

# THE CAPSTONE PROJECT - 1



## EDA ON AIRBNB BOOKING ANALYSIS

PROJECT DONE BY  
**PAVANKUMAR KASHYAP**

# INDEX



- **INTRODUCTION**
- **PROBLEM STATEMENT**
- **DATA COLLECTION AND DESCRIPTION.**
- **DATA DEFINING.**
- **DATA PREPARATION.**
- **DATA MANIPULATION.**
- **EXPLORATORY DATA ANALYSIS.**

# INTRODUCTION



- AirBnB is one of the organisations which has made it easier for the passengers to get accomodation at places across the globe keeping in mind the cost, location and affordability.
- The organisation has made it easier for the customers and service providers to experience greater productivity and has enculcated the idea of short term renting of personal and commercial properties.

# INTRODUCTION

- As the business and services of the organisation expands, it is essential to analyse the data collected in an orderly manner to better the understanding of the organisation about the needs of the consumers. The platform has promoted efficiency and good faith among the participants of the services and made it possible for people to book their stays from their hands.
- The goal of the project is to analyse the data collected from 2008 to observe numerous factors affecting the business and consumer behaviours across the service years. The data consists of various fields of which some important ones are id, host id, neighbourhood, neighbourhood group, room type, number of reviews and reviews per month.

# PROBLEM STATEMENT

- The Airbnb (“Air Bed & Breakfast”) is a service that lets property owners rent out their spaces to travellers looking for a place to stay.
- Airbnb are usually cheaper than hotels for longer stays & larger groups.
- For this project, we have analysed Airbnb's New York City (NYC) data of 2019. NYC is not only the most famous city in the world but also top global distribution of visitors drawn to its museums, restaurants, commerce.
- In this “Airbnb Booking Analysis” project we’ll analyze the whole data and this data set contains information like types of listing, price of listing, number of checks in a month, price for a night, etc.

# PROBLEM STATEMENT

- The main objective from this “Airbnb booking analysis” to analyze and research about their factors that can be affected in security, business and bookings.
- Other like number of booking per year, business growth, types of rooms that are usually booked, their prices, number of reviews per year, availability of rooms.

# SOME QUESTIONS WE HAVE TO ANALYZE

1. Location wise Room distribution
2. Maximum nights spend in respect to room types
3. Top 5 hosts based on their turnover in Airbnb listing
4. Average price of each room type with respect to neighborhood groups
5. Top 10 neighbourhood having highest number of apartments
6. Top 5 hosts with respect to number of reviews
7. Top 5 expensive locality in airbnb listings

# DATA COLLECTION AND DESCRIPTION

- According to our given data we have **48895** Rows and **16** columns with non null values in most of the columns but,as we can see there are some null values or missing data in our columns such as **last review ,reviews per month**.

## DATA DESCRIPTION :

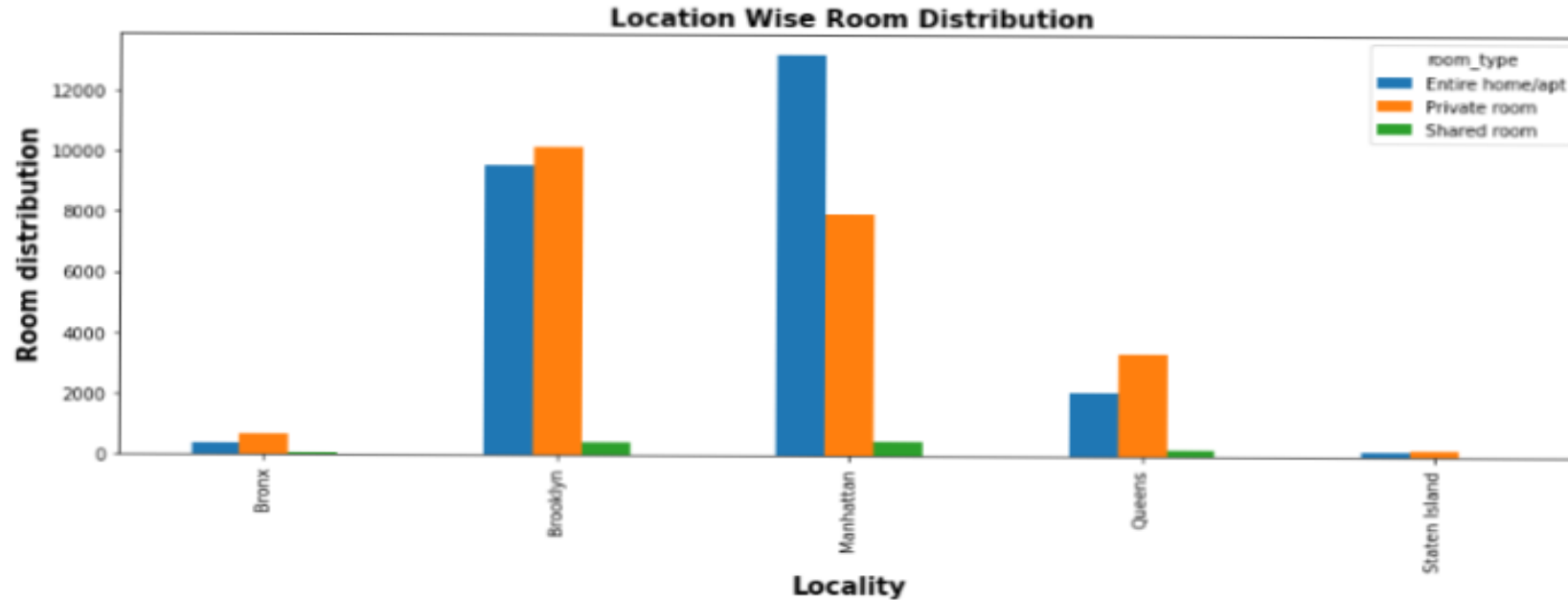
- **id** : a unique id identifying an airbnb lisitng
- **name** : name representating the accomodation
- **host\_id** : a unique id identifying an airbnb host
- **host\_name** : name under whom host is registered



# DATA COLLECTION AND DESCRIPTION

- **neighbourhood\_group** : a group of area
- **neighbourhood** : area falls under neighbourhood\_group
- **latitude** : coordinate of listing
- **longitude** : coordinate of listing
- **room\_type** : type to categorize listing rooms
- **price** : price of listing
- **minimum\_nights** : the minimum nights required to stay in a single visit
- **number\_of\_reviews** : total count of reviews given by visitors
- **last\_review** : date of last review given
- **reviews\_per\_month** : rate of reviews given per month
- **calculated\_host\_listings\_count** : total no of listing registered under the host
- **availability\_365** : the number of days for which a host is available in a year.

# 1. Showing the relation between different neighbourhood group and room types



With the help of this data ,we can understand the distribution of different **room type** with respect to different **neighbourhood group**.

- According to the above given figure,we can understand **Manhattan** holds highest number of entire apartments in comaprison to other type of rooms.

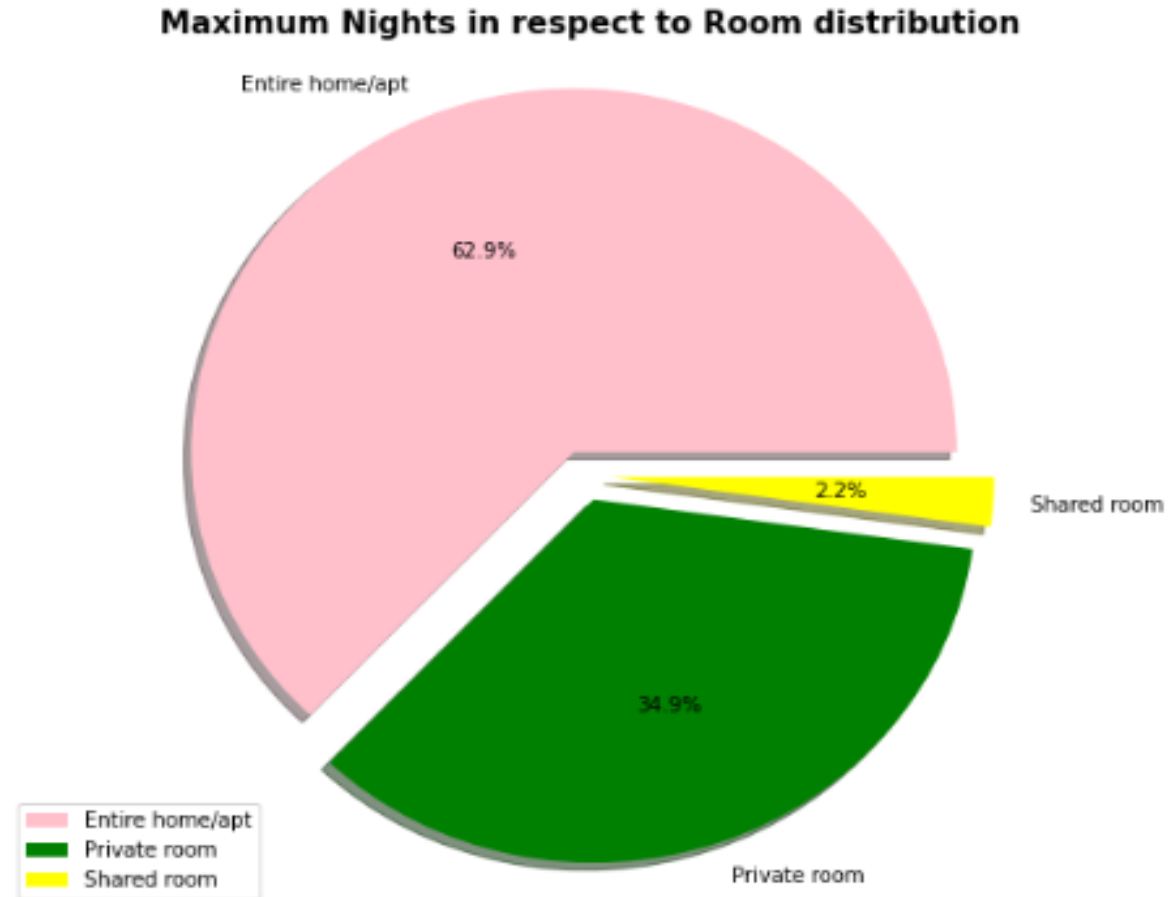
- **Brooklyn** holds the highest number of private rooms as compare to private rooms in **MANHATTAN**.
- **State island** and **Bronx** holds least number of Entire Private and Shared rooms as compared to other neighbourhood group.

## **Will the gained insights help creating a positive business impact ?**

- Yes, the gained insights will help create a positive business impact as the data helps the business to understand which location requires greater input to onboard new service providers to expand its reach and provide better feasibility options to clients.

## 2. Maximum nights spend by the guests in each type of rooms.

- Entire home/Apt
- Private room
- Shared room



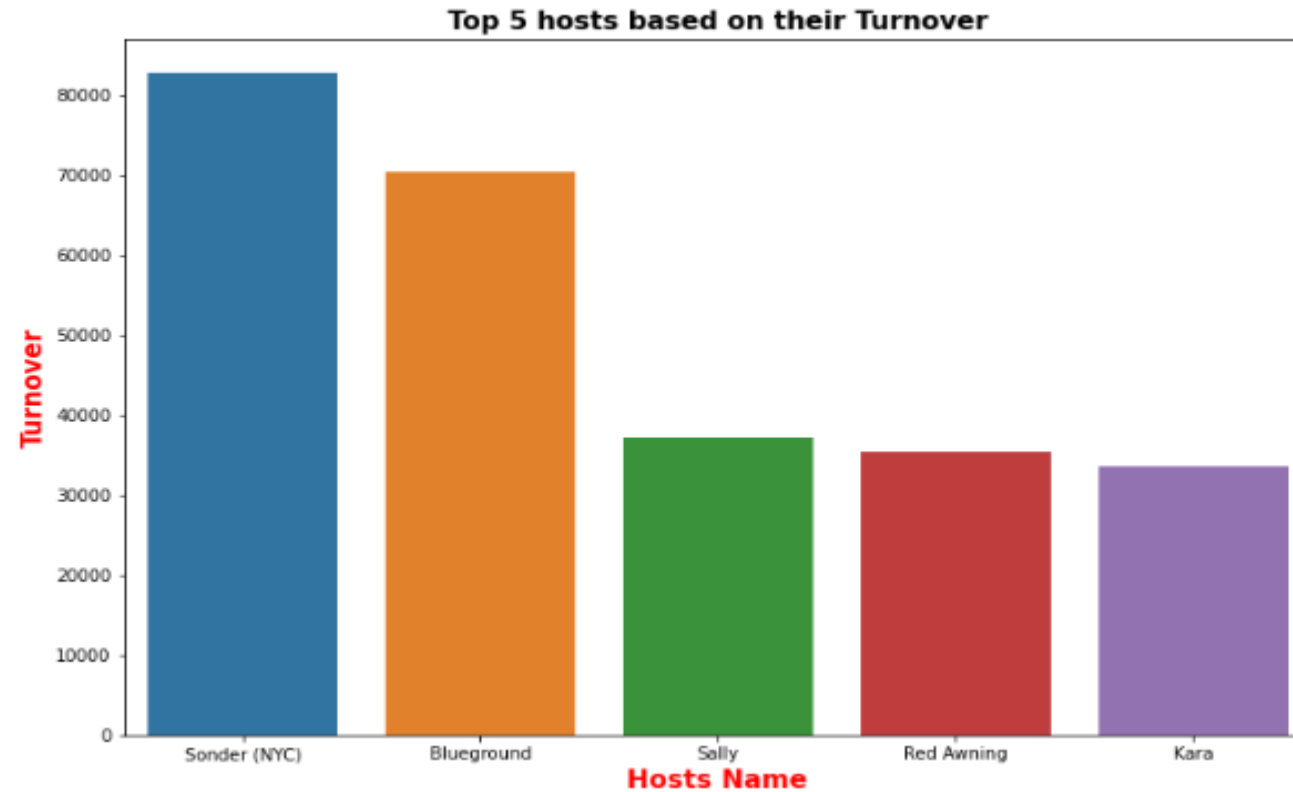
From above figure,we can see that most customers preferred to stay in **Entire Apartments** in comparison to other room types.

- 62.9% of customers preferred to stay in entire home apt.We can easily understand that the price is high for Entire room as it has high space area , and have different facilities provided by host for the convenient of guests
- 34.9% of customers preferred to stay in Private rooms.
- 22% of customers preferred to stay in Shared rooms.

### **Will the gained insights help creating a positive business impact?**

- The gained insights would definitely benefit the business as it provides efficient distribution of choices made by customers and would help the organisation to enhance the services offered in the near future.

### 3. Top 5 host based on their turnover



What is/are the insight(s) found from the chart ?

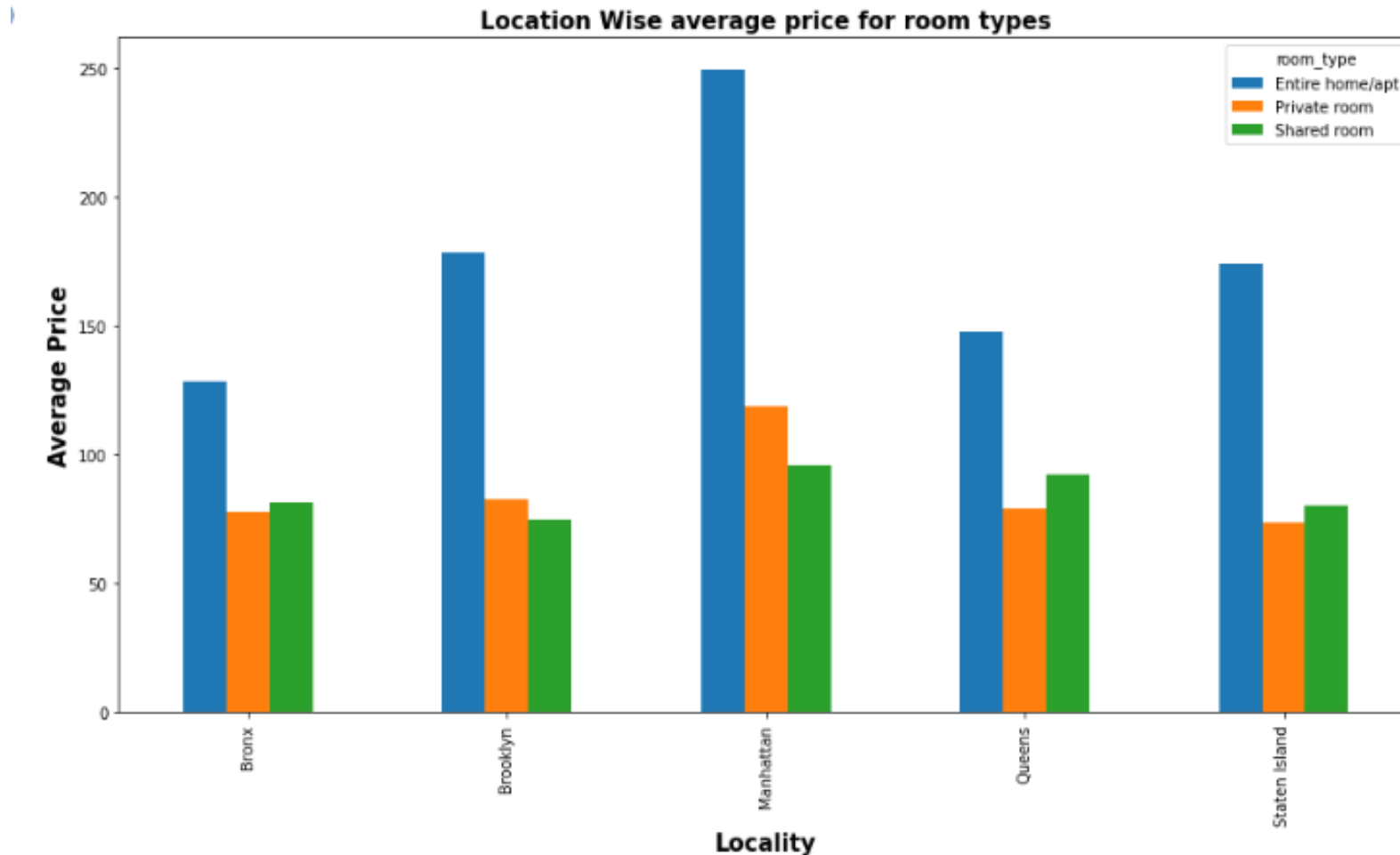
The insights gained are as follows:

- Sonder(NYC) has the highest turnover.
- There is a strong need for reforms in the business to increase turnover in certain areas so that the reach and demand can increase exponentially.

## **Will the gained insights help creating a positive business impact?**

- The gained insights helps the business to analyse and compare the performance of hosts providing services and to judge further potential of growth and efficiencies which would take the business to new heights.

## 4. Showing the relation between different neighbourhood group and average price with respect to room types.





- Manhattan has the highest average price among all the room types available across all locations.
- The average price for shared room is greater than private room in Bronx, Queens and Staten Island.

### **Will the gained insights help creating a positive business impact?**

- The average price distribution helps the business analyse the inconsistency of prices offered at various locations. The organisation might also consider providing certain discounts to attract customers avail their services.

- Manhattan has the highest average price among all the room types available across all locations.
- The average price for shared room is greater than private room in Bronx, Queens and Staten Island.

### **Will the gained insights help creating a positive business impact?**

- The average price distribution helps the business analyse the inconsistency of prices offered at various locations. The organisation might also consider providing certain discounts to attract customers avail their services.

## 5. Top 10 neighbourhood group having maximum number of apartments



The graph is about the top 10 neighbourhood having highest number of apartments.

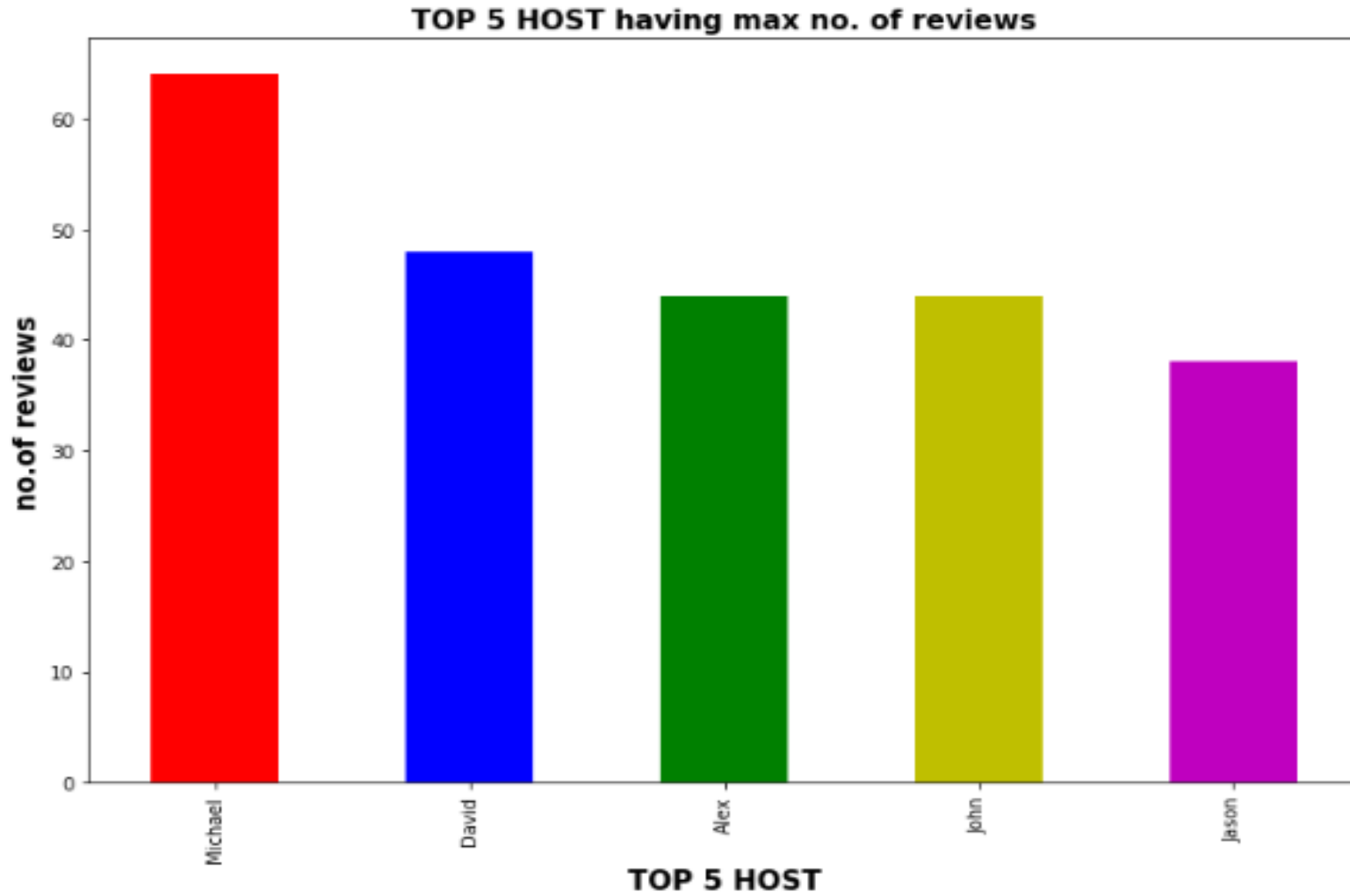
- **Williamsburg** in neighbourhood holds highest number of apartments in comparison to other neighbourhood.

- **Midtown and Crown heights** holds least number of apartments in comparison to other neighbourhoods.

## **Will the gained insights help creating a positive business impact?**

- The gained insights would help business perceive goals relating to maximum growth and profit accumulation with respect to maximum number of apartments available in the neighbourhoods. The data would help business provide various extra amenities and services to top serving neighbourhoods.

## 6. Top 5 host according to maximum number of reviews



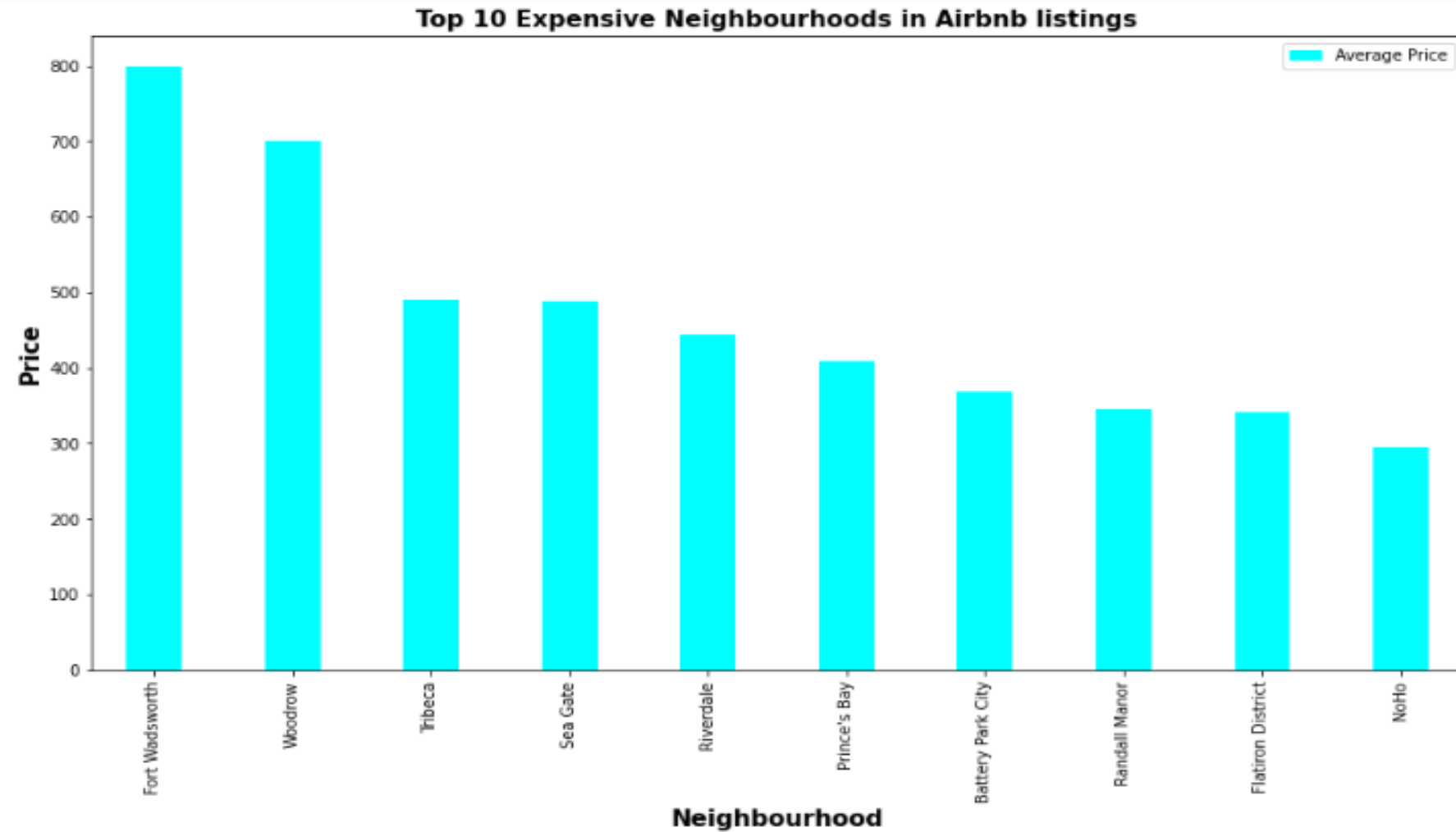
The graph provides insights about top 5 hosts having maximum number of reviews.

- The host Michel has received highest number of reviews among all summing up to 64.
- After the host securing maximum number of reviews, it is observed that there is a sudden fall in number of reviews gained thereafter.

### **Will the gained insights help creating a positive business impact?**

- The gained insights help the business understand the best hosts providing continual services and better opportunities for growth. The top hosts may provide valuable knowledge and experience to enhance the scope of business.

## 7. Top 10 expensive neighbourhoods in airbnb listings



In the above figure, we can see the top 10 expensive neighbourhood with the respect to price.

- **Fort wadsworth** holds the highest price i.e 800USD in airbnb listings which is quite expensive.

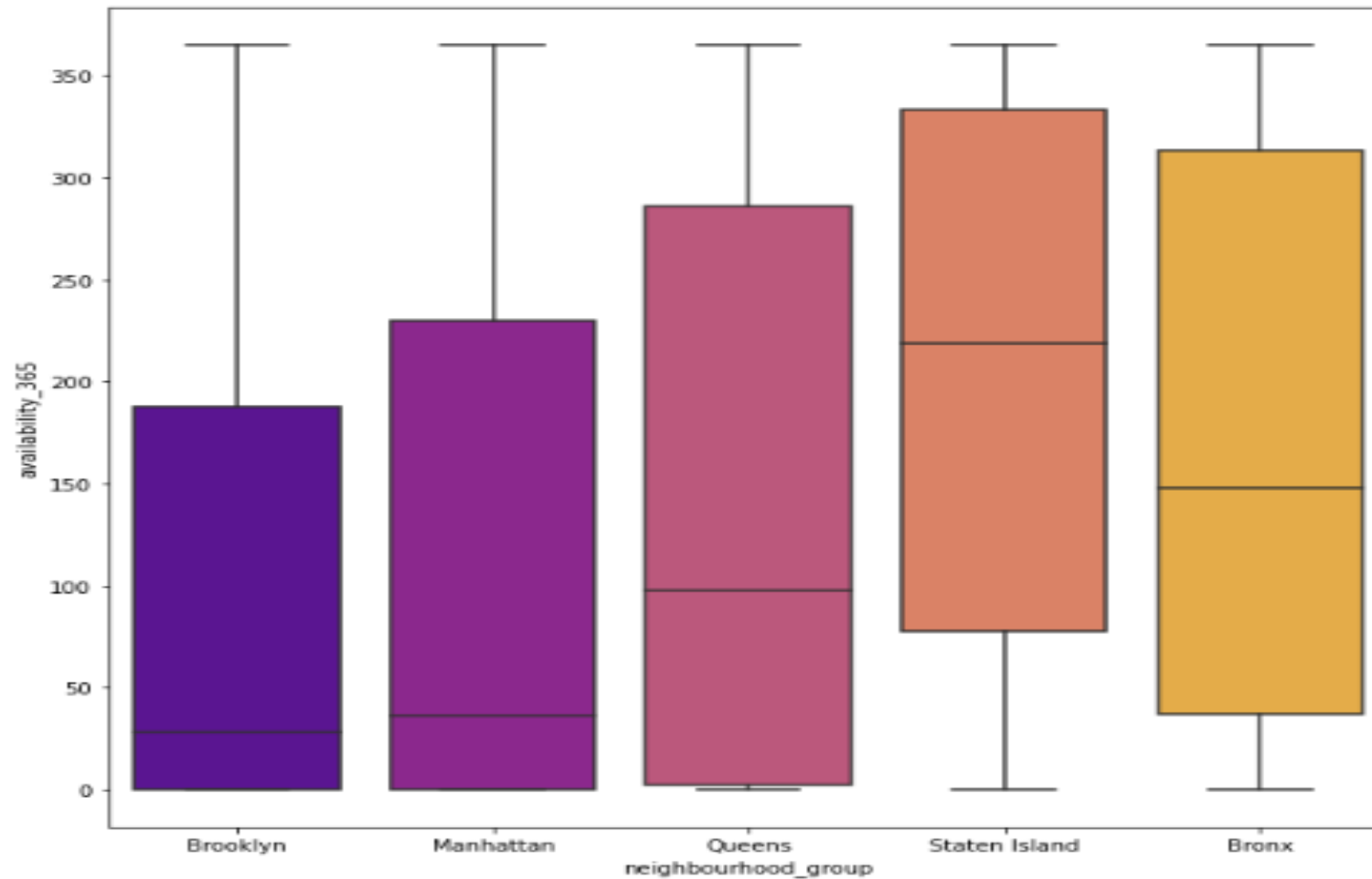
- **NoHo** in neighborhood have the least price among top 10 i.e 296 USD in airbnb listings which is affordable for the tourists.

**Will the gained insights help creating a positive business impact?**

- The gained insights would help the business control price brackets over the most expensive neighbourhoods to enable more customers to avail the services at better rates.

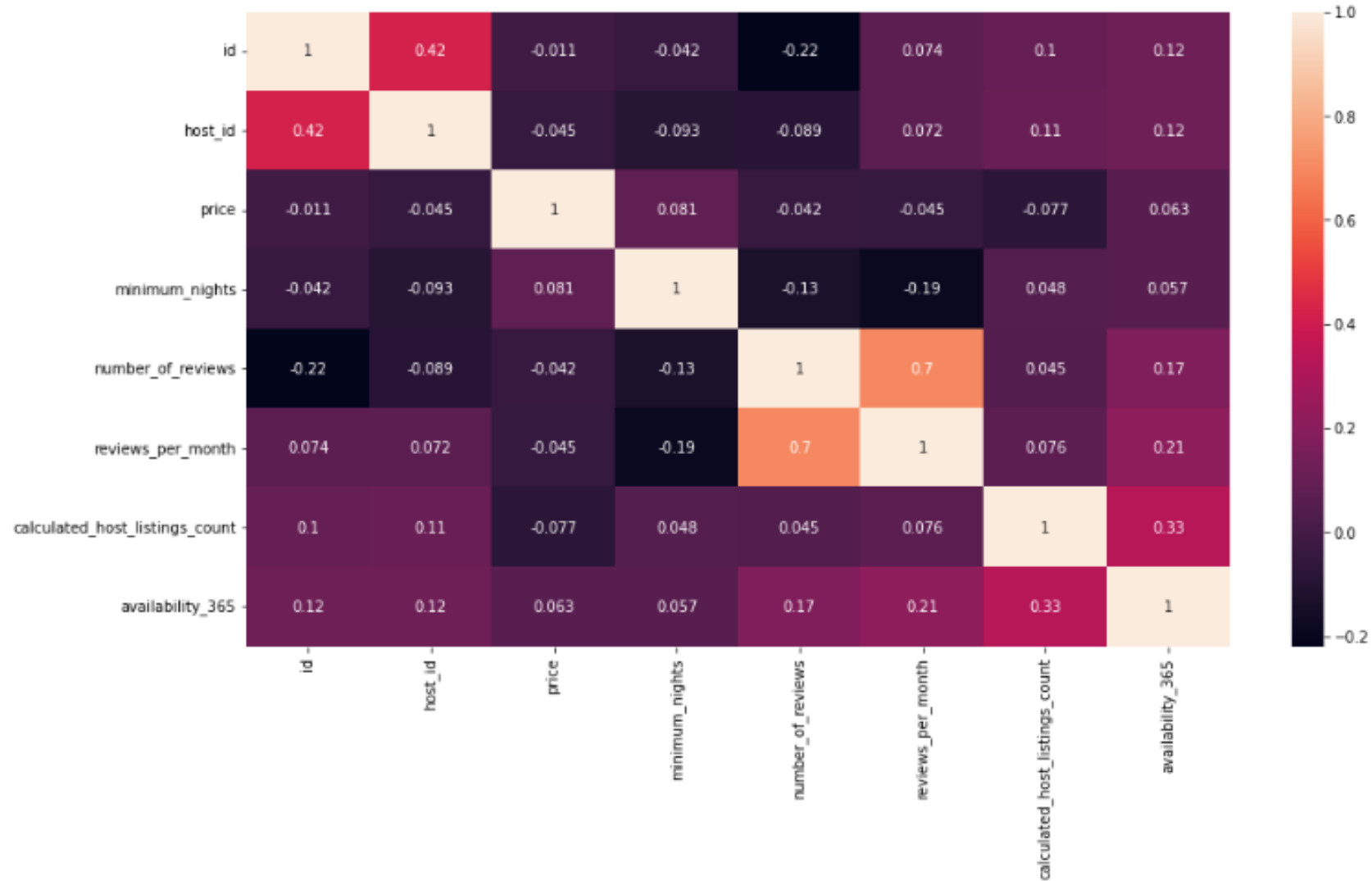


## 8. Relation between neighbourhood group and availability of room in our Airbnb Dataset.



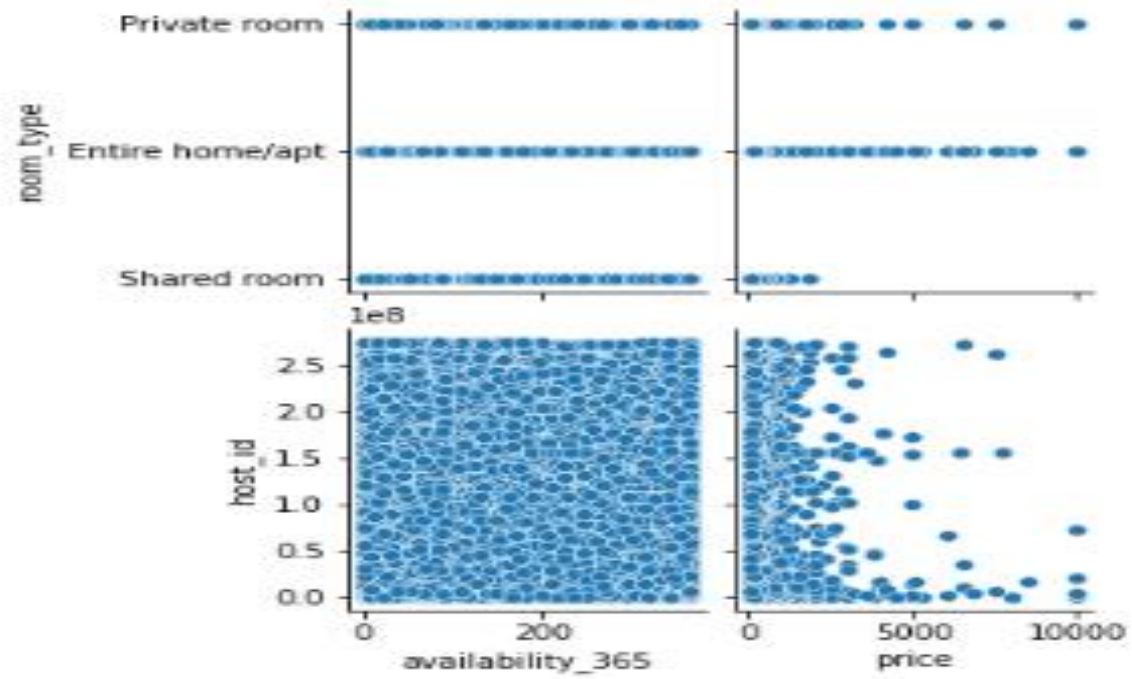
- The above box plot shows minimum and maximum values of availability across the year for all neighbourhood groups through vertical extremities.
- The horizontal line inside the box shows the median values of availability for a neighbourhood group.
- The vertical maximum of a box exhibits that 25% of the values are greater than the current point and the vertical minimum of a box exhibits that 25% Of data is less than the current value.
- The visualization helps the business analyse the availability of services across the year while comparing within the neighbourhood group and also comparing with other neighbourhood groups. From the above graph we can state that Staten Island has the highest availability.

# Correlation Heatmap visualization code



- The heatmap above shows variance across multiple variables, revealing any patterns, displaying whether any variables are similar to each other, and for detecting if any correlations exist.
- The rightmost cell is the value scale which determines cell colouring.
- It can be observed that the heatmap depicts correlation between numeric column variables present in the table and plots them in a 2-dimensional graph.

## Pair Plot visualization code



# CONCLUSION

- After exploring and analysing through data and visualization, we obtained some interesting insights into the Airbnb domain....
- The neighbourhood group "Manhattan" has the most expensive bookings compared to the other neighbourhood group. We can say this based on the neighbourhood vs listings and neighbourhood vs price, the chart and graph clearly shows us that Manhattan has the highest number of Airbnb bookings and expensive because it receives average price of \$194.8.!
- Manhattan is also considered as the best location based on the graph of neighbourhood group vs number of reviews. Why Manhattan is best, most expensive and most trafficked location? So, we did some research and found out that why Manhattan is the best location, because it is closest to the famous city hotspots like Time Square, Empire Street, Central Park, etc and have very convenient transportation services.

# CONCLUSION

- Busiest Host = Based on the different graphs like neighbourhood vs price, neighbourhood group vs Airbnb listings and neighbourhood group vs number of reviews, the host(host id = 219517861 and his name = Sonder NYC) who has 327 listings is considered as the busiest host in NYC and he belongs to the Manhattan. We get to understand that Manhattan and brooklyn provides the most houses and rooms. Hence, we can imagine that not only the host who has 327 listings but if we consider the top 10 busy hosts, they all are from the area of Manhattan and Brooklyn.
- According to our analysis, we noticed some difference in traffic among different areas. Manhattan, Brooklyn and some part of Queens have the high traffic of airbnb bookings because they got some famous city hotspots for exploring, travelling or for business purposes, and also they have very convenient transportation. Also there could be many reasons such as clean and hygienic houses and rooms, best amenities provided by hosts, and many more...

**THANK YOU**