

Capstone Project

Airbnb Bookings Analysis

by –

Anshu Gangwar
Sourav Chowdhury

How Airbnb Works



Airbnb is an Billion dollar online marketplace for renting out homes/villas/private rooms.connects people who want to rent out their homes with people who are looking for accommodations in specific locales. The website charges a commission(3 to 20 percent) for every booking.According to Airbnb's latest data, it has in excess of 5.6 million listings, covering more than 100,000 cities and towns and 220-plus countries worldwide.

Founders: Nathan Blecharczyk, Joe Gebbia and Brian Chesky.

The 3 customer segments defining Airbnb's Business Model

Hosts:

Hosts are the people who own property and want to make some money by renting out their available space.

They can create a listing for their property on Airbnb, add property details and set their own rent, check-in, check-out time etc.

Hosts can accept or reject a booking after reading the reviews of the traveller or after going through his social profiles.

Travelers:

Travellers are the people who book the listed available spaces from local hosts.

Travellers have the option to search for a property by filtering them according to rent, amenities provided, location etc.

Travellers can instant book a space by paying through the portal.

Freelance Photographers:

Airbnb has a vast network of freelance photographers in all major cities of the world who go to a location and click high-definition photographs of the property.

The high quality photographs get more responses and the freelance photographers are paid by Airbnb directly.

5 Step Model of Airbnb

1. Hosts list out their property details on Airbnb along with other factors like pricing, amenities provided etc.
2. Airbnb sends a professional photographer(if available) to the property location in order to take high quality photographs.
3. Travellers search for a property in the city where they wish to stay and browse available options according to price, amenities etc.
4. Booking is made through Airbnb where traveller pays the amount mentioned by host and some additional money as transaction charges.
5. Host approves the booking. Traveller stays there and finally Airbnb pays the amount to the host after deducting their commission.

Problem Statement

Explore and analyse the data to come up with the answers of following questions :

What can we learn about different hosts and areas?

What can we learn from predictions? (ex: locations, prices, reviews, etc)

Which hosts are the busiest and why?

Neighbourhood groups with minimum number of nights to stay and people preferring which room type?

Problem We Faced

- Columns with more number of null values.
- Data mining was Challenging.
- Plotting the graphs took computation time.

```
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
```

How to Solve the Problem

The Problems stated in above slide were solved by taking help from different sources:

- as we removed only those column null values, having least number of null.**
- Used different sources so as to plot the graph and do the data mining :**
 - Geeks for Geeks**
 - Stackoverflow**
 - W3Schools**

Impact we Made

In this project we have shown using different charts how the different factors are affecting the business growth .

- Listings from Different Hosts and Areas.**
- Predictions for different places and number of customers.**
- How Number of reviews is giving the various information regarding the dataset.**
- Busiest and the most Demanded place, So as to increase Business Opportunity.**

Reading Input dataset

To load the dataset in the colab notebook, first we mounted the notebook with google drive. And then read the dataset using pandas built in function read_csv.

```
from google.colab import drive
```

```
drive.mount('/content/drive')
```

```
df = pd.read_csv('/content/drive/MyDrive/Capstone Project - EDA/Airbnb NYC 2019.csv')
```

Python Modules/Packages/Libraries

Following are the Python Libraries Used in the project done on Colab notebook.

import numpy as np # For numerical computing

import pandas as pd # For Data Processing

import matplotlib.pyplot as plt # For plotting graphs

Get Basic information of Data

To start with the project , first we collect the basic information related to no. of rows, head,tail.Using following methods.

df.head()

Displays the initial 5 rows of dataset.

df.tail()

Displays the bottom 5 rows of dataset.

df.isnull().sum()

To get null values in each column.

Dropped the null values so as to do basic cleaning of data

DATA SUMMARY

Data set name :- Airbnb Booking Analysis

Shape :-

- Rows - 48895
- Columns - 16

Important columns :

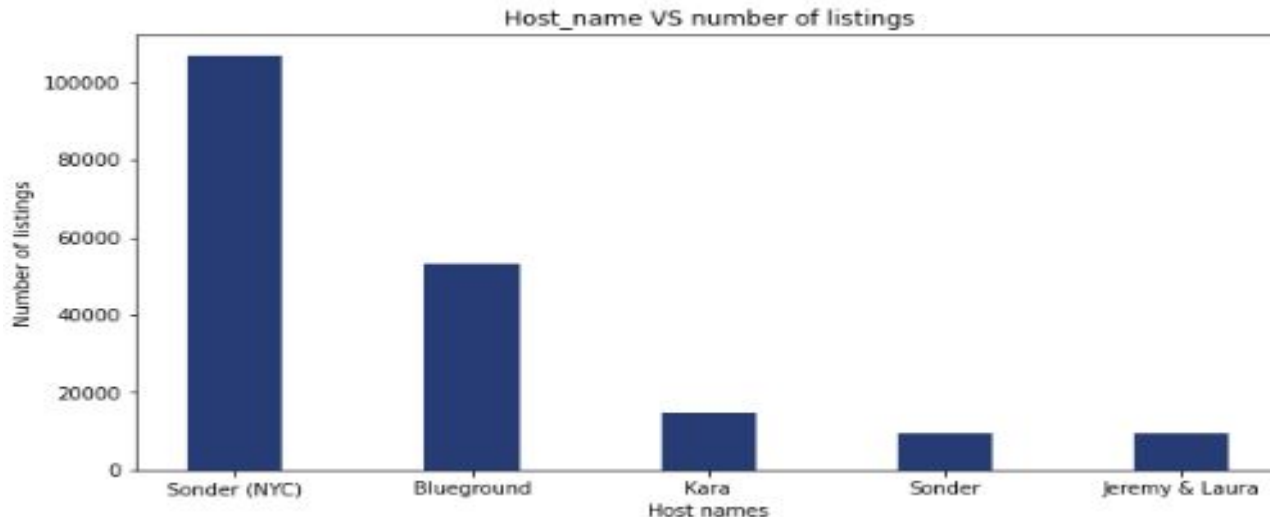
- Neighbourhood_group
- Neighbourhood
- Room_type
- Price
- Minimum_Nights
- number_of_reviews

Variable Breakdown

- 1. Name : Name of the rental property.**
- 2. Host name : Name of the owner of the property.**
- 3. Neighbourhood : Area to which property belongs.**
- 4. Neighbourhood group : Grouped areas consisting various neighbours.**
- 5. Room type : Type of room it could be Private, Shared and Entire home/apt.**
- 6. Price : The rental price of the property.**
- 7. Minimum nights : Minimum no. of nights the property can be booked.**
- 8. Number of reviews : Reviews received from customers.**
- 9. Calculated host listings count : No. of listings by host.**
- 10. Latitude/Longitude : LAT/LONG of the particular listing.**

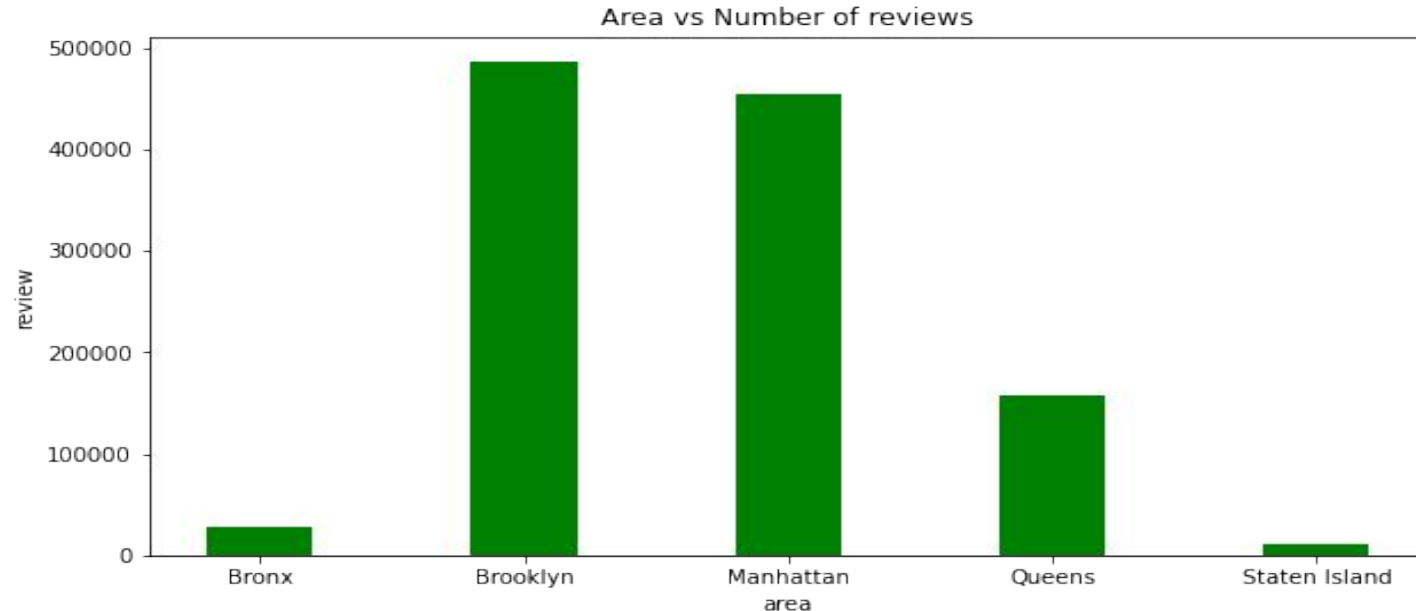
Most number of Listings

- Manhattan, Brooklyn are the top most business generated area for Airbnb.
- Sonder (NYC), Blueground, Michael, David have hosted maximum number listings.
- Highest Listings number comes from Manhattan and hosted by Sonder (NYC).



Most Preferred Neighbourhood groups

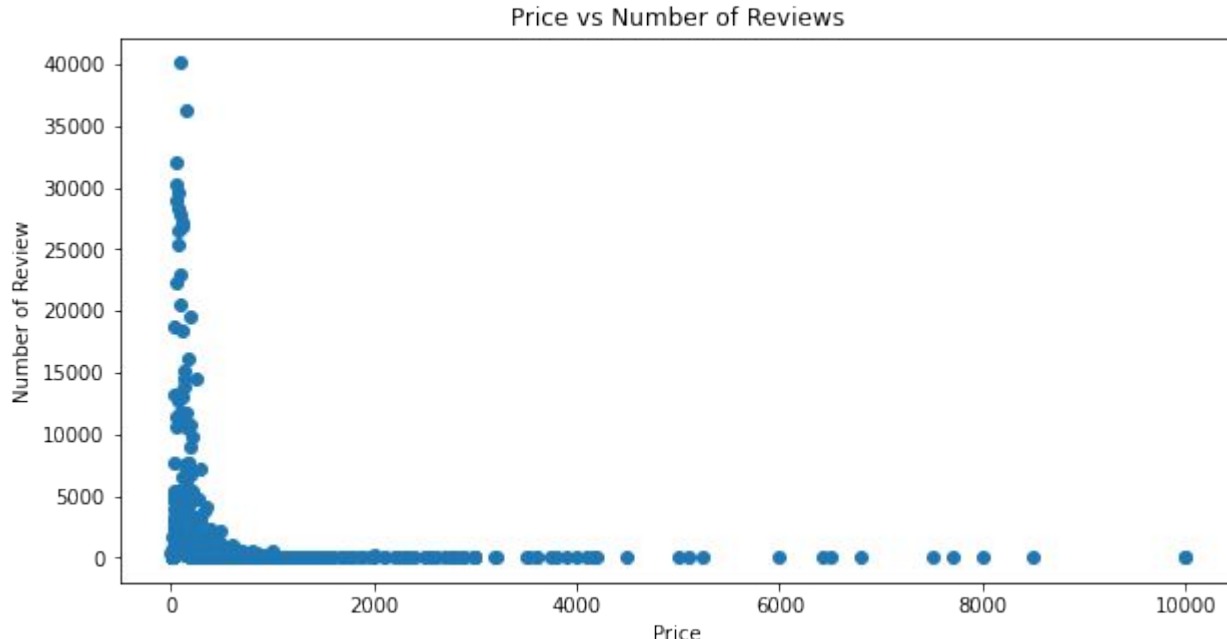
Most preferred location to stay can be estimated by the number of customers visiting that place. If the place has more reviews means maximum number of customers stays there. Hence from the following graph it can be predicted that Brooklyn is the most preferred neighbourhood group and Staten is the least one to stay. means that these neighbourhood has most number of customers.



Price affecting the Number of customers

Price of any commodity or service affects a lot on Business , thus setting a best suited price is always a challenge, let us see how the graph tells about the same. Number of reviews gives the estimation of number of customers.

Here the Scatter Plot tells that most people prefer to stay in place where price is less.

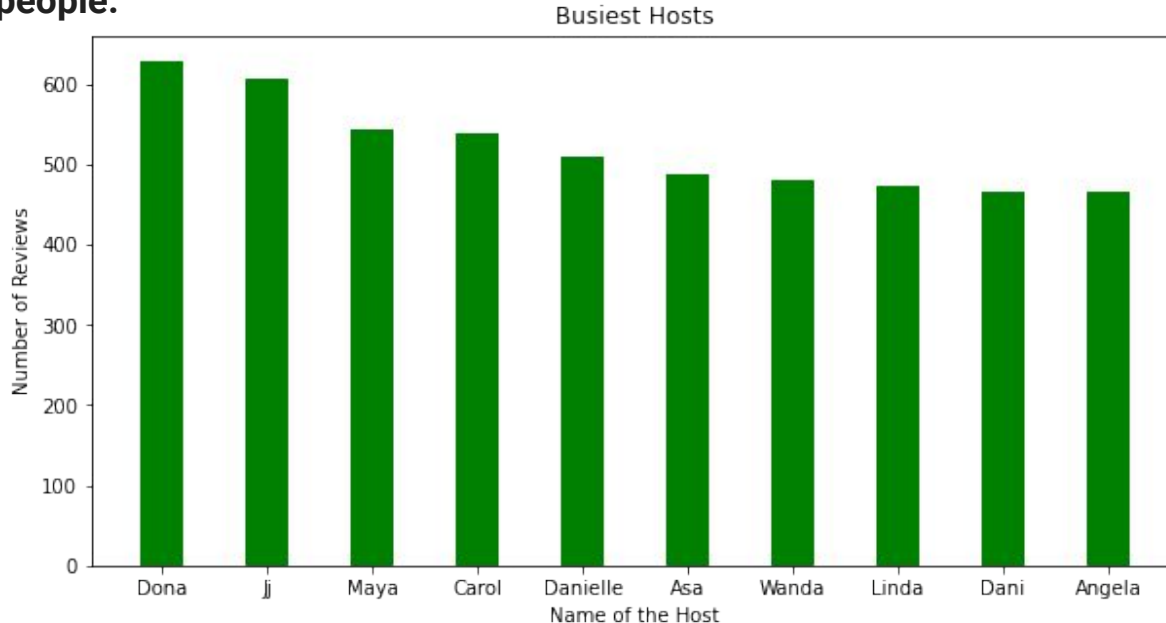


Busiest Host Name

Busiest hosts are those who have maximum number of customers, thus more number of reviews. From the below analysis we have found that the :

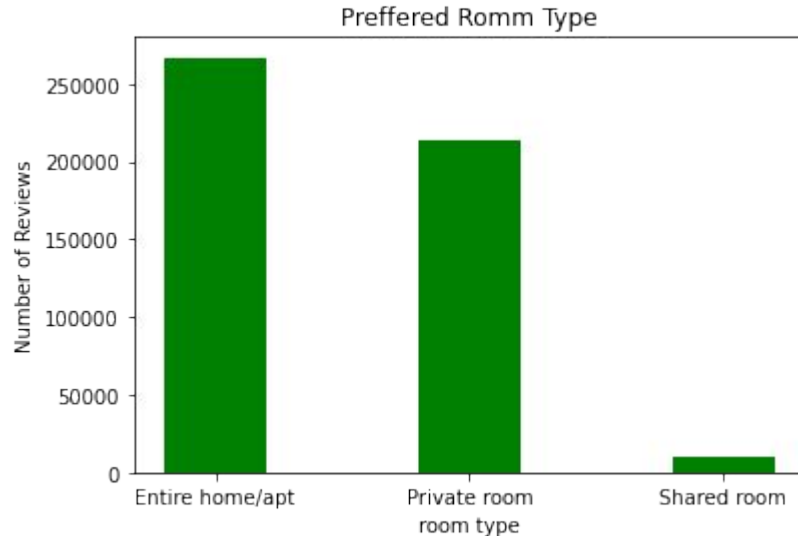
Busiest Hosts are : DONA > JI > MAYA > CAROL

Because these hosts listed room type as Entire home and Private room which is preferred by most number of people.



Most demanded room type

In this analysis report we will analyze the below observation: people are preferring Entire home/apt or private room which are present in Brooklyn, Manhattan and Queens, and which are in less price.



neighbourhood_group	room_type	minimum_nights
Manhattan	Entire home/apt	138081
Brooklyn	Entire home/apt	62393
Brooklyn	Private room	56098
Manhattan	Private room	43463
Queens	Private room	17263

Conclusion

- **Manhattan and Brooklyn are the top most booking generated areas for Airbnb.**
- **High price hotels available in Manhattan, Queens and Brooklyn.**
- **People prefer to stay in low price hotel.**
- **Sonder (NYC), Blueground, Michael and David are the top most busiest host.**
- **People prefer to stay in Entire home/apt or Private room.**

Thank You !

Anshu Gangwar
Sourav Chowdhury