# EDA on Airbnb Booking

**Technical Documentation by Rajvee Sharma**

**Data science trainee at Alma Better**

**Abstract**

*Airbnb is an American Company and was founded in 2008 by two friends Joe Gebbia and Brian Chesky are the founders of Airbnb that allow travelers to rent spaces from host or people who have vacant spaces in their houses. While relatively new, Airbnb’s current worth is $25.5 billion. Over the few years Airbnb have been generated much attention from customers. More than 60 billion individuals have used Airbnb in 34000 cities and in 190 countries. Furthermore, there have been around 2 million listings worldwide.*

**Problem Statement**

*Airbnb is a two-sided marketplace that sought to match people that owned real estate properties with people interested in renting short-term lodging. My main aim is to find out what are the crucial factors that attract our customers to use Airbnb services. For this purpose I would like to analyze the room type, location ,price range, listing counts etc and find out how they are inter related .*

## Main Goal

*The main goal of analysis is to provide insights about data generates from Airbnb. Airbnb has strong competitive position in rapid growth industry and to achieve market growth rate with an overall 80% market share from existing market. I want to deep dive in data for the factors of influence on Airbnb booking. This project will helps you to find out important and crucial insights which create positive impact on listing count on Airbnb and its growth.*

# Introduction

*When booking travel accomodation, there are many sources used. As I analysed and reached to conclusion that people tend to look for accommodation that that have good security in the area that are clean and safe. For analysis of mentioned data I have used and my business case is to find best insights from given data that are being used in future growth of company. This dataset has around 49000 observations and 16 columns in it :*

|  |  |
| --- | --- |
| **First column** | **Second column** |
| Host id | It is id number of provider |
| Host name | Name of that particular host |
| Number of reviews | Reviews given by customer to that host |
| Availability 365 | Availability of that particular room in a year |
| Room type | Type of room it is |
| Price | Price of that particular type of room |
| Host listing count | Calculate the total listing in certain period of time |

# Steps involved:

# Raw tables-

*It has information such as host id , host name, number of reviews, availability in a year, room types, price and many more. And the main and the important is the number of listings because listings generates a lot of data that can be analysed and used for security, business decisions, understanding of customers and hosts behaviour and performance on the platform, guiding marketing initiatives and implementation of innovative additional services by company.*

**Data preparation and cleaning-**

*Data preparation is the process of cleaning and transforming raw data prior to processing and analysis. It is an important step prior to processing and often involves reformatting data, making corrections to data and combining of datasets to enrich the data. Data cleaning is the process of detecting and correcting (or removing) or incorrect data from record set, table, datasets and refers to identifying incomplete, inaccurate or irrelevant parts of datasets and replacing, modifying, or deleting the dirty data from it.*

# Null value treatment-

*This dataset contain large number of null values which might tend to disturb the accuracy hence I dropped them in the beginning of my project so that I can get better result.*

# Removing duplicates-

*I don’t want to count certain entries more than once when I analyse the data so I need to remove duplicate entries from my datasets.*

# Feature standardisation-

*My main motive through this step was to scale our data into a uniform format that would allow us to utilize the data in better way. The basic goal was to enforce the level of consistency or uniformity to certain practices within selected environment.*

# Exploratory Analysis and Data Visualisation-

*In statistics, exploratory data analysis is an approach to analysing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the format modelling or hypothesis testing task. Data visualisation is the graphic representation of data. It involves producing images that communicate relationships among the represented data to viewrs of the images. This communication is achieved through the use of systematic mapping between graphic marks and data values in creation of visualisation. This mapping establishes how data values will be represented visually, determining how and to what extent the property of size and colour, will change to reflect changes in value of datum*.

# Conclusion-

*The dataset contain immense possibilities to improve business value and have a positive impact. Many other possibilities can be explored using this Airbnb dataset. This dataset deliver insights to understand customer demand better and thus help providers (hosts) to provide better services. Dataset can also provide information that how many number of listings generated per room type. And which room type and average price preferred by customers more.*

# Reference-

* *Towards data science*
* *Geeks for geek*
* *Kagle.com*
* *Other sources used*