Capstone Project

AirBnb-Booking Analysis

Presented By

Abhinav Singh

Tanay Tupe

Ana Khan

Introduction

1. Background and Business Goal
2. Data Explanation and Preparation
3. EDA and Data Visualization with business insights
4. Conclusion

Background and Business Goals

Airbnb is an American Company which hosts an online marketplace and hospitality service, for people to lease on rent or short term lodging. It currently covers more than 100,000 cities and 220 countries worldwide.

The company's name comes from "air mattress B&B."
How Airbnb works for guest?
Airbnb does not own properties. It acts as an intermediary between those who want to rent out space and those who are looking for space to rent.

Airbnb's mission is to help create a world where you can belong anywhere and where people can live in a place, instead of just traveling to it.

Data Explanation

The data set we have used here is “New York Airbnb dataset”.This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values

The data set consisted of :

Host Information:

- Host ID
- Host Name

Airbnb Property Description :

- Property name
- Room type
- Neighbourhood
- Availability
- Reviews
- Location

Let’s Get Familiar With Each Series Present In The Data Set:

id: Unique ID for each Airbnb (Numeric)
name : Distinct name for each Airbnb, 16 null values.
host_id : Unique ID assigned to each Airbnb owner or host of the property(Univariate)
host_name : Name of the Airbnb Property owner, 21 null values(Univariate)

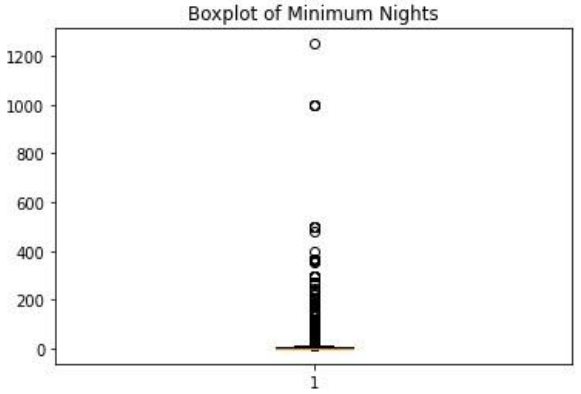
neighbourhood_group : Depicts the presence of different Airbnbs in 5 distinct neighbourhood groups. (Categorical)
neighbourhood : Depicts the presence of different Airbnbs in 221 distinct neighbourhood cities. (Categorical)
number_of_reviews Total number of reviews currently listing have. There are 1 million + reviews till date(Univariate)

latitude:"latitude" It represents the measurement of distance north or south of the Equator.(Continuous)
longitude:"longitude" It illustrates the measurement of east or west of the prime meridian.(Continuous)
Room_type: It has three different types of stay options.i. e. ‘Entire home/Apt’ , 'Private room' ,'Shared room'. (Categorical)
Minimum_nights: It illustrates minimum number of nights for booking a particular listing/Airbnb(Continuous)

price: Price details of all listings in US Dollars.(Univariate)
last_review : Date on which last review was received by property(Univariate)
reviews_per_month: Average number of reviews listing have per month(Univariate)
calculated_host_listings_count: Total counts of listings per host. (Continuous)
availability_365 : Number of days when listing is available for booking in the year.(Numeric)

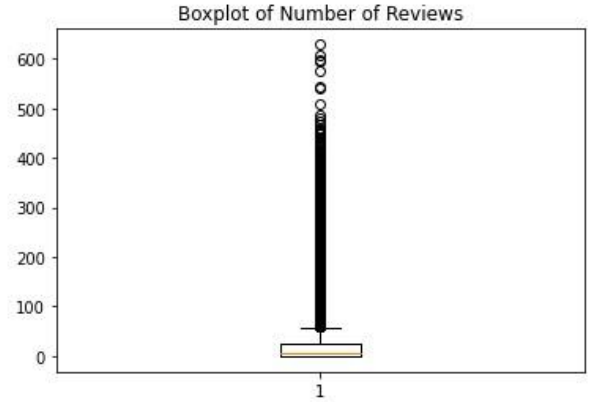
Data Preparation

- Last review and host name column dropped
- We replaced null values for number of reviews with 0
- We replaced null values for name of property with “not available”
- Check for duplicate values and eliminate, if any.



Checking Outliers

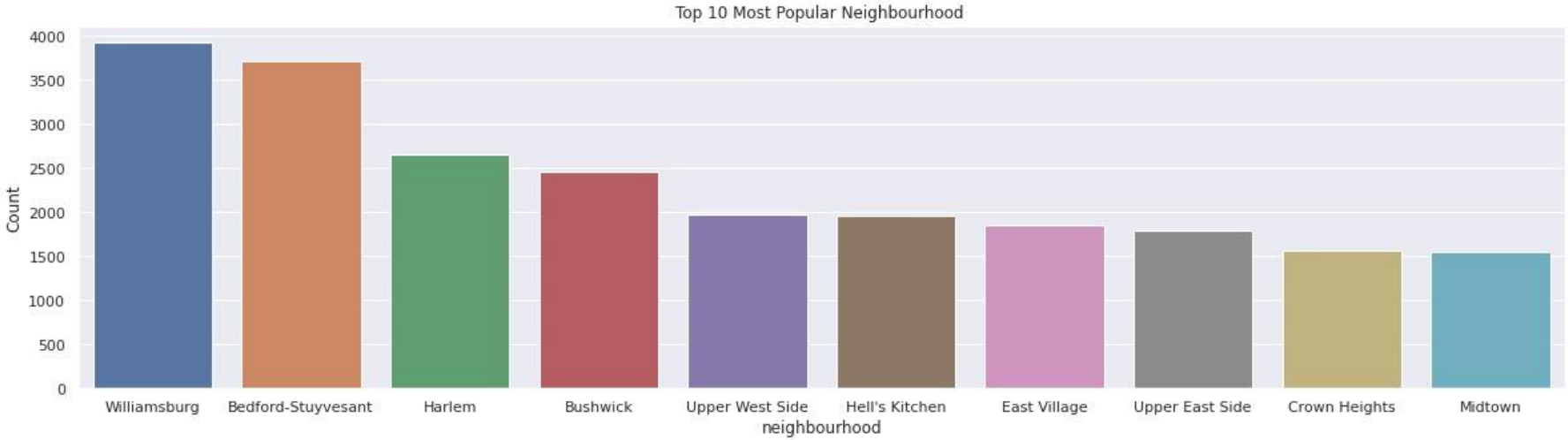
- Number of outliers are really high for the given variables.(Minimum Nights, Number of reviews)



EDA and Data Visualization

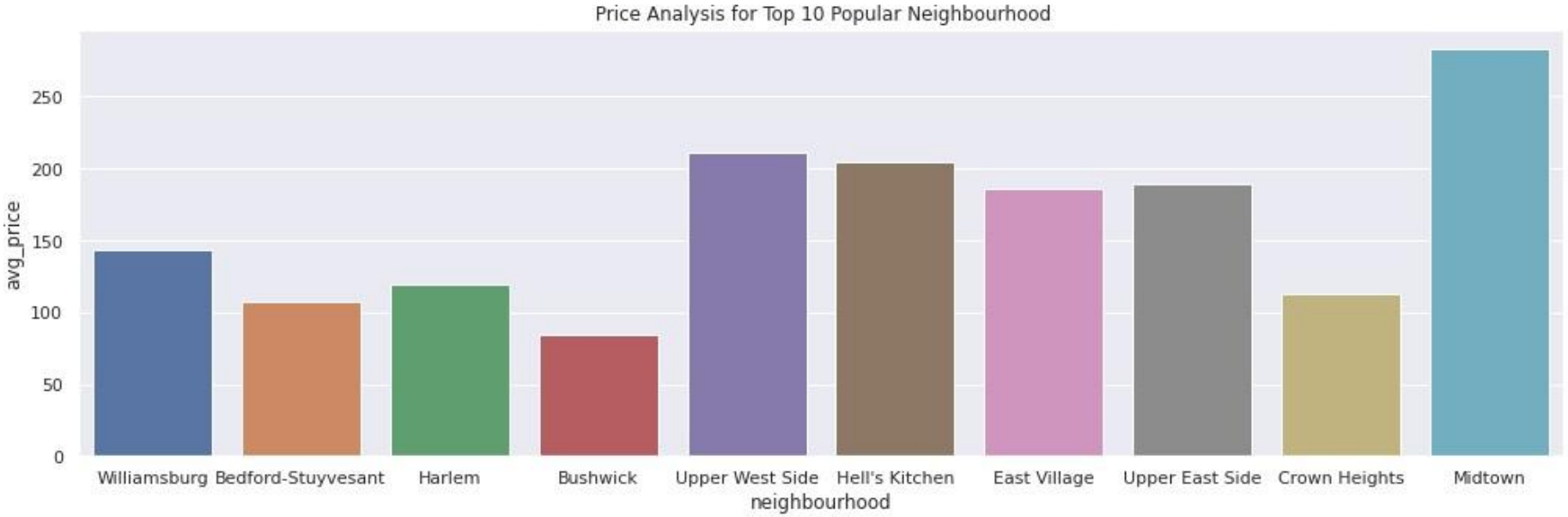
- Spatial Data Analysis
- User Review Mining

Spatial Data Analysis



- "***Williamsburg***" and "***Bedford-stuyvesant***" are almost equally popular lying within the count range of ***3500*** and ***4000***

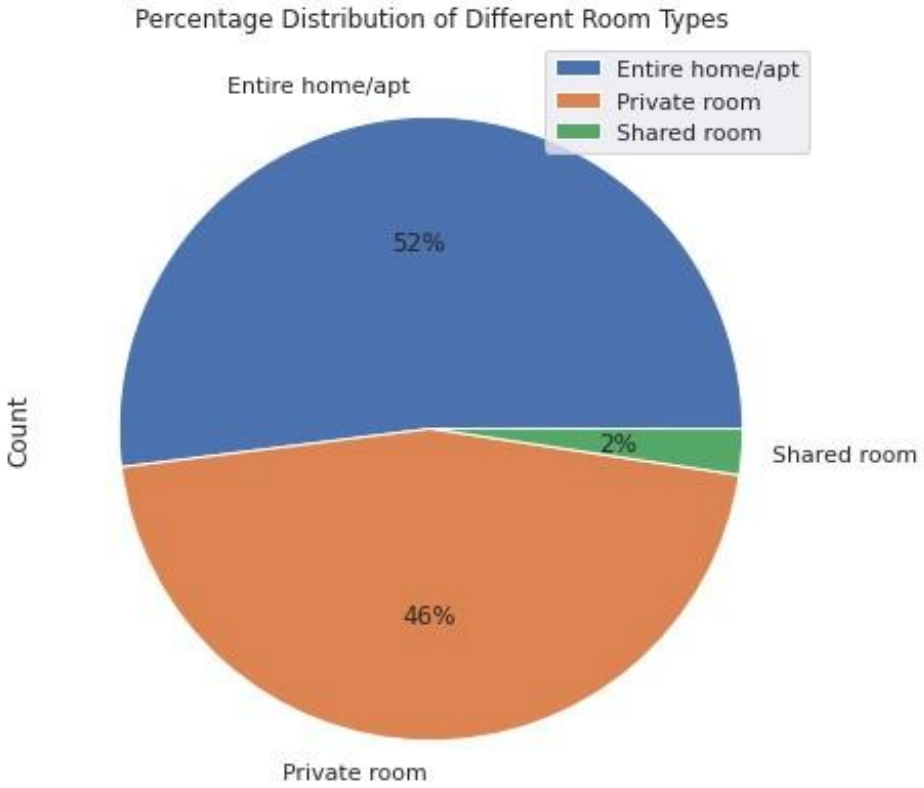
Spatial Data Analysis



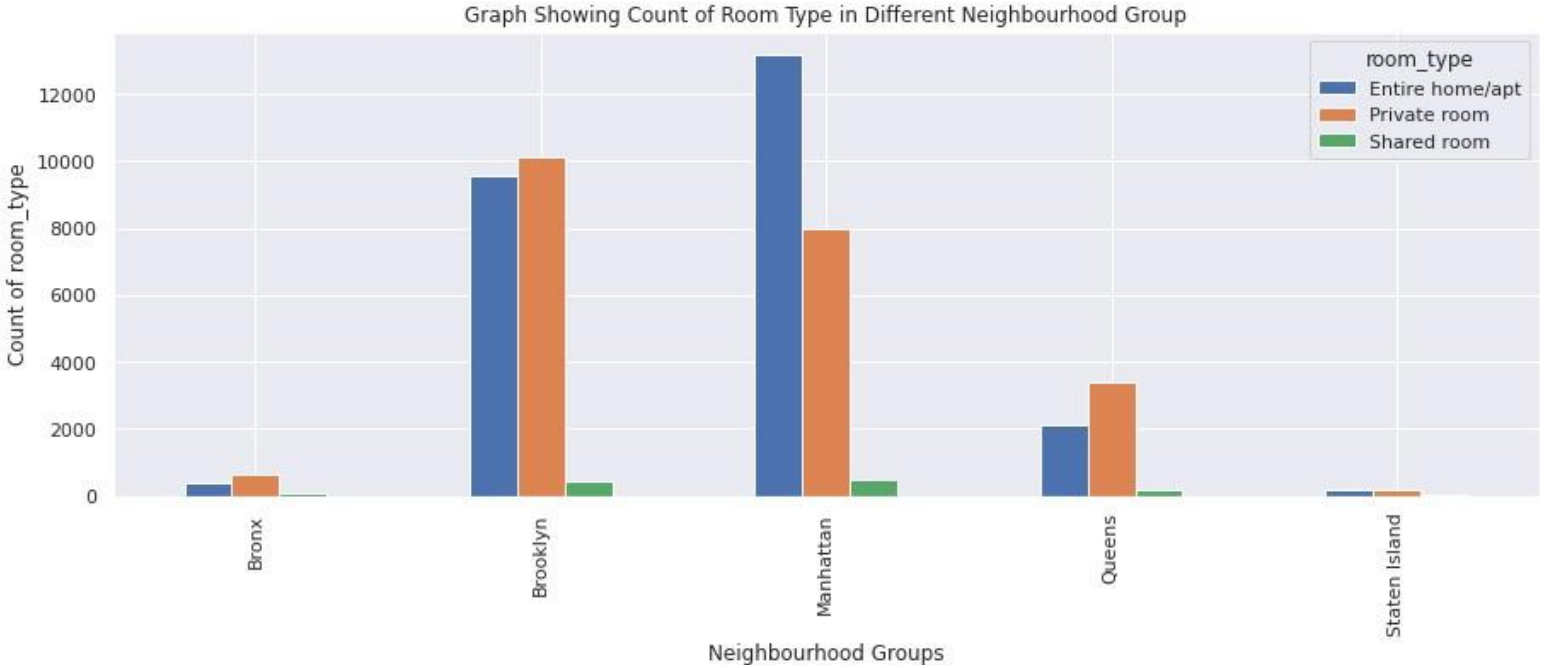
- ***Low*** average price for neighbourhood like "***Williamsburg***" , "***Bedford-stuyvesant***", "***Harlem”***

Types Of Listings

- Entire home /apartment type is the most in demand room type followed by Private type.
- Shared Room type is the least preferred.



Types Of Listings

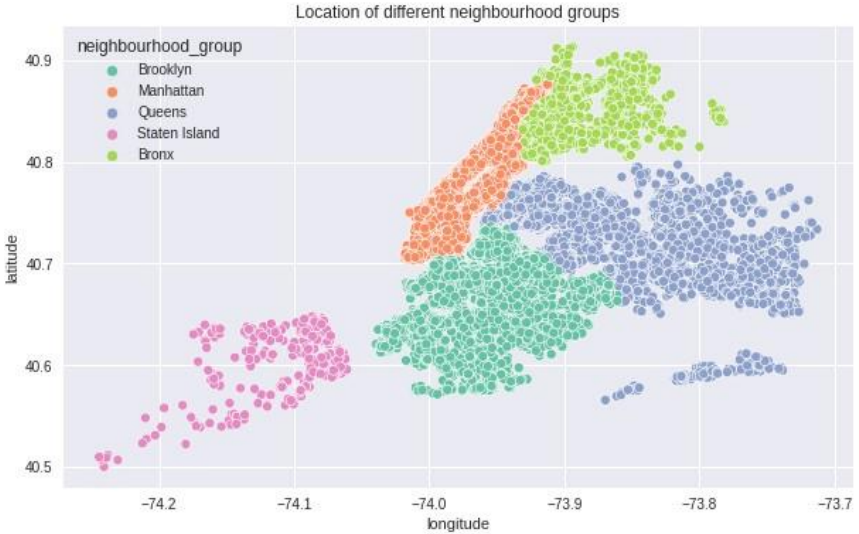


- **Manhattan** has maximum number of airbnbs
- "***Entire Home/apt***" and "***Private Rooms***" are way greater in count compared to "***Shared Rooms***"

Price Analysis for listings:

- **“Manahattan”** has the highest aggregate price for each room type.
- **Entire room/apt *lowest* aggregate price** is in **'Bronx'**
- **Private room 'Staten Island'** is having ***lowest* aggregate price**.
- **Shared room 'Brookyln'** is having ***lowest* aggregate price**.

Airbnbs plotted over NYC

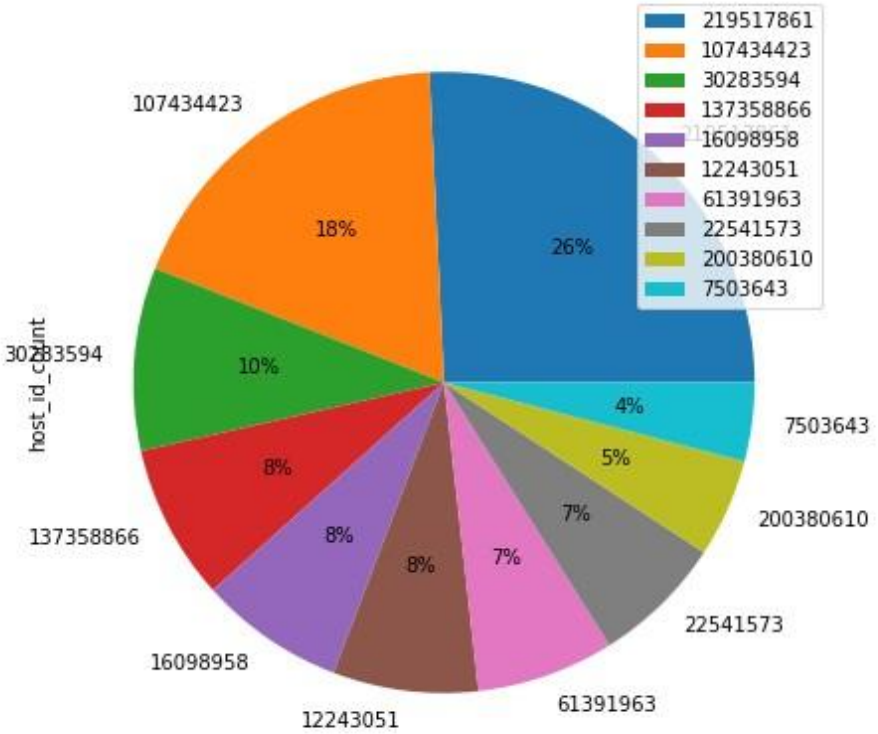


- The **ratio** of number of airbnbs to area is **highest** for **Manhattan**.

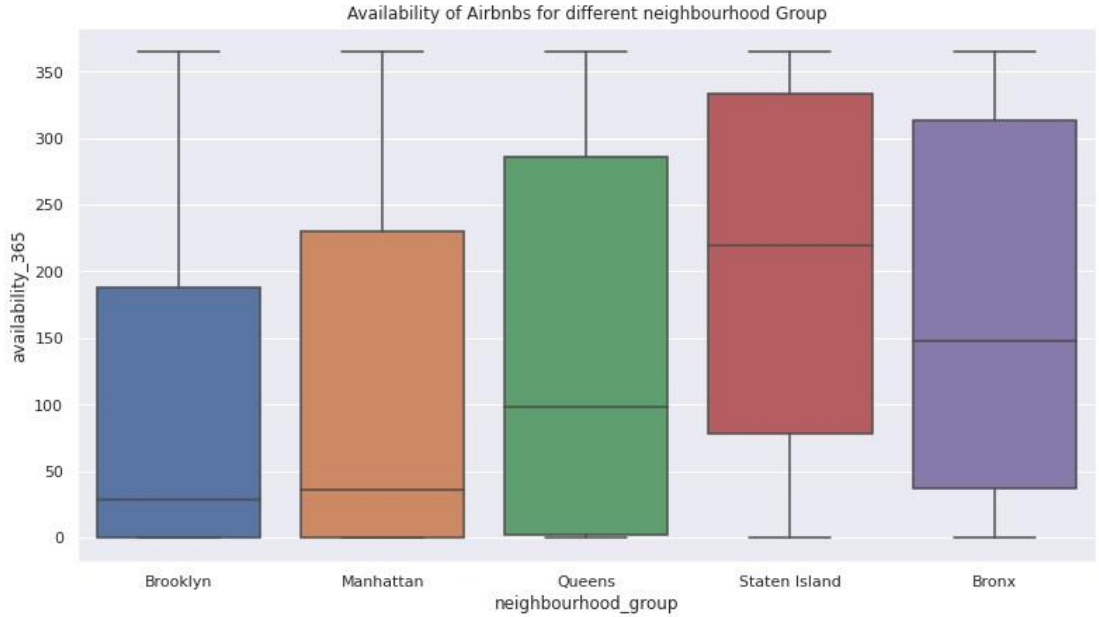
- Total listings with reviews less than 10 are around **29k.**
- "***60%***" of listings got less than **10** reviews by the visitors.
- Around **10k** airbnbs got zero reviews by the visitors.

Top Host Id

- 26% of airbnbs are owned by host_id number 219517861.
- 18 % of airbnbs are owned by host id number 107434423.

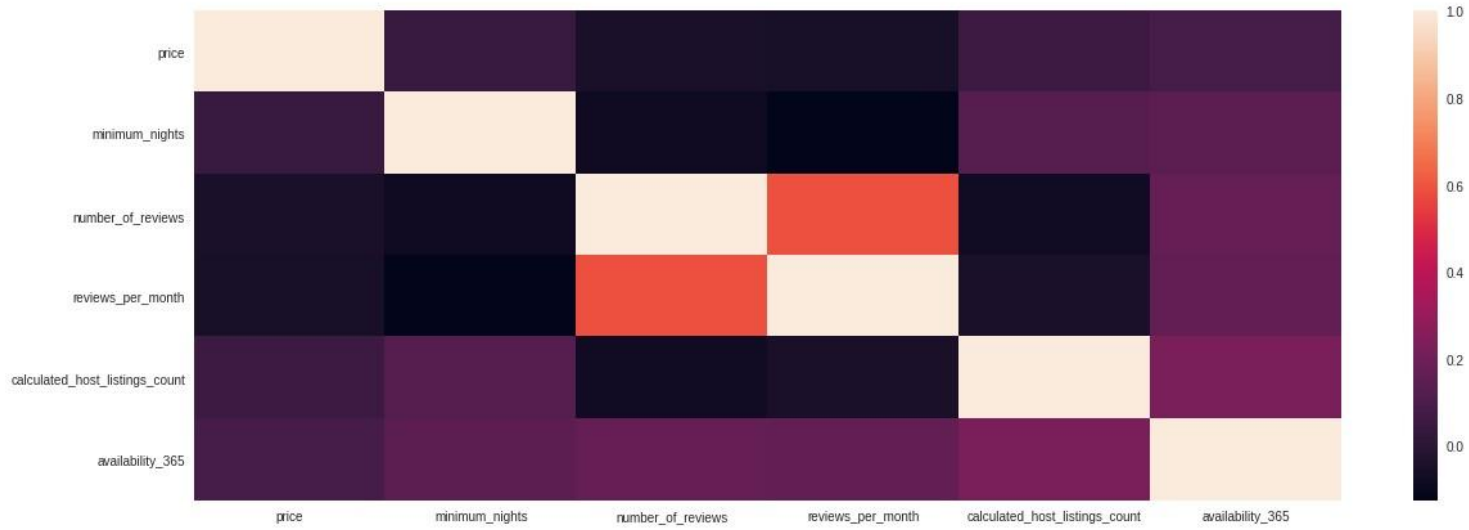


Average Availability of Airbnbs



- In **Staten Island ,** mean room availability, in days throughout the year, is over **200 days**(6.5 months approx)
- Next to this we have **Bronx,** with mean around **150 days**(5 months approx)
- **Manhattan** and **Brooklyn** have low availability of rooms throughout the year compared to other neighbourhood.

Correlation Between INT and Float Values:



- Variables are **slightly** correlated with each other.
- **number_of_reviews** and **reviews_per_month** are **much likely** correlated with each other.

CONCLUSION

1. "Williamsburg" and "Bedford-stuyvesant" are most popular neighbourhoods. Average price for these neighbourhood is quite lower in comparison to other neighbourhood cities, which could be the reason behind customers might preferred to stay in these neighbourhoods.

2. Among all the 5 neighbourhood groups “Manhattan” has the maximum numbers of listings, also average price over different room types is quite higher than other groups, so we can conclude that Manhattan might be posh area to stay.

3. In our dataset, 98% listings are of Entire home/apt or Private room, and 2% listings are of Shared room, which we can say that most visited customers might use airbnb for family stay purpose and only few customers belong to solo traveller.

4. In our dataset, out of around 49K, approx 13K airbnb listings have minimum nights as 1, which gives us an idea that customers might prefer to book airbnb for short time period.

5. It is quite surprising that 60% of listings in dataset have number_of_reviews less than 10, which seems very low numbers of reviews.