INSIGHTS AND INFERENCES FROM AIRBNB ANALYSIS

By – Deepali Aggarwal and Tanishq Bhalla

AGENDA AT HANDA

- Objective of this Case study

- Background Information

- Key Findings from Analysis

- Appendix-
Data Sources
Data Methodology
Data Model assumptions

OBJECTIVE

- Improve the strategies to revive the business in the post-COVID period.

- Understanding key insights from Airbnb NYC business in pre-COVID period.

- Estimate the customer preferences for the post-COVID period travel.

BACKGROUND

- COVID-19 pandemic affected Airbnb business due to travel restrictions in many countries.

- The revenue took largest hit in NYC (New York City) in the Q2 and Q3 of the year 2019.

- Now that the travel restrictions are lifted so the business should be operated in a way to recover the loss.
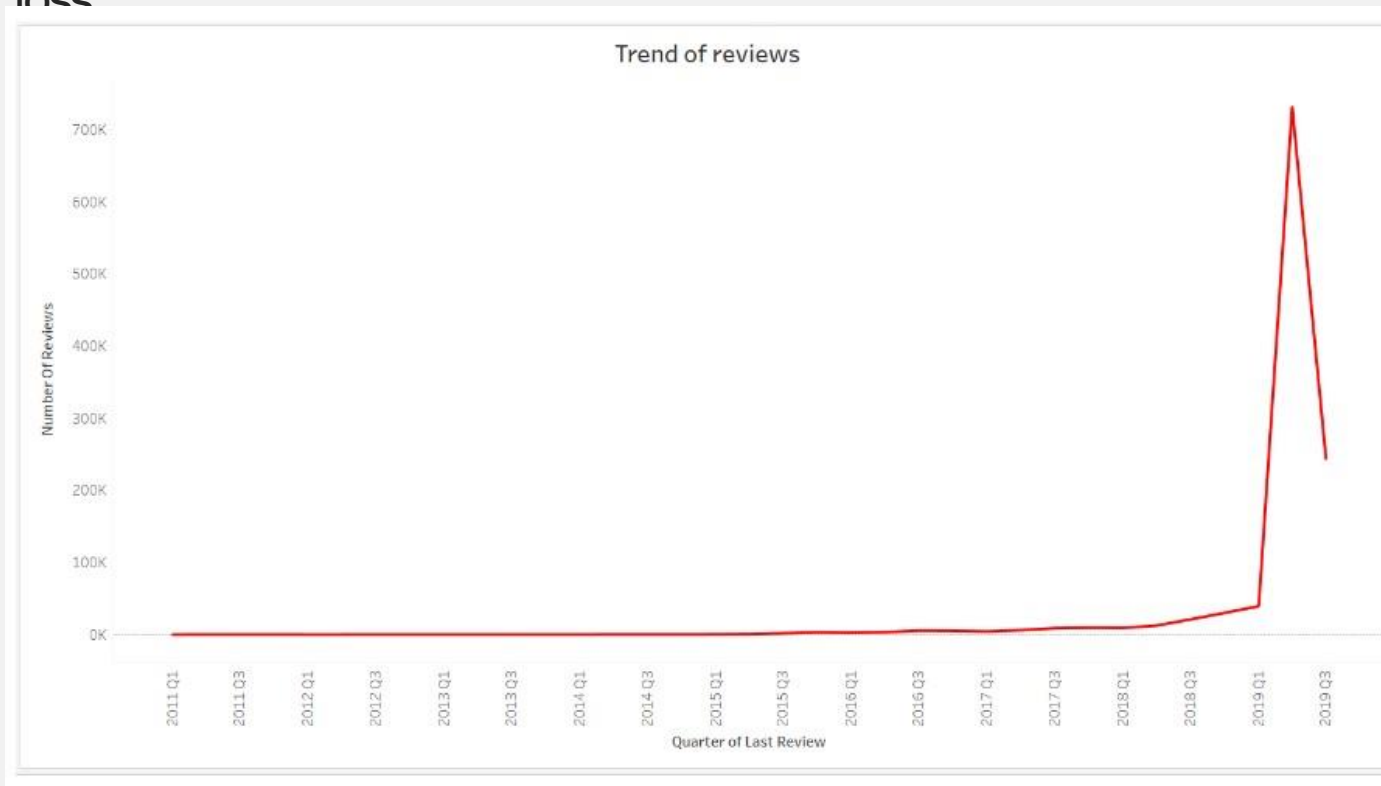
# OBJECTIVE OF THE ANALYSIS OF THE NEW YORK CITY DATA

- Customer preferences (neighborhood), ratings and reviews of different hosts (Room type and Name) of Airbnb in were analyzed.

- Dataset was cleaned and analyzed to derive key insights and inferences from the Pre-COVID period.

- Insights provided would be used to take decisions for the business in NYC Airbnb for travel in post-COVID period.

DATA PREPARATION AND DATA CLEANING

- The missing values present in some of the columns were treated using python

- Various statistical methods were used to analyse the accuracy and consistency of data and to check if the discrepancy in data can be corrected eg describe, info, shape, etc.

- Correlation of different features in the dataset were evaluated and no significant relationships were found when analysed using pair plots and heatmaps.

- Cleaned data was exported to tableau where it was visualized to get insights from analysis. Customer preference became clearer by using various charts like bar, clustered bar, tree maps, text tables, etc.

CUSTOMER'S PREFERENCES FOR ROOM-TYPE

- Only 1.69% belong to the "SHARED ROOM" type, which clearly shows the priority of customers.

- Room type "ENTIRE HOME/APT" and "PRIVATE ROOM" have a major stake in the listed properties and account for 98.31% of the total in the pre-covid period. Following the trend, it should remain the same for post-covid period too as customers are taking more precautions and necessary steps to avoid close contacts

- The SHARED ROOM can be converted to PRIVATE ROOM with multiple occupancy, this would increase the revenue and footfall and would also reduce wastage of resources as the shared rooms can be utilized now
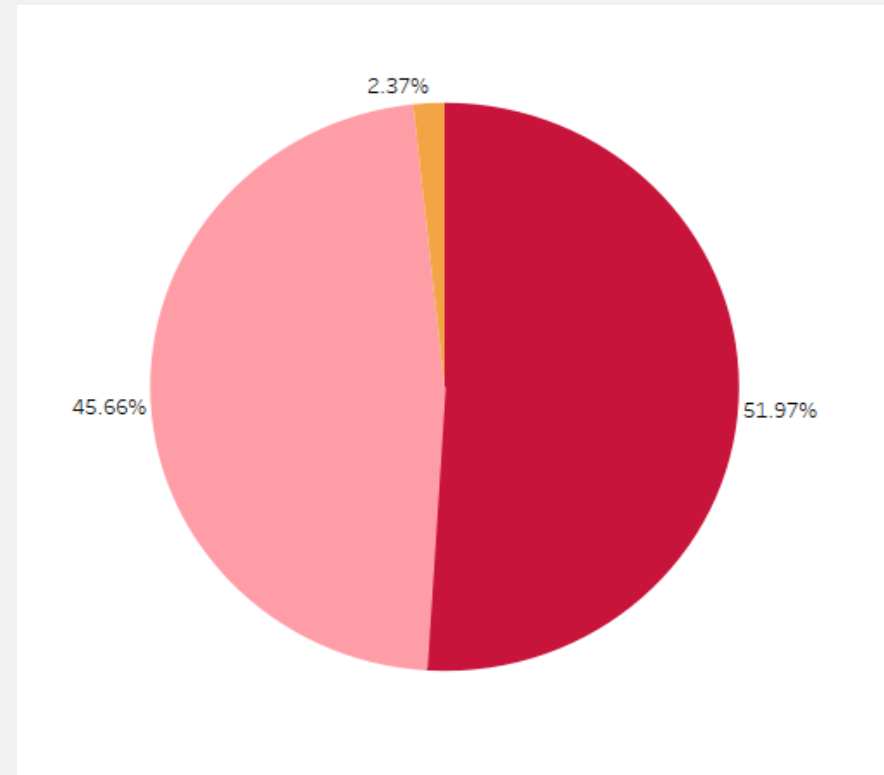
CUSTOMER'S PREFERENCES FOR LOCALITIES (NEIGHBOURHOODS)

- Williamsburg, Bedford-Stuyvesant, and Harlem show the highest number or reviews as well as listing in NYC neighbourhood

- Cannot see a proper trend with respect to the number of reviews

- Upper East Side, Midtown, Financial District, Upper West Side and Williamsburg have comparatively lesser reviews than the number of listings present

What are the most popular localities in New York currently?
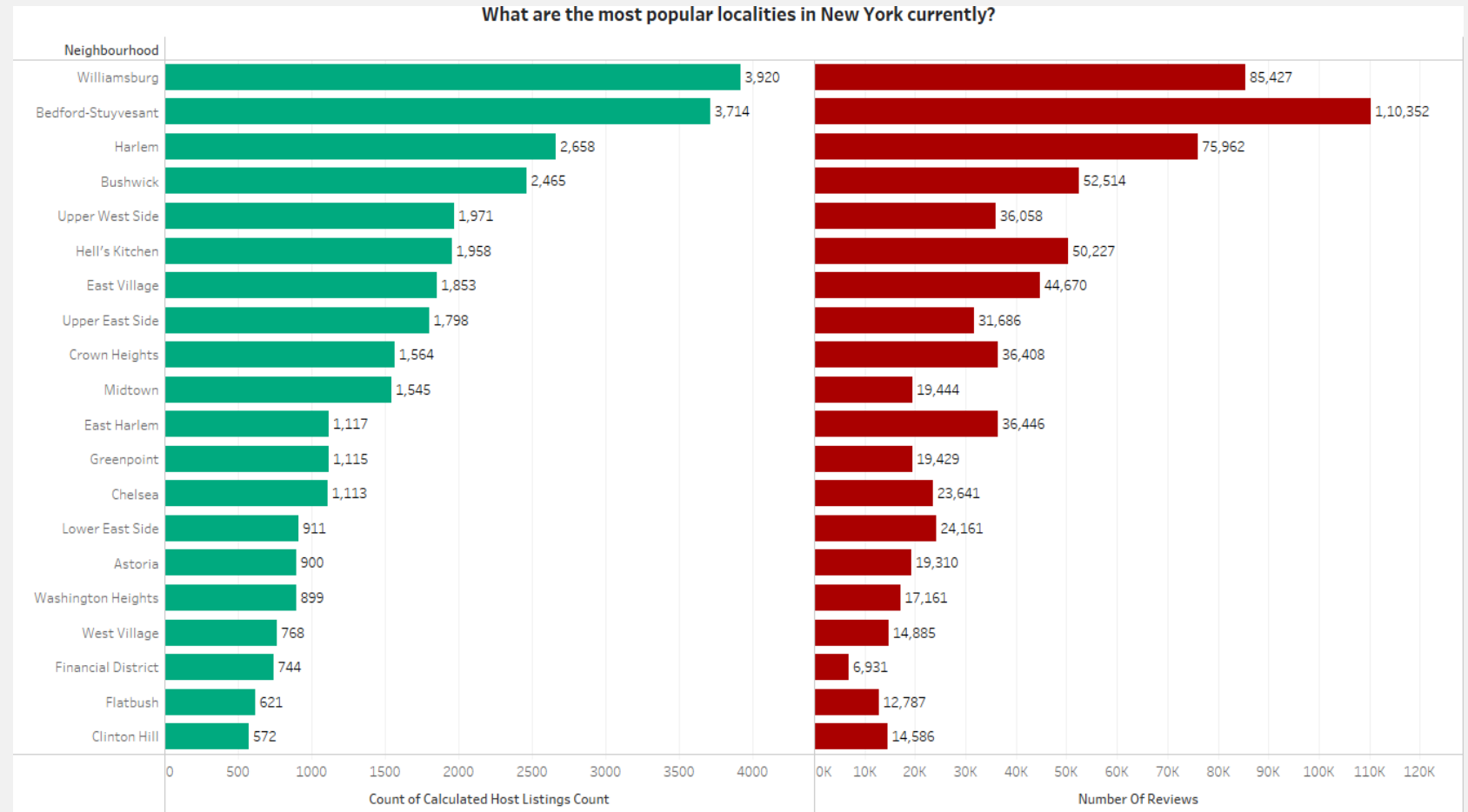
CUSTOMER PREFERENCES OF PROPERTIES IN NYC

- You could find properties listed with the same name across neighbourhood

- According to the number of reviews properties in Harlem and Jamaica are the most popular and can be seen frequently in top records. These are followed by East Elmhurst

What are the most popular properties in New York currently?
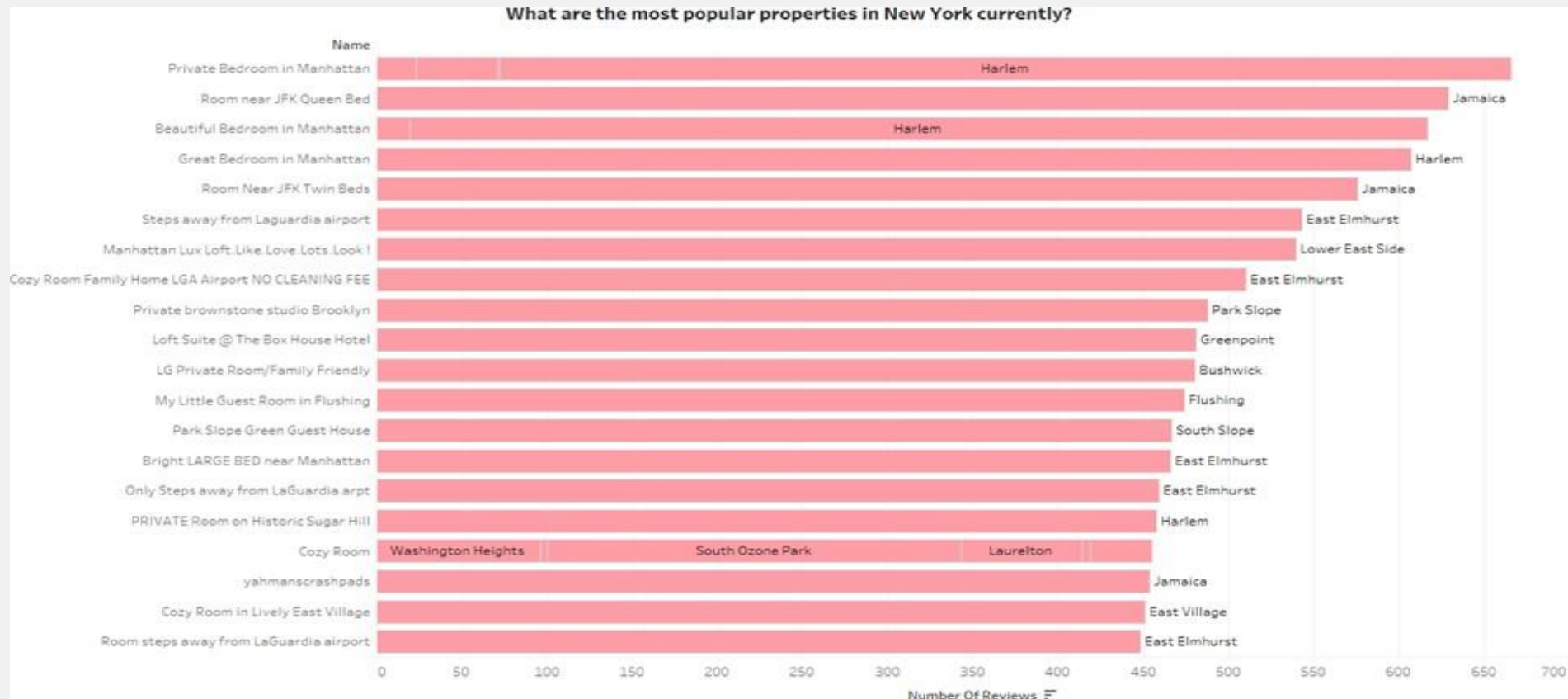
CUSTOMER PREFERENCES OF PROPERTIES IN NYC

- A 1-10 days of minimum nights per stay is preferred by customer. During the pre-covid period it showed a spike of 30,60 and 90 days

- Properties that charge high and also have higher number of minimum nights per stay have low number reviews, possibly because there is less free movement and also becomes expensive

- It is a possibility that due to the current status (covid times), customers would book a property that is available for 30, 60 or 90 days as they want less contact and henceforth lesser travel
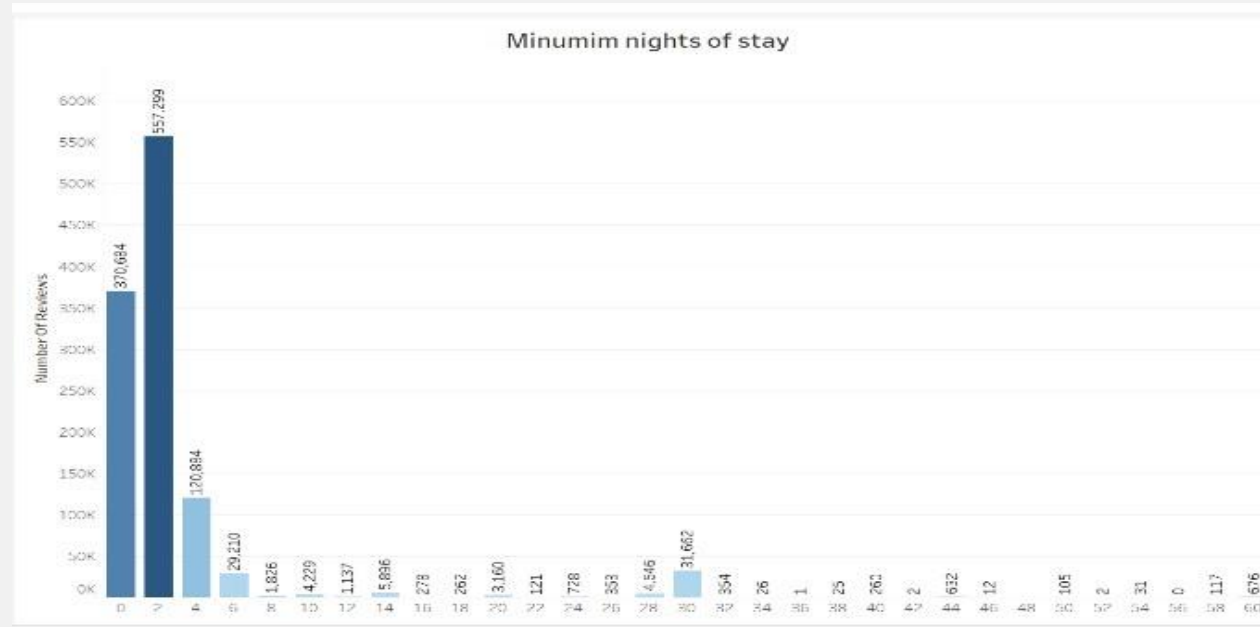
APPENDIX: DATA SOURCES

Column	Description
id	listing ID
name	name of the listing
host_id	host ID
host_name	name of the host
neighbourhood_group	location
neighbourhood	area
latitude	latitude coordinates
longitude	longitude coordinates
room_type	listing space type
price	
minimum_nights	amount of nights minimum
number_of_reviews	number of reviews
last_review	latest review
reviews_per_month	number of reviews per month
calculated_host_listings_count	amount of listing per host
availability_365	number of days when listing is available for booking

```
Categorical Variables:
    - room_type
    - neighbourhood_group
    - neighbourhood

Continous Variables(Numerical):
    - Price
    - minimum_nights
    - number_of_reviews
    - reviews_per_month
    - calculated_host_listings_count
    - availability_365
- Continous Variables could be binned in to groups too

Location Varibles:
    - latitude
    - longitude

Time Varibale:
    - last_review
```

Variable Categories

APPENDIX: DATA METHODOLOGY

- Performed analysis on Airbnb data using Python and Tableau.
- Importing data
- Doing Quality Check on data
- Cleaning the data:
  - Treating Missing Value
  - Data Removal and Imputation
- Data Analysis:
  - Bivariate Analysis using Heatmap and Pair plot
- Exported cleaned data in csv file
- Imported data in Tableau
- Data Visualization on cleaned data set

APPENDIX- DATA ASSUMPTION

- Data prior to Covid-19 was achieving required revenue
- The properties that have not been reviewed are the ones that are unpopular or popular properties are ones that have at least one review
- Assuming that AirBnB has no plans to expand his horizon even in NYC and are moving forward to post-covid period with same neighbourhood groups
- Strategies are made assuming lesser restrictions on travel in post-covid period, eventually leading to increase in travel

THANK YOU!