**Peer-graded Assignment: Restaurant Capstone (Week 1)**

# DATA SECTION

For this project we need the following data:

a) Toronto Neighbourhood data that contains Borough, Neighbourhoods along with their latitudes and longitudes

1. Data Source: Geospatial Co-ordinates CSV File
2. Description: This data set contains the required information. And we will use this data set to explore various neighbourhoods of Toronto.

b) Toronto Postal Code Data that contains Postal codes for neighbourhood present withing Toronto City.

1. Data Source: [*https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M*](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
2. Description: By using this API we will get all the Postal Code information for all the neighbourhood within Toronto City.

c) Indian restaurants Toronto city.

1. Data Source: Foursquare API
2. Description: By using this API we will get all the venues in the Toronto. We can filter these venues to get only Indian restaurants.

# METHODOLOGY

1. We will first use, the data **Beautiful Soup Package** to scrape the Postal Code information of Toronto neighbourhood from the Wikipedia page, and store the same information within **Pandas** dataframe.
2. Then data cleansing will be done to remove unassigned values in Borough and Neighbourhood.
3. We will then import the contents of Geospatial Co-ordinates csv file into the dataframe. We will then use the two data frames created, to create one single Dataframe.
4. Foursquare API will help to get the venues within all the neighbourhoods of Toronto city.
5. Filter out the Venues of all Indian Restaurant present in Toronto city.
6. We will then collect the likes, rating , tips etc for each of the Indian restaurant and then sort them according to their values, again through Four square API .
7. The comparison data and the ranking will be plotted on Bar Chart graphs using matplotlib library, Seaborn library.
8. We will Visualize the data of top Neighbourhood based on the venue categories in a Toronto City map using *Folium Package*.
9. Finally, we will cluster these top Neighbourhood based on the venue categories and use K-Means clustering Machine learning technique to build a model using Elbow point method.