

# Capstone Project Submission

## Instructions:

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

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

**Team Name:** The Scouts

**Team Member:** Ishan Barway, Jasmine Sulekha Nag, Nitin M Sakhare, Roshan Alte

### Contribution Role:

- **Ishan Barway :** ([barwayishan@gmail.com](mailto:barwayishan@gmail.com))
  - Data Wrangling.
  - What can we learn about different hosts and areas?
  - Visualizing the hosts with most listings and plot graphs.
  - What can we learn from predictions? (ex: locations, prices, reviews, etc.).
  - Average pricing based on apartment category in different neighbourhood.
  - Conclusion.
- **Jasmine Sulekha Nag :** ([sulekhajasmine@gmail.com](mailto:sulekhajasmine@gmail.com))
  - Data Wrangling.
  - Area map Analysis (plot the scatter plot of map).
  - Number of reviews given by customers to the neighborhoods hotels.
  - Overall price exploration.
  - Conclusion.
- **Nitin M Sakhare :** ([nmsakhare@gmail.com](mailto:nmsakhare@gmail.com))
  - Data Wrangling.
  - Is there any noticeable difference of traffic among different areas and what could be the reason for it?
  - Room type based on neighborhoods groups.
  - Conclusion.
- **Roshan Alte :** ([roshanalte4151@gmail.com](mailto:roshanalte4151@gmail.com))
  - Data Wrangling.
  - How many properties are available for more than 100 days?
  - Getting the values where availabilities of room are more than 100 days.
  - Conclusion.

### Please paste the GitHub Repo link.

GitHub Link:- <https://github.com/ishanbarway/Airbnb-Bookings-Analysis>

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

**Airbnb, Inc.** (Air\_Bed\_and\_Breakfast.com) was founded in 2008 in San Francisco, California. It operates as an online broker for hotel booking, home & apartment for rentals, based on requirements and charges a commission from each booking.

This EDA will to analyze and investigate given data sets and summarize their main characteristics, often employing data visualization methods to find the best possible answers. This dataset has approx. 49000 observations in each 16 features. To proceed with the analysis, firstly we have to mount the Google Drive. In order to access the Google Drive. Then we create the path to the CSV file. The next step is to import Python libraries such as NumPy, Pandas, matplotlib, seaborn. Data wrangling should be done as this data set is really messy. I had to clean up all the null values that is present in the dataset by replacing them with some values to get the proper result.

Then we start to explore the data based on the questions we got to perform in this EDA project. Once the datasets are cleaned and free of error, they can then be analyzed. A variety of techniques can be applied such as exploratory data analysis- understanding the messages contained within the obtained data and descriptive statistics finding the average, median, etc.

Then I start to check host percentage in the area using group by method to get the access off neighbourhood group and the host count, by sorting them descending. Using matplotlib I created a graph of this data. Similarly, for active locality in all the neighbourhood group I did the same and got the graph based on that data. For the second question I again use group by method to get access to room type, neighbourhood group, price. For this data this time I choose line graph to explain it in better way. in the 4th position I added image off New York City for better explanation of the location of the property.

For the conclusion, I checked every observation I got and mix them together to get the short and detailed conclusion possible. Here are the conclusions:

- Manhattan is the most expensive Neighbourhood, with less land area but has highest number of Hotels as well as highest customer count.
- Staten Island has larger land area then Manhattan but still has less hotels and customer count.
- Most of the people prefer Entire home/apt or private rooms, very few prefer shared rooms.
- Brooklyn and Manhattan has the most hotels with high review count but Hillside Hotel, Queens has the highest review count in all