



Insights from Airbnb, NYC dataset

Submitted by : Rithesh Kamath



Agenda



- Objective
- Background
- Data Preparation & Cleaning
- Data Analysis
- Appendix:
 - Data Assumptions
 - Data Methodology



Objective

- 
- Understand key insights from Airbnb NYC business from the provided dataset & Improve strategies to revive the business as the restrictions have started lifting.
 - To prepare for the next best steps that Airbnb needs to take as a business



Background



- For the past few months, Airbnb has seen a major decline in revenue. Now that the restrictions have started lifting and people have started to travel more, Airbnb wants to make sure that it is fully prepared for this change.
- The different leaders at Airbnb want to understand some important insights based on various attributes in the dataset so as to increase the revenue

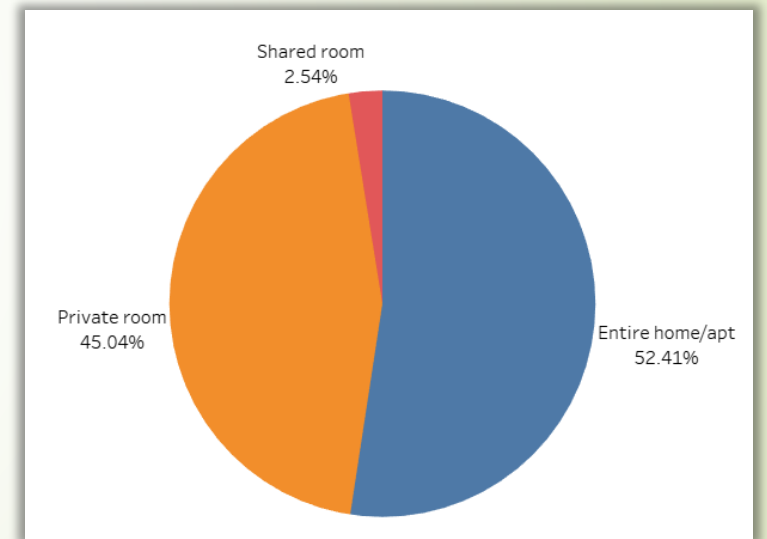


Data Preparation and Cleaning

- The data was cleaned from any missing values and outliers
- The data was analyzed using simple statistical analysis of the data
- The visualizations were derived using the Excel & Tableau tools to understand key inferences from the analysis

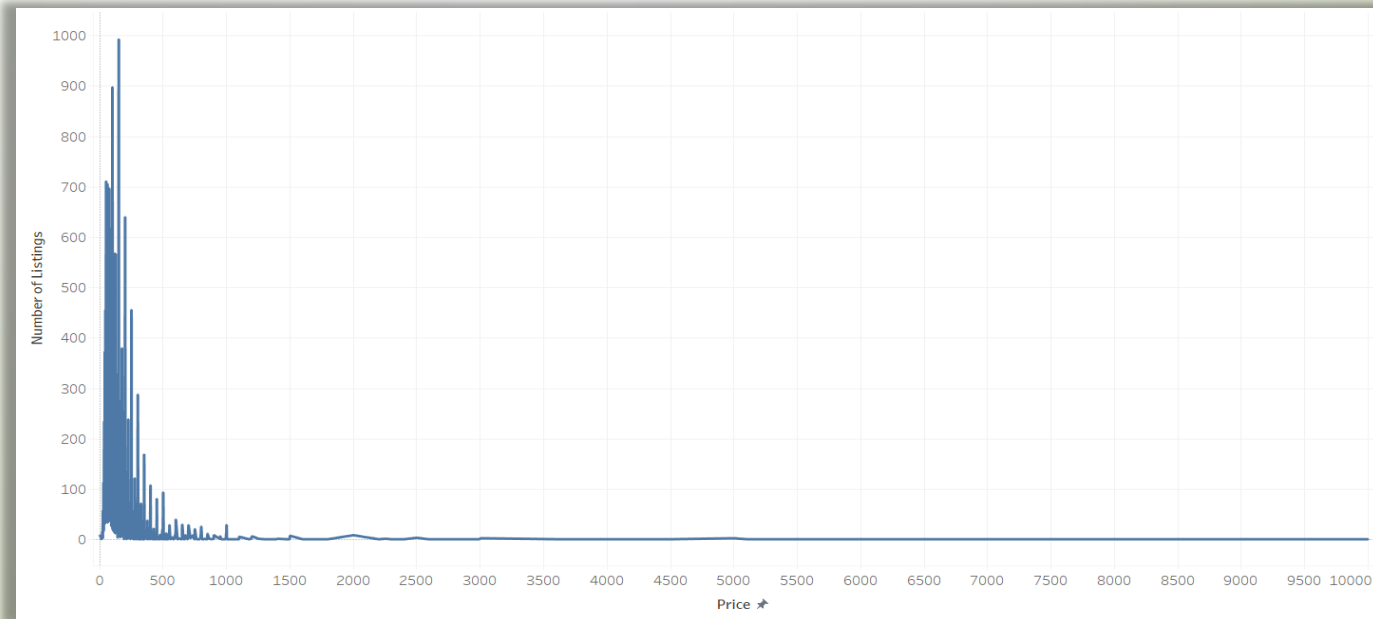
Customer preference of the three Room types

- “Entire home/apt” and “Private room” remain to be preferred over “shared rooms” by Airbnb hosts
- “Entire home/apt” and “Private room” together account for around 97.45% of the listed properties
- “Shared rooms” only account for around 2.54% of the total listed properties



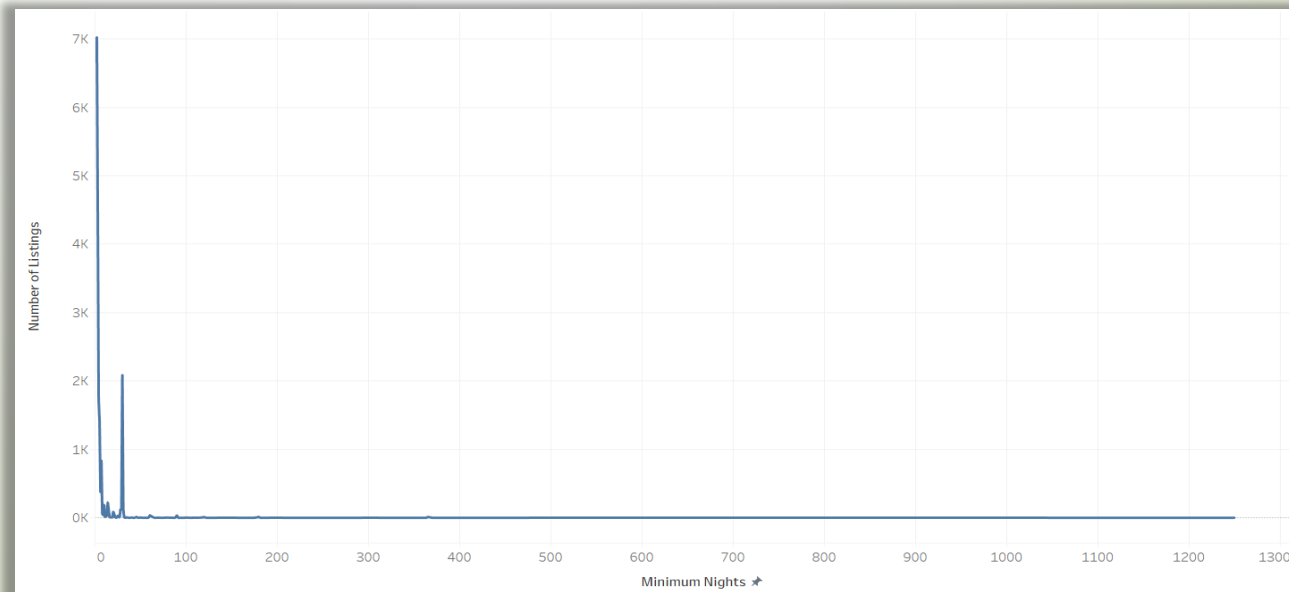
Preference for Price

- Right skewed distribution of price vs listings
- Most of the listings are average to low in price and only a few listings are highly priced
- Some listings are unreasonably expensive which should be treated as outliers & should check as to how is it possible that someone is allowed to list a property so expensive on the platform



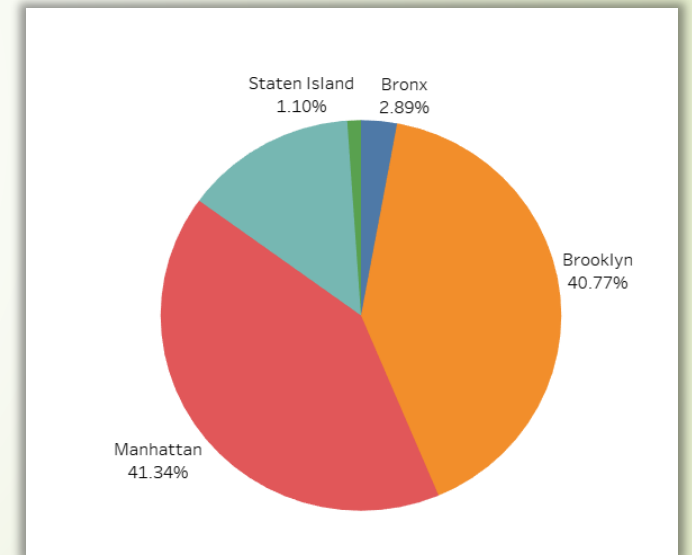
Preference for Minimum Nights

- Right skewed distribution of Minimum nights vs listings
- Most of the listings have average to low number of minimum nights & very few properties have a very high number of minimum nights that needs to be booked.
- A particularly interesting observations is the spike at 30 minimum nights which means that there is a huge latent demand for properties where people prefer to stay for 30 days or nights



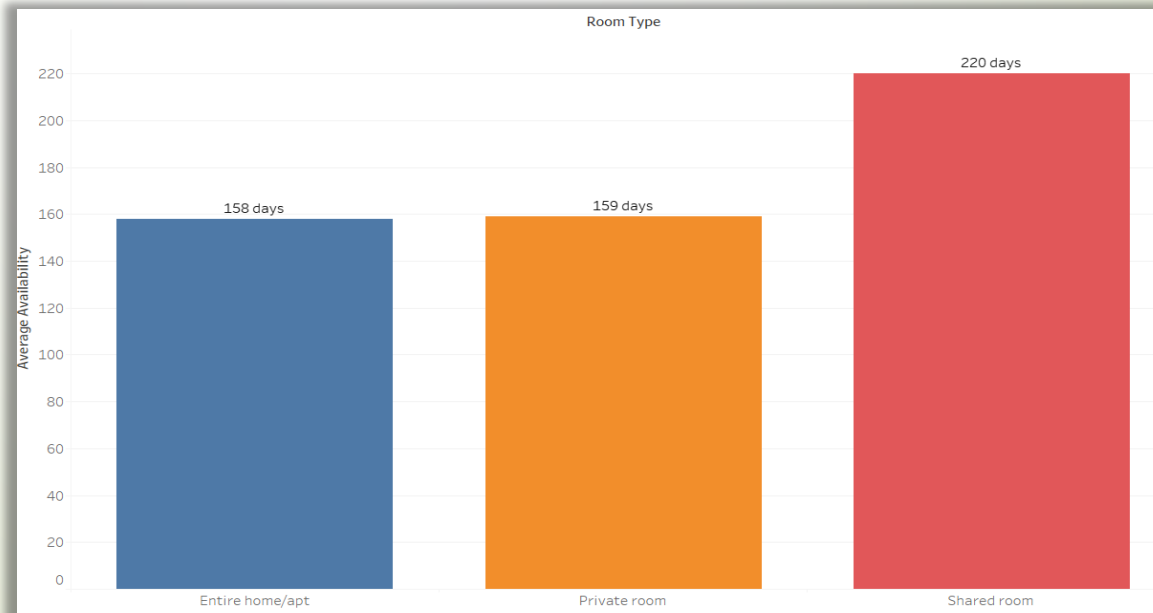
Properties in NYC areas

- Majority of the properties are located in “Manhattan” and “Brooklyn” neighbourhoods
- “Manhattan” and “Brooklyn” together account for around 82.11% of the listed properties
- The least number of properties are located in “Staten Island” which account for just 1.10% of the listed properties



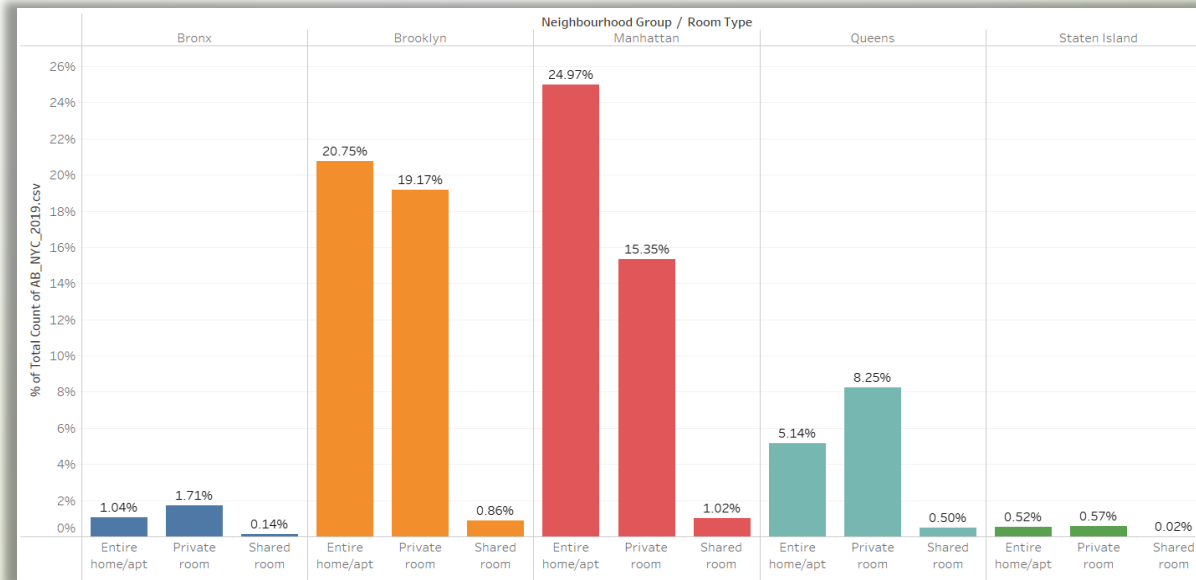
Availability of Room types in a year

- “Shared Rooms” on an average are available for the maximum period in a year
- “Shared Rooms” are available on an average for around 220 days in a year
- “Private Rooms” and “Entire home/apartment” on an average are available for only around half a year



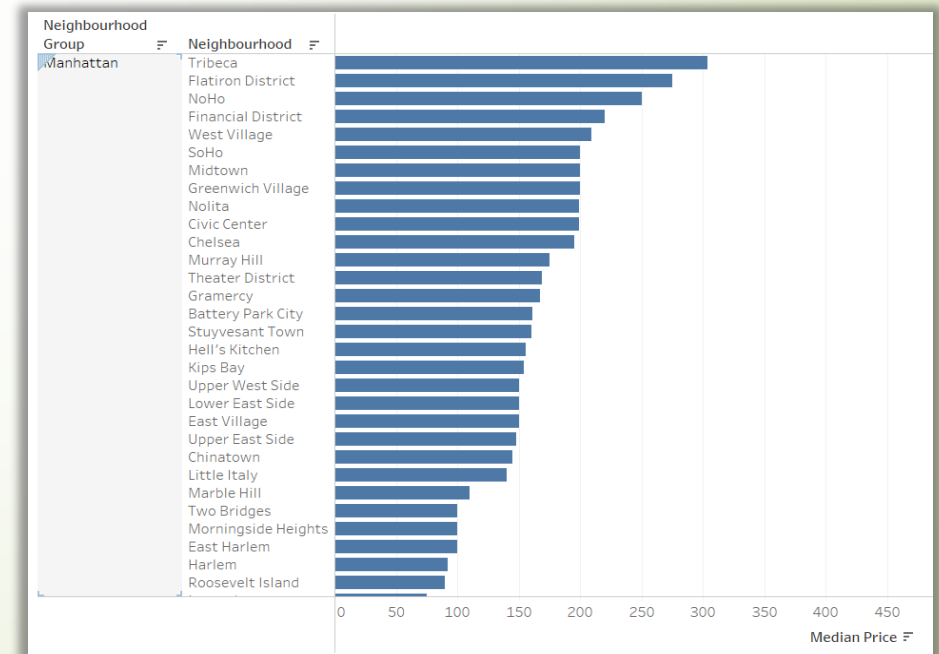
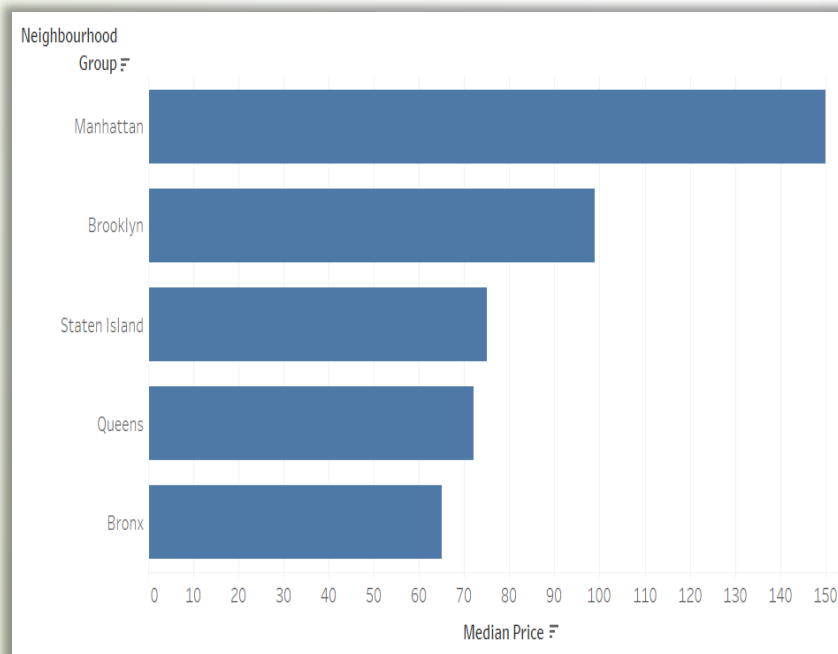
Customer preference of properties in NYC Areas

- Majority of the listings are from “Manhattan” & “Brooklyn” with the Room types being “Entire Home/apt” & “Private Room” which account for almost 80.24% of the listed properties.
- Very small percentage of users go for “Shared room”



Price of properties in NYC Areas

- “Manhattan” has the highest average price which means expensive real estate overall.
- “Tribeca” in Manhattan seems to be the poshest neighbourhood which is at a granular level of understanding





Appendix – Data Assumptions

- Number of bookings per listing was not provided in the dataset, so we have assumed that “number_of_reviews” columns corresponds to number of bookings. The data was analyzed using simple statistical analysis of the data
- Null Values have been assumed to have not much of an impact on our analysis, hence null values were dropped from the columns - last_review, number_of_reviews and availability_365



Appendix – Data Methodology

- A combination of excel & tableau has been used to get insights for our data analysis & visualization.
- Since there were some extreme values in column “Price”, we have used Median (average) for analysis purpose & not mean
- Please find the detailed Methodology Documents by clicking [here](#).



Thank You!