**CAPSTONE PROJECT: BATTLE OF NEIGHBOURHOODS**

**INTRODUCTION/BUSINESS PROBLEM:**

Food, irrespective of the place, is loved by people of all races and religion. It is something considerably basic needed for survival and hence starting a business-like restaurant or investing in one which satisfies that need can be lucrative if done correct. So before opening a restaurant there are few factors to be considered like the location, the cuisine type, the locality, and the local competition, permits, expenses like construction, salary etc. Few things that can be solved using data science are the location, the cuisine type and in which borough to start. Ideally, for a restaurant to be a successful one, it must grab a lot of people’s attention, entice them to enter and assuming that the service is good eventually it would become a successful one. Hence the location is the crux of the business, for the sake of the capstone project, I have selected a US state which is famous for the prestigious Universities located here. I have selected the Massachusetts state, where millions of students from almost all-around the world come here and pursue higher studies and jobs. Since it is a global hub for students, we can expect people from different countries having various preferences in the cuisine type to be living there. The goal of the project will be to identify a place suitable for the restaurant, the type of restaurant and the cuisine type.

**DATA:**

1. A csv file containing the zip, zip code name, city, state, county name of Massachusetts can be downloaded from the following link.

<https://www.downloadexcelfiles.com/us_en/download-list-us-zip-codes-massachusetts-state#.XukY8WgzZPY>

2. For the venue details we can use the foursquare api

<https://foursquare.com/developers/apps>

3. For the latitude and longitude data we use a library name pygeocode,

Input: a list of pin code (5 digits)

Output: A data frame with the following columns: - postal code, country code, place name, state name, accuracy, county code, latitude and longitude, community name and community code

After few processing we have a perfect final version of the data frame with postal codes, place name, state name, country code, county code and their respective latitude and longitude