**A Recommender System for Groceries Contractor**

1. **Business Problem**

* Taking the example of a groceries contractor in one of the boroughs of Toronto (Scarborough)
* The contractor provides many diffenet places such as: Different types of Coffee shops, Restaurants, Brewery and Café, Bakery with fresh and high quality ingredients
* He intends to construct a warehouse to store all the ingredients he buys from farmers in Scarborough, such that he can provide ingredients to even more customers
* However, it is difficult to determine where should the contractor set up his warehouse at because if the warehouse is located near to those famous restaurants, then the ingredients can be delivered to the restaurant early and in time before the restaurant opens in the morning. Hence allowing the contractor will gain more reputation as a reliable contractor and possibly gain more customers in the future and thus improving his earnings
* On the other hand, if the contractor set up his warehouse in nearer the farmers, he may not be able to deliver the ingredients to the restaurants as early
* In addition to the above example, the question of which neighbourhood in Scarborough is the be a best location for the contractor to set up the warehouse in should be considered as well
* Hence, finding the right neighbourhood for the contractor to set up his warehouse is the main objective of this project
* This is done my coming up with a recommender system which will produce a sorted list of neighbourhoods in which the first element of the list will be the best suggested neighbourhood, allowing the contractor to determine the best location to set up his warehouse

1. **Data We Need**
2. We will need geo-locational information about Scarborough and the neighbourhoods in that Scarborough. Thus, the latitude and longitude numbers of Scarborough are required for locating it on the map. This will be provided by the contractor. The Postal Codes that fall into Scarborough are required as well. The Postal Codes will then by used to find the neighbourhoods in Scarborough.
3. We will need data about different venues in different neighbourhoods of Scarborough. To obtain these information, we will use "Foursquare" locational information. Locational information for each venue means basic and advanced information about that particular venue. Basic information includes the precise latitude and longitude and the distance of a particular venue from the centre of the neighbourhood. Advanced information includes the category of that venue, whether this venue is a popular one and maybe the average price of the services of this venue.

A typical request from Foursquare will provide us with the following information:

[Postal Code] [Neighborhood(s)] [Neighborhood Latitude] [Neighborhood Longitude] [Venue] [Venue Summary] [Venue Category] [Distance (meter)] [M1L] [Clairlea, Golden Mile, Oakridge] [43.711112] [-79.284577] [Tim Hortons] [This spot is popular] [Coffee Shop] [592]

**3. Body**

**3.1: Identifying Neighbourhoods inside "Scarborough"**

**- We will use Postal Codes of different regions in Scarborough to find the list of neighbourhoods. We will obtain our information from** [**https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M**](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)**and then process the table in this site.**

**- Images from the data frames and the maps will be depicted in the presentation. In this report, I will only focus and present my strategy and how I manged to recommend the best location for the contractor the set up his warehouse**

**3.2: Connecting to Foursquare and Retrieving Locational Data**

**- After finding the list of neighbourhoods, we connect to the Foursquare API to gather information about venues inside each neighbourhood.**

**- For each neighbourhood, radius is set to 1000 meter 🡪 This means that Foursquare will to find venues that are at most 1000 meter away from the centre of the neighbourhood is selected in Foursquare**

**3.3: Processing the Retrieved Data and Creating a Data Frame for all the Venues inside the Scarborough**

**- When all the data is retrieved, the raw data will be processed to according to our desirable features for each venue**

**The main feature that is considered is the category of the venue because it is the key feature that determines whether the location of warehouse is successful.**

**Then, the column "Venue's Category" will be One-hot encoded, resulting in different venues having different feature-columns.**

**- Next, all the restaurant columns will be combined to form a single column "Total Restaurants"**

**All the food joint columns will also be combined to form a single column "Total Joints".**

**- The assumption that different restaurants use the same ingredients will be taken.**

**Such an assumption is made as detailed information about different venues is not accessible for me.**

**- Now, the dataset is ready for machine learning and statistical analysis uses**

**4. Applying K-Means Clustering**

**- K-means Clustering will be used to cluster neighbourhoods**

**5 clusters are used as it is enough to determine the best location for the contractor the set up his warehouse due to the fact due the size of** Scarborough

**- After clustering, I will update the dataset and create a column representing the group for each neighbourhood**

**5. Decision Making and Reporting Results**

**- I will focus on the centres of the clusters and compare them using the "Total Restaurants" and "Total Joints" columns.**

**- The group which its centre has the highest "Total Sum" will be our best recommendation to the contractor.**

**(Note: Total Sum = Total Restaurants + Total Joints + Other Venues)**

**6. Results**

**Based on this analysis, the best neighbourhood for the contractor to set up his warehouse would be ‘Agincourt’ with Postal Code ‘M1S’, Neighbourhood Latitude ‘43.7942003’ and Neighbourhood Longitude**

**‘-79.26202940000002’**

By - TJQ