DATA SCIENCE CAPSTONE PROJECT

*Sakshi Chahande*

**Requirements of Project Report**

1. Introduction where you discuss the business problem and who would be interested in this project.

2. Data section where you describe the data that will be used to solve the problem and the source of the data.

3. Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

4. Results section where you discuss the results.

5. Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

6. Conclusion section where you conclude the report.

**Introduction**

I am a student pursuing my B.Tech in Computer Engineering in Mumbai, India. My idea for this project is to identify the most visited places in a particular city within India, and compare them, in order to determine which city is more suitable to live in. I believe this project is useful for people looking to move to a new city to expand their career, for furthering their education or to improve their lifestyle.

**Interested Audience:**

The idea behind this project is to provide a streamlined analysis for students and working people alike, to determine what which area, in which city would be most suited for their needs if they have to live there.

**Data Selection**

I am comparing India's two largest cities, Mumbai and Delhi.

I have obtained the data-set from the following link:

https://data.gov.in/resources/all-india-pincode-directory-contact-details-along-latitude-and-longitude/api

This dataset contains the postcodes of the post offices of each locality in India. I have extracted two data frames from this dataset and applied the GeoPy API to find and integrate the latitude and longitude for each area. This has helped me create the map using Folium API. I further plan to find the different venues in a particular area and compare the two using the FourSquare API.