Airbnb Data Analysis Report

RAHUL Y M
Data Science Trainee,
AlmaBetter, Bangalore

Abstract:

Many of the successful companies today are playing the part of brokers in their corresponding market region.

We see companies connecting restaurants to customers, connecting shareholders to the stock market, connecting house owners to people who need houses on rent.

All in all, the model where the business owner provides the platform for connecting demand and supply is taking over as the “New Mantra” in the companies.

Now, when these companies work on their success metrics the main, variable for improvement is the efficiency of its service providing platform, so that it can connect the right customers with the right service providers
In the given report we have Analysed the Airbnb Data and tried to find out hidden facts and insights from which we have found reasons for hidden trends and concluded some Business solutions

Keywords-*Data Analysis, Pandas, Booking Prices, Customer Review, listings*

1.Problem and Data

- We have been provided the listing Data from Airbnb. Whenever a property owner wants to offer their services through Airbnb, they have to list their property and its details. These millions of listings generate a lot of data - data that can be analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services, and much more.

- This dataset has around 49,000 observations in it with 16 columns and it is a mix of categorical and numeric values. If we go through the fields present in the dataset :

- **Id**-Unique Id of each property
- **name**-Name of properties, which also describes the property
- **host_id-**ID of the host, this is unique for all the hosts.
- **host_name**- Name of the host , Names may be similar for two hosts.
- **neighbourhood_group**-have 5 different regions of New York
- **neighbourhood**- these are the subregions of neighbourhood_group , where one neighbourhood belongs to only one neighbourhood group.
- **Latitude** contains the coordinates of Latitude for that location.

- **Longitude** contains the coordinates of longitude for that location
- **room_type** – there are three types of property available ,
- **price**- price of stay (we have taken it for one night basis).
- **minimum_nights**- this is the minimum number of nights for which you have to book a host's place
- **number_of_reviews**-This will be the number of reviews each property has got,we can also use this to infer about the number of customers considering the same customer behaviour across regions.
- **last_review**-Date of last review that is posted for the property.
- **reviews_per_month**- Number of reviews for each property per month
- **calculated_host_listings_count**-Number of listing for each property
- **availability_365**-Number of days each property is available out of 365 days of the year

2. Introduction

Analyzing the hospitality industry's data from the past few years, we can clearly see that vacation rentals are now growing much faster than traditional hotels.Gone are those days when the only type of stay consisted of traditional hotels.

Fast forward to the present, bungalows, private rooms, shared occupancies, and flats are collectively becoming the new face for short-term rental bookings, in addition to traditional hotels. Airbnb is an American company that operates an online market space for homestays and vacation rentals. As the hosts and travelers benefit from the BnB model, the hotel industry considers Airbnb to be a competitive threat.

Nevertheless, for the hosts and travelers to get the best matches corresponding to their respective requirements and to get an edge over competitors in the hospitality market, it is necessary to perform data-driven research and harness insights that will help in increasing the market value and revenue of Airbnb.

3. How does Airbnb Make Money?

Airbnb is a platform business model making money by **charging guests a service fee between 5% and 15% of the reservation**, while the commission from hosts is generally 3%. For instance, on a $100 booking per night set by a host, Airbnb might make as much as $15, split between host and guest fees.

4. Importance of Customer Experience

Customer expectations are higher than ever before, and business success is becoming more and more dependent on a company's ability to meet or exceed those expectations with nearly every interaction. Customer experience is the summary of all the interactions a company has with a customer over time. While understanding how to consider and choreograph those interactions is undeniably complex, at its core, customer experience is about something quite simple: relationships.

5. Steps involved:

- **Exploratory Data Analysis :**

  After loading the dataset we performed this method by first taking a Snapshot of the whole data.This process helped us figuring out which columns have null values and the datatype of each column. It gave us a better idea of which are key columns on the basis of which we can group our data ,and for which columns will need some cleaning or feature engineering.

- **Null values Treatment:**

  Our dataset contains some of the null values in Host information and Reviews Data which might tend to disturb our Plots and Aggregation we performed hence we took their values as 0 for reviews in our project in order to get a better result.

- **Making some categorical columns :**

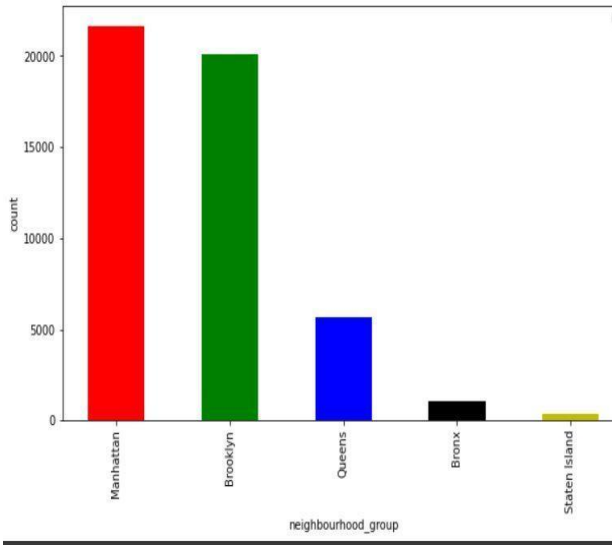
  We have converted the minimum night column to a category based feature on normal understanding of time.
  Reason being ,we wanted to see which region has more number of 1 day booking implying business trips mostly.

- **Understanding Correlation among fields:**

  Here we have plotted the correlation of fields in a heatmap,which then explains that no two fields are highly correlated; only the number of reviews and reviews per month are correlated with each other and that is (corr=0.55) quite reasonable.

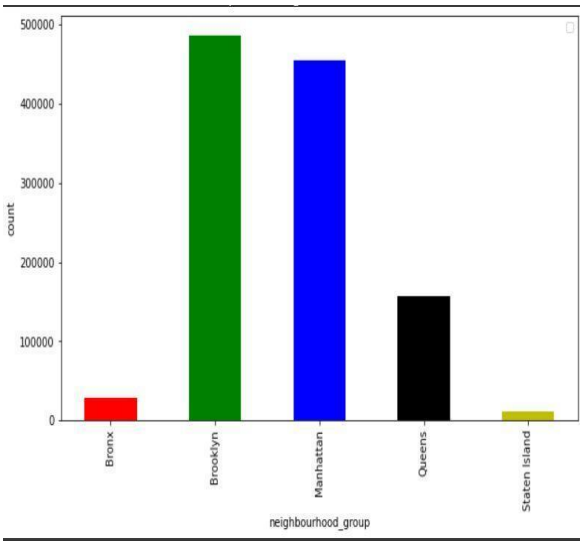
Analysis and Insights-

- **Comparison of number of host in each neighbourhood group:**
  This is a plot between the 5 unique neighborhood groups vs the total number of properties listed corresponding to the neighborhood groups.
  We can infer that Manhattan and Brooklyn have the most listed properties. A reason for this could be that Manhattan and Brooklyn belong to New York's city center whereas Queens, Bronx and Staten Island are away from the city center.



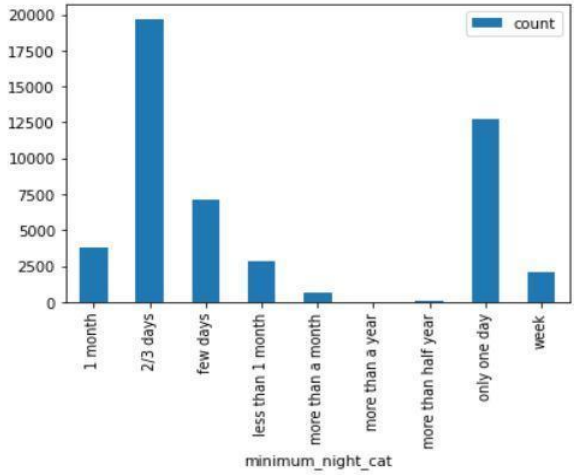
- **Comparison of number of reviews in each neighbourhood group**

More number of reveiws are available for Brooklyn and Manhatan and Very less for Staten Island and bronx



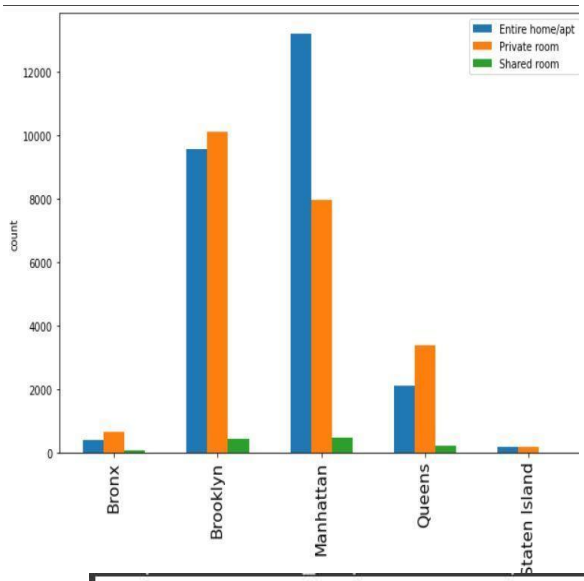
- **Converting minimum night stay to categorical column**

Most of the host provide only one day minimum night stay
we can say that most of the one night stay occurs in manhatten and in brooklyn these trips based on thier minimum night stay can be buisness trips

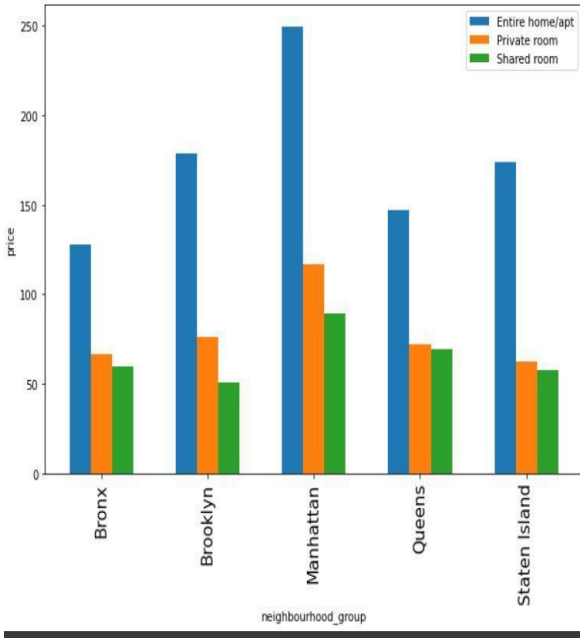


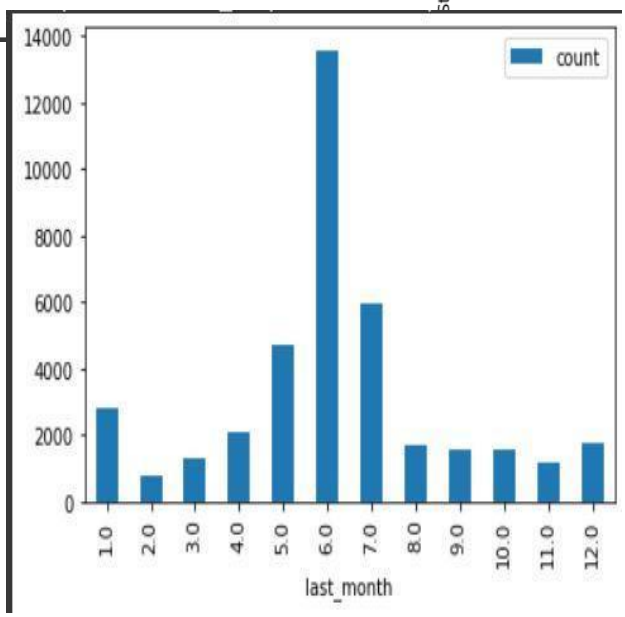
-
- **Analysis of last month of review**

From here we can infer that assuming people give review after leaving the service most people leave/Change the peoperty in the 6 th month of the year



- **Count of number of Properties in each type of room across different groups**





behaviour across each group and for all three type of room.

- **Mean price for each type of room across different neighbourhood groups**

Upon plotting for each neighbourhood group we can see the

- **Identify High Demand areas where AIRBNB should increase supply**

neighbourhood_group	neighbourhood	room_type	availability_365	host_id	price
Manhattan	Morningside Heights	Entire home/a	0	150	130
Manhattan	Marble Hill	Entire home/a	0	5	90
Manhattan	Kips Bay	Private room	0	112	99
Bronx	Fieldston	Private room	0	7	64
Brooklyn	Cobble Hill	Private room	0	22	81
Manhattan	Greenwich Village	Private room	0	99	110
Manhattan	Gramercy	Shared room	0	5	69
Brooklyn	Downtown Brooklyn	Entire home/a	0	52	161

Aggregating the median of available days, count of host_id's and median of price over a group of neighborhood groups, neighborhood and room types, we see that even though total hosts and prices are high, availability is 0 for these are 0 in high demand areas like Manhattan and Brooklyn. From a previous plot and this table, we can infer that customers in these high demand areas prefer to stay in either private rooms or Entire home/apt as compared to shared rooms.
There are two insights we can infer from the above:

There is a gap in demand and supply for room types : private rooms, Entire home/apt in high demand areas. Since the price for such listings is also high, we can introduce some initiative to get more private rooms or entire home type Airbnb's registered in these neighborhoods which will help in increasing the overall revenue.

Since customers in these neighborhood groups prefer to stay in private or entire home type Airbnb's, we can also push for an initiative where some of the shared Airbnb's could be converted to the other two room types to get more crowd.

- **Marketing + Discounts (People can travel extra to stay here)**

Groups	neighbourhood	room_type	availability_365	host_present	price
Brooklyn	East Flatbush	Shared room	365	24	34.5
Queens	South Ozone Park	Shared room	365	8	30
Bronx	Eastchester	Private room	364	6	68
Queens	Ridgewood	Shared room	361.5	10	31.5
Bronx	Eastchester	Entire home/apt	358.5	6	202.5
Brooklyn	Borough Park	Shared room	357	16	35
Staten Island	Randall Manor	Private room	355	9	55
Queens	Corona	Shared room	354	24	27.5
Staten Island	Port Richmond	Private room	353	6	48.5
Bronx	Clason Point	Private room	349	11	70
Queens	Sunnyside	Shared room	343	25	30
Brooklyn	Brighton Beach	Shared room	343	7	50
Queens	Bellerose	Private room	342.5	10	70
Staten Island	West Brighton	Entire home/apt	342	5	99

We can infer that it is expected there will be fewer customers for areas like Bronx and Staten Island, since it is away from the city center.
We can introduce some discounts on listings in these neighborhood groups which might attract a crowd and might help in revenue generation.

8. Conclusions:

Coming to the End!!!

i) Manhattan and Brooklyn are the most crowded suburbs of New York City as compared to Staten Island, Queens and Bronx.
ii) Rate of the listings across all the 5 unique neighborhood groups follows a similar trend for Entire home/apt room type.

iii) Customers prefer to stay in Entire home/apt or private room types as compared to shared room types.

iv) Maximum customers stay in Airbnbs for 2 to 3 days.

v) The highest count of customers is in the months of May, June and July.

vi) Converting shared room type of Airbnbs in Manhattan and Brooklyn to entire home type or private room type will improve revenue generation.

vii) Registering more client Airbnbs of the Entire home/apt and private room types might help in increasing the revenue.

REFERENCES:

- *MachineLearningMastery*
- *GeeksforGeeks*
- *Analytics Vidhya*
- *AirBnb site*