CAPSTONE PROJECT - I

AIRBNB BOOKING ANALYSIS

Team:

Shrinidhi Choragi

Nakshith D N

Contents

Introduction

Since 2008 Airbnb has expand on travel possibilities and presented a more unique and personalised way of experiencing the world.
It is an American company that facilitates an online marketplace for lodging, primarily homestays for vacation, rentals, and tourism activities.

Problem Statement

The objective of the project is to perform an exploratory data analysis, data pre-processing, data cleaning & imputation, and in the end, apply different Data Visualization techniques to get meaningful insights from the given data.

Explore and analyze the data to discover the following key understandings.

- What can we learn about different hosts and areas?
- What can we learn from predictions? (ex: locations, prices, reviews, etc)
- Which hosts are the busiest and why?
- Is there any noticeable difference in traffic among different areas and what could be the reason for it?
- What is average revenue per host and how does it vary for different neighbourhood groups
- Depict the price distribution among neighbourhood groups.

Dataset

This dataset describes the listing activity and metrics in NYC, NY for 2019.

This dataset has around 48895 observations in it with 16 columns and it is a mix of categorical and numeric values.

The features of the dataset are

Id: Identity number of the property listed

Name: Name of the property

Host_id: Id number of hosts registered on Airbnb

Host_name: Name of the host registered

Neighbourhood_group: Names of neighbourhood groups in NYC

Neighbourhood: Names of neighbourhood present in neighbourhood groups

Latitude:Coordinate of latitude of the property listed

Longitude: Coordinate of longitude of the property listed

Room_type: Type of room listed by host

Price: Rent of the property listed

Minimum_nights: the minimum number of nights customer can rent the property

Number_of_reviews: Number of customers that have reviewed the property

Last_review: Date when the property was last reviewed.

Reviews_per_month: Number of reviews per month

Calculated_host_listings_count: Number of listings done by particular host

Availability_365: Number of days the property is available

Data Cleaning

The Percentage of missing values in the dataset is found to be: 2.57%

Handling missing values:

- *host* and *host name*:

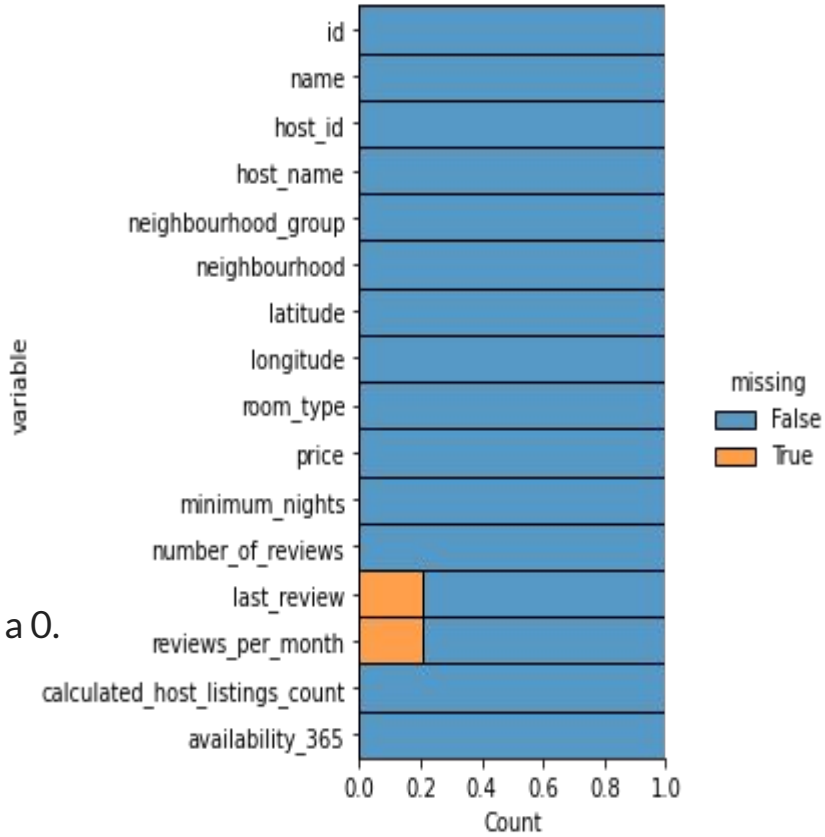
imputed with a dummy variable.

- *"review_per_month"*:

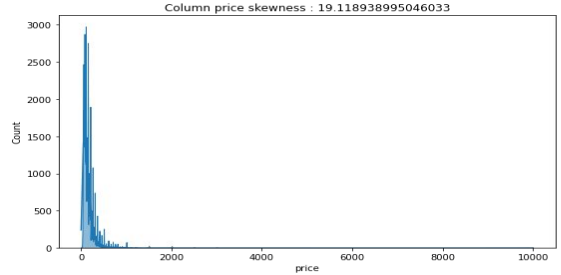
imputed with 0.0 for missing values since "*number_of_review*" of the corresponding column has a 0.

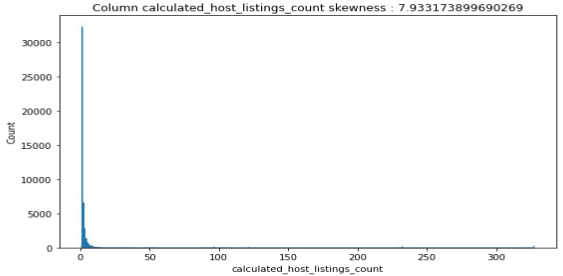
- "*last_review*" :

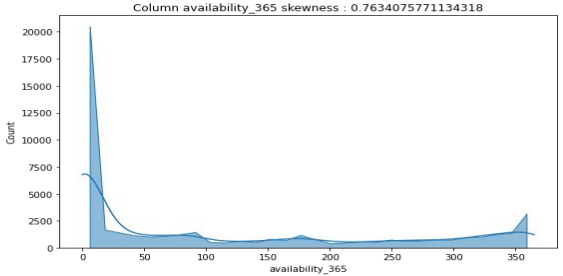
It's dropped.

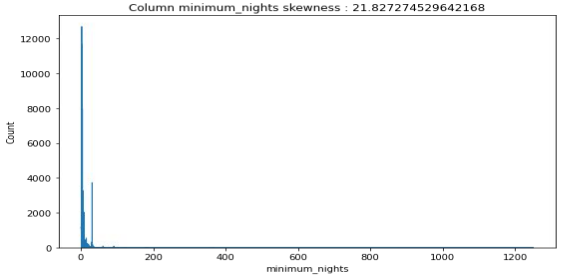


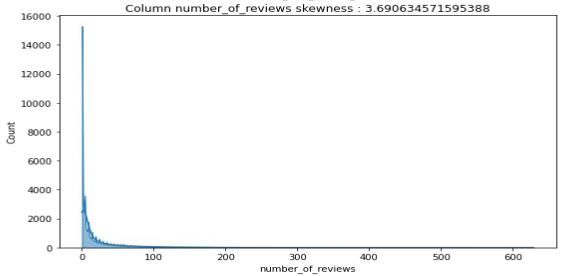
Descriptive Statistics

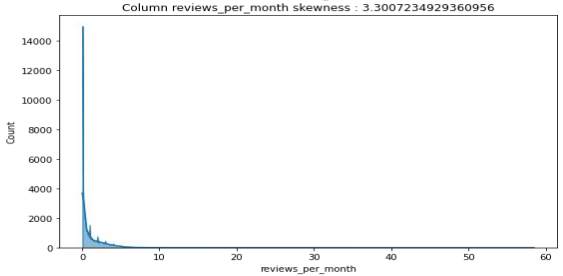






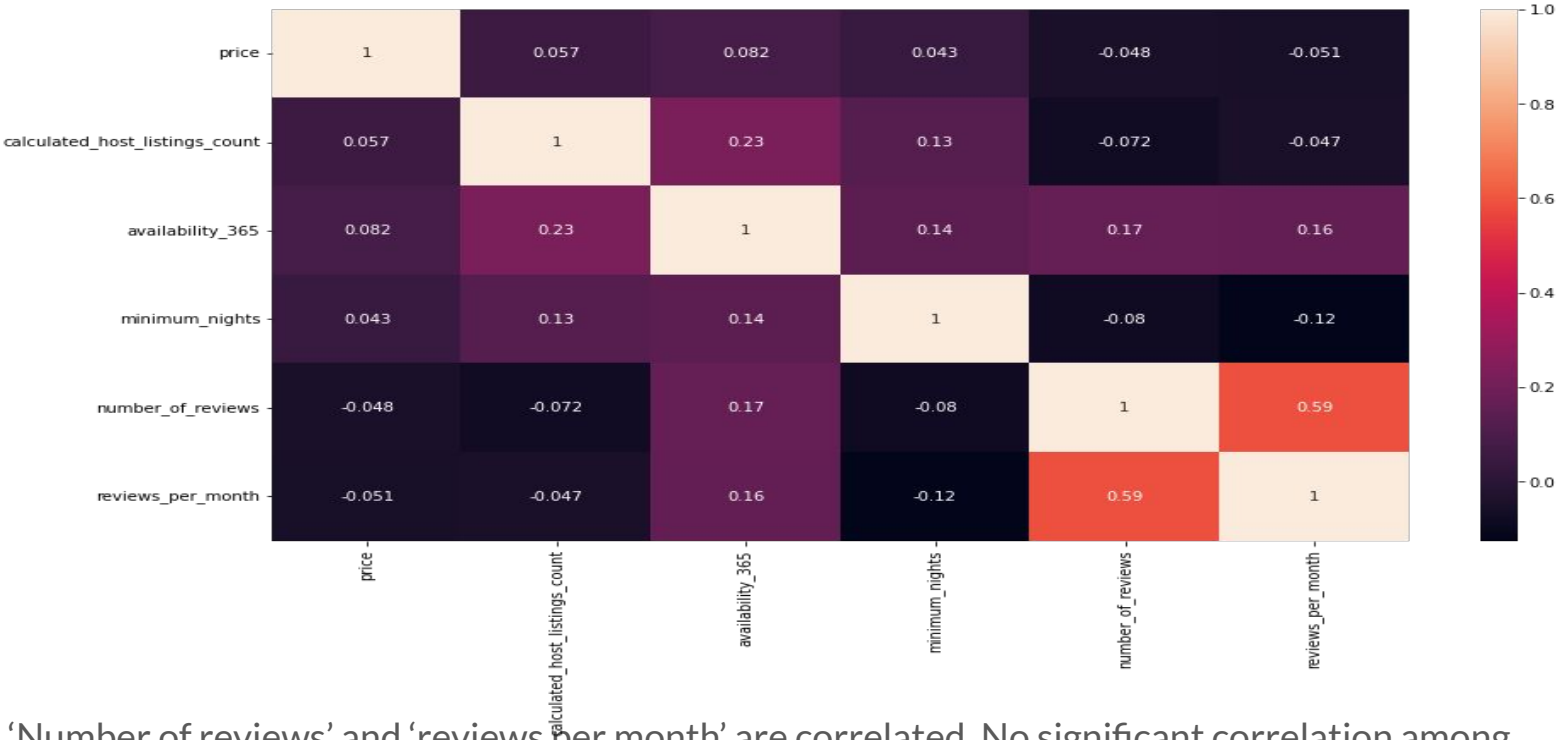






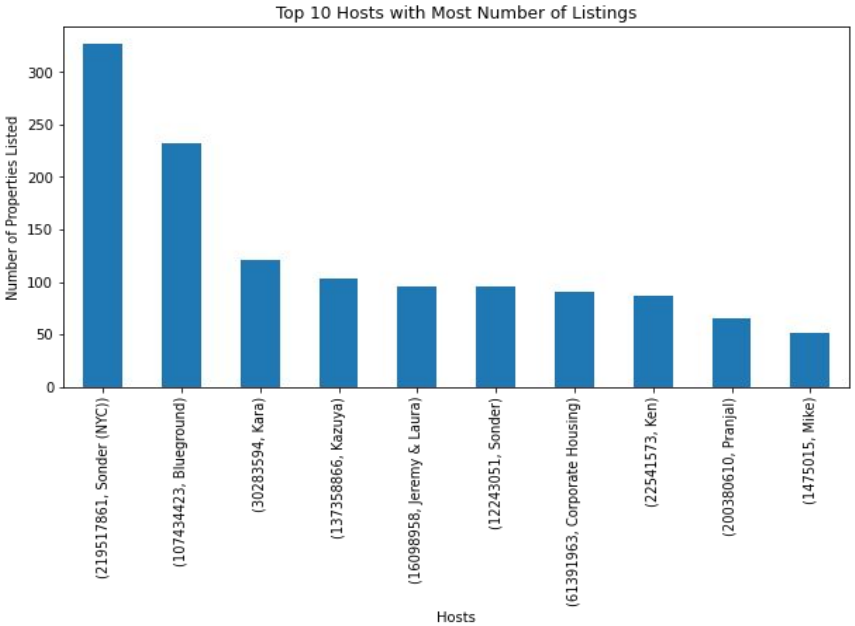
Most of the features are positively skewed

Correlation Analysis

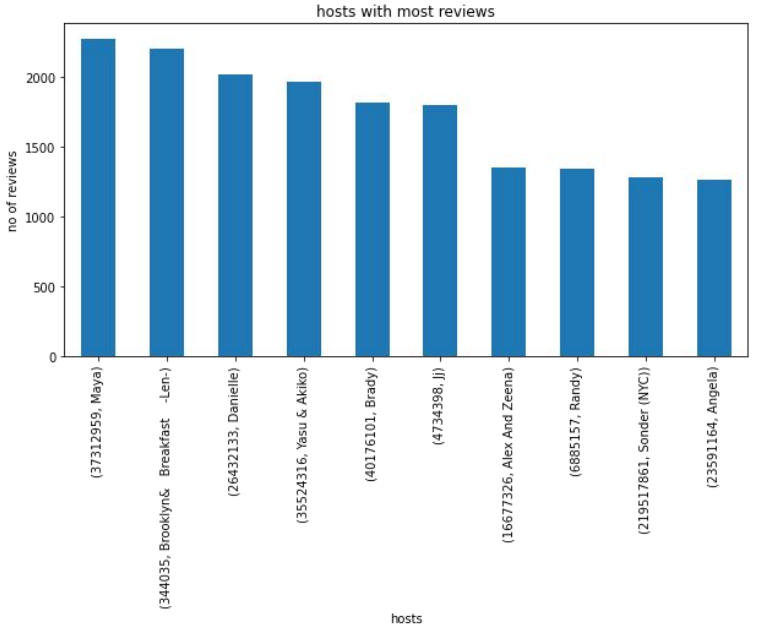


‘Number of reviews’ and ‘reviews per month’ are correlated. No significant correlation among others.

Hosts

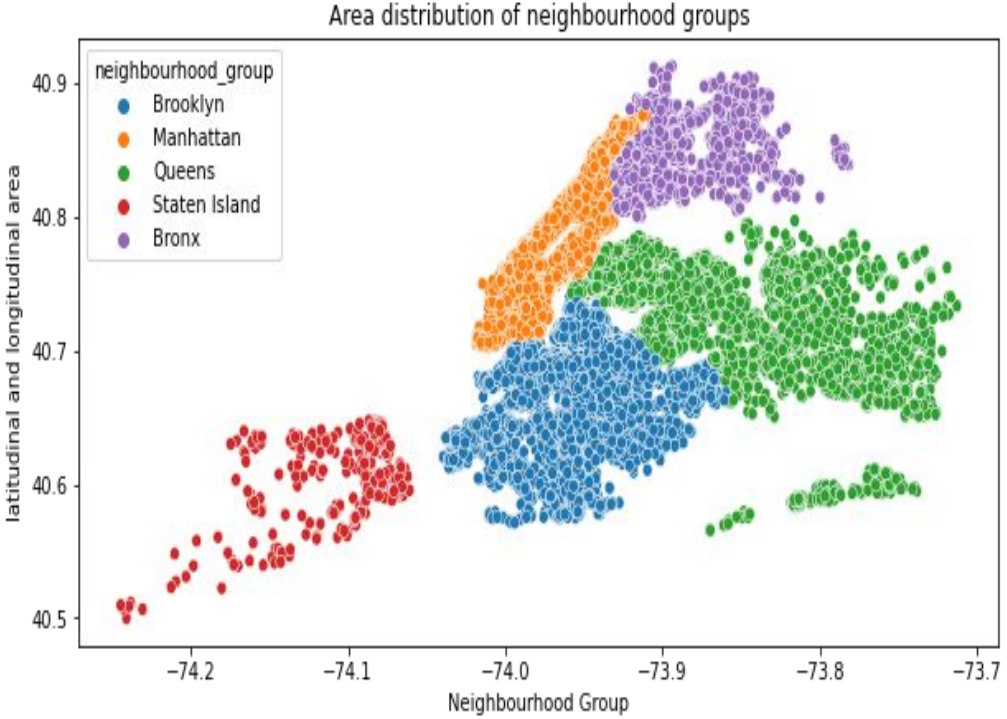


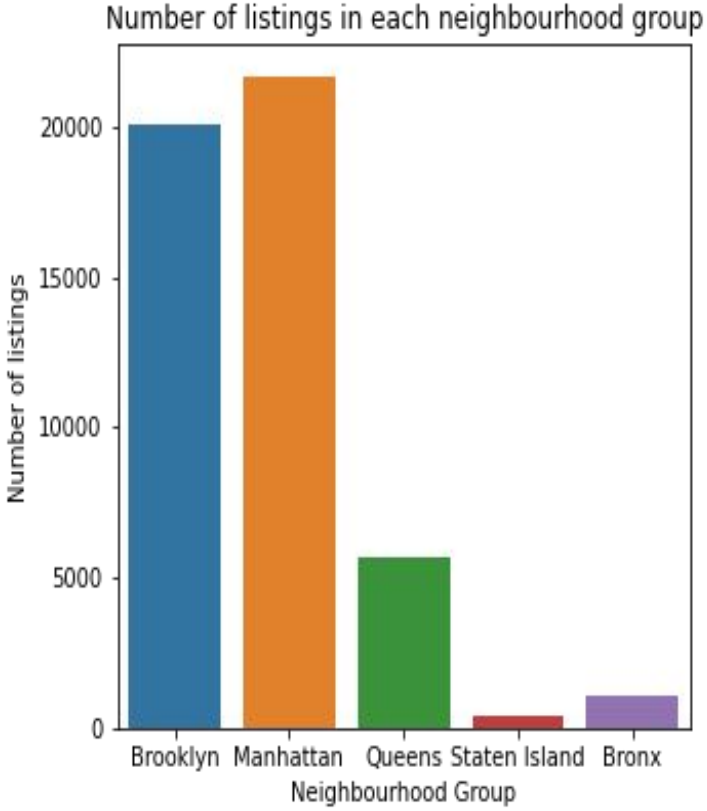
Sonder(NYC) has the most listings



Maya has received the highest no of reviews

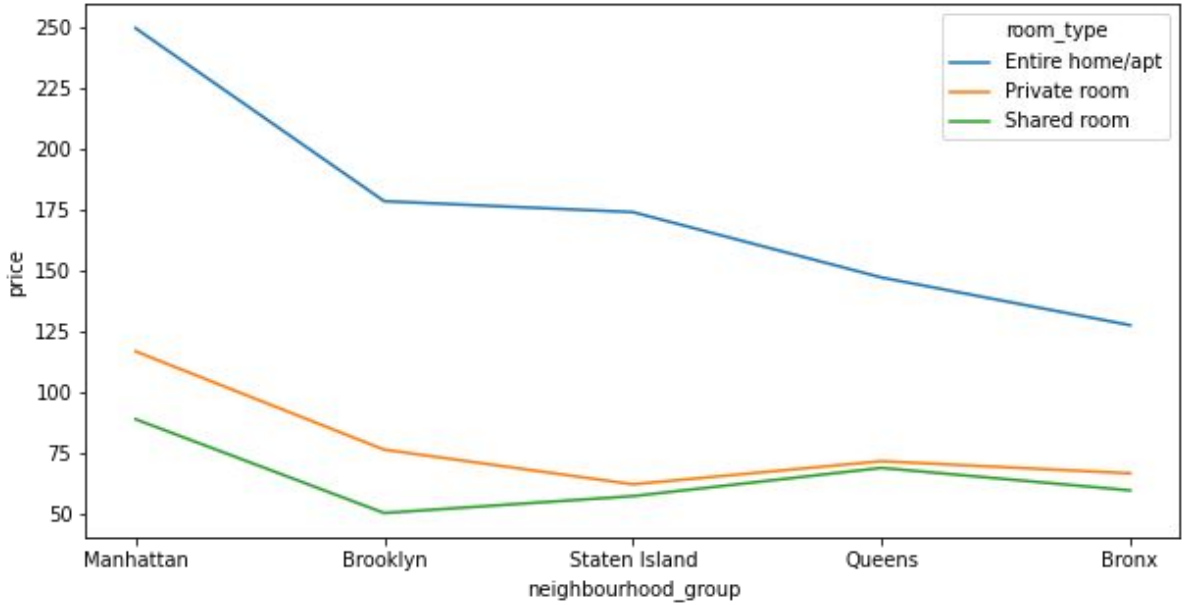
Neighbourhood groups

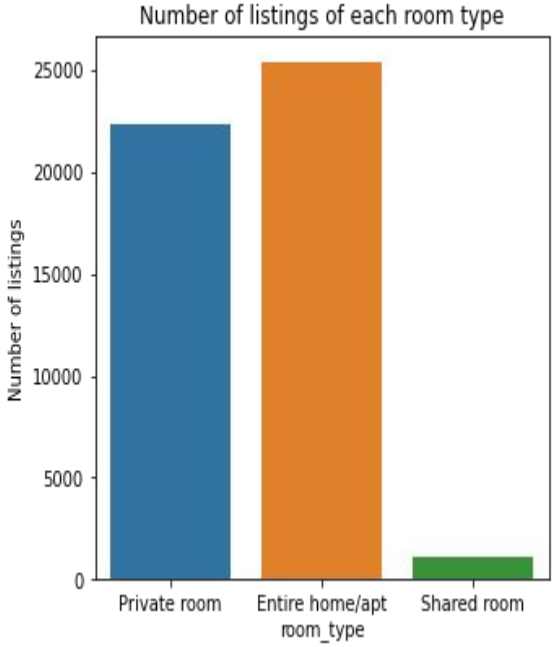




Manhattan has the highest listings.

Price Distribution

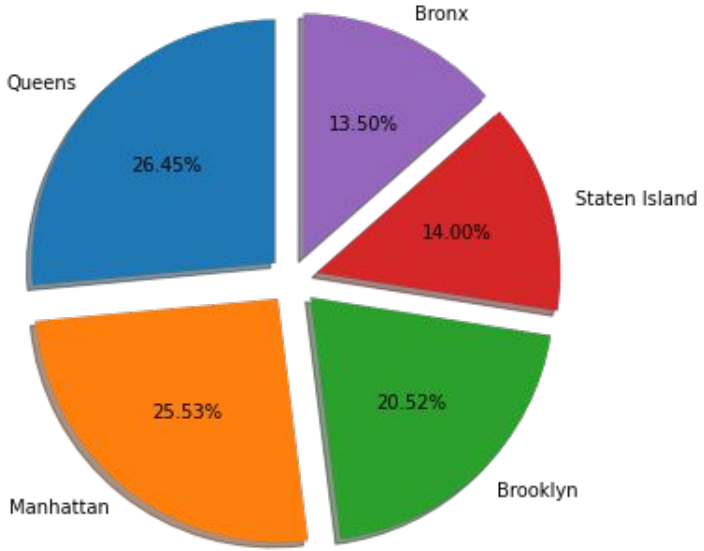




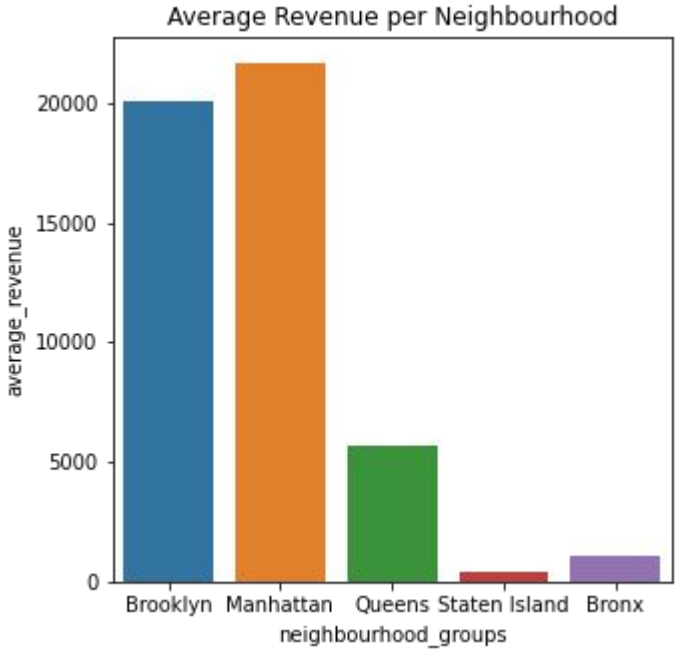
Manhattan has the highest avg price for all room Types. Most private rooms are of medium price, hence being affordable and opted by many. Entire home/apt room type has the highest avg price and most listed type in all neighbourhood groups.

Reviews

Number of reviews in each neighbourhood group

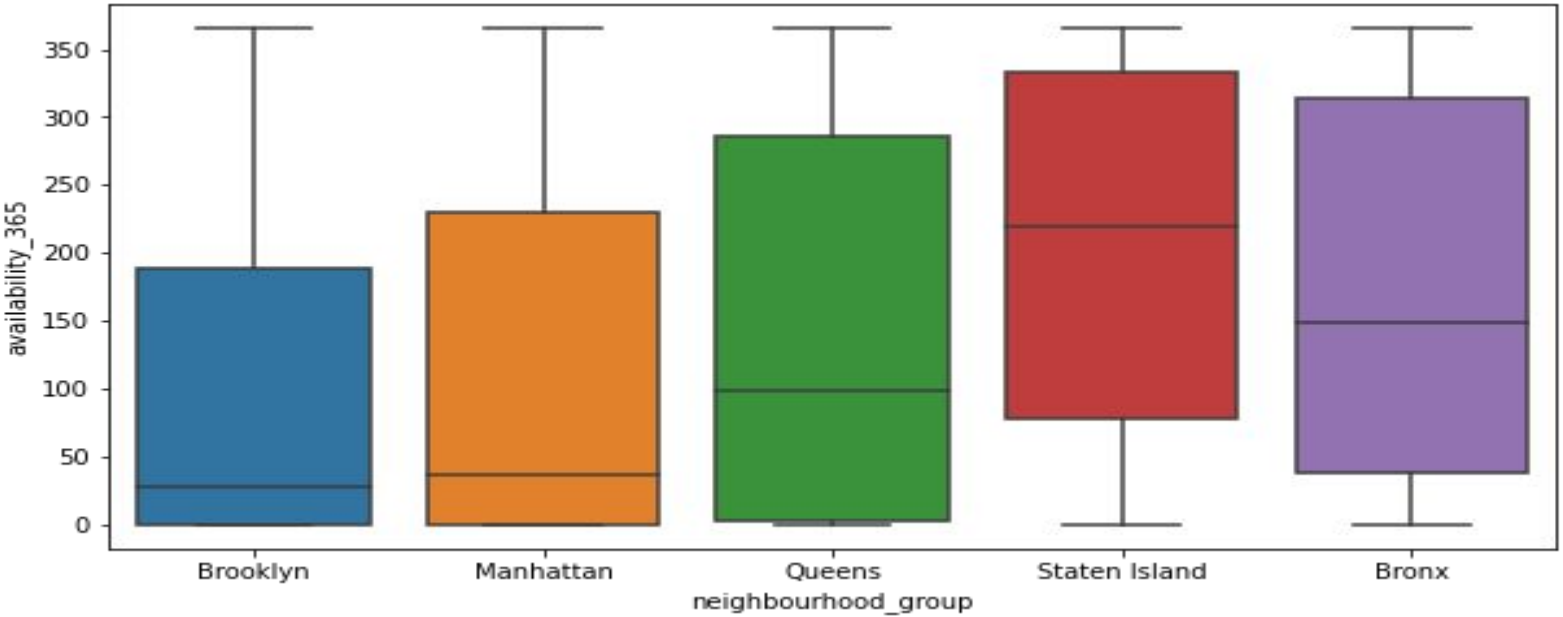


Average revenue



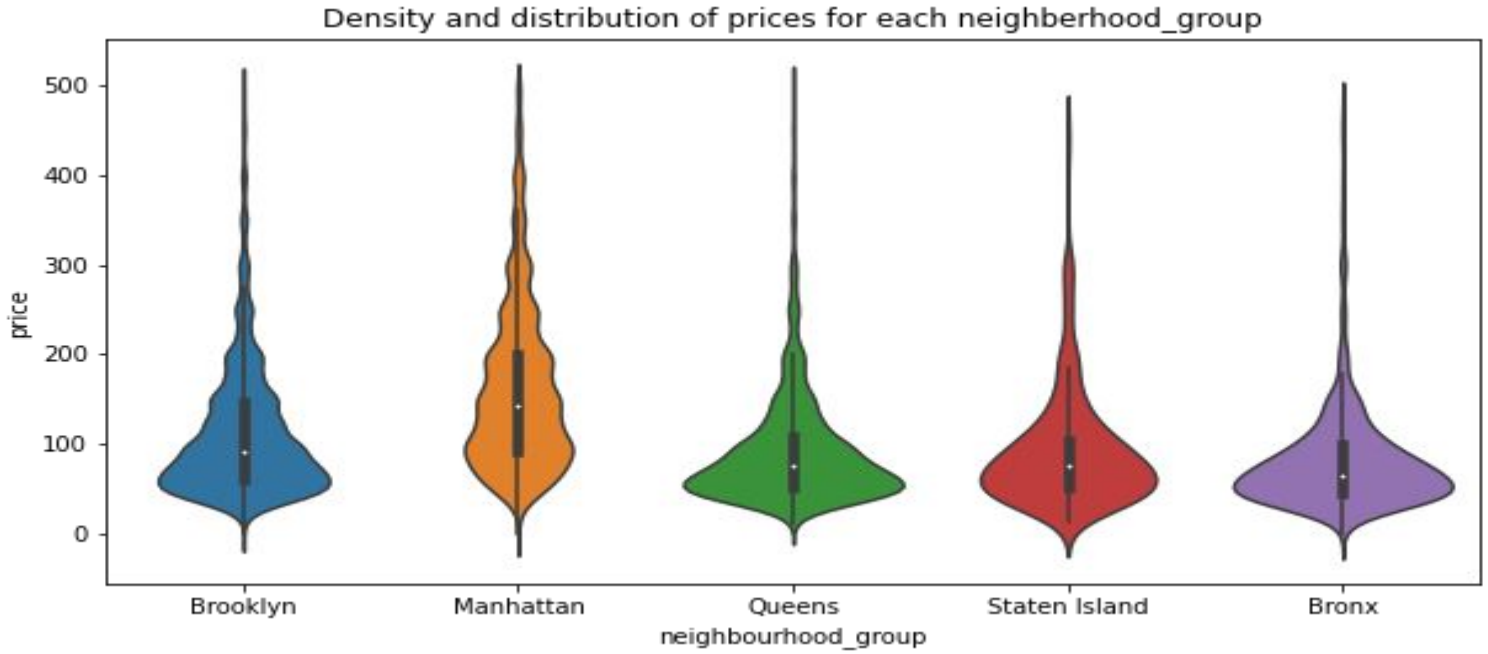
Queens has the highest no reviews

Availability of Rooms



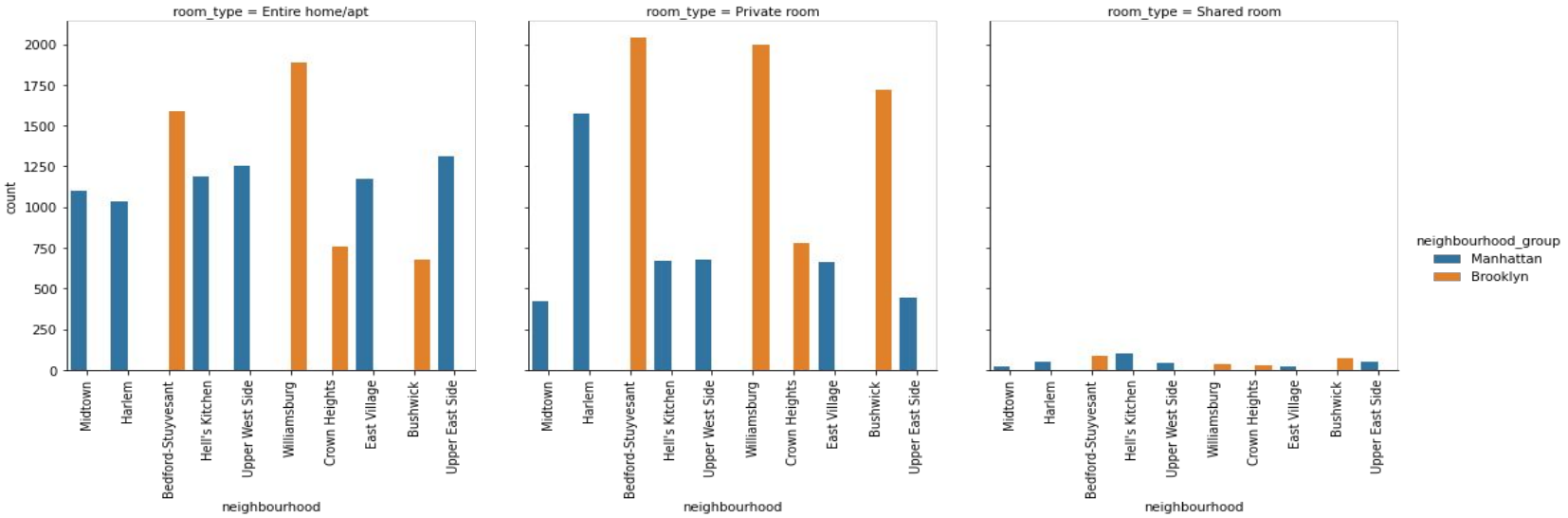
The mean availability of Brooklyn shows that it Is the least available and hence, the busiest

Price Density Distribution



Brooklyn has an avg price of 80$, Manhattan has the highest avg price around 150$

Listings In Top Neighbourhoods



For top 10 neighbourhoods Manhattan and Brooklyn are the busiest destinations with most listings. 'Shared room' type listing is barely available among the 10 most listing-populated neighbourhoods.

Conclusion

The exploratory data analysis for Airbnb dataset has been successfully done and the following inferences have been made from the obtained visualizations and also from the dataset,

- Data cleaning, data preparation is done and correlation of features is checked..
- The problem objectives are met successfully and in-depth analysis has been done in the dataset
- With proper visualizations the problem objectives are explained thoroughly and systematically.

Thank you