**Airbnb Bookings Analysis**

Vaibhavkumar Gupta, Bhavik Verma,

Priyanka Pal, Shayan Somanna, Dilkhush Sharma

Data Science Trainees

**Almabetter, Bangalore**

**Abstract:**

Airbnb is an American enterprise that operates an internet market for lodging, in most cases homestays for holiday rentals, and tourism activities.

We were provided with a data set of Airbnb NYC (2019). Our analytics can be used to make better business decisions, understand customer and host behaviour and performance on the platform, target your marketing initiative, implement innovative add-on services, and more.

1. **Problem Statement**

Since 2008, guests and hosts have used Airbnb to expand travel options and present a more unique and personal way of experiencing the world. Today, Airbnb has become a unique service, used and recognized around the world. Analysing data from millions of Airbnb listings is a critical enabler for the company. These millions of entries create a huge amount of data.

The data we will analyse is from Airbnb NYC (2019). Our main analysis goals are across four propositions that can be summarized as Host Learnings, Areas,

Price, Ratings, Locations, etc. but we are not limited to that, we will also try to explore some more ideas.

1. **Understanding the variables**

* **id**: Unique listing ID
* **name**: Name of the listing
* **host\_id**: Unique 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**: price in dollars
* **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

1. **Introduction**

Airbnb is an internet marketplace that connects folks that need to farm out their homes with people who are trying for accommodations in specific locales.

Airbnb revolutionized the hospitality industry. Before 2008, travellers would likely have booked a hotel or hostel for their trip to another city. Today, many of these people are turning to Airbnb.

Today, Airbnb is a unique service that the entire globe uses and recognises. For the business, data analysis on the millions of listings offered through Airbnb is essential. These millions of listings generate a tonne of data, which can be analysed and used for a variety of purposes, including security, business decisions, understanding customer and provider (host) behaviour and performance on the platform, directing marketing initiatives, putting into practise cutting-edge extra services, and much more.

1. **Python Libraries used**

* Pandas
* Matplotlib
* Seaborn
* Folium
* Klib

1. **Graphs used for data visualization**

* Count Plot
* Bar graph
* Heatmap
* Box Plot
* Maps

1. **How pricing works?**

Airbnb reservation will cost you the nightly rate indicated by the host plus any additional fees or expenses decided by the host or by Airbnb.

1. **Type of Fees**

Airbnb service fee: Airbnb charges a guest service fee that funds round-the-clock community assistance and general smooth operation.

Some hosts charge a cleaning fee to cover the cost of maintaining their place.

Some hosts charge an extra guest fee for each visitor over a predetermined number.

Security deposit: Using Airbnb offline fees function, hosts who manage their listings using software that is connected to the API can set a security deposit.

Value Added Tax (VAT, JCT, and GST) is levied against visitors from certain nations.

Local taxes are assessed according on where the host's property is located.

Final price mostly includes all the type of fees that are mentioned above.

1. **Approach Used**

The approach we have used in this project is defined in the given format-

**1) Loading our data:**  In this section we just loaded our dataset in colab notebook and read the csv file.

**2) Data Cleaning and Processing:** In this section we have tried to remove the null values and for some of the columns we have replaced the null values with the appropriate values with reasonable assumptions.

**3) Analysis and Visualization:** In this section we have tried to explore all variables which can play an important role for the analysis. In the next parts we have tried to explore the effect of one over the other. In the next part we tried to answers our hypothetical questions.

**4) Future scope of Further Analysis:** There are many apartments having availability as 0 and date of last\_review is very old, which can mean that they must have stopped their business, we can find the relation with neighbourhood with these apartments if we could dig much, various micro trends could be unearthed, which we are not able to cover during this short duration efficiently. There are various columns which can play an important role in further analysis such as number of reviews and reviews per month finding its relation with other factors or other grouped factors can play an important role.

1. **Challenges Faced**

* While doing the analysis we found out that 36% of the data has 0 availability in the availability\_365 column, which is an extreme case. But we didn’t have other relevant required data so we couldn’t alter this column.
* Further we found out that there were many listings whose price was 0, which is not normal. So, we filled these values by the respective median price and updated the price column.
* While getting host\_name with highest listings we found out that there are many hosts whose name are same so we went by host\_id as this is unique, host\_name is not unique.
* There are many listings whose date of last review is very old this can mean that they must’ve stopped their business then those listings are of no use to us for doing analysis in present. But this assumption can also be wrong so we didn’t alter this column.
* There were many outliers in price column of some hosts which weren’t benefitting the host as well as the customer.
* The biggest challenge that we faced is finding busiest hosts. If we try to find the busiest hosts by only number of reviews then this may be not the correct metric, because we don’t the current status of the host having highest number of reviews. For example, if we check that one host are x number of reviews which is highest but when we check the date of last\_review and find out that the reviews are very old than the current date, then we can infer that business is currently shutdown so how can we take such hosts into consideration for knowing the busiest hosts. Ideally busiest host should be that one whose occupancy is almost full or full.

1. **Conclusions**

* The host Sonder (NYC) has the most listings on Airbnb in NYC.
* The area around Williamsburg has the most listings.
* The most expensive listings in NYC are in the Upper West Side, Astoria, and Greenpoint neighbourhood.
* The Bedford-Stuyvesant neighbourhood have the most reviews overall and monthly, respectively.
* Manhattan and Brooklyn neighbourhood group have the most listings. The quantity of listings in the Staten Island and Bronx neighbourhood group is relatively very low.
* Private Room or Entire Home/Apartment rentals make up the majority of Airbnb listings in New York City. People who want to stay in an entire home or apartment are likely to stay longer, whereas those who prefer to stay in a private room are probably going to stay for less time than those who prefer to stay in an entire home or apartment.
* The price field has several rows with values of 0, suggesting an error that Airbnb should fix.

Keeping the listing's price high and having no availability doesn't help the host because the customer is willing to pay the amount, but what is the benefit if there are no rooms available even after that?

* The host, Maya, has received the most reviews overall.
* The average cost of every room type in Manhattan is higher than the cost of every room\_type in the other neighbourhood group. The average cost of every room type in the Bronx neighbourhood group is lower than the cost of every other neighbourhood\_group.
* To learn more about the busiest hosts, refer to the answer to Question 12 in the Colab notebook.