**AKSHAD GOYANKA**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Introduction & Problem Statement**

Airbnb, Inc. is an online marketplace for long & short term rental accommodations founded in 2008 by Brian Chesky, Nathan Blecharczyk, and Joe Gebbia.

This Project will include all the processes that are required to perform data analysis - Data Discovery & Understanding, Wrangling , visualisation & statistical analysis when required.

**Analyse the Airbnb NYC 2019 dataset to discover key insights that can help in understanding customer and host behaviour, inform business decisions, and identify trends in the listings.**

EDA or exploratory data analysis - where we will analyse the data using charts & statistics [if necessary]. This can help us find valuable insights that we can use to make decisions that will help our organisation to increase their profits.

**Dataset Discovery**

The dataset contains around 49000 rows & 16 columns

ID - Identity number for each listing

name - name of property

HostID - identity number for host/property owner

Host\_name - name of host/owner

neighbourhood group - a set of neighbourhood clubbed together

neighbourhood - locality in the city

latitude & longitude - geographical coordinate system

room\_type - type & orientation of room

price - rate charged by customer per night

minimum\_nights - minimum number of nights stayed by customer

number of reviews - number of reviews given by customers

last review - date when last review was given

reviews per month - number of reviews given by customer per month averagely

calculated\_host\_listings\_count - total count

availability\_365 -availability around the year

**Steps Involved**

**1] Data Discovery -** this step involves importing the dataset, & gaining some preliminary information. Used functions like describe() , info() to gain insights on the variables. Also found unique values per variable.

2] **Data Cleaning** - this step involves cleaning the data so that it becomes fit for analysis. Includes imputing missing values , removing useless columns. For imputing , we used median [for numerical variables] & ‘null’ [for categorical variables]

3] **Exploratory data analysis** - This step involves creating graphs & statistics to gain insights from the dataset. Various graphs such as bar graphs, pie chart, scatterplots, histograms have been used in this project. The entire process is divided into 2 phases -

3.1] **Univariate Analysis** - Univariate analysis explores each variable in a data set, separately. It was done on neighbourhood group & neighbourhood, Count of room type , Latitude , longitude, minimum nights, availability 365.

3.2] **Bivariate Analysis** - Avg Price per neighbourhood group & neighbourhood, Price vs latitude,Price vs longitude,Avg price per room\_type,Price vs minimum nights, avg Availability per room type, Availability per neighbourhood group [& neighbourhood].

3.3] **Correlation Heatmap -** Correlation heatmap is the best for analysing correlations between all possible variables.Theres not much correlation between these variables , i.e. they are independent.

**3.4]**  **Pair Plot** - Pair plot draws all possible graphs for given variables.

## 

## 

## **Solution to Business Objective**

Most of the listings are in brooklyn or manhattan. If you want to rent out flats , then brooklyn & manhattan are the best places to do it

Entire homes & private rooms are the majority of listings. Seems like private rooms & entire homes are in demand

Majority of listings lie between latitudes - 40.65 - 40.75. If you want to advertise your flat you can do it in these locations

The majority of listings lie in longitude ranges - 74-73.9. this can help us to find where we should advertise our flats to rent

The majority of listings have 0 to 10 min nights, there's also a considerable number of 30 minimum nights. 0 to 10 minimum nights & 30 nights package seems to be the industry standards

The majority of listings are available for 0-100 days per year , mosIf you're starting a rental business , then it's advisable to make it available for weekends.t likely weekends [52 weeks per year i.e. 52 weekends i.e. around 100 days]. There are also a lot of listings that are available almost the entire year.

Staten Island & Manhattan are the most expensive , then Bronx & brooklyn. If you want to set up a high end apartment for rent theses are the places to do so - Staten islands , manhattan

the prices & number of listings seems to be higher between latitudes 40.6 - 40.8. For advertising and setting up your rental business , the above latitude range is the best

Majority of listings lie between longitude -74 to -73.9. The graph shows that the longitude range above has the highest number of renting opportunities & its the best for advertising.

# 

# **Conclusion**

* *Summarising the key insights and their implications for Airbnb: Pricing strategies might need to be neighbourhood-specific.*
* Marketing efforts could focus on promoting less popular room types or areas.
* Hosts with multiple listings contribute significantly to the platform's supply.
* Reviews are crucial for understanding listing popularity and guest satisfaction
* The dataset contains around 49,000 observations and 16 columns, with a mix of categorical and numeric values.
* The columns likely include information such as listing ID, name, host ID, neighbourhood, room type, price, minimum nights, number of reviews, last review date, reviews per month, number of listings per host, availability, and more.
* From the above data analysis process we can deduce that there are certain factors that impact the properties of listings , like geographical locations , availability , minimum nights etc. The above visualisations displays these very factors & how they impacts the listings.