**IBM DATA SCIENCE PROJECT**

**Done By: Zain**

**INTRODUCTION: BUSINESS PROBLEM**

The city of Toronto is packed with restaurants, nightlife, and different ethnicities and types of cuisines. An entrepreneur wants to open a Middle Eastern Restaurant in Toronto but doesn’t know which neighborhood to open it in. The things that need to be taken into consideration are population of every neighborhood, the best neighborhoods where the habitants would likely eat in this type of cuisine, and where the competition is limited (such as the number of Middle Eastern restaurants in each neighborhood). As Middle Eastern food is popular among the Arab community, so this entrepreneur might think of open the restaurant in an area where the Arab community resides.

The objective of this project is about finding the best neighborhood in the city of Toronto to open a Middle Eastern Restaurant. By using data science methods and tools along with machine learning algorithms such as clustering, this project will aim to provide a solution to answer the business question: Where should the entrepreneur consider opening a Middle Eastern restaurant in Toronto?

**TARGET AUDIENCE:**

Entrepreneurs or chefs who want to open a Middle Eastern restaurant in Toronto would find this project very interesting and informative.

**DATA:**

To solve this problem, we will need the following data:

1. List of Neighborhoods in Toronto, Canada using the FoursquareAPI.
2. Latitudes and Longitudes of these neighborhoods.
3. Venue data related to Middle Eastern Restaurants. This will help us find neighborhoods that are more suitable to open the restaurant in.
4. Look into the demographic data of a specific area. For example: areas with the majority of Arabs people would be good for opening a Middle Easter Restaurant.

**Extracting Data:**

* The scrapping of Toronto Neighborhoods via Wikipedia:

<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

* Getting Latitude and Longitude Data of these neighborhoods via Geocoder package.
* Using Foursquare API to get venue data related to these neighborhoods.

**Demographic data from the City of Toronto's open data:**

<https://www.toronto.ca/ext/open_data/catalog/data_set_files/2016_neighbourhood_profiles.csv>

We will use the features Ethnic Origin and Neighborhood Information (from the above data file) for each neighborhood, in order to cluster the neighborhoods of Toronto.

Examples of Data from the Dataset:

Neighborhood information data:



We can see that:

* We have the **name of each neighbourhood** in each column name (starting at position 6)
* The **neighbourhood number** (also called CDN number) in the first row (starting at position 6)

Ethnic Origin Data:



We can see that:

* We have the **name of each neighbourhood** in each column name (starting at position 6)
* We have the **name of each ethnic origin** in the Characteristic column
* The **number of people living in each neighbourhood**, associated to each ethnic origin name.

**List of venues by neighbourhood using the FoursquareAPI**

In order to obtain the list of venues, and especially the list of middle eastern restaurants same as the one we want to open, we are going to request FoursquareAPI with an Explore query.  
The documentation for the Explore query can be found here :

* <https://developer.foursquare.com/docs/api/venues/explore>

We query FoursquareAPI suplying the neighbourhood's information (coordinates calculated with the **Geocoder** package), the radius of scan(500m radius), and the limit of number of venues we want to retrieve(limit of 100 venues).