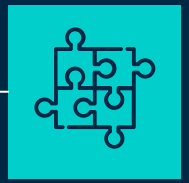
Airbnb CASE STUDY

Audience: Data Analysis Managers and Lead Data Analyst

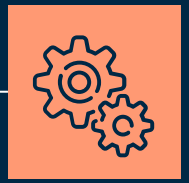
Sreehari Katageri
Vishesh Shroff
Iranna Chatti

DS_Cohort_C39

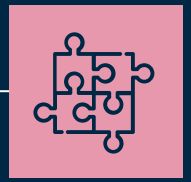
TABLE OF CONTENTS



01 Objective



02 Background



03 Data Cleaning & Preparation

04 Insights & Recommendations

05 Appendix

01 Objective

- Understand customer preferences from 2019 pre covid data and estimate preferences post covid.

- Recommend strategies to grow Airbnb's business and gain more traction post covid.

- Analyse the key insights from the Airbnb NYC 2019 data in pre-COVID situation.

Background 02

- During COVID-19 all types of businesses were affected across the world, which has also affected Airbnb's business due to travel restrictions.

- New York city being the largest city, contributes the most revenue to Airbnb, therefore it made a big loss in 2020

- Its been a while since the lockdown has been uplifted, the situation currently should be ideal for Airbnb to recover some income.

03 Data Cleaning & Preparation

Data Understanding

Below are the list of attributes and data gathered for the Analysis

- **Id:** A unique number identifying an Airbnb listing
- **Name:** Name of the listing
- **Host_id:** A unique number identifying an Airbnb host
- **Host_name:** Name of the person who hosted the listing
- **Neighbourhood_group:** Grouping of the region/ city where the listing resides
- **Neighbourhood:** Its is the Region/city where the listing resides
- **Latitude:** Latitude coordinates of the listings
- **Longitude:** Longitude coordinates of the listings
- **Room_type:** one of the "Entire home/apt", "Private room" or "Shared room"
- **Price:** price of the listing
- **Minimum_nights:** The minimum stay for a visit, as posted by the host.
- **Number_of_reviews:** The number of reviews that the listing has received from the customer. For this exercise we assume all these reviews are positive. It can also be considered as number of visits.
- **Last_review:** latest review date
- **Reviews_per_month:** The number of reviews that a listing has received per month.
- **Calculated_host_listings_count:** The number of listings for a particular host.
- **availability_365:** The number of days for which a particular host is available in a year.

Data Cleaning & Preparation

1. There are 16 columns & 48,895 records in the Airbnb dataset.
2. There are 6 Categorical columns of Object datatypes.
3. There are 10 Numerical values of int & Float datatype, among which id, host_id is converted to STR datatype & last_review is converted to date datatype.
4. We do not impute name & host_name missing values as the corresponding IDs are present.
5. Last_review & reviews_per_month missing values are imputed by MODE & MEAN of those columns respectively.
6. Tools used for Data Cleaning & Analysis are Python & Excel for Data cleaning, Tableau for Analysis.

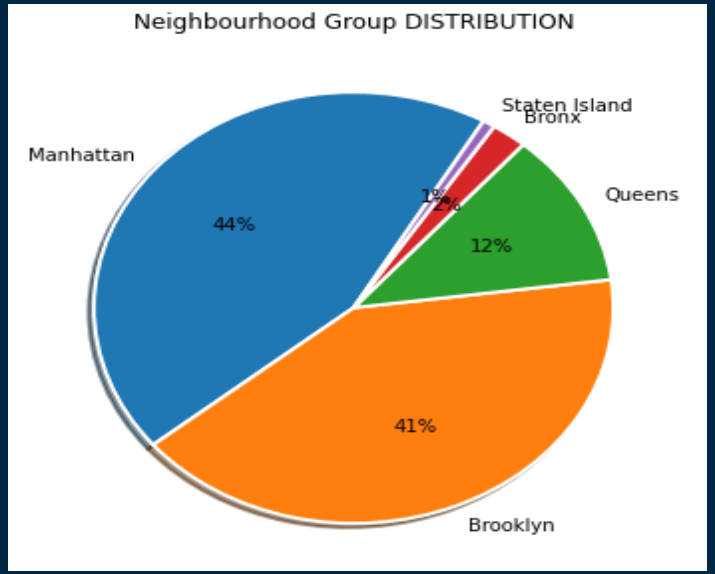
```
# Convert datatype
Airbnb_data["id"] = Airbnb_data["id"].astype(str)
Airbnb_data["host_id"] = Airbnb_data["host_id"].astype(str)
Airbnb_data["last_review"] = pd.to_datetime(Airbnb_data["last_review"])
```

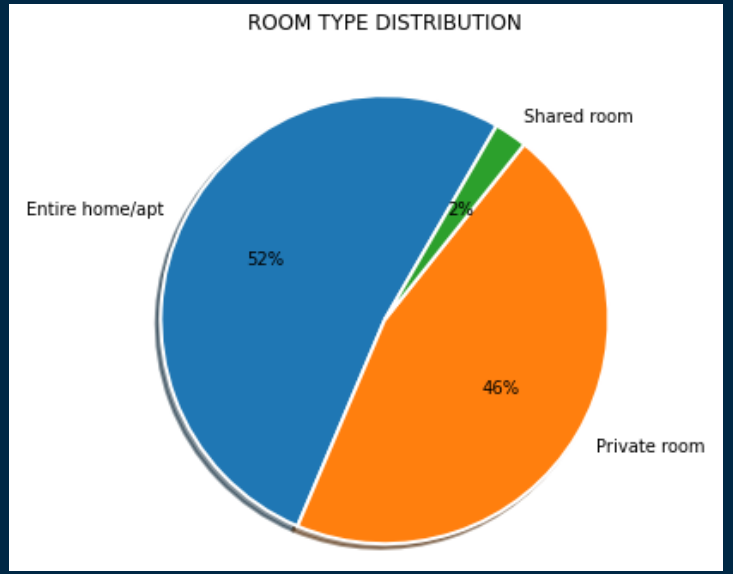
```
# Impute missing values by mode
Airbnb_data.last_review.fillna(Airbnb_data.last_review.mode()[0],inplace = True)

# Impute missing values by mean
Airbnb_data.reviews_per_month.fillna(Airbnb_data.reviews_per_month.mean(),inplace = True)
```

04 Insights & Recommendations

Distribution by Categorical Variables

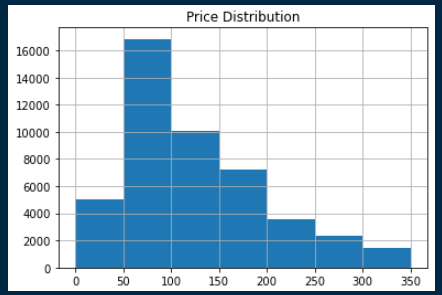


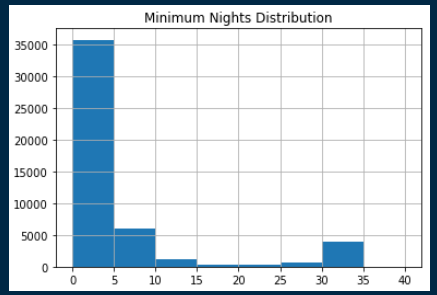


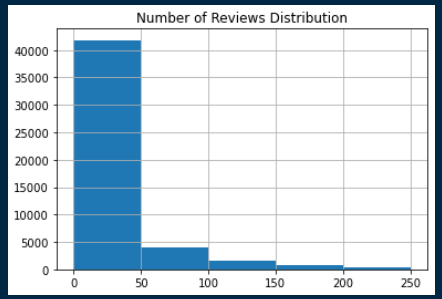
Manhattan is the city having most of the listings with 44% of entire dataset. The least being in **Staten Island** only 1%. **Brooklyn** consisted on 41% listing, 12% in **Queens** and 2 % in **Bronx**.

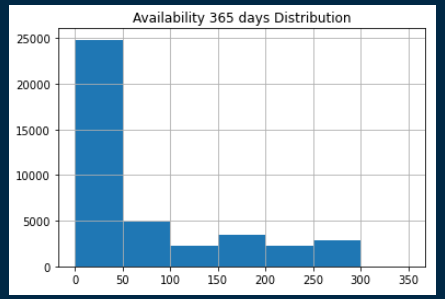
Entire home/apt has 52% and **private rooms** has 46% of the overall listing. Which explains that overall in New York shared rooms are less preferred and majority of people who opt Airbnb, travel with family or group of friends.

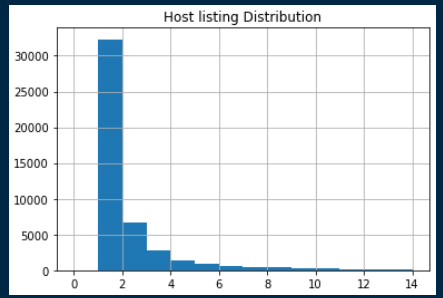
Distribution by Numerical Variables









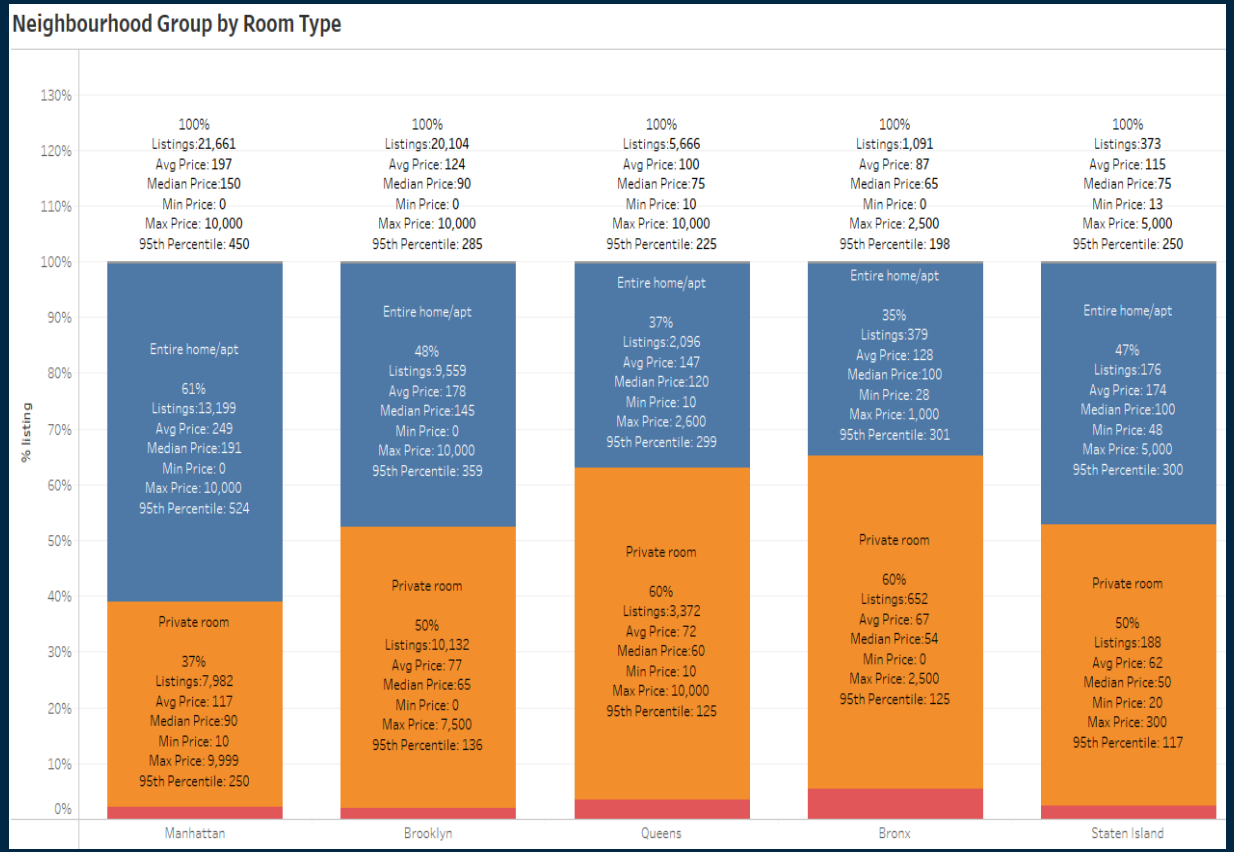


Looking at the overall distribution irrespective of any categorical variable:

- Majority of the listings fall under the Price less than 200$
- Majority of the listings offer minimum stay between 1 to 5 nights, followed by 10 to 15 & 30 to 35 nights.
- Majority of the listings received reviews between 0 to 50 reviews.
- Majority of the listings are available between 1 to 50 days a year, very less listings are available through out the year.
- Majority of the Host has only one listing.

NOTE: After checking the quantiles, some outliers were identified and removed from Analysis.

Customer Preference



by Room Type:

- The Room Type "Entire home/apt" and "Private rooms" are preferred more over the shared rooms by Airbnb hosts offering rentals in NYC and they take the majority of the portion among the listed properties in NYC (~97%).
- Shared Rooms are less listed (~2%) on Airbnb NYC, as it could be that customers less preferred shared rooms in the past.

by Neighbourhood Group:

- Properties in Manhattan are most expensive compared to other areas, while in Bronx are least expensive.
- Manhattan has higher % of listing in "Entire home/apt" properties compared to all other groups.
- Brooklyn has higher % of listing in "Private rooms" properties compared to all other groups.

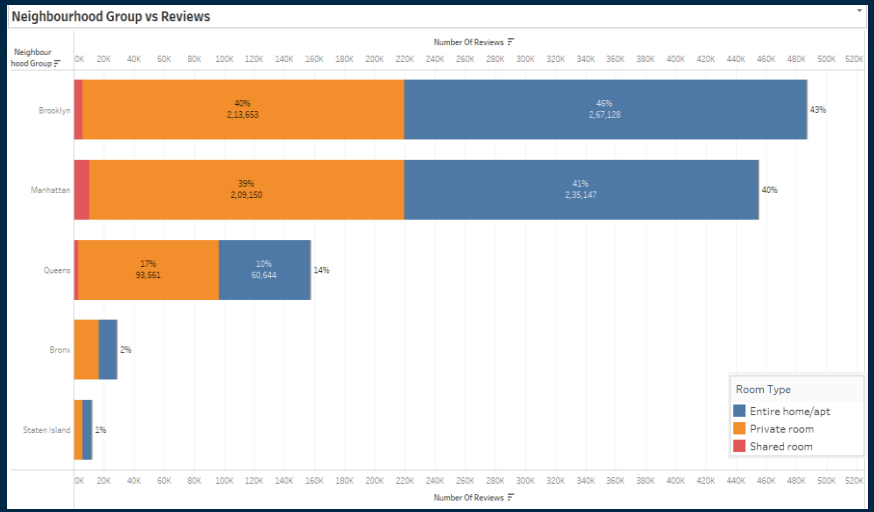
Customer Preference

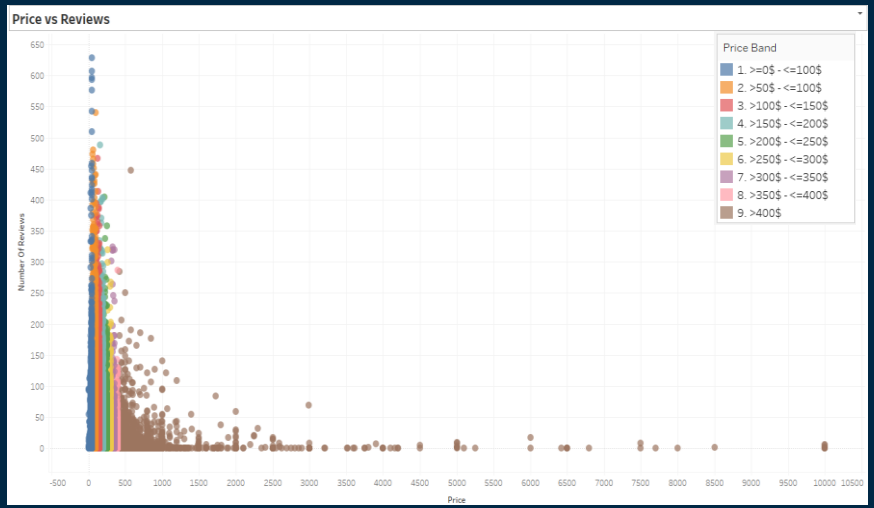
by Reviews:

- Majority of Reviews is for Brooklyn though Manhattan has many listings. This could be because, though Manhattan and Brooklyn has major listings, Compared to Manhattan, Brooklyn has better listings with price range 50$ to 100$ (27% Manhattan vs 41% Brooklyn)

by Price:

- Assuming all reviews are positive, as the Price increase, reviews decrease. This shows that customer prefer most budget friendly properties.
- Reviews are more for the price range 0$ to 150$.





Recommendations

- Looking at the preference by price & reviews, customers look for the properties less than 200$, hence Airbnb should look for some budget friendly properties in Manhattan either by acquiring budget friendly properties or by offering some discounts/ coupons.
- Since the price at Bronx are cheaper compared to other Neighbourhood groups Airbnb should focus on acquiring more properties in this area.
- Manhattan has more of Entire home/ apt (61%) and less of Private room (37%). Hence Airbnb need to look at increasing Private room properties.
- Since majority of listing are available between 1 to 50 days a year (~50% listings), Airbnb should make sure at least 50% of the listings are available almost through out the year to keep the Business up and running all the seasons of the year.

Appendix 05

APPENDIX –ASSUMPTIONS & METHODOLOGY

Assumptions:

- We assumed that the Airbnb Business prior to COVID-19 was achieving the desired Revenue.
- The Companies Strategies are considered based on the local Govt bodies on uplifting the COVID travel restrictions and the constraints enforced on the travel.
- Considered "Number of Review" to be Positive.

Methodology :

Data Preparation & Cleaning: There were 10 Numerical values of int & Float datatypes among which id, host_id was converted to str datatype & last_review was converted to date datatype. Though name and host_name have missing values we have retained the null values as the corresponding IDs are present. last_review & reviews_per_month had null values which was imputed by mode and mean.

Data Analysis: We have used Python and Tableau for analyzing the data through visualization. We identified the outliers in the numeric variables and accordingly excluded the extreme/ outlier values from our analysis. For Analysis we have used bar chart, Stacked bar chart, Pie Chart, Pareto Chart, Scatter Plot and Histogram.

NOTE: Methodology Document along with this presentation would provide much more details.

Thank You