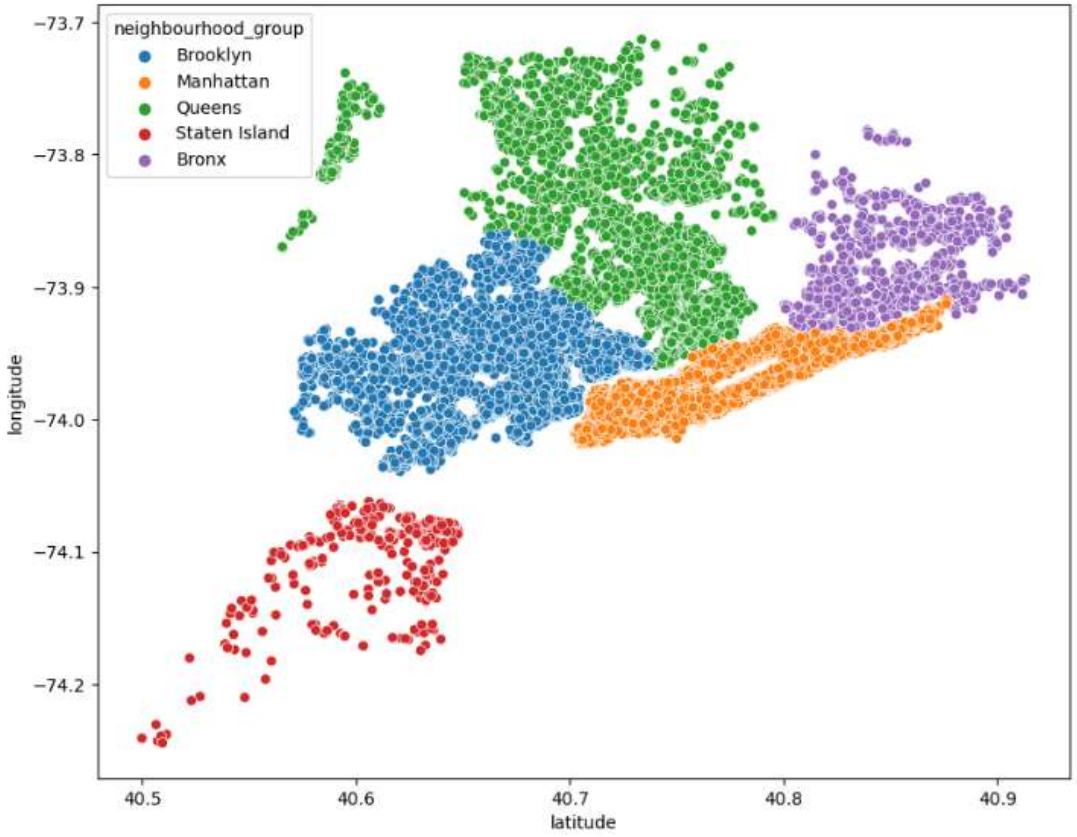
Storytelling Case Study: Airbnb, NYC

- Anuj Srivastava, Karthik Raj Narayanan, Malini S



AGENDA

- Objective
- Background
- Dataset preparation
- Key insights
- Inferences
- Appendix - Data Sources
  - Data Methodology
  - Assumptions

OBJECTIVE

- Airbnb is an online marketplace that connects homeowners looking to rent out their properties with **individuals seeking short-term accommodations** that **prioritize safety**.

- Airbnb follows a peer-to-peer model, enabling hosts to list their properties and customers to book directly, with the goal of maximizing revenue in **high-demand areas**.

- Airbnb **focuses on popular locations** and partners with well-known hosts to increase visibility and drive bookings for less popular properties

•.

BACKGROUND

• Airbnb has experienced a ***significant drop in revenue in recent months*** due to pandemic.

• Travel ***restrictions are now being lifted***, and people are beginning to travel more.

• Airbnb aims to ensure it is ***well-prepared for this shift*** in the travel landscape.

•.

Dataset preparation

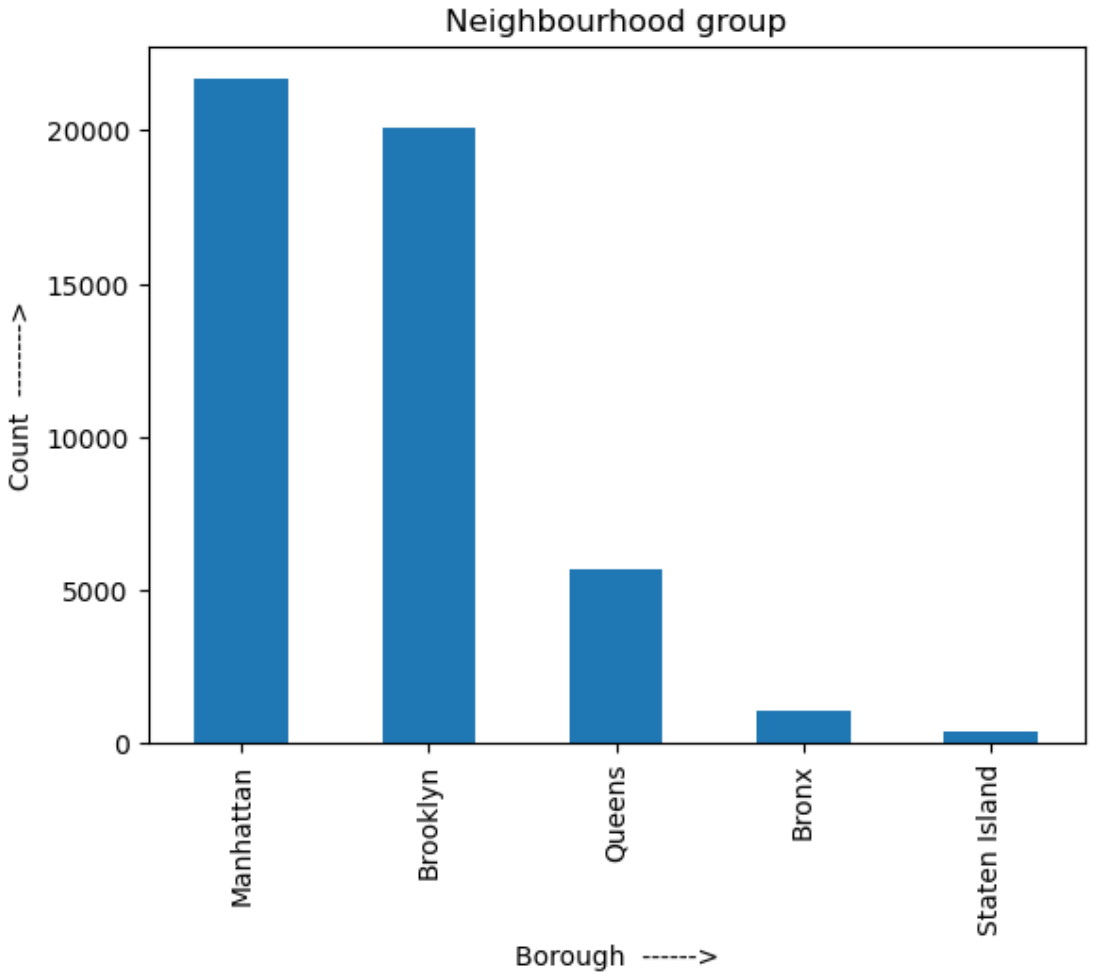
• Load the dataset into a DataFrame, and cleaned the data by handling missing values through imputation for both numerical and categorical columns.

• Conducted univariate analysis on categorical and numerical variables using visualizations, followed by handling outliers using Interquartile Range (IQR) methods.

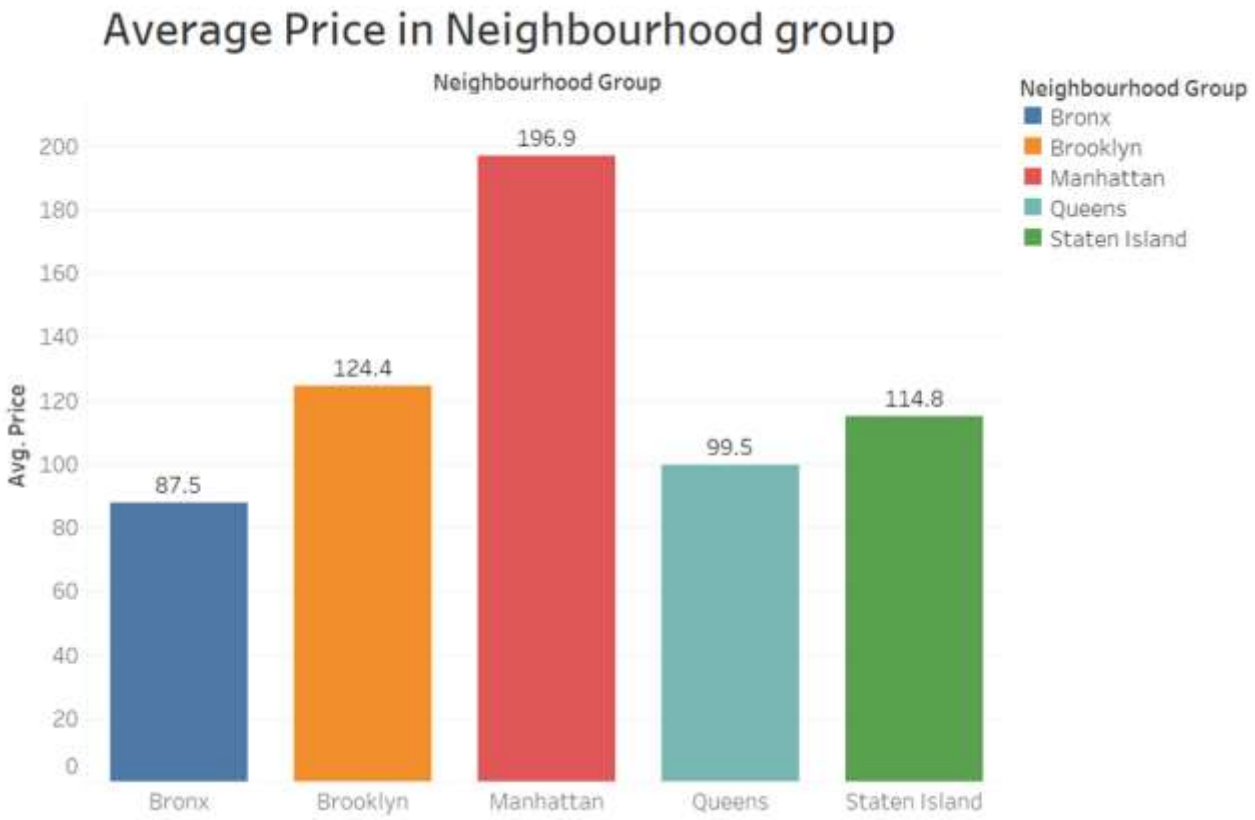
• Performed bivariate analysis in Python, with additional bivariate and multivariate analysis in Tableau to derive the meaningful insights.

High Costs Fail to Deter Customers from Manhattan and Brooklyn Boroughs

Univariate Analysis – neighbourhood_group

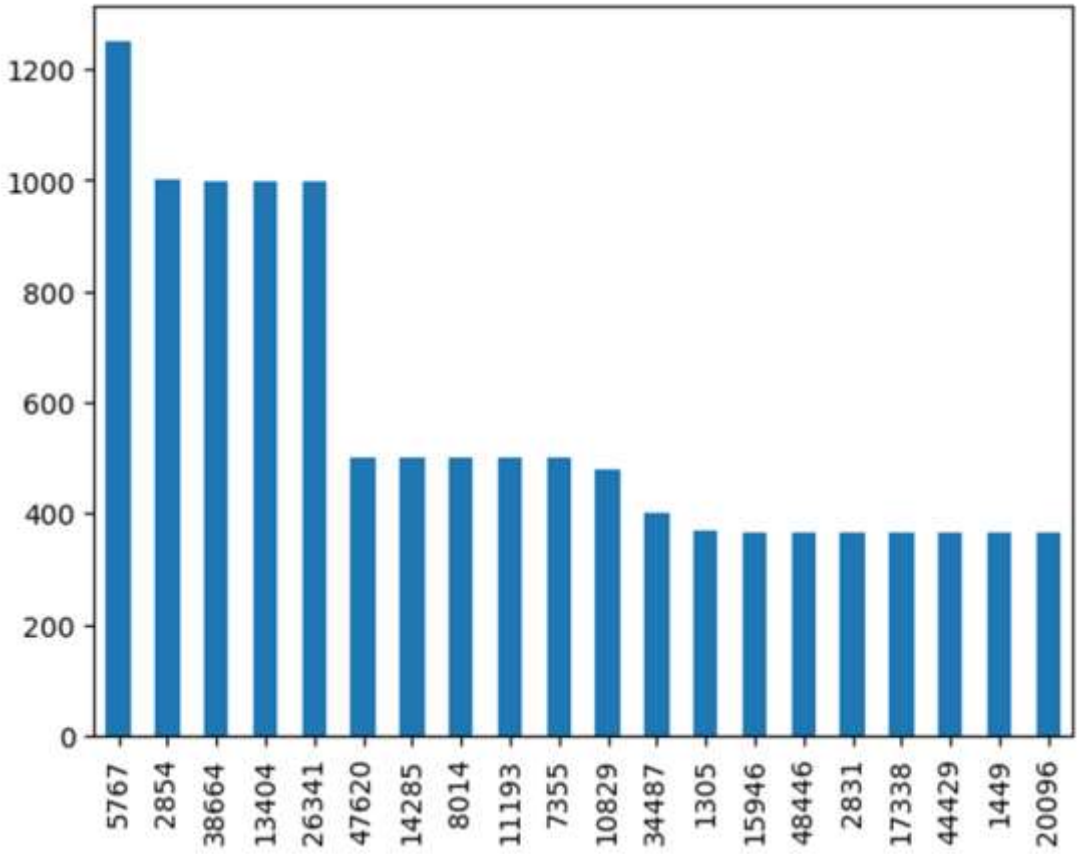


Bivariate Analysis – neighbourhood_group Vs Price

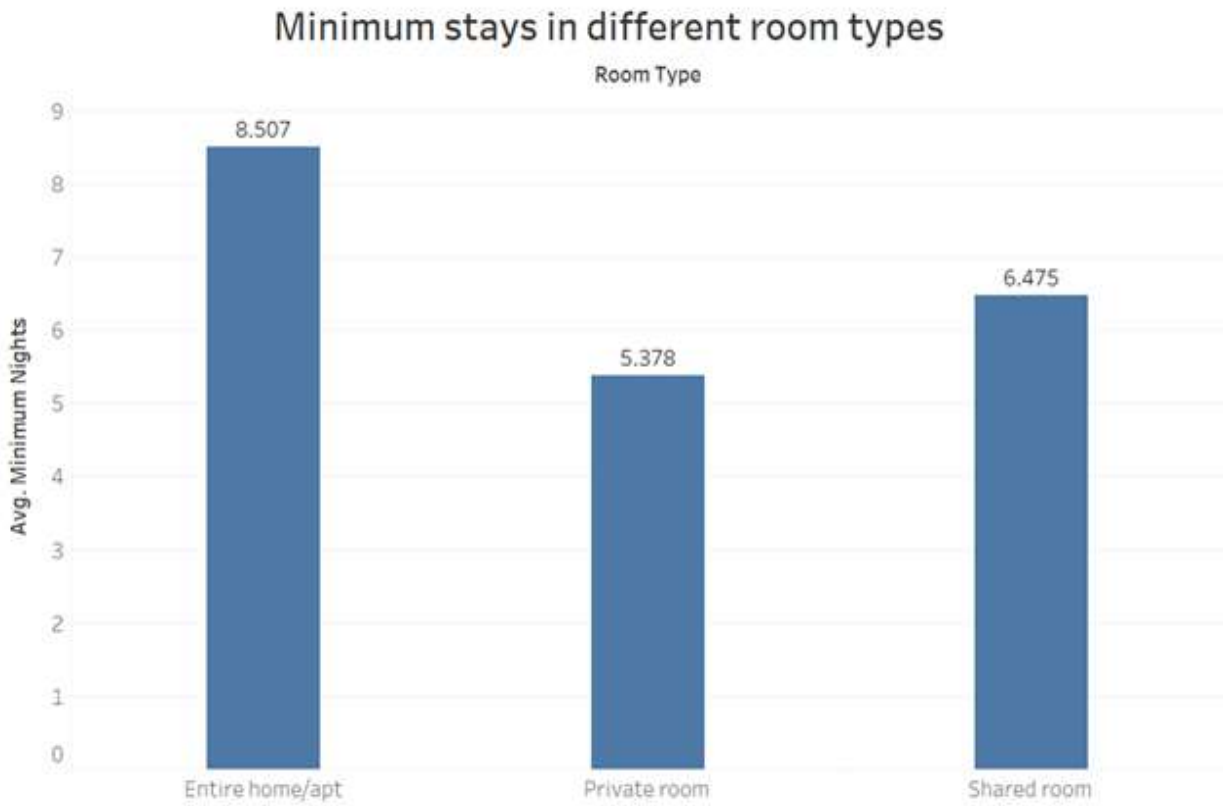


Highest Minimum Stay for Safety with Room type

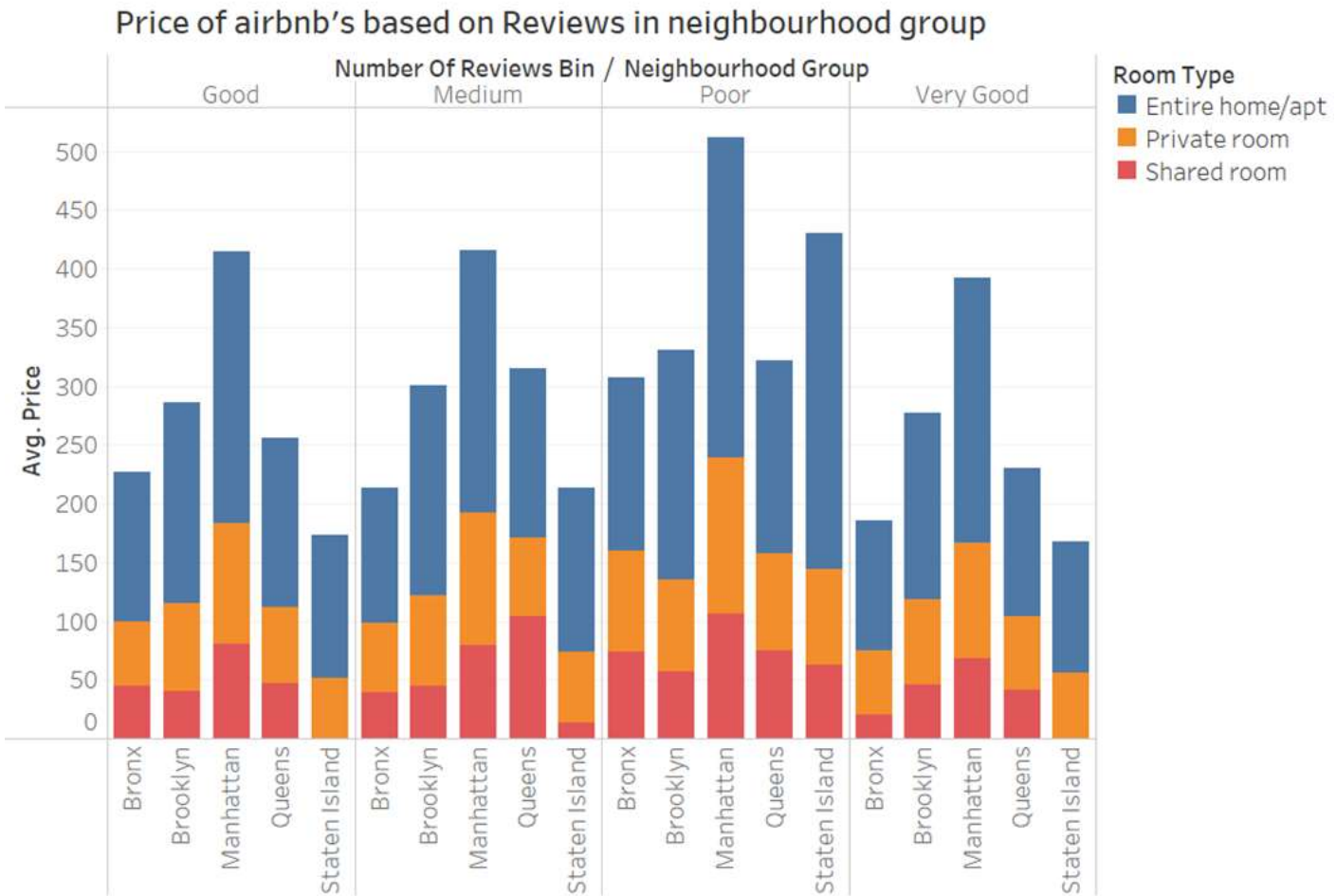
Univariate Analysis – minimum_nights



Bivariate Analysis – minimum_nights Vs room_types



Reviews drive customer choices

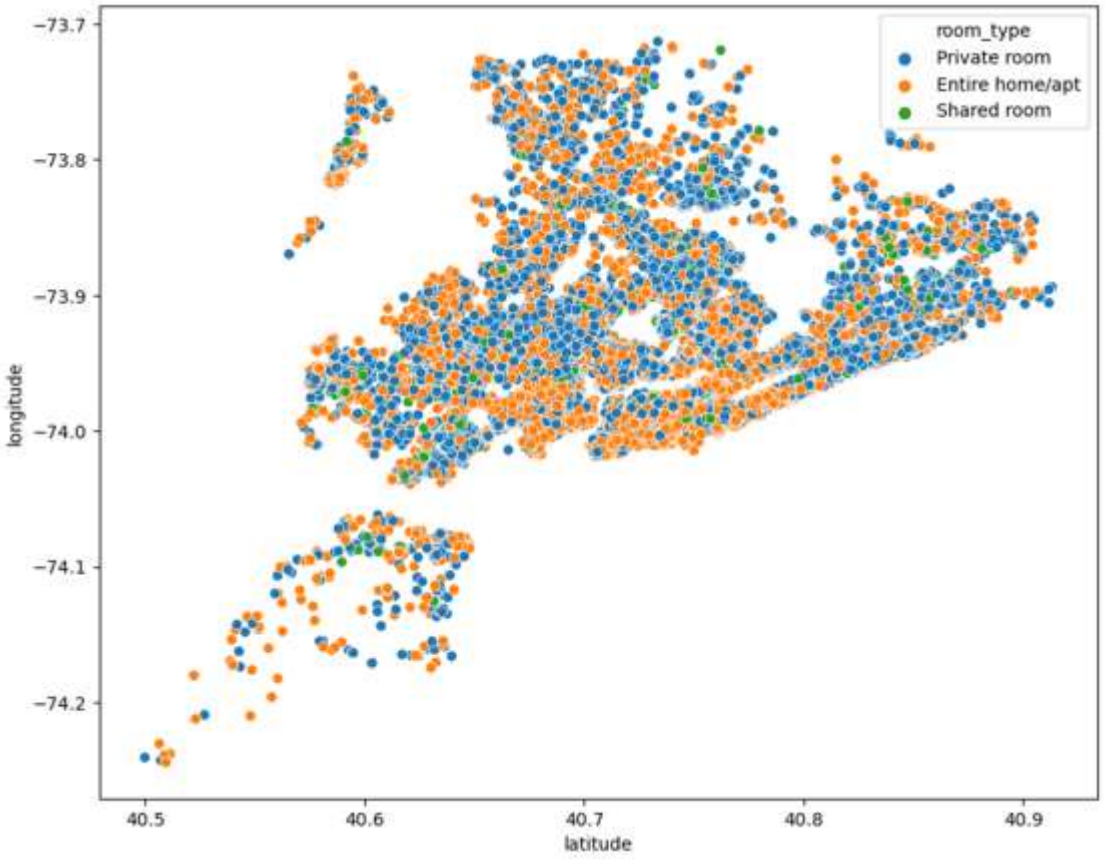


Price of airbnb's based on Reviews in neighbourhood group

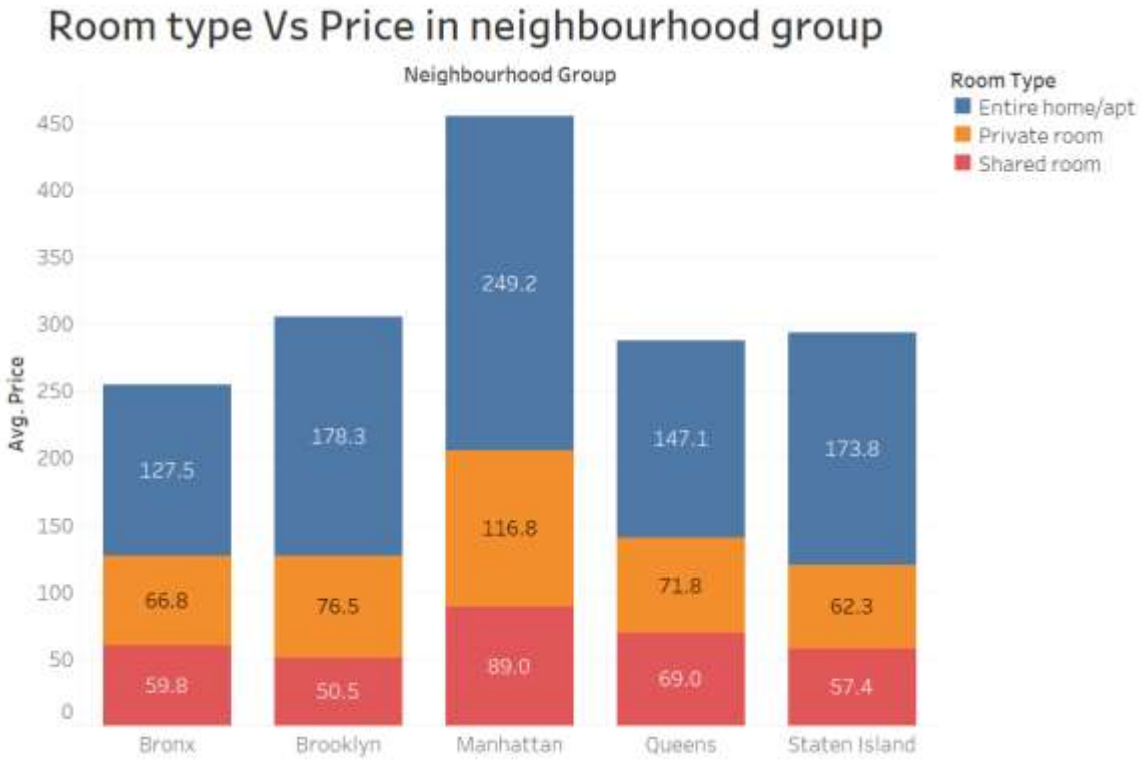
		Number Of Reviews Bin			
Room T..	Neighbourh..	Good	Med..	Poor	Very..
Entire home/ apt	Bronx	127.2	115.0	147.3	111.0
	Brooklyn	170.4	179.7	195.2	158.8
	Manhattan	232.1	223.0	273.2	225.9
	Queens	144.4	144.3	164.5	126.5
	Staten Island	121.7	139.2	286.4	112.3
Private room	Bronx	54.1	59.5	85.7	54.4
	Brooklyn	75.0	76.2	78.8	72.3
	Manhattan	102.7	113.5	132.7	98.2
	Queens	65.1	67.6	82.6	62.1
	Staten Island	51.3	61.1	81.4	55.7
Shared room	Bronx	45.0	39.4	74.0	20.0
	Brooklyn	40.2	45.2	56.7	46.0
	Manhattan	80.2	79.2	106.0	68.3
	Queens	46.9	103.6	75.2	41.7
	Staten Island		13.0	63.0	

Customer Preference for Home Apartments: Prioritizing Safety and Privacy

Bivariate Analysis neighbourhood_group Vs room_type

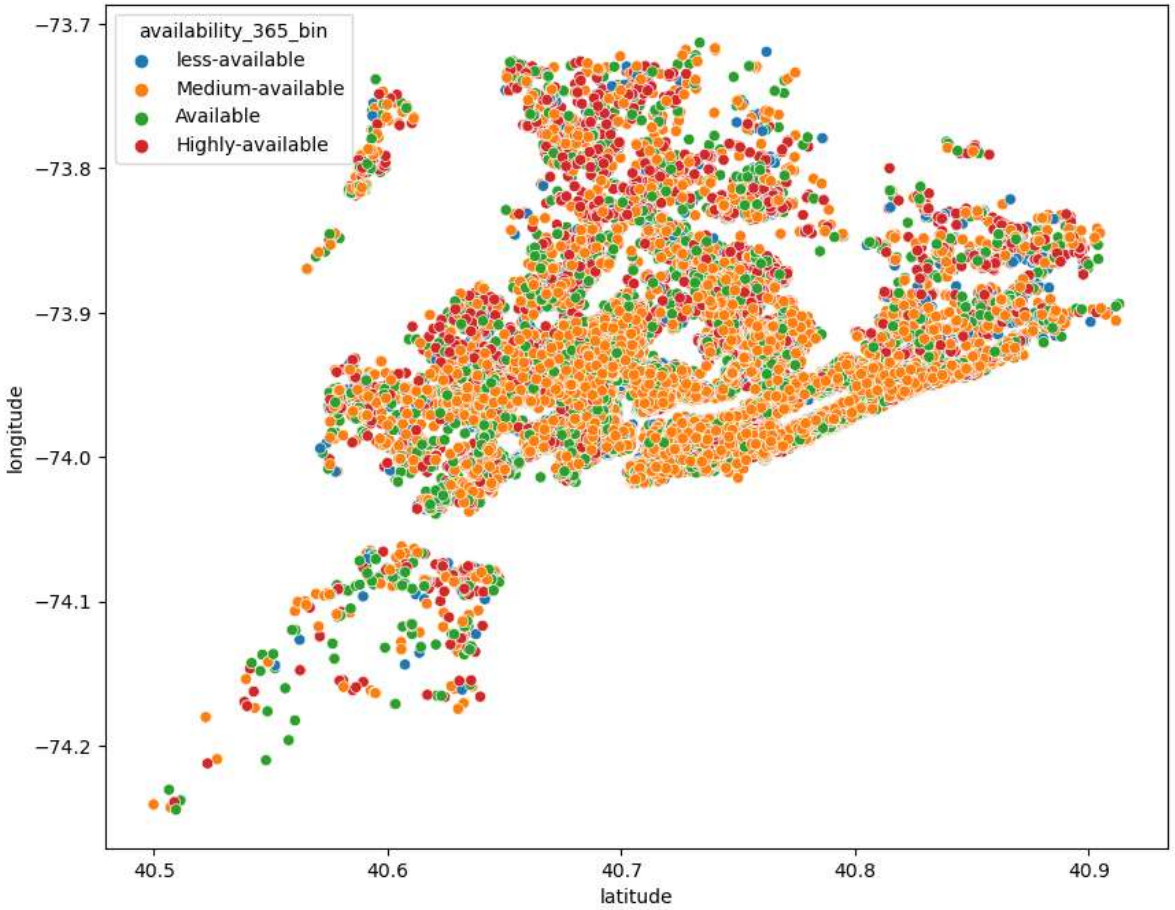


Bivariate Analysis neighbourhood_group Vs room_type

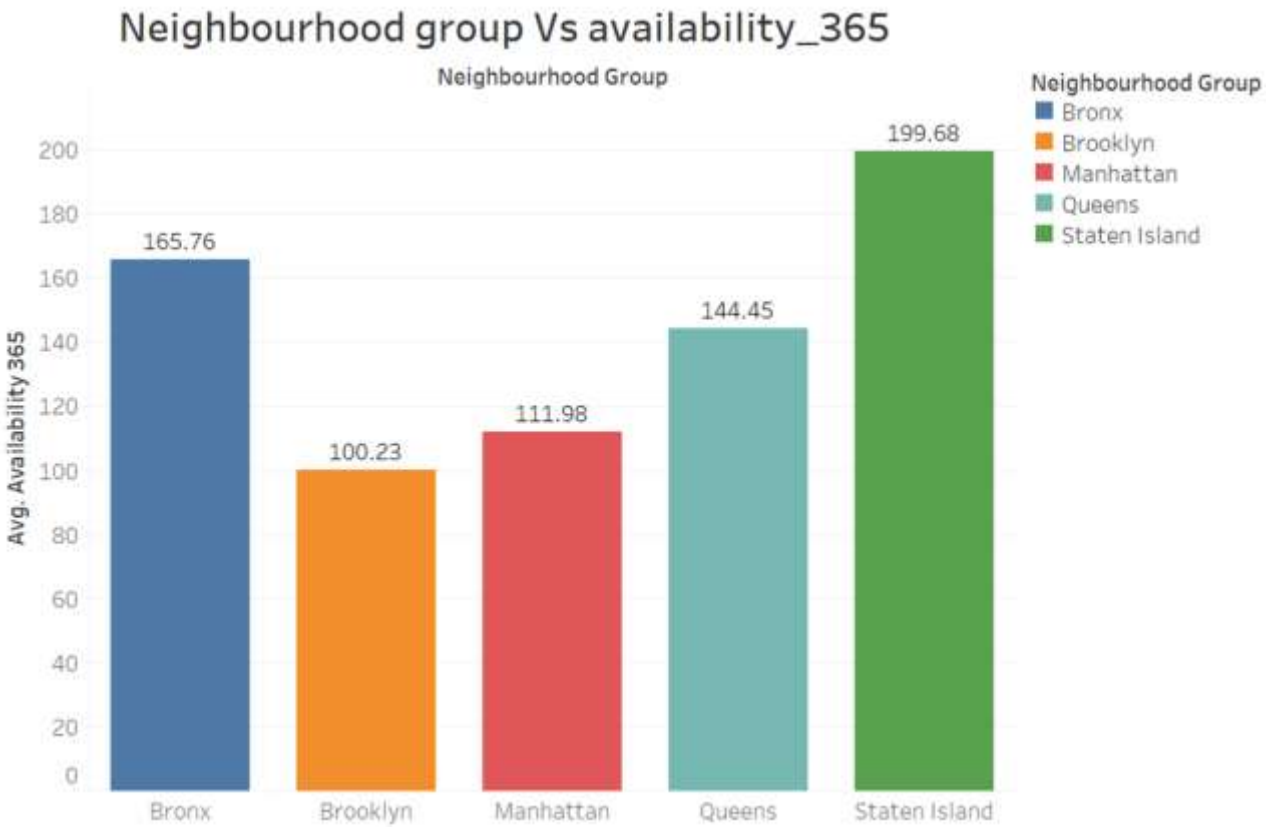


Maximizing Airbnb Availability: Off-Peak Seasons, Midweek Stays, and Early Bookings

neighbourhood_group Vs availability_365

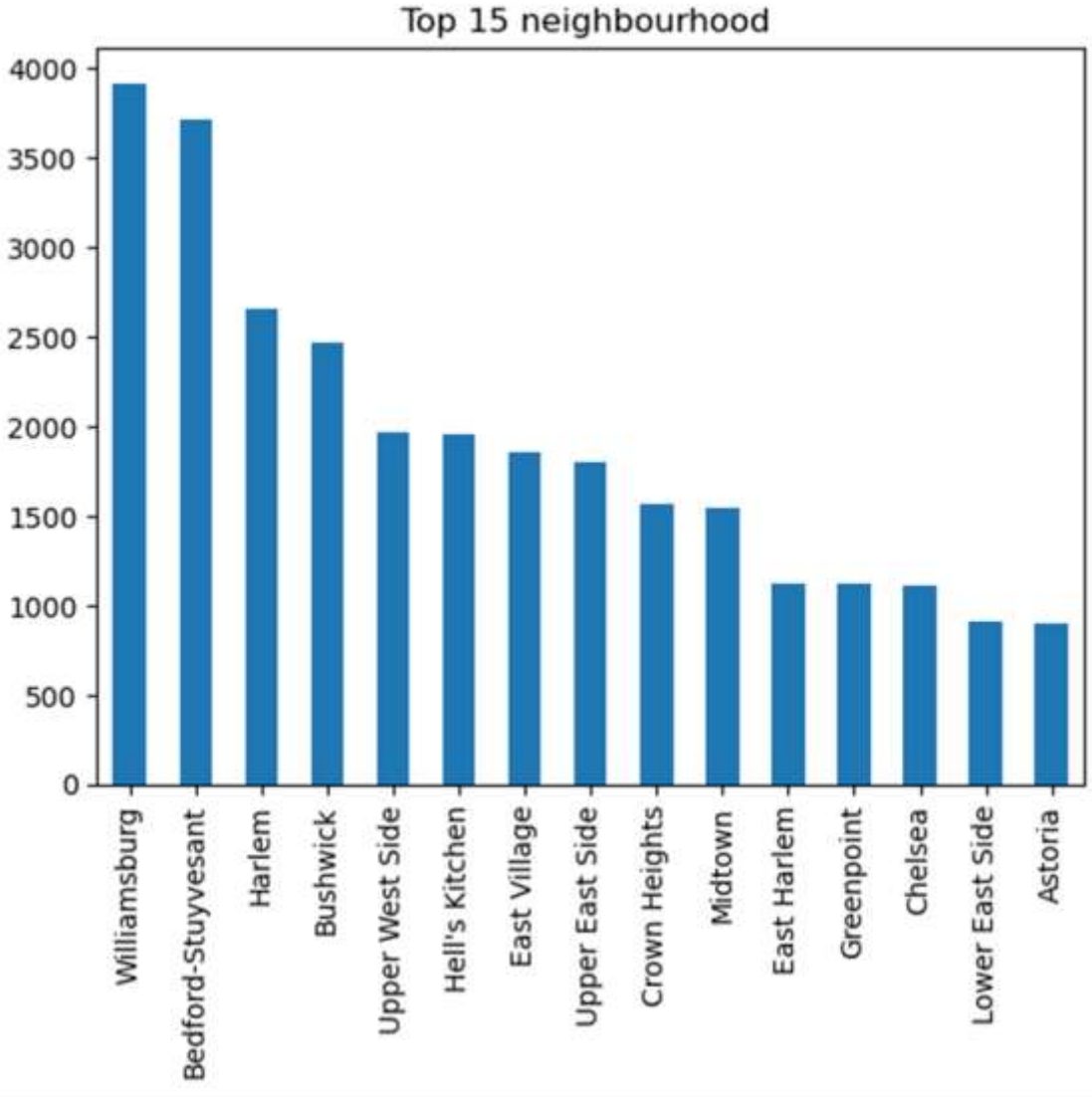


neighbourhood_group Vs availability_365



Locality Preferences of Customers: Key Factors Driving Their Choices

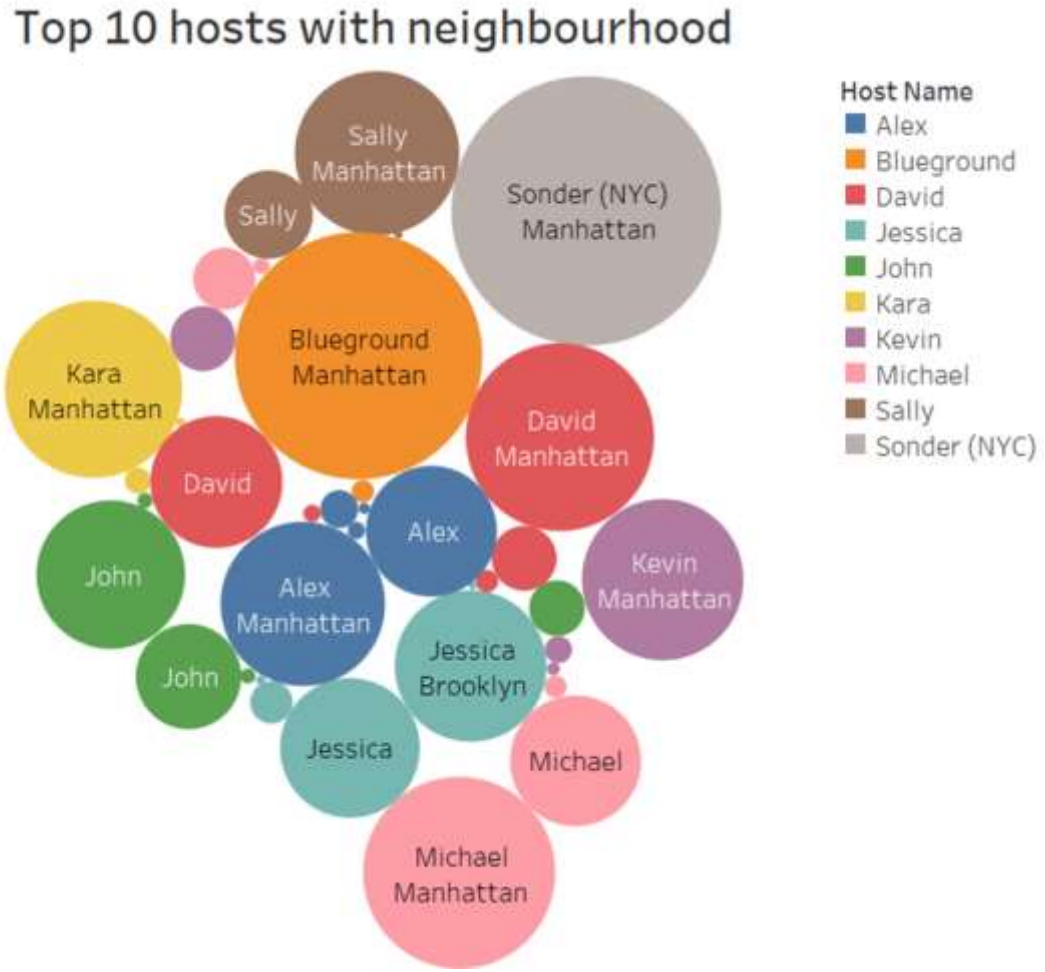
Univariate Analysis - neighbourhood



- The top 6 locality preferences are :
  1. Williamsburg
  2. Bedford-Stuyvesant
  3. Harlem
  4. Upper West Side
  5. Hell's Kitchen,
  6. East Village.

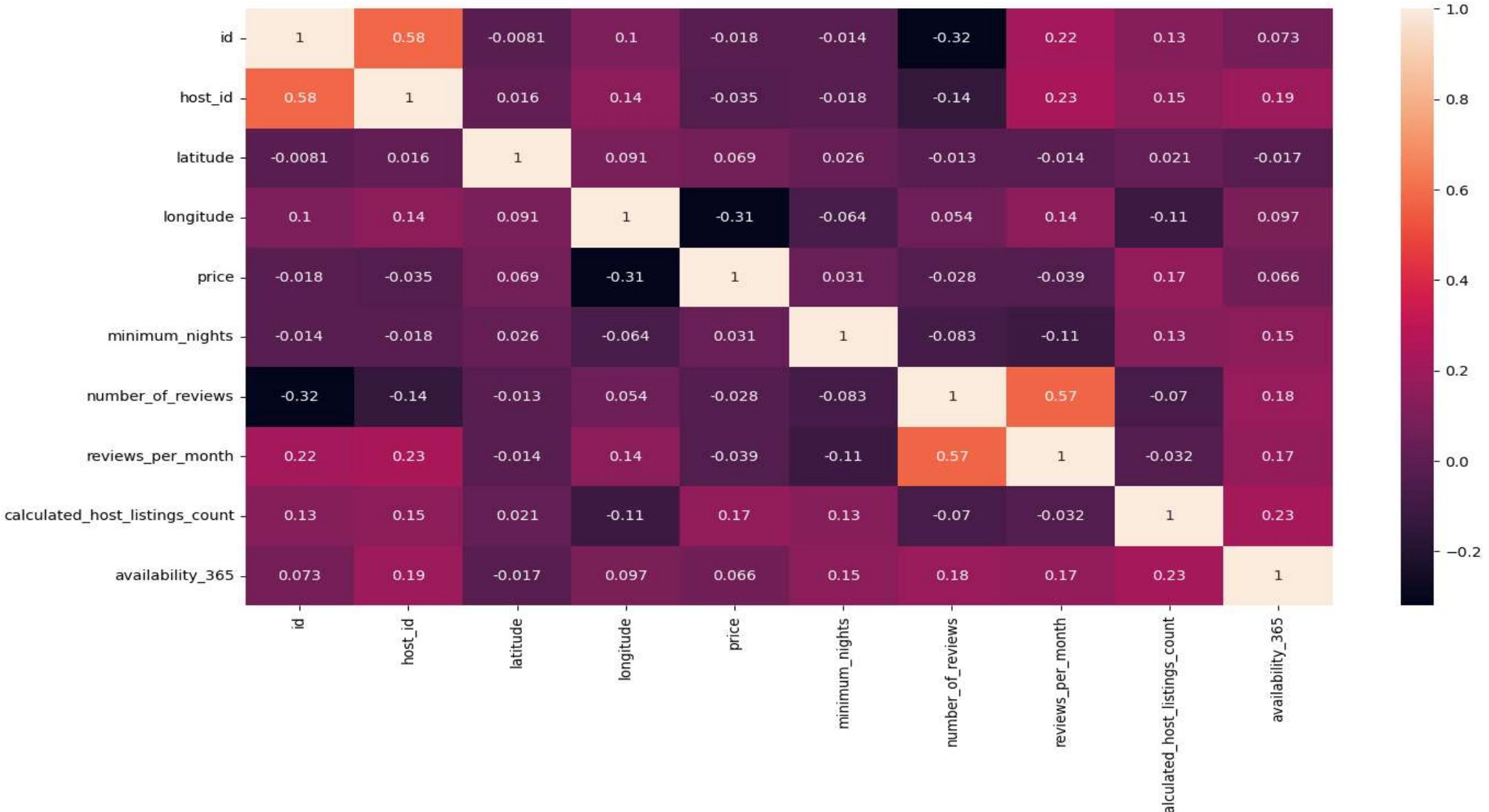
- Williamsburg, Bedford-Stuyvesant, Bushwick, neighbourhood is situated in Borough Brooklyn whereas Harlem, Upper West Side, Hell's Kitchen located in borough Manhattan.

Top 10 hosts with Neighbourhood group



- Sonder (NYC) and Blueground from Manhatten are the top hosts followed by David and Alex.

Multivariate Analysis – Heat map



Inferences:

- Customers favor Entire ***Homes/Apartments*** over Private and Shared Rooms due to ***safety and privacy concerns***. To increase bookings for Shared Rooms, special offers could be considered.

- Based on price and reviews, ***Manhattan*** and **Brooklyn** attract a higher number of customers despite their high costs.

- The top 6 locality preferences are Williamsburg, Bedford-Stuyvesant, Bushwick, neighbourhood is situated in **Borough Brooklyn** whereas Harlem, Upper West Side, Hell's Kitchen located in borough ***Manhattan***.

- ***Staten Island*** has the highest availability for booking. Hence, management can communicate with famous hosts who have listed their properties in Staten Island ***to help boost attraction for unpopular properties.***

Appendix: Data Sources

Here is a snapshot of our data dictionary.

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

data sources: (for downloading the dataset)
https://learn.upgrad.com/course/5674/segment/53344/324214/981584/4903670#:~:text=New%20York%20Airbnbs%20Dataset

Appendix – Data Methodology

Step 1: Import necessary python libraries

Step 2: Reading the dataset

- A. Conversion of data into dataframe
- B. Examining the dataframe

Step 3: Data cleaning

Handling missing values

- A. Segregation of numerical and categorical column
- B. Imputation on numerical and categorical column

Step 4: Univariate Analysis

- A. Categorical variables
- B. Numerical variables

Step 5: Inferences from univariate analysis

Step 6: Handling outliers

Step 7: Bivariate Analysis and Multivariate Analysis

- A. Bivariate Analysis carried out in python
- B . few other Bivariate Analysis carried out using Tableau

Appendix – Data Assumptions

Assume the data is from a reliable source, analysed, knowing that the accuracy and validity of the insights and conclusions will be well-supported.