

Capstone Project

EDA - Airbnb Bookings Analysis (NYC)

Team Members

Gaurav Rajgor
Shrishti Dixit
Shubham Agarwal



airbnb

New York



About Airbnb Company



- Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Based in San Francisco, California, the platform is accessible via website and mobile app.
- Airbnb does not own any of the listed properties; instead, it profits by receiving commission from each booking. It currently covers more than 100,000 cities and 220 countries worldwide.
- The company was founded in 2008 by Brian Chesky, Nathan Blecharczyk and Joe Gebbia. Airbnb is a shortened version of its original name, AirBedandBreakfast.com.

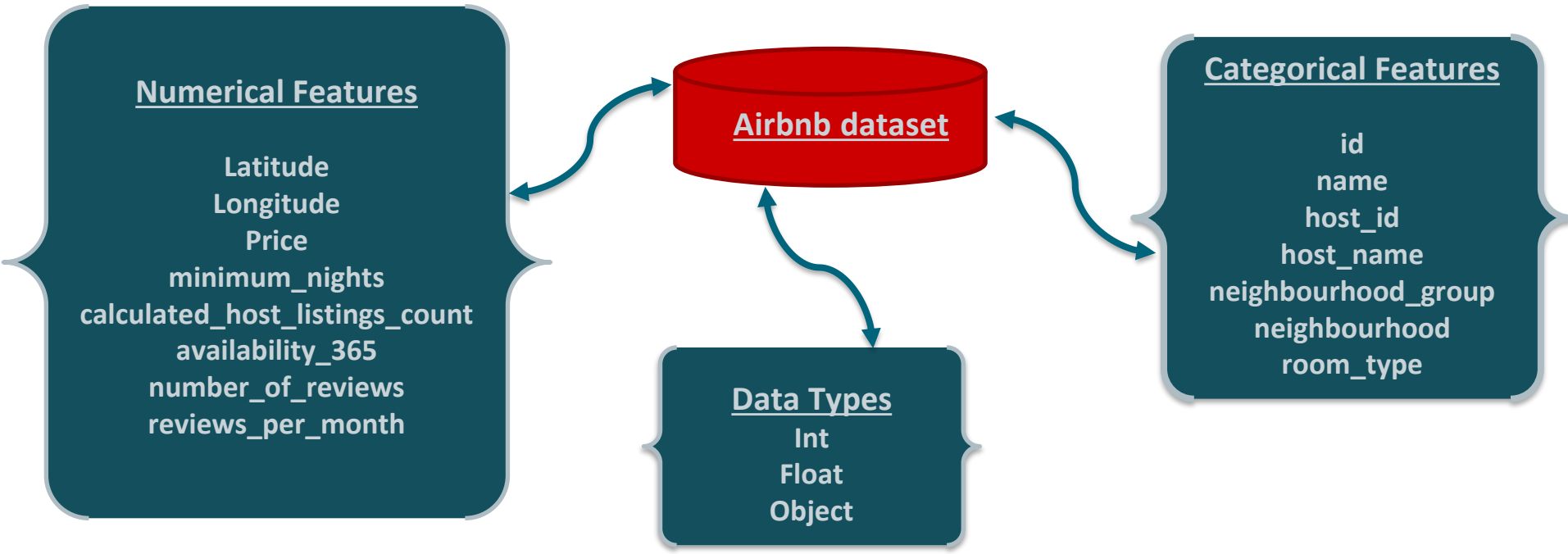
Data summary

- The dataset has 16 different features with more than 49,000 Observations and it is a mix between categorical and numeric values.
- **Id** : unique reference number for each different.
- **Name** : name of different hotels of various neighborhood groups.
- **Host id** : It is id given to specific host and there are 1270 host id available in given dataset.
- **Host name** : name of host hosting different.
- **Neighborhood group** : It represent location, given data set contain 5 different locations.
- **Neighborhood** : It represent specific areas where the listing is located.
- **Room type** : It represent category of room type being listed.

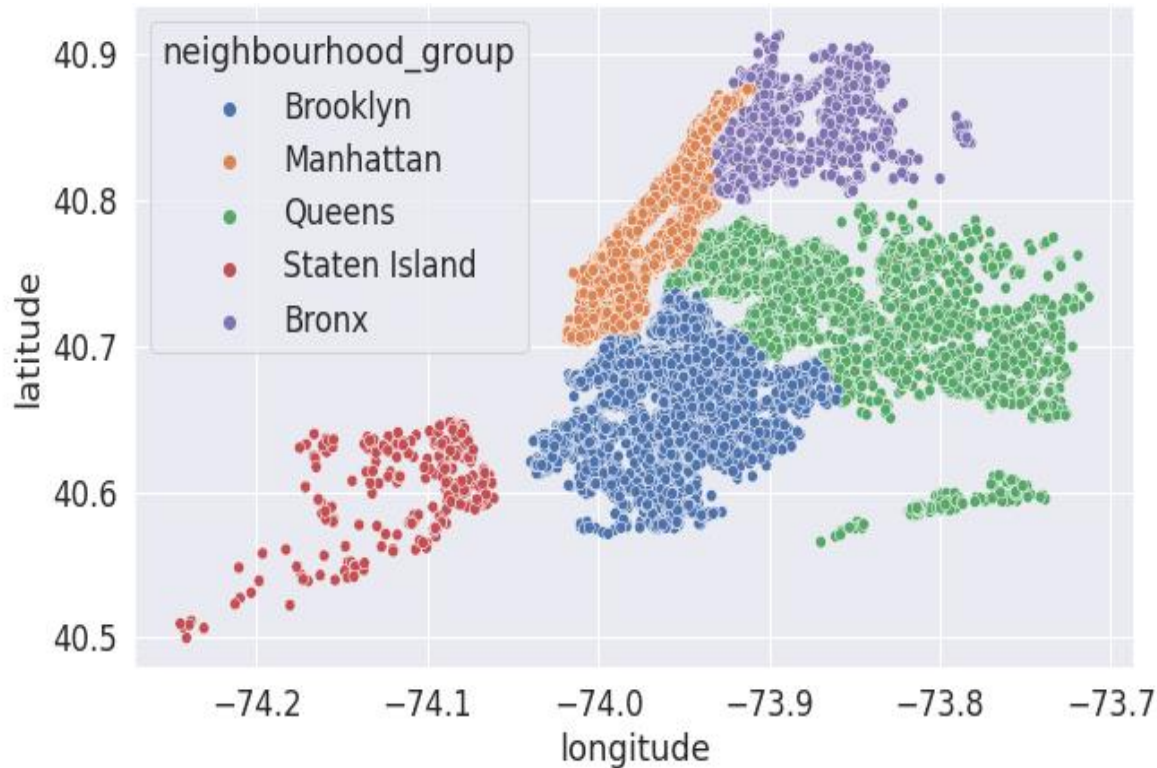
- **Latitude:** The geographical co-ordinates of respective listings.
- **Longitude :** The geographical co-ordinates of respective listings.
- **Minimum nights :** It represents number of nights spend by customer in given listing.
- **Number of reviews :** It represents the number of reviews for listings.
- **Reviews per month :** count of reviews getting per month of a particular hotel.
- **Last review :** date of last review got by a customer to a particular hotel.
- **Availability 365 :** It represents number of days in year for which given property is available for rent.
- **Price :** It represent rate for given room type in given location for one night.
- **Calculated host listings count :** It represents total number of listings made by a specific host. In some cases, the properties are same but some of the other features differ like(Room type).

Distinguishing Features

- The Airbnb dataset contains a mixture of numerical features and categorical features. Below is the distinguished view of the features that we have used in our analysis :-

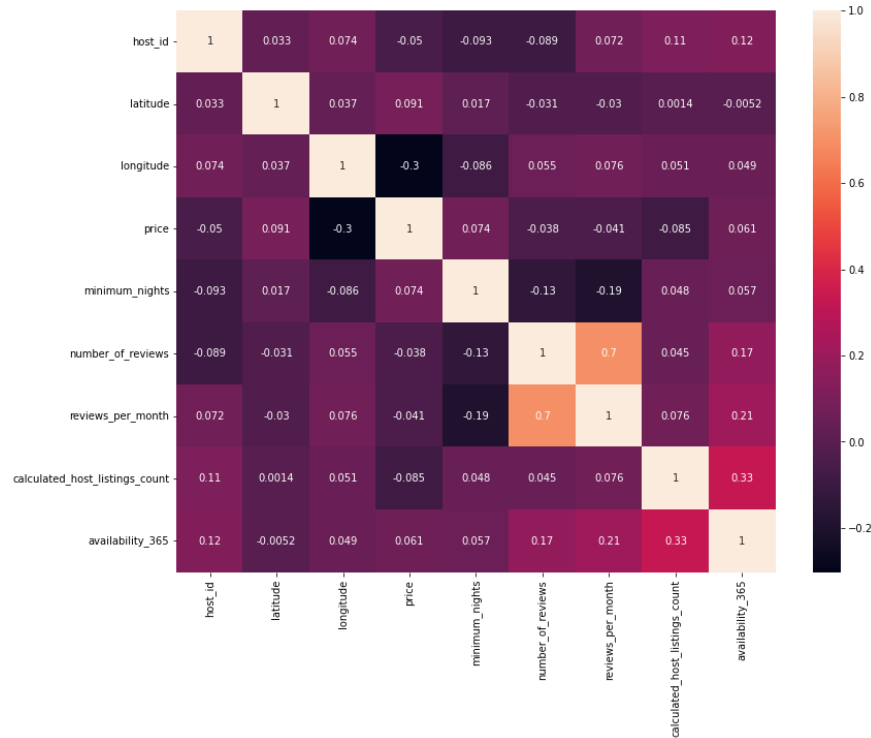


Location of neighborhood groups



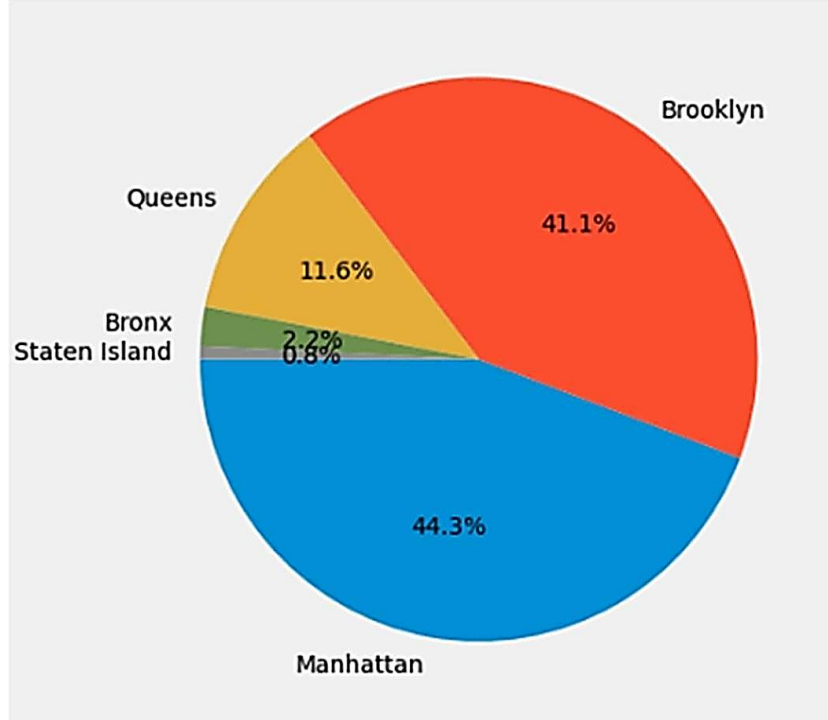
Majority of traffic(number of listings) are observed in **Queens** and **Brooklyn** followed by **Manhattan**.

Correlation between data



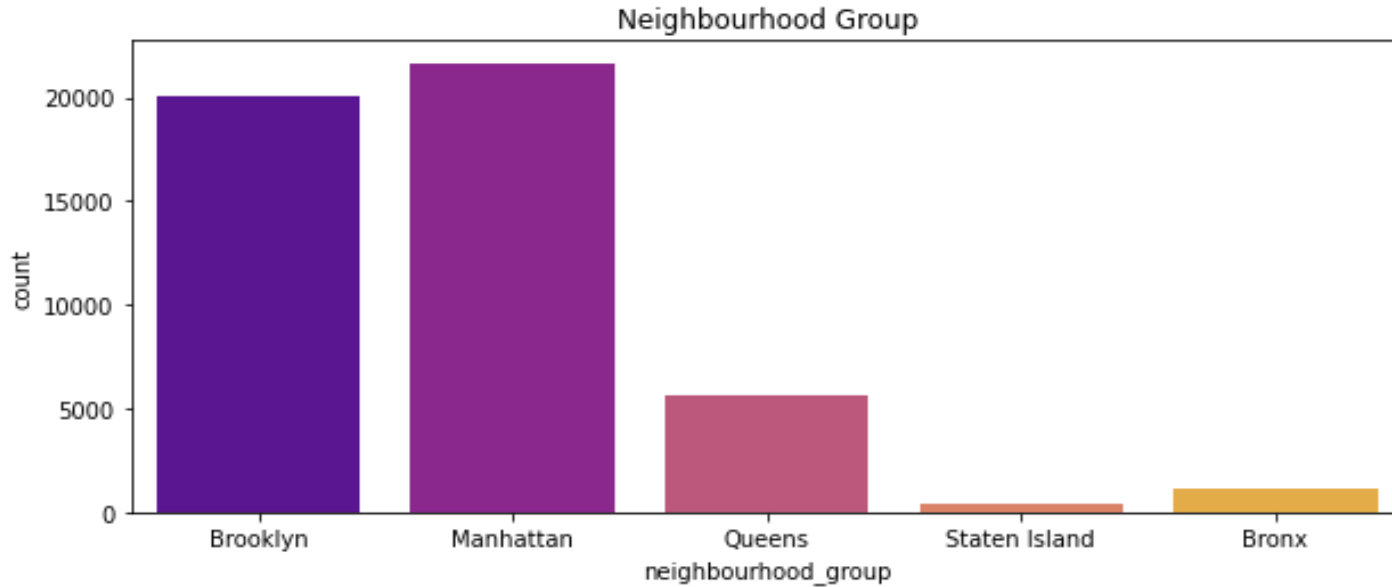
- **No strong correlation** is observed between given different features of dataset.
- **Number_of_reviews** and **availability-365**, **calculated_host_listings** and **availability_365** are **weakly correlated**.

Most Host Listing Areas



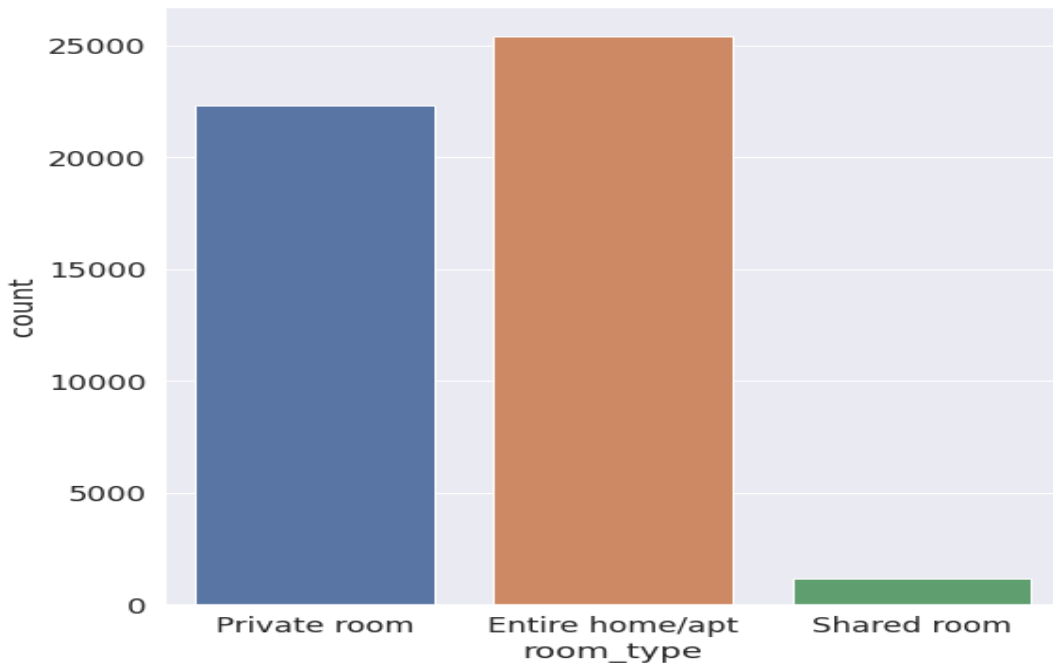
From pie chart it is observed that maximum number of listings are found in **Manhattan(44.3%)**.

Most Host Listing Areas



This Shows that most of Airbnb Listings in **New York** are near **Brooklyn** and **Manhattan**.

Preferred Room types



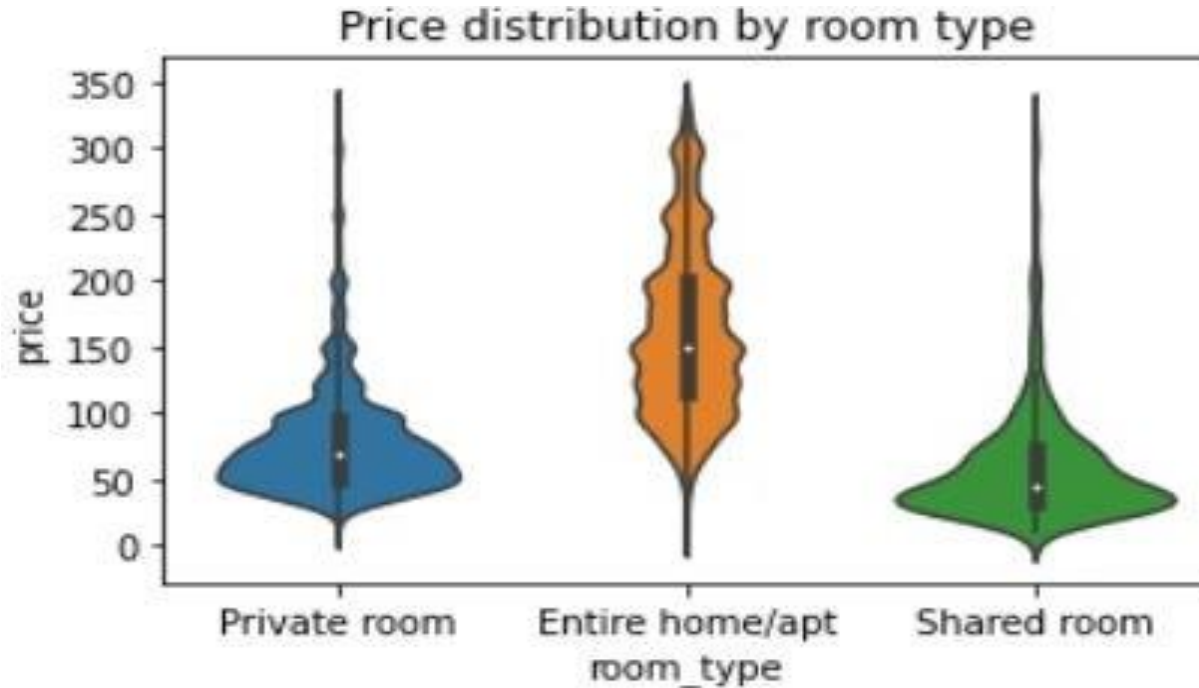
We observe that **Entire home / Apartment** has highest share, followed by **Private rooms**, and least preferred is **Shared rooms**.

Price distribution as per Neighborhood Group



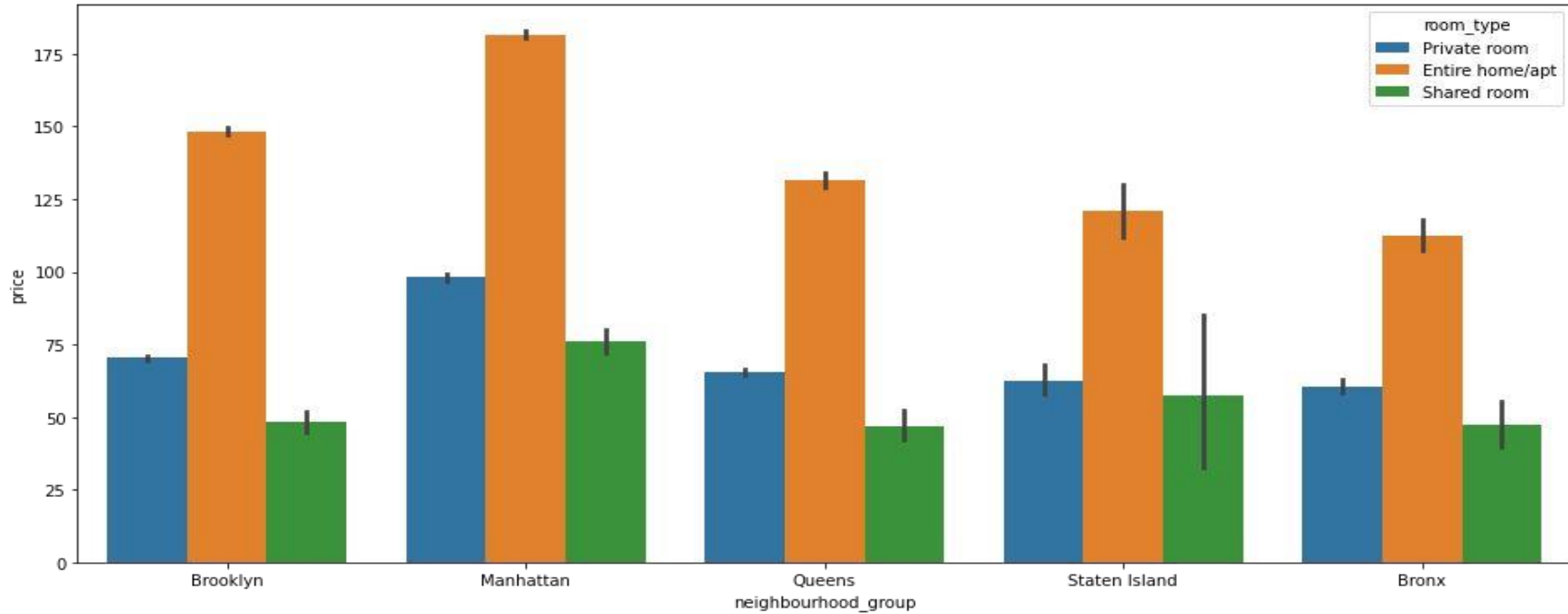
It's observed that **Manhattan** is most expensive location followed by **Brooklyn**.

Price distribution as per Room type



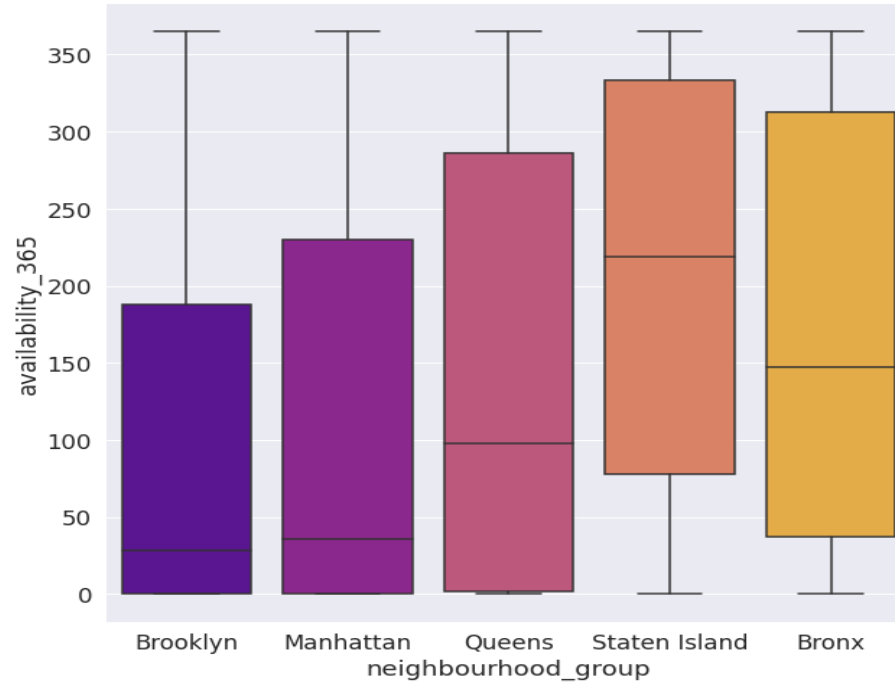
Private rooms and **Shared rooms** price range are more centered around their mean.

Average Pricing For Room Types



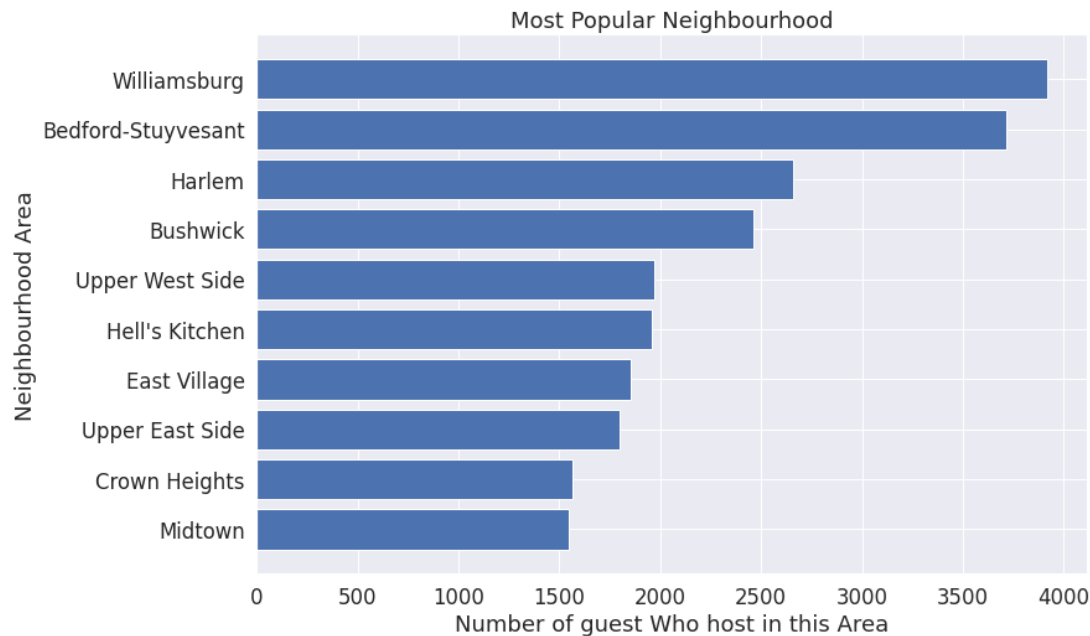
- **Entire Home/apt** is the most expensive property in **Manhattan** and **Brooklyn** among all the property types and Neighborhood Groups.
- Entire home/apt have **\$ 181.644693** avg pricing in **Manhattan** while same property type stands for **\$ 148.218656** in **Brooklyn**.

Availability as per Neighborhood Group



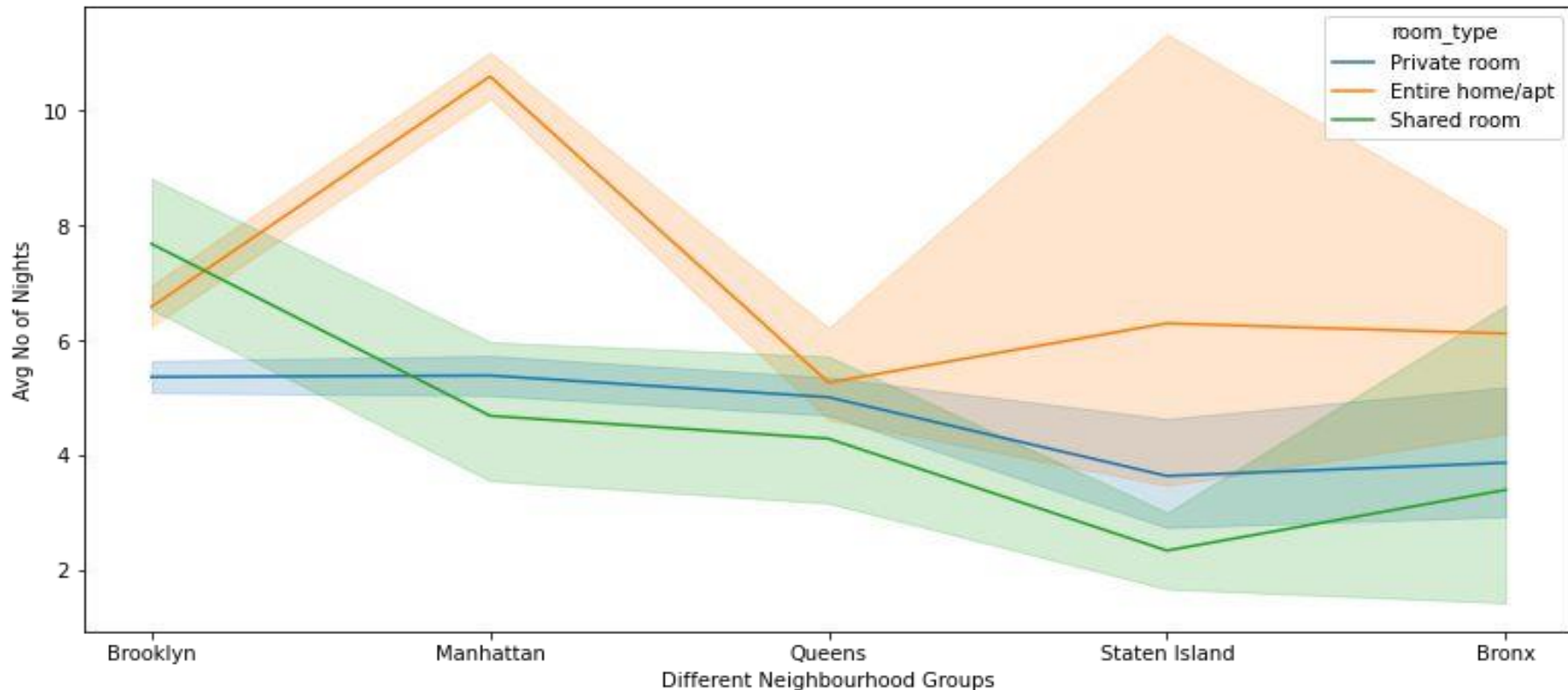
The graph shows relationship between **Availability of room** and **Neighborhood group**.

Popular Neighborhood



The graph shows relationship Top 10 popular **Neighborhood** areas with maximum number of hosts.

Average Property Booking Duration



- Average booking duration for **Entire home/apt in Manhattan** is higher than any other property type in any Neighborhood Group.

Recommendation

1. Pricing and Booking rate is good for **Entire home/apt** in **Manhattan** and **Brooklyn** So, Increasing the properties in these 2 near locations will be beneficial.
2. Increase the property listing in the price range **50 \$ to 100 \$**.
3. Company should give priority to **Entire Home/Apt** and **Private Room** as most of the customers prefer these 2 property type.
4. **Minimum number of nights** should be decreased (especially for Entire Home/Apt).
5. Reviews plays an important role for the host to get more bookings (and revenue), so company should do something so that more customers can review the properties.

Project & Dashboard links



- **GitHub Repository:**
<https://github.com/GauravRajgor/Capstone-Project-EDA-AirBnB>
- **Google Collab Notebook :**
https://colab.research.google.com/drive/1AHWchR-k_fSAeVUv2RgvPUkTB50FJXc?usp=sharing
- **Tableau Dashboard :**
https://public.tableau.com/views/AirBnbNYC2019-Dashboard/AirBnB?:language=en-US&:display_count=n&:origin=viz_share_link

Thank You!

Reach out me at:

- LinkedIn: www.linkedin.com/in/gaurav-rajgor
- GitHub Id : <https://github.com/GauravRajgor>