

# Capstone Project Submission

## EDA on Airbnb NYC Dataset

### Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

### **Team Member's Name, Email and Contribution:**

Name – Akshat Mishra

E-mail – [Mishra.akshat05@gmail.com](mailto:Mishra.akshat05@gmail.com)

As I am doing this project individually so all the steps involved in this project are done by me.

### **Please paste the GitHub Repo link.**

GitHub Link: - <https://github.com/Link/to/Repo>

**Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**

### **EDA on Airbnb NYC Dataset**

#### **About Airbnb Business -**

Airbnb is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in specific locales. Airbnb offers people an easy, relatively stress-free way to earn some income from their property.

#### **Airbnb Dataset -**

This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values.

#### **Problem Statements –**

- What are the different room types in the whole dataset.?
- Different Room types on Neighborhood Groups
- Top 10 neighborhoods having maximum listings and their room types.
- Which room type is the most expensive.? And what is the average price of each room category.?
- What is the average price of rooms of all the neighborhood groups.?
- Name of different neighborhood groups and number of listings associated with them.?
- Hosts with maximum number of listings.
- Which are top neighborhoods with respect to average price/day?

## Approach to solve –

I followed a stepwise approach to get the desired information using this data set.

**Step 1** – Arranged every important thing to start like – importing libraries, reading the data etc.

**Step 2** – Getting basic info like number of rows and columns.

**Step 3** – Finding null values to check if it will impact the analysis or not and cleaning the data according to the need.

**Step 4** – Started to explore stepwise on each and every column to find the answer of all the problem statements.

## Conclusion –

On the basis of exploration and analysis, following points can be considered

### **"Neighborhood group Manhattan": Most preferred location is Manhattan"**

Why?

- In the whole Airbnb data set, most preferred room type is "Entire home/ apt". And the highest number of this room category is the found in Manhattan. Which means people have more option for their desired room category in Manhattan that's why they prefer Manhattan.
- Average room price on the basis of neighborhood group, is highest in Manhattan. But if look at the bigger picture and try to figure out the most expensive rooms on the basis of neighborhood, we found that the most expensive rooms are in "Fort Wadsworth" which is in Staten Island and that is least preferred place.

### **"Neighborhood Group Staten Island": Least preferred place is "Staten Island"**

Why?

- Most preferred room category is "Entire home/apt" and Staten Island has the least number of rooms of this category.
- The average per day price on the basis of neighborhoods, in Staten Island there are very expensive neighborhoods.

Rest of the findings are given in the notebook and presentation.