

# Airbnb (EDA) Capstone Project

**Abhishek Kumar**  
**Data science trainee,**  
**AlmaBetter, Bengaluru**

## Abstract:

Abstract overview of the notebook Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. Today, Airbnb became a one-of-a-kind service that is used and recognized by the whole world. Data analysis on millions of listings provided through Airbnb is a crucial factor for the company. 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.

## Introduction:

Airbnb is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in that locale. It currently covers more than 81,000 cities and 220 countries worldwide.

For this project we are analyzing Airbnb's New York City (NYC) data of 2008. Right from choosing the quarters to offering support around the clock until we part, it brings all sorts of convenient measures right into our hands. **Airbnb** is one such wonderful platform that serves as an online marketplace to help people find accommodation during their trip as per their needs. Airbnb serves over **220 countries and regions** across the world today.

Data analysis on thousands of listings provided through Airbnb is a crucial factor for the company. Our main objective is to find out the key metrics that influence the listing of properties on the platform. For this, we will explore and visualize the dataset from Airbnb in NYC using basic exploratory data analysis (EDA) techniques. We have found out the distribution of every Airbnb listing based on their location, including their price range, room type, listing name, and other related factors. We have analyzed this dataset from different angles and have come up with interesting insights. This can help in making strategic data-driven decisions by the marketing team, finance team and technical team of Airbnb

## Analysis of Data Quality

We had to perform a few imputations and transformations on our dataset for us to create the desired visualizations. There were no major inconsistencies or mismatches in the data, but most of the columns/features we were interested in did

not contain data in the required format and hence were manipulated in a way that their meanings are retained.

## **Problems:**

**There are major steps to getting started on Airbnb:**

- **What can we learn about different hosts and areas?**
- **What can we learn from predictions? (Ex: locations, prices, reviews, etc)**
- **Which hosts are the busiest and why?**
- **Is there any noticeable difference of traffic among different areas and what could be the reason for it?**

**How do prices of listings vary by location? What localities in NYC are rated highly by guests?**

Manhattan has the most expensive rentals compared to the other boroughs. Prices are higher for rentals closer to city hotspots. Rentals that are rated highly on the location by the host also have higher prices. There are a few outliers in Bronx, Staten Island and Brooklyn that defy the above hypothesis.

**How does the demand for Airbnb rentals fluctuate across the year and over years?**

The demand (assuming that it can be inferred from the number of reviews) shows a seasonal pattern - demand increases from January to October, then drops slightly in November and December. In general, the demand for Airbnb listings has been steadily increasing over the years.

**Are the demand and prices of the rentals correlated?**

Average prices of the rentals increase across the year, which correlates with demand. However, the prices show a spike in December as opposed to slowing demand in this month, which is counterintuitive. Prices are higher on average on Fridays and Saturdays, compared to the other days of the week.

**What are the different types of properties in NYC? Do they vary by neighborhood?**

There are more than 20 different types of listings in NYC. The ratio of the type of listings to total numbers varies by borough. Staten Island and Brooklyn tend to have property types that are larger and can accommodate more people.

### **What makes a host a Super host?**

Ratings and Response rates tend to have a direct correlation with a host being 'promoted' to the status of the Super host. However, there are other factors too that makes someone a super host as not all hosts with high ratings and response rates were superb hosts.

### **Conclusion:**

Manhattan is the most focused place in New York for hosts to do their business  
Customers pay the highest amount in Brooklyn, Queens and Manhattan that is 10,000 and lowest amount is 10\$.

For the three types of room type (i.e., Entire home, Shared room, & Private room) average price of the entire home is around \$157, for shared room is around 60, and for private room is around 75.

'Entire home/apt' room type has the highest number of listings of 52% and 'Shared Room' is the least listed room type at only 2.4% in total.

People stay for a longer duration of time in Private rooms in Brooklyn and Manhattan. Words such as 'bedroom', 'cozy', 'private', 'apartment' and 'spacious' are used more frequently than words such as 'park', 'near', 'village' and 'heart'. Count of listings by top 10 hosts is almost 2.5% (1270 listings) of the whole dataset. Top three host base on their turnover are Sonder (NYC), Red awning, Henry and best host is Sonder (NYC)

More customer preferred Manhattan location for night stay then Brooklyn \*63.2% of customers spend the night in Entire home and 1.6% spend the night in Shared room.