Capstone Project - 1

Airbnb Bookings Analysis

Team Members

Rupesh Sharma

Namandeep Singh Bisht

Contents

Introduction

- Airbnb is a platform that connects people who wish to rent out their rooms or apartments with people who are looking for lodging in specific areas.

- The objective of our project is to extract information on traffic and costs in different areas using Airbnb's NYC data to help businesses make better judgements that will improve profitability, market expansion and customer experience.

About the dataset

The Airbnb dataset is about listing activities in New York City, and it helps to perform data-driven exploration and acquire bits of knowledge that will aid in recognising patterns and increasing Airbnb's market value and income.

It consists of 48,895 rows and 16 columns.

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	reviews_per_month	calculated_host_listings_count	availability_365
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	9	2018-10-19	0.21	6	365
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	2019-05-21	0.38	2	355
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	0	NaN	NaN	1	365
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	270	2019-07-05	4.64	1	194
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	9	2018-11-19	0.10	1	0

Data Dictionary

Column	Data Type	Description
id	Integer	Unique id of each property
name	Object	Name of the property
host_id	Integer	Id of the host (unique for each host)
host_name	Object	Name of the host(two hosts may have same name)
neigbourhood_group	Object	Consists of five different regions of New York
neighbourhood	Object	Consists of sub-regions in each neighbourhood group
latitude	Float	Consists of latitude coordinates of the location
longitude	Float	Contains the longitude coordinates of the location

Data Dictionary

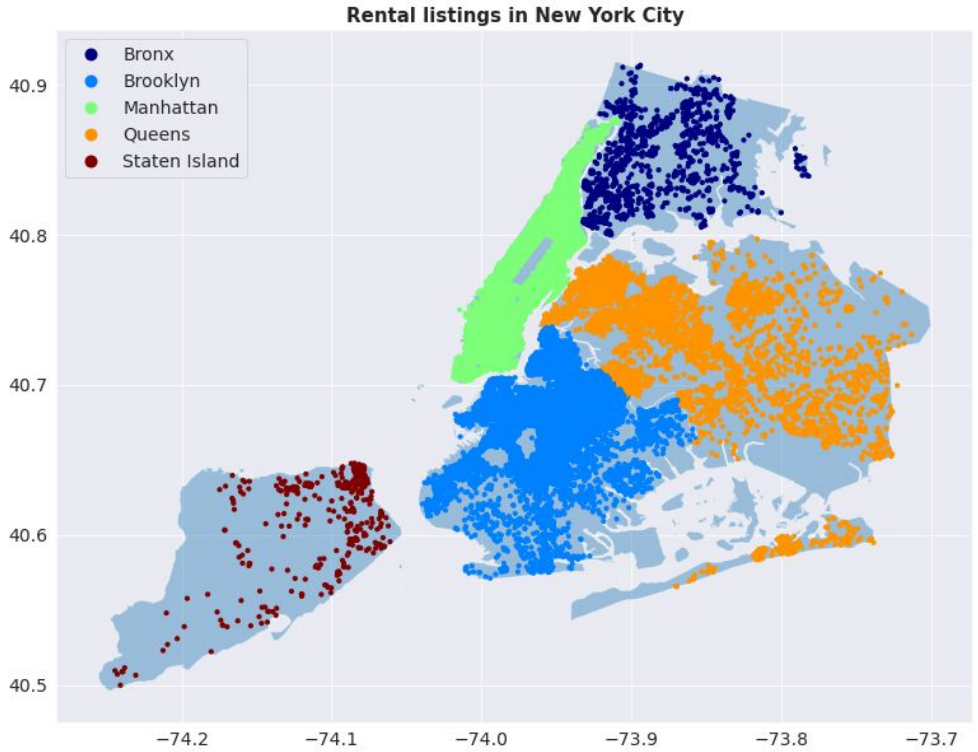
Column	Data Type	Description
room_type	Object	Types of property listed by host
price	Integer	Rent price of the listing
minimum_nights	Integer	Minimum number of nights to be booked for listing
number_of_reviews	Integer	Number of reviews of each property
last_review	Date	Date of last review of the property
reviews_per_month	Float	Number of reviews received per month for each property
calculated_host_listings_count	Integer	Total number of listings for specific host
availability_365	Integer	Number of days each property is available out of 365 days

Data Cleaning

- The dataset contains null values in name, host_name, reviews_per_month and last_review columns.
- The rows with null values in the name and host_name columns were eliminated.
- The null values in the reviews_per _month columns were replaced with 0.
- The month and year were extracted from the last_review column while ignoring the null values for our analysis. Later we dropped last_review column.
- With the support of domain knowledge, we increased the minimum listing price from $0 to $10.

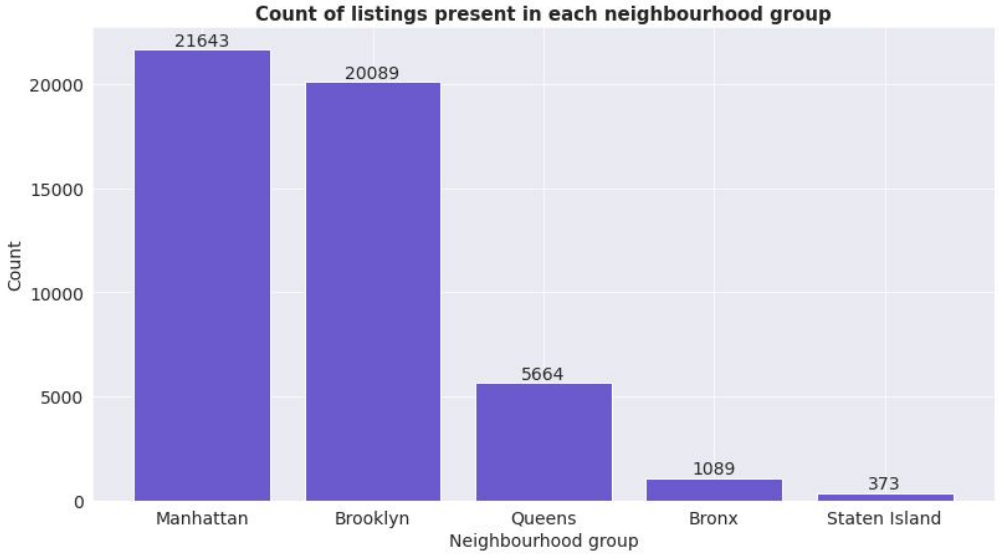
Data Analysis & Visualisation

- The right-hand visualisation depicts the marked location of the listings on a map of New York City.



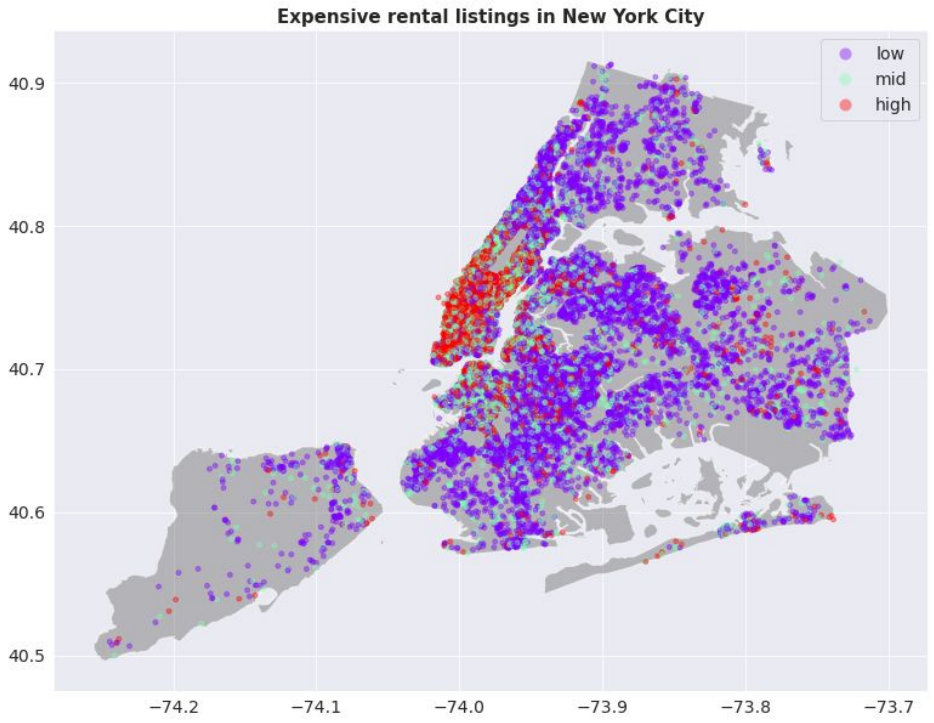
Data Analysis & Visualisation

- The right-hand countplot shows the number of listings present in each neighbourhood group.
- The plot indicates that Manhattan and Brooklyn have the most listed residences.



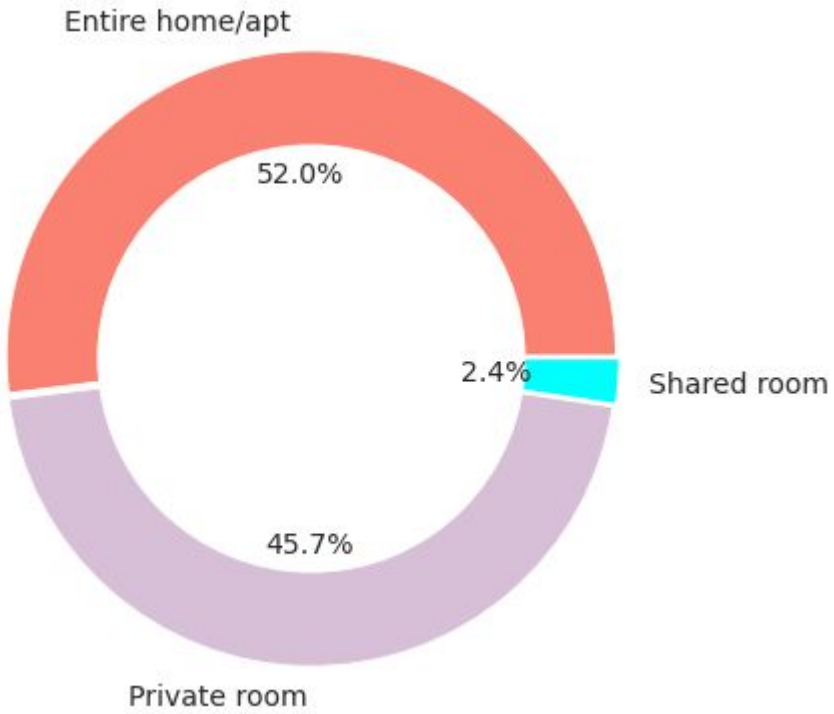
Data Analysis & Visualisation

- The map on the right depicts the most costly listings in order of cost.
- We can deduce from this visualisation that the most costly postings are in Manhattan.



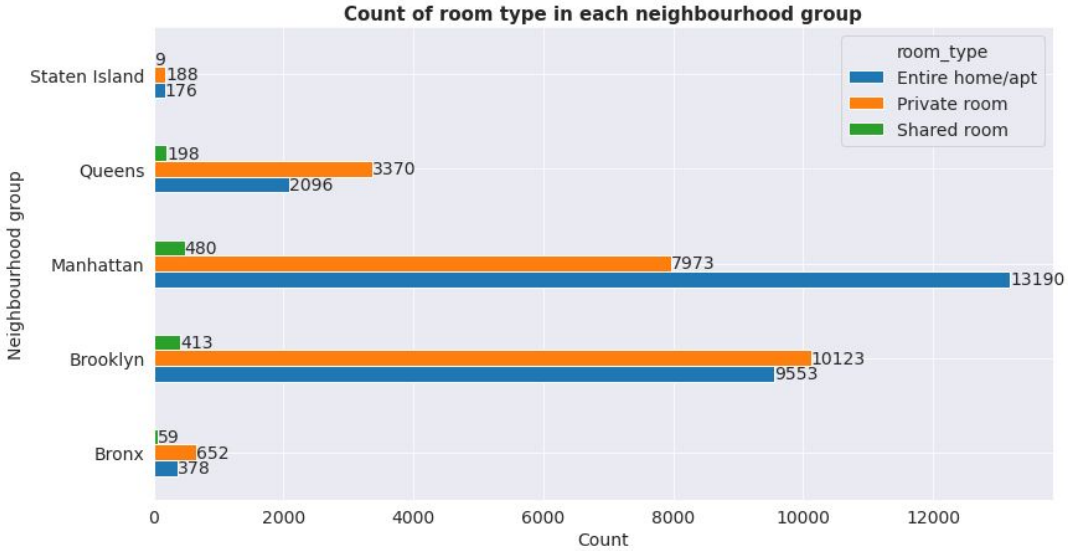
Data Analysis & Visualisation

- The donut chart on the right represents the type of room listed by the hosts.
- Most of the Airbnb hosts rent Entire home/apartment and Private room.



Data Analysis & Visualisation

- The barplot on the right shows the number of different types of rooms in each neighbourhood.
- We may deduce from the bar chart that Manhattan accounts for half of the listings renting out full houses and apartments.
- Other neighbourhood group prefer private rooms.

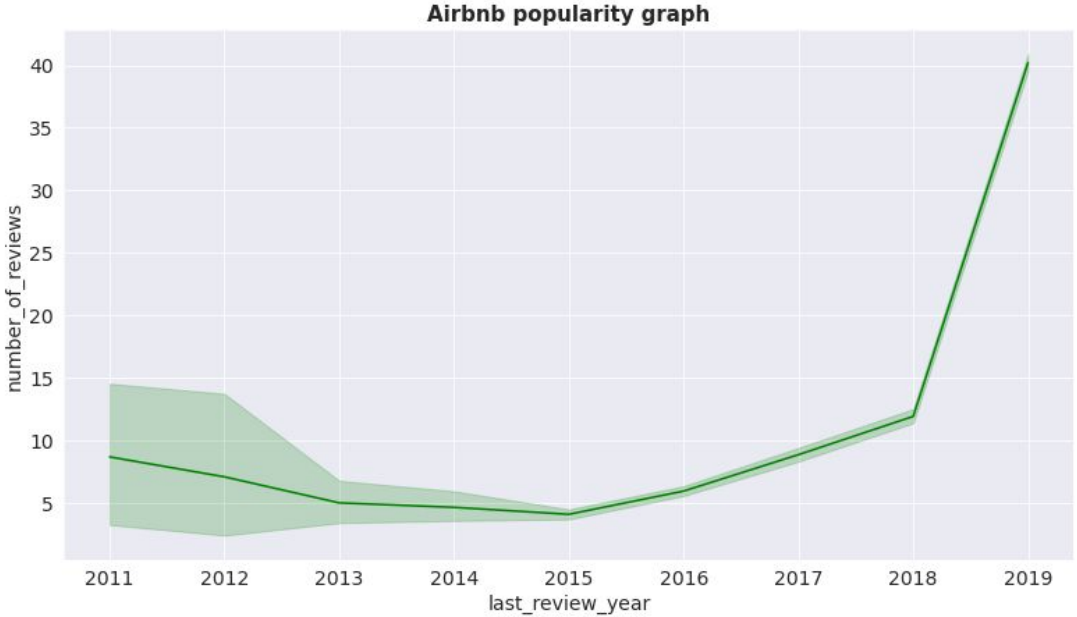


Data Analysis & Visualisation

- The stacked bar chart depicts the median cost of different accommodation categories in each suburb.
- We may conclude from the stacked bar chart that entire houses/apts are pricey in each neighbourhood group.
- In Brooklyn and Queens, the prices of private and shared accommodations are comparable.
- For each accommodation category, the Bronx and Staten Island have similar price ranges.

Data Analysis & Visualisation

- The plot on the right shows the popularity of Airbnb over the last few years.
- We can see from the visualisation that since 2015, Airbnb has experienced a surge in popularity.
- Our analysis is based on the assumption that the more reviews a Airbnb listings have received in recent years, the more popular it becomes.

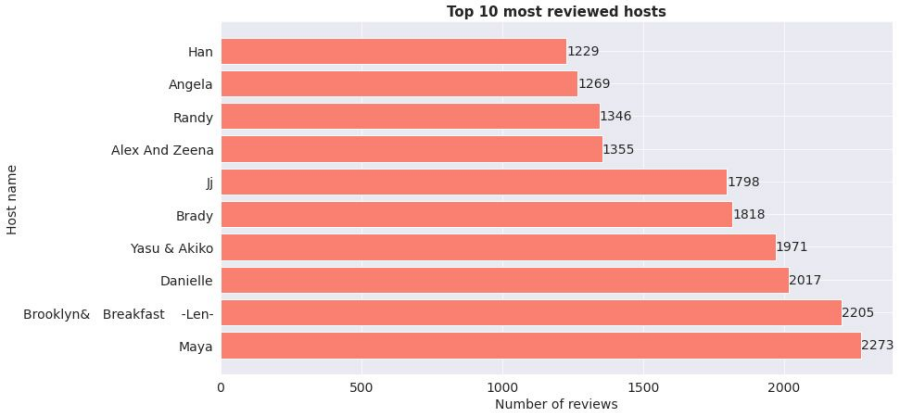


Data Analysis & Visualisation

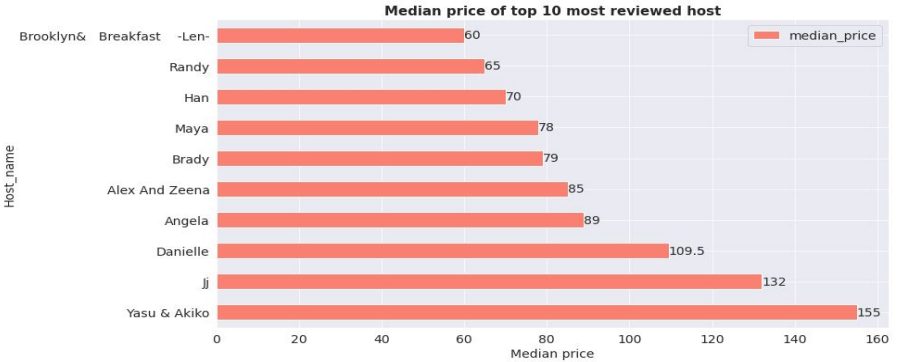
- The right-hand plot depicts the busiest months for Airbnb bookings.
- According to our analysis, the months of May, June, and July are the busiest for Airbnb.
- In this scenario, we assume that guests leave reviews while they are enjoying their stay or after they have left.

Data Analysis & Visualisation

- The right-hand plot depicts the most reviewed hosts.

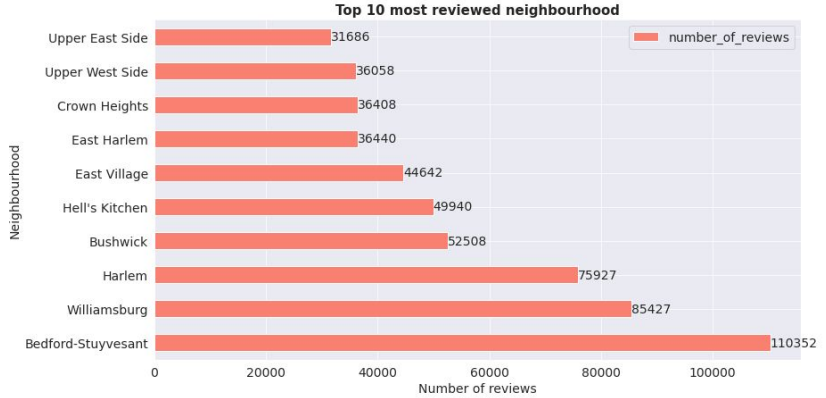


- The right-hand plot depicts the median price of most reviewed hosts.

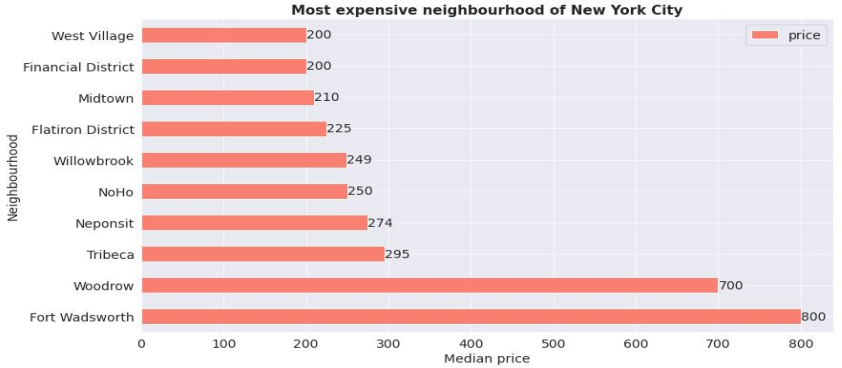


Data Analysis & Visualisation

- The right-hand plot depicts the most reviewed neighbourhood.

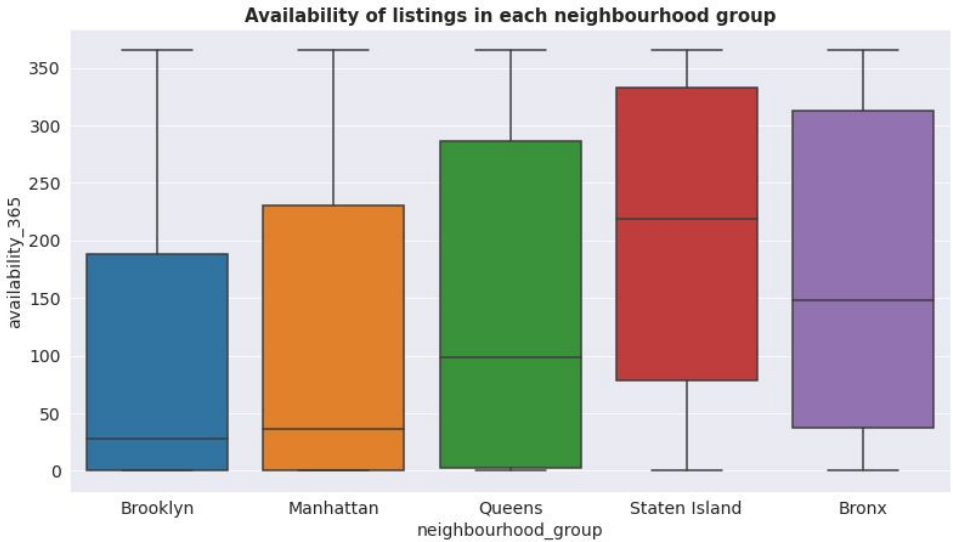


- The right-hand plot depicts the most expensive neighbourhood.



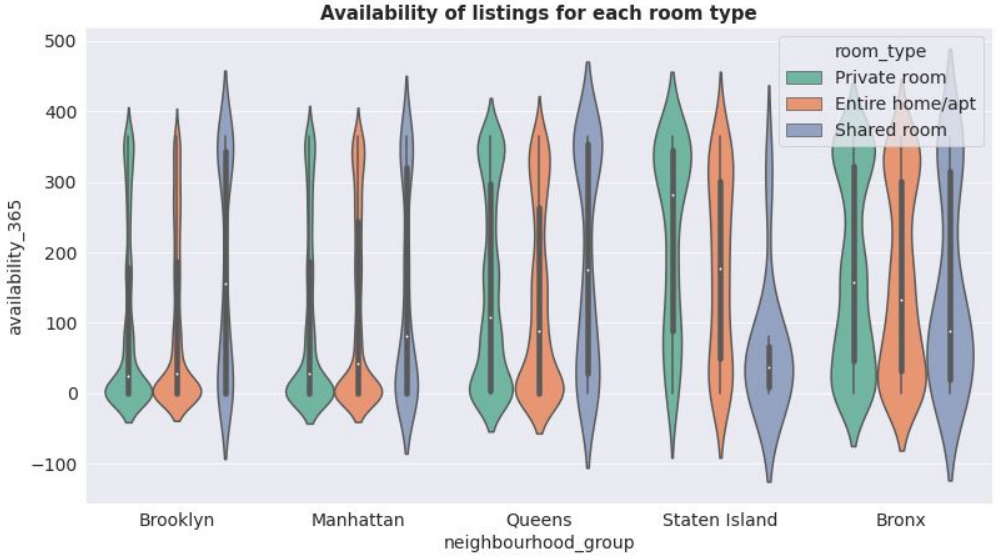
Data Analysis & Visualisation

- The boxplot on the right, depicts the availability of listings in each neighbourhood group.
- The plot indicates there are only a handful open listings in Brooklyn and Manhattan due to high demand.



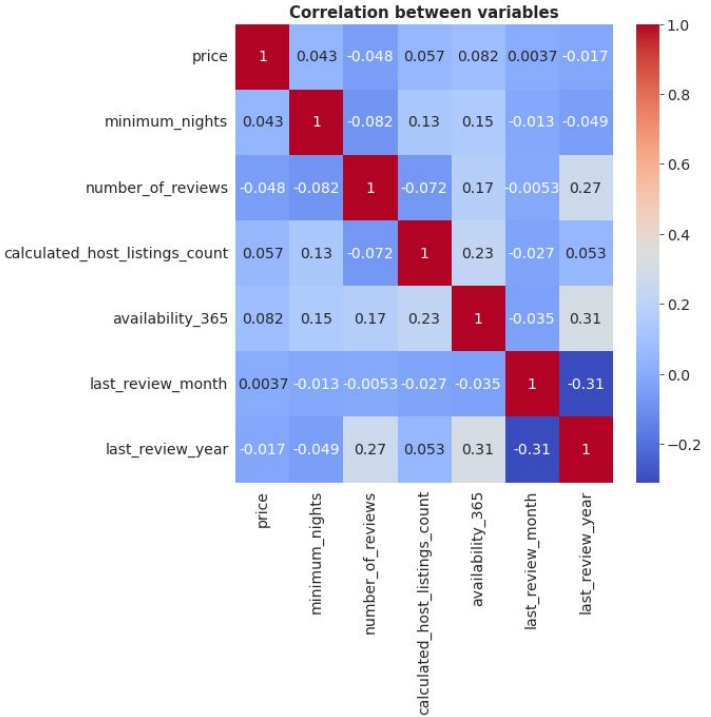
Data Analysis & Visualisation

- The violinplot depicts the number of listings available for each room category in each borough.
- In Staten Island and the Bronx, there is a higher demand for shared rooms.



Data Analysis & Visualisation

- A heatmap depicts the correlation between the features.
- There is no strong correlation between variables.



Challenges Faced

- Interpretation of all features and selection of those that will be useful in our analysis.
- Taking care of NaN values for the last review feature.
- Filtering out important trends and visualisations from those that aren't.
- Choosing the right sort of visualisation to best communicate data.

Conclusion

- Manhattan and Brooklyn are the most crowded boroughs of New York City as compared to Staten Island, Queens and Bronx so increasing the number of hosts in Manhattan and Brooklyn will help in generating more revenue.
- Manhattan's listings are most expensive so offering special discounts in the peak season might help in attracting more customers.
- Optimising the price of listings in other boroughs using our analysis also helps in increasing the business.
- Customers prefer to stay in Entire home/apt or private room types as compared to shared room types so Airbnb should strategize on making more entire and private rooms available.

Conclusion

- Minimum stay for Airbnb customers in New York city is 2 to 3 days, trying experimenting on reducing it to 1 day might help.
- Most bookings are made in the months of May, June and July so to counter the demands of customers special arrangements could be made.
- By pushing the popular (most reviewed) hosts listings forward Airbnb can increase its user's experience.
- Ask for suggestions from most reviewed hosts and as well as customers to enhance their experience.
- Advertising more about the famous sites in Queens, Bronx and Staten Island to tourists can boost revenue in these regions as well.

Thank You