Team Member's Name, Email and Contribution:

1. Vinayak Marathe: vinmarathe100@live.com

2. Riya Patel: riyapatelrp8308@gmail.com

3. Muskan Kasere: muskankasere.mk@gmail.com

Contributor's Role :

1. **Vinayak Marathe :**
    - Worked on Problem Statement with Business Objective
    - Datasets Reading and Mapping the NYC datasets using Folium Library
    - Data Wrangling :
        - ❖ Dataset Cleaning and Dropping
        - ❖ Handling Null Values
        - ❖ Finding Top Host, Top Neighbourhood and Most Demanded Room Type
    - Worked on Some Data Visualizations (Pie Charts, Violin Plots, Distribution Plots) and their Insights
    - Worked on Competitive Advantages and Business Goals of Airbnb.
    - Presentation Preparation

2. **Riya Patel :**
    - Worked on Variables and Statistical Descriptive Analysis
    - Deep Dive into the Relationships of Variables and Understanding the Variables
    - Data Wrangling :
        - ❖ Visualizing Missing Values
        - ❖ Analysing the Expensive and Cheapest Listings of the NYC
        - ❖ Analysing Less Reviewed Listings by the Customers
        - ❖ Analysing the Average Price of the Listings
    - Visualizing the Data and Different Types of Variables using Bar Plots, Categorical Plots, Box Plots, WordCloud Visualization and Correlation Heatmaps.
    - Worked on Different Insights of the Every Visualization of the Dataset
    - Worked on Technical Documentation of the Project

3. **Muskan Kasere :**
    - Worked on Unique Values and Data Explanation
    - Data Wrangling :
        - ❖ Analysed the Availability of the Listings throughout the year
        - ❖ Analysed the Demanded Locality/Neighbourhood group with Maximum Number of Bookings
        - ❖ Analysed reviews, available properties and average price.
    - Data Visualization using Histograms, Scatter Plots, Point Plots and Cat Plot
    - Worked with latitude and longitude to visualize distribution.
    - Analysed Overall Results of the Project /Conclusion
    - Worked on Summary of the Project

GitHub Repo link.

1. Vinayak Marathe: https://github.com/v1git12
2. Riya Patel: https://github.com/riyapatelrp
3. Muskan Kasere: https://github.com/MuskanKasere

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 is a brief identify for "**Air Bed and Breakfast**", is a online hospitality company focused on short time length homestays or rental platform. On Airbnb, users can book a nearby place to stay according to their convenience in more than 34000+ cities in some 200+ countries. Airbnb used to start in 2008 by Brian Chesky and Joe Gebbia, specifically based in San Francisco California. Airbnb platform is reachable via mobile app and website.

In our project, there are 16 columns which have hosts, host_name, neighbourhood group, neighbourhood, room type etc. and 48895 rows which have all the information about Airbnb's hosts. These datasets also have some null values which are handled by us. We can also see the overall details of NYC using the map.

As the first step, we are performing data wrangling over the raw data. In this we have two things: "Know Your Data", you can just check the shape, size and types of the data & "Understanding Your Variables", where you can go deep dive into the relationships and nature of the variable. Further we deeply analyse data and then visualize the data, Storytelling & Experimenting with charts to understand relationships between variables.

In the **Data Wrangling process**, we have analysed different use cases based on locality, reviews, room type etc, from which we can clearly infer that Manhattan Neighbourhood group has the most expensive and luxurious entire home/apartments with maximum number of reviews. Rather than it, most of the number of bookings took place in Williamsburg Location. We have also analysed the top 10 earners in terms of hosts and average price per day of the listing with respect to neighbourhood.

In the **Data Analysis session**, we perform code to get insights of your data like top hosts with most listings, which property and room type is cheapest, average price of listing in Airbnb, properties with less reviews, maximum number of booking, etc.

Next session is **Data Visualization**, in this we show different types of charts so that anyone can easily understand the relationship between variables and get the best insights. In this we have used the Pie Chart, Dist. and Point Plot on Dependent Variable, Word Cloud for Univariate then Bar Plot, Correlation Heatmap and Histogram, Violin Plot, Boxplot, Scatter Plot and Cat Plot for Bivariate variables.

From all these charts, we get some information like 44% of customers prefer Manhattan, 50% customers prefer entire rooms and many more.

So the conclusion is that, it has been found that Most of the Bookings took place for the "Williamsburg" of around "27%" followed by "Bedford-Stuyvesant", "Harlem" which has "25%" & "18%" respectively. Additionally, we also find-out the Top Earners (Host), relationship between neighbourhood group and Prices, Price comparison in terms of Room Type, Preference of Guests with respect to Room Type. Furthermore, we have also analysed Maximum Number of Bookings, Customer Reviews and many more.