

Predictive Analytics on Airbnb listed host information and approach to Business

Nandita Sharma
Msc in Data Analytics
National college of Ireland
X18185100

Abstract- Airbnb is nowadays in very trend, it is becoming first choice of the individual over the hotel for booking accommodation. It offers leisure apartment, the entire house, private room, shared room to their guest with all amenities such as location, varieties, breakfast, kitchen kinds of stuff so that they even cook by itself all at a reasonable price. In this research paper main aim is to analyse host details, a person who list their own property at the Airbnb platform and help them to set prices and aware them about the area where they can offer good services to their upcoming host. This paper will emphasize listing activity and metrics in detail using predictive analysis of Machine learning.

Keywords—*Bookings, cancellation; machine learning; predictive modelling; Exploratory data analysis; revenue management; prototyping*

I. INTRODUCTION

New York City ranked first choice for international travellers compared to the US in 2010. It was found that 48.7 million total visitors along with 9.7 million international travellers made spending of a total \$31 billion. And this number is kept on increasing for the coming years. Tourist has some expectations while choosing a particular place. They look for better place, different cultures, architecture, events, shopping, and landscape, etc. Overall expenditure becomes very expensive while booking flights, hotel, transportation costs, and food expenses, etc. Nowadays especially hotels prices are increasing tremendously. People prefer a nice restaurant with a good locality at a cheaper rate while booking a hotel online. Thus the website like Airbnb attracts guests by providing good location and excellent amenities in the traditional manner at

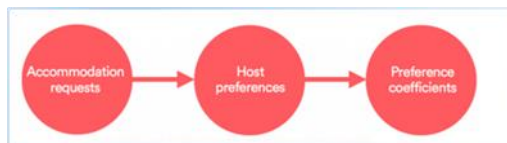
a reasonable cost. Airbnb is a platform that provides people the opportunity to rent or share their occupancy which is also called a shared economy. It offers the best price to both the tenant and the host [1]. Cuomo, a governor of New York state signed a bill into law which announced it illegal to rent an apartment for less than thirty days on Airbnb platform. Later another “Anti-Airbnb Advertising Law” was made which states that if a host is not present it is illegal to rent out the apartment on Airbnb. The result was observed illegal Entire house or apartment number dropped to 5.9 percent from October to number. In order to hide illegal “Entire house/apartment” listing, they have changed its names to “Private room” or “shared room”.

Dataset is taken from Kaggle. The dataset tells about the listing activity and different metrics in New York City for the year 2019. The chosen dataset contains information of host, geographical availability and different important metrics to predict and make conclusions.

These listings generate data in huge amount which can be used to analysed, helps to make a business decision and can be used for security. Data also provides an understanding of guest and hosts’ behavior. Different marketing strategies can be made in a required area along with the implementation of new additional services and much more.

II SCOPE

The scope is to find an optimum match between guests and hosts to continue the growth of Airbnb. Its simple work like a request is generated by the guest on the Airbnb platform and on the basis of host preferences they decided and decide on lending to the guest. The host is the people who list their property with details such as date of availability, cost, check-in and out the timing on Airbnb website. The traveller who planning for trip they book their accommodation by looking at Airbnb search engine, they filter according to their preferences such as price range, location, and amenities, etc. It can be said here Airbnb act as medium between host and guest. Freelancer photographer got hired and paid by Airbnb to click good quality photo of the apartment. Once the booking is made and confirmed by host, the additional charges and commission come into play. This is how Airbnb makes money.[2]



It is also called a revenue model. It has two ways:

- **Commission from Hosts**
10% commission is charged by Airbnb, when is booking done.
- **Transaction fee from Guests**
Addition 3% of booking amount is charged from the traveller once the booking is confirmed.

Airbnb advertises on platforms like social media, advertised through digital marketing and also performed promotional offers to obtain new customers and retain an existing customer. The system also faces come problem such as :

a. Trust issues

Airbnb does a verification process for both host and traveller. So that they can be built their trust among them. It also tells people to sign up with a Facebook account for better transparency. It provides an insurance policy if a case of wrong happens.

b. Guest Retention

They keep giving offers and promotion codes to their travellers to retain them. And it also encourages their host to take a vacation and chose Airbnb to stay by offering good discounts.

III. GOALS OF PROJECT

The main goal of the project is to know more about listing, analysed different geographical availability and make a prediction from raw data.

This project will focus on different listed host and areas in New York City. It will also emphasize who is the busiest host and what parameter involved in keeping them so.

The main aim is to find out learning from prediction considering different parameters such as price, reviews, location, etc.

IV. ETHICAL ISSUE

Some of the factors such as ethical concerns must be taken into consideration while dealing with data. Data science and another related field of data analyzed the data to solve the problem and develop meaning full insights. Its main aim is to benefits businesses and consumers. However, there are chances of ethical concerns such as privacy loss of people while performing data mining [3]. Hence it is very important to understand a few principles of the ethical and legal framework such as:

1.Privacy:

It is one of the major ethical considerations. Privacy is one's personal freedom. Many marketing companies, an online website has the ability to gather information about a customer without knowing them. Then the company uses its information to target potential customers to offer service. But this process causes loss of personal freedom. Hence it is advisable to keep the identity of customer privacy. Private data can be audited based on a legal requirement. Also if data is collected from a person with their permission and exposed to any other business or individual for their own use is considered as

privacy loss. Using the name of an individual is a privacy concern. Hence the names of hosts would not be taken into consideration while performing this research [3].

2. Informed consent:

Customers must have a clear view, how their data is being used by companies or being analyzed by others.

3. Confidentiality:

The set of rules involved in keeping data private. It is another ethical consideration must be taken care of. It clearly states that individuals should not disclose their client information to a third party without the client's permission or clear legal justification. Companies share or sell customer information with their affiliate and partners which is unethical somehow. Its company responsibility to keep data confidential. Sharing name, phone no and other personal details violate the law of confidentiality. Airbnb made data publicly available for the user to analysed it for the project [3]. However, name of the listed host in Airbnb dataset is unethical here. Hence considering ethical consideration "names", phone number of hosts is kept confidential in this research project.

4. Taxes & Legality of Practice: Its deals with legal practice are being followed or not. The government sets some rules and regulations, considering it should be in business it must be obeyed. In many cities, permission must be taken by people whoever are renting the house or apartment. However, the typical Airbnb hosts or Uber drivers are not aware of how to acquire permits or do not feel that they require permission for their own property or assets such as a house, apartment or car. It leads to especially at local and state levels, are perturbed about the dubious legality of Airbnb and Uber transactions, as well as the lack of taxes raised form them. Hence Airbnb also has taken a step to educate local host about legislation draft and comply them with local laws and rules [3].

V. STRATEGY

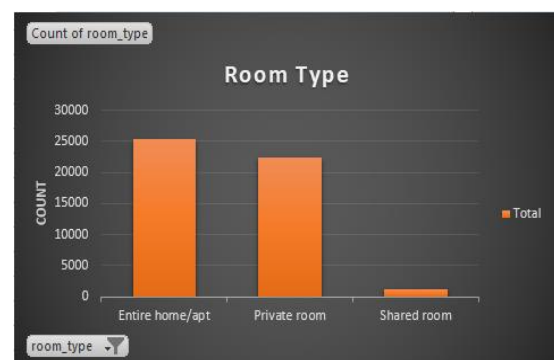
It has the following scope :

- Airbnb provides a great vacation rental service but they don't have much control over the market. It allows future entrepreneurs to build businesses in such a way so that they

should be controlled over market size [4]. Hence Airbnb provides great scope for new entrepreneurs to start with.

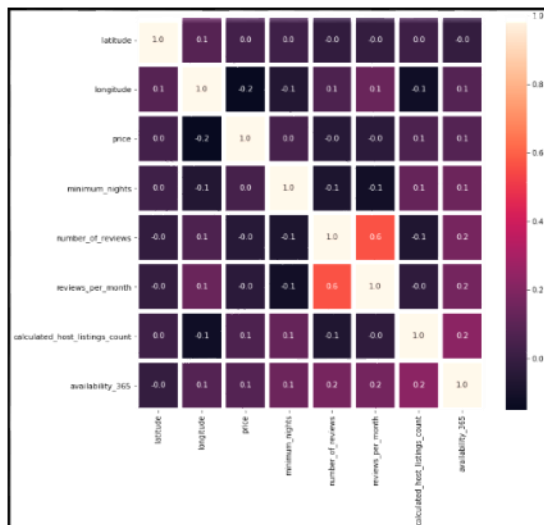
- As there is no monopoly system of the company being followed here, it all based on who provides the best service to customers will stand out [4].
- It would also provide the platform for local people to rent there places for a shorter period and earn some money from it.
- The "sharing economy" brings the exact version of Airbnb by saving money and your time as well.
- In the off-season, demand falls down so reviews also go low, hence, at the time of off-season listing don't get much reviews almost receives zero reviews and leads to unavailable due to misclassification [5]. And there is a limitation in supply measures.
- As some statistics published by Airbnb indicate, Airbnb does not promote significant additional occupancy, but nonetheless, Airbnb may bring benefits to local tourism economies.
- This data has some limitations as the analysis is only limited to one country [5]. Therefore the results of one country will not the same for another country as a country have different social economics characteristics.

VI. PRELIMINARY VISUALISATION



This graph represent total number of room type registered with Airbnb for New York City. The room type are private, shared room, entire home or apartment. Count for shared room is least among others. Overall it can be seen through the graph that host are mostly

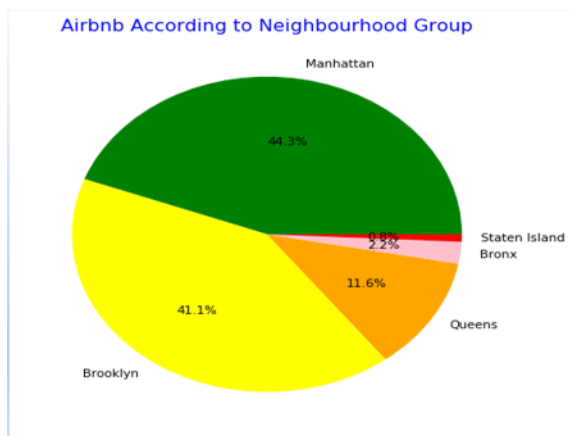
renting Entire home or apartment rather than renting private room or shared room.



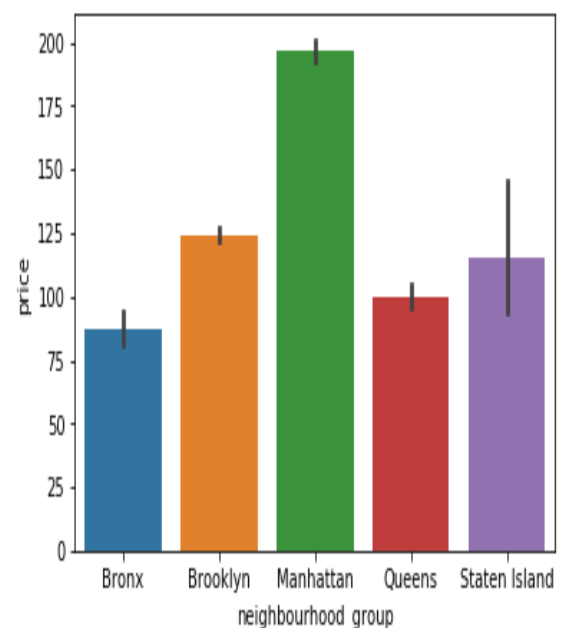
It is correlation graph used to represent correlation between variable in dataset. Its shows how they are related to each other.



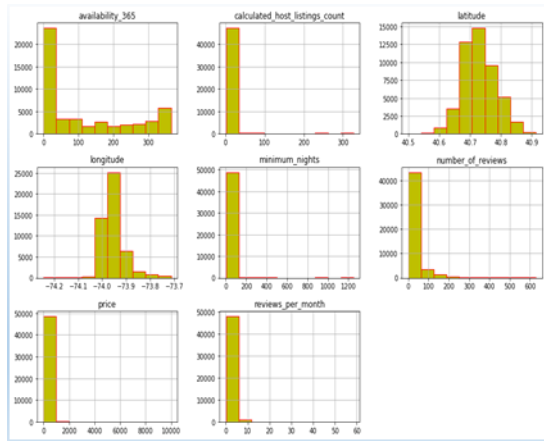
Its is 3D diagram which tell average of mini night spent in neighbourhood group of New York. Manhattan City has highest minimum price for listing with \$8.6 per night as compare to another cities. On another hand Bronx has least price for listing with \$4.5 per night. Brooklyn comes second in listing price after Manhattan.



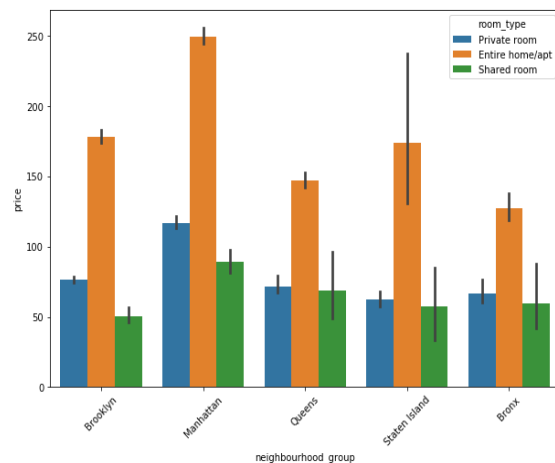
Its is pie chart. It is used to illustrate percentage of city in neighbourhood group of New York City. The percentage is high for Manhattan with 44.3%, however staten Island has 0.8% which is lowest. Staten island can be taken into consideration to setup business by Airbnb.



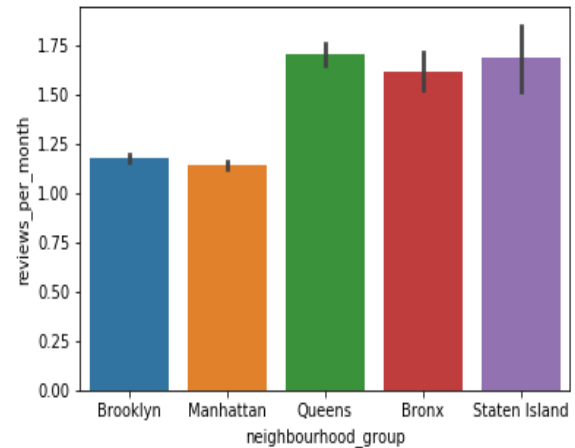
This graph tell price list for neighbourhood group. Manhattan city has high price for room with maximum price of appox 200. Also Staten island is wonderful city comes in south most of New York city. It has maximum price of 125 but sometime it goes to 150 also in high season.



It is distribution of variable points showed through histogram graph. For reviews, minimum night, host and price showed left side distribution. However only latitude point is making correct normal distribution. Its also tell price lie between 0 and 2000 range.



This bar chart represents room type in city of NYC along with price. All room type for listing in Manhattan is highest among others city. Majority of host prefer to list their entire home or complete apartment in all the city. And on second number Brooklyn comes with first room type majority apartment then second private room and shared room.



This histogram represents reviews given by guest per month. Although price for listing in Manhattan is highest but reviews received by Queens is more as compare to others cities.

VII. APPLICABLE TECHNIQUES

Reasonable prices always attract the guest. Yang Li has used a Multi-scale clustering algorithm in his research paper to provide a reasonable price option for the tenant [5]. They consider the first geographic information and then house rental information. Also LRNN model has used to predict the increasing price of rent. Airbnb is widely used in many countries, “Moloud A , Kuan-Hua” have concentrated on the type of sharing economy on Airbnb. Human behavior is being used in his research to analysed a similar pattern of human behavior while making their choice for Airbnb [1]. Through this, they have analyzed the kind of product customers will prefer and the reason behind liking them. Analyzing the behavior of a human in sharing economy (like Airbnb) allows the companies to know what decision they have taken previously while booking, as a result, it would help the new individual to make decisions based on previous one’s behaviour and experiences. Different approaches like RMRate and statistical analysis is used in this paper to detect common behaviors [6].

The opinion mining technique is being used in “Wararat.S” research paper in order to improve the services for the customer [2][6].

In this process extraction of information is preformed from the text processing. This analysis aim was to provide opportunities to develop and improve to company or business. The classifier model is used by researchers to classify positive and negative. A decision tree is used to class label i.e positive or negative. And naïve Bayes is used to generate a probability value rating and it classified correctly 94.37 percent compare to decision tree [2].

VII. REFERENCES

- [1]. M. Abdar, K. Lai and N. Y. Yen, "Crowd Preference Mining and Analysis Based on Regional Characteristics on Airbnb," *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)*, Exeter, 2017, pp. 1-6.doi: 10.1109/CYBCONF.2017.7985771, Available: IEEE Xplore,<https://www.ieee.org/>
- [2]. N. Antonio, A. de Almeida and L. Nunes, "Predicting Hotel Bookings Cancellation with a Machine Learning Classification Model," *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Cancun, 2017, pp. 1049-1054. doi:10.1109/ICMLA.2017.00-11,Available: IEEE Xplore, <https://www.ieee.org/>
- [3]. E. J. Sujdak, "Ethical issues with target marketing on the Internet," *Proceedings International Symposium on Technology and Society*, Stamford, CT, USA, 2001, pp. 105-113. doi:10.1109/ISTAS.2001.937728, Available: IEEE Xplore, <https://www.ieee.org/>
- [4].G. Zervas, D. Proserpio and J. W. Byers,"The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry" First Published October 1, 2017 Research Article, October 1, 2017; Issue published: October 1, 2017 <https://doi.org/10.1509/jmr.15.0204>
- [5].Y. Li, Q. Pan, T. Yang and L. Guo, "Reasonable price recommendation on Airbnb using Multi-Scale clustering," *2016 35th Chinese Control Conference (CCC)*, Chengdu, 2016, pp. 7038-7041.doi:10.1109/ChiCC.2016.7554467,Available: IEEE Xplore, <https://www.ieee.org/>
- [6]. W. Songpan, "The analysis and prediction of customer review rating using opinion mining," *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, London, 2017, pp. 71-77. doi:10.1109/SERA.2017.7965709, Available: IEEE Xplore, <https://www.ieee.org/>
- [7]. Z. Li, L. Liu and C. Li, "Analysis of customer satisfaction from Chinese reviews using opinion mining," *2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, 2015, pp. 95-99.doi: 10.1109/ICSESS.2015.7339013, Available: IEEE Xplore, <https://www.ieee.org/>