**A Recommender System for Groceries Contractor**

**Part 1: Identifying Neighborhoods inside "Scarborough"**

We will use Postal Codes of different regions inside Scarborough to find the list of neighborhoods. We will essentially obtain our information from <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M> and then process the table inside this site. Images from dataframes and also from maps will be provided in the presentation. Here we only present our strategy and how we got the mission accomplished.

## Part 2: Connecting to Foursquare and Retrieving Locational Data for Each Venue in Every Neighborhood

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 1000 meter. It means that we have asked Foursquare to find venues that are at most 1000 meter far from the center of the neighborhood. (I think distance is measured by latitude and longitude of venues and neighborhoods, and it is not the walking distance for venues.)

## Part 3: Processing the Retrieved Data and Creating a DataFrome for All the Venues inside the Scarborough

When the data is completely gathered, we will perform processing on that raw data to find our desirable features for each venue. Our main feature is the category of that venue. After this stage, the column "Venue's Category" wil be One-hot encoded and different venues will have different feature-columns. After On-hot encoding we will integrate all restaurant columns to one column "Total Restaurants" and all food joint columns to "Total Joints" column. We assumed that different resaturants use the Same raw groceries. This assumption is made for simplicity and due to not having a very detailed dataset about different venues.

Now, the dataset is fully ready to be used for machine learning (and statistical analysis) purposes.

## Part 4: Applying one of Machine Learning Techniques (K-Means Clustering)

Here we cluster neighborhoods via K-means clustering method. We think that 5 clusters is enough and can cover the complexity of our problem. After clustering we will update our dataset and create a column representing the group for each neighborhood.

# Decision Making and Reporting Results

Now, we focus on the centers of clusters and compare them for their "Total Restaurants" and their "Total Joints". The group which its center has the highest "Total Sum" will be our best recommendation to the contractor. {Note: Total Sum = Total Restaurants + Total Joints + Other Venues.} This algorithm although is pretty straightforward yet is strongly powerful.