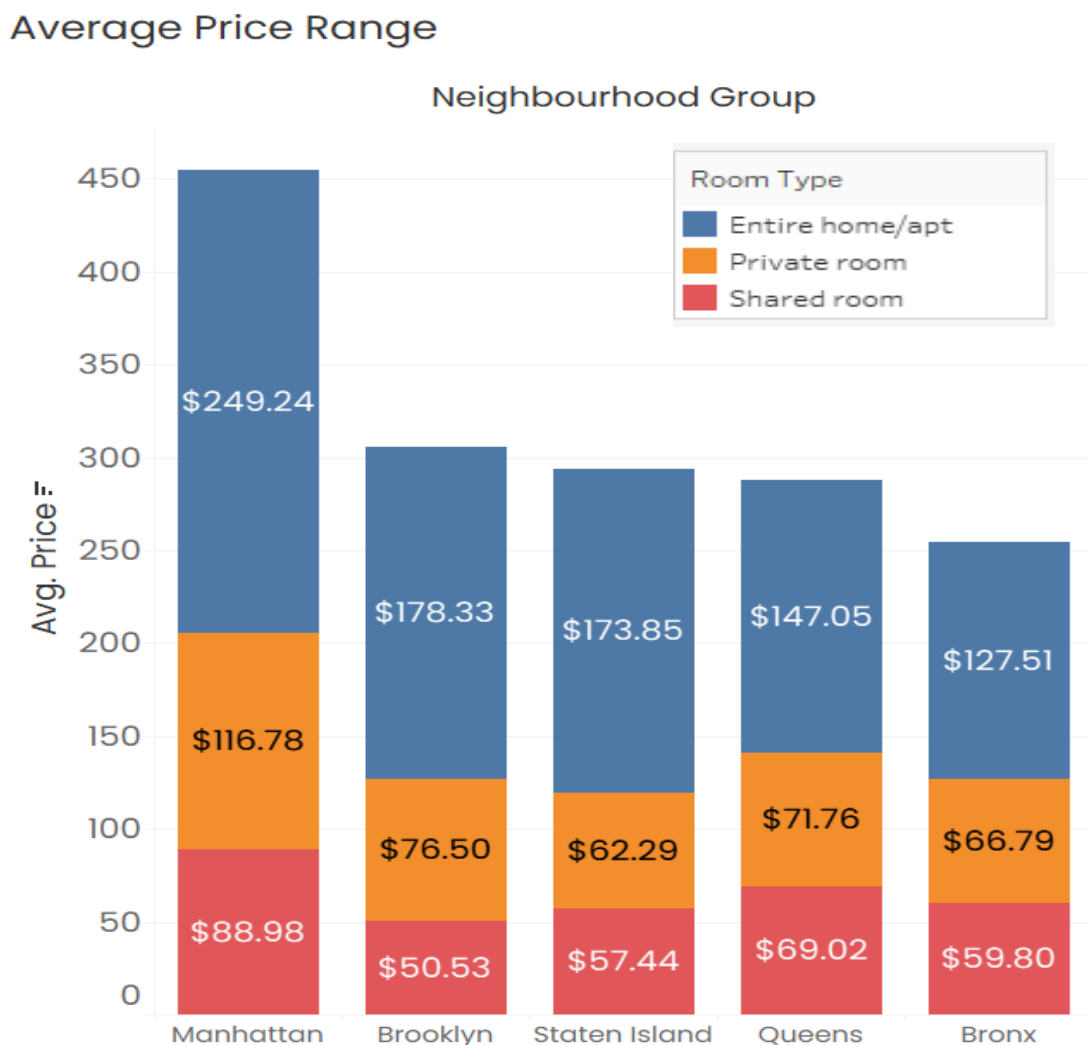
## 3. Customer Preferences of Minimum Nights Per Stay

Minimum Nights Per Stay



- 12.7k customers prefer to stay for a day on their visit closely followed by 11.6k customers who stay for 2 nights.
- Customers also prefer to stay for 3-7 nights.
- Customer preferences for night stay rise on 14 nights, 30 nights.

#### 4. Average Price Range of the Neighbourhood Groups



- Manhattan is the only neighbourhood group that has the most expensive average price of approx. \$200 on the properties listed.
- Brooklyn and Staten Island have an average price of \$124 and \$114 resp. which is less expensive compared to Manhattan.
- Queens and Bronx have an average price of \$100 and \$88 resp. which is the most affordable neighbourhood in NYC.



## 5. Count of Top 10 Host Listing in the Neighbourhood Group

### Top 10 Host Listing



- Sonder(NYC) has the highest number of rooms listed in Manhattan – 327 followed by Blueground at 230.
- Manhattan is the most listed neighbourhood group followed by Brooklyn and Queens.
- Bronx has only 1 listing whereas Staten Island does not fall in the Top 20 category.

## 6. Neighbourhood vs Avg Price – Neighbour Group

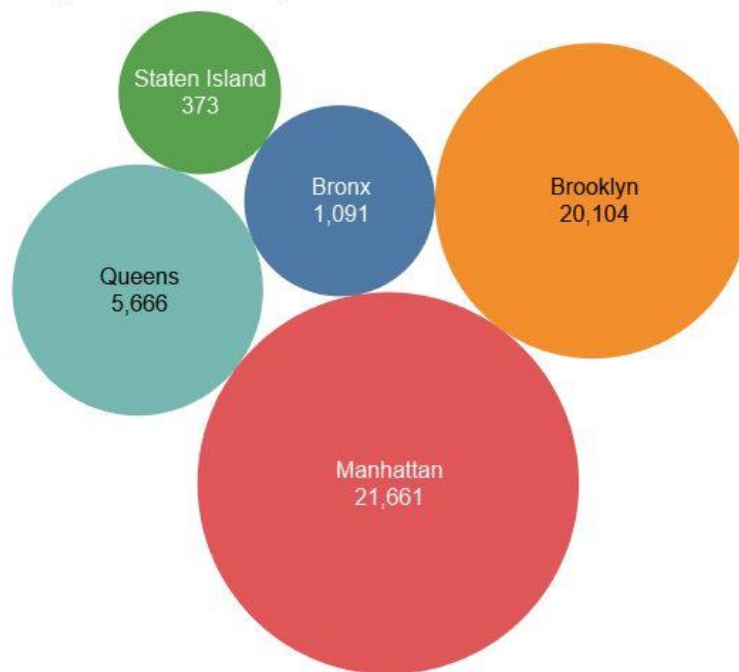
- Top two neighbourhood with highest average price of properties – Fort Wadsworth & Woodrow are in Staten Island.
- Manhattan has 12 neighbourhood in the top 20 leading with Tribeca & Battery Park City.
- Brooklyn, Queens & Bronx have only 1 neighbourhood each – Sea Gate, Riverdale & Neponsit respectively in the Top 20 category.

Neighbourhood vs Avg Price – Neighbour Group



## 7. Host listings Vs Neighbourhood Group

Host listings Vs Neighbourhood Group



Neighbourhood Group

- Bronx
- Brooklyn
- Manhattan
- Queens
- Staten Island

- Created a bubble chart between Host listings neighbourhood group just to see which neighbourhood group has the most host listings.
- Taken count and distinct count of host listings and count.
- Manhattan has the highest count of host listings around 21661 and Staten Island has the least count of host listings.