AIRBNB NYC Case Study – Methodology

Authored by: Archana Lahiri,
Lavanya Korripadu
& Akshita Sharma

(Image courtesy: https://en.wikipedia.org/)

Introduction

Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Airbnb provides platform for hosts to accommodate guests with short-term lodging and tourism-related activities.

New York City is the most diverse and populated city in the United States. The city consists of 5 borrows: Manhattan, Brooklyn, Queens, the Bronx and Staten Island, all of which were “grouped” together into a single city. It is widely recognized as the global center for the financial services industry. It is also the heartbeat of the American media, entertainment (along with California), telecommunications, and law and advertising industries.

Business Objective:

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.

Assumption:

As we are not aware about the nature of reviews, we have assumed that the properties, which received higher number of reviews, have a better customer liking.

Data Source

Provided with Airbnb New York City Listings Dataset till 2019 (48895 Rows * 16 Columns)

Column	Description
Id	Listing ID
Name	Name of Listing
Host_id	host ID
Host_name	Name of Host
Neighbourhood	Neighbourhood_group - Location
Neighborhood	Neighborhood - Area
Latitude & Longitude	Map co-ordinates
Room_type	Listing space type
Price	Price of listing
Minimum_nights	Amount of nights minimum
Number_of_reviews	number of reviews
Last_review	Lastest review
Reviews_per_month	number of reviews per month
Calculated_host_listings_count	no. of listings per host
Availability_365	no. of days when listing is available for booking

Column Name	No. of Rows and Column Datatype	
id	48895	int64
name	48879	object
host_id	48895	int64
host_name	48874	object
neighbourhood_group	48895	object
neighbourhood	48895	object
latitude	48895	float64
longitude	48895	float64
room_type	48895	object
price	48895	int64
minimum_nights	48895	int64
number_of_reviews	48895	int64
last_review	38843	object
reviews_per_month	38843	float64
calculated_host_listings_count	48895	int64
availability_365	48895	int64
dtypes: float64(3), int64(7), object(6)		

Presentation – I:

Objective:

The presentation will focus mainly on the following points:

1. Get a better understanding about Airbnb listings with respect to various parameters
2. Understand the customer preferences
3. Understand the customer booking trend

Exploratory Data Analysis:

To understand some important insights, we have explored the following questions:

1. How are the Airbnb listings spread out in NYC?
2. What type of rooms do customers prefer?
3. What could be the ideal number of minimum nights to increase customer bookings?

Based on customer review:

1. Most preferred neighborhood
2. Most preferred room type
3. Who are the Hosts who have the highest listings w.r.t Neighborhood?

Methodology:

In the case study we have used Jupiter notebook to perform initial analysis of the data and Tableau for data analysis and visualization.

Initial Analysis using Jupiter Notebook: Data Set Used: AB_NYC.2019.csv

Number of Rows: 48895

Number of Columns: 16

Importing the libraries ¶

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
%matplotlib inline
import seaborn as sns
```

Loading Data

```
df = pd.read_csv("AB_NYC_2019.csv")
df.head()
```

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

```python
df.shape
```

```
(48895, 16)
```

```python
#checking type of every column in the dataset
df.dtypes
```

```
id                                  int64
name                               object
host_id                             int64
host_name                          object
neighbourhood_group                object
neighbourhood                      object
latitude                          float64
longitude                         float64
room_type                          object
price                               int64
minimum_nights                      int64
number_of_reviews                   int64
last_review                        object
reviews_per_month                 float64
calculated_host_listings_count      int64
availability_365                    int64
dtype: object
```

```
#looking to find out first what columns have null values
df.isnull().sum()
```

```
id                                0
name                             16
host_id                           0
host_name                        21
neighbourhood_group               0
neighbourhood                     0
latitude                          0
longitude                         0
room_type                         0
price                             0
minimum_nights                    0
number_of_reviews                 0
last_review                   10052
reviews_per_month             10052
calculated_host_listings_count    0
availability_365                  0
dtype: int64
```

```
#dropping columns that are not significant or could be unethical to use for our future data exploration and predictions
df.drop(['id','name','last_review'], axis=1, inplace=True)
#examing the changes
df.head()
```

	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	reviews_per_month
0	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	9	0.21
1	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	0.38
2	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	0	NaN

- We removed the columns like Id, Name, Last Review which was not giving much information.

```
#replacing all NaN values in 'reviews_per_month' with 0
df.fillna({'reviews_per_month':0}, inplace=True)
```

```
#verifying changes
df.reviews_per_month.isnull().sum()
```

```
0
```

```
#examining the unique values of n_group
df.neighbourhood_group.unique()
```

```
array(['Brooklyn', 'Manhattan', 'Queens', 'Staten Island', 'Bronx'],
      dtype=object)
```

```
#examining the unique values of neighbourhood
len(df.neighbourhood.unique())
```

```
221
```

```
#examining the unique values of room_type
df.room_type.unique()
```

```
array(['Private room', 'Entire home/apt', 'Shared room'], dtype=object)
```

Step 2: Data Wrangling :

- Checked the Duplicate rows in our dataset and no duplicate data was found.
- Checked the Null Values in our dataset. Columns like name, host_name, last review and review_per_month have null values.
- We've dropped the column name as missing values are less and dropping it won't have significant impact on analysis.
- Checked the formatting in our dataset.
- Identified and review outliers.

Step 3: Data Analysis and Visualizations using Tableau:

We have used tableau to visualize the data for the assignment. Below are the detailed steps used for each visualization.

1. **Top 10 Host :**
   - We identified the top 10 Host Ids , Host Name with count of Host Ids using the tree map.

2. **Preferred Room type w.r.t Neighbourhood group:**

- We created a pie chart for understanding the percentage of room type preferred w r t neighbourhood group
- We added Room Type to the colors Marks card to highlight the different Room Type in different colors and count of Host Id to the size

3. **For Variance of price with Neighbourhood Groups:**
    - We used a box and whisker’s plot with Neighbourhood Groups in Columns and Price in Rows.
    - We changed the Price from a Sum Measure to the median measure.

4. **Average price of Neighbourhood groups:**
    - We created a bubble chart with Neighbourhood Groups in Columns and Price column in Rows.
    - We added the Neighbourhood Groups to the colors Marks card to highlight the different neighbourhood Groups in different colors. Also Put Avg price in Label.

5. **Customer Booking w r t minimum nights:**
    - We created the bin for Minimum nights as shown below.

- The bins were used to display the distribution of minimum nights based on the amount of ids booked for each neighbourhood group.

6. **Popular Neighborhoods:**
    - We took neighbourhood in rows and sum of reviews in column and took neighbourhood groups in color.
    - We used filter to show Top 20 neighbours as per the sum of reviews.

7. **Neighbourhood vs Availability:**
    - We created a dual axis chart using bar chart for availability 365 and line chart for price for top 10 neighbourhood group sorted by price.

- <u>**Inference:**</u>

1. In line with our earlier observation, we see the maximum reviews in listings for Manhattan & Brooklyn, implying that more bookings happen in these neighborhoods. The higher number of customer reviews imply higher satisfaction in these localities.

2. Also, we see the maximum reviews in room types ‘Entire home/apt’ & ‘Private rooms’.
We can safely infer that; customers do not prefer ‘Shared rooms’.

Presentation – II:

Objective:

The presentation will focus mainly on the following points:

1. Get a better understanding about Airbnb listings with respect to various parameters
2. Understand the pricing relation to various parameters

3. Recommendations to improve quality of new acquisitions and customer experience.

Exploratory Data Analysis:

To understand some important insights we have explored the following questions:

1. Customer preference for neighborhood & room type
2. Property demand based on minimum nights offered
3. Price range preferred by customers
4. Understanding Price variation w.r.t Room Type & Neighborhood
5. Understanding Price variation w.r.t Geography
6. Top reviewed properties

Methodology:

1. **Preferred Room type w.r.t Neighbourhood group:**
   - We created a pie chart for understanding the percentage of room type preferred w r t neighbourhood group
   - We added Room Type to the colors Marks card to highlight the different Room Type in different colors and count of Host Id to the size

2. **Customer Booking w r t minimum nights :**

- We created the bin for Minimum nights as shown below.

- The bins were used to display the distribution of minimum nights based on the number of ids booked for each neighbourhood group.

3. **Neighbourhood vs Availability:**

- We created a dual axis chart using bar chart for availability 365 and line chart for price for top 10 neighbourhood group sorted by price.

4. **Price range preferred by Customers:**

- We have taken pricing preference based on volume of bookings done in a price range and no of Ids to create a bar chart. We have created bin for Price column with interval of $20.

5. **Understanding Price variation w.r.t Room Type & Neighbourhood:**

- We created Highlights Table chat by taking Room Type in rows & Neighbourhood Group in column.
- We took the average price in colour Marks card to highlight the different Room Type in different colors.

6. **Price variation w r t Geography:**

- We used Geo location chart to plot neighbourhood , neighbourhood Group in map to show case the variation of prices across.

7. **Popular Neighborhoods:**

- We took neighbourhood in rows and sum of reviews in column and took neighbourhood groups in color.
- We used filter to show Top 20 neighbours as per the sum of reviews.

Inference:

1. Manhattan, Brooklyn and Queens have the most liked properties (most reviewed).
2. the most reviewed property "Private Bedroom in Manhattan", though it appears to be steeply priced still has managed to get the maximum number of reviews making it the most favorable property in NYC.

Recommendations Consolidated:

1. Promotion of shared rooms with targeted discounts to increase bookings.
2. More number of hosts & listings with monthly rental duration (30-60-90) can be acquired. We see a good potential in the 30-day rental window. Manhattan & Brooklyn have higher number of 30-day bookings compared to the others; these areas can be further targeted.
3. Weekly or bi-weekly rentals can also be acquired, as these can be used customers stranded in NYC for quarantine purposes.
4. New acquisitions and expansion can be done in the price range of $40 - $190 as it satisfies both parameters of volume of customer traffic and customer satisfaction.
5. New acquisitions can be explored to acquire 'private rooms' in Manhattan and Brooklyn and 'entire homes' in Bronx and Queens.
6. Brooklyn has an average price of $124. As there are already many listings available in Manhattan, Brooklyn can be considered for expansion.
7. Increasing acquisitions and new properties in coastal regions can increase customer bookings.