

### Brief Introduction (Cover Page)

## Part 1: Problem Description

There is a groceries contractor in one of the boroughs of Toronto (Scarborough)

For example, if the warehouse is close to those old and famous restaurants, then the vegetables and other groceries would be delivered to the restaurant in the right time and there would be no delay so the restaurant cooks can start their job from the morning and the Quality of Service will be high and this contractor will gain more reputation and income.

The contractor should build this warehouse where it is closest to its customers.

## Part 2: Data We Need

1- We will need geo-locational information about that specific borough and the neighbourhoods in that borough. We specifically and technically mean the latitude and longitude numbers of that borough. We assume that it is "Scarborough" in Toronto. This is easily provided for us by the contractor, because the contractor has already made up his mind about the borough. The Postal Codes that fall into that borough (Scarborough) would also be sufficient for us. In fact we will first find neighbourhoods inside Scarborough by their corresponding Postal Codes.

2- We will need data about different venues in different neighbourhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. By locational information for each venue we mean basic and advanced information about that venue. For example there is a venue in one of the neighbourhoods. As basic information, we can obtain its precise latitude and longitude and also its distance from the centre of the neighbourhood. But we are looking for advanced information such as the category of that venue and whether this venue is a popular one in its category or maybe the average price of the services of this venue. A typical request from Foursquare will provide us with the following information:

[Postal Code] [Neighbourhood(s)] [Neighbourhood Latitude] [Neighbourhood 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]

Type *Markdown* and LaTeX:  $\alpha 2 \alpha 2$

Some Notes about "Foursquare": <https://foursquare.com/>

*Foursquare is a local search-and-discovery service mobile app which provides search results for its users (Wikipedia).*

*Founded: New York City, New York, U.S*

*Users: 60 million*

*Date launched: March 11, 2009*

*Employees: Over 200*

*Founders: Dennis Crowley, Naveen Selvadurai*

*Owner: Foursquare Labs, Inc.*

## Main Article

### Part 1: Identifying Neighbourhoods inside "Scarborough"

We will use Postal Codes of different regions inside Scarborough to find the list of neighbourhoods. We will essentially 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 inside this site. Images from data frames 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 Neighbourhood

After finding the list of neighbourhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighbourhood. For each neighbourhood, 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 centre of the neighbourhood. (I think distance is measured by latitude and longitude of venues and neighbourhoods, and it is not the walking distance for venues.)

### Part 3: Processing the Retrieved Data and Creating a DataFrame 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" will be One-hot encoded and different venues will have different feature-columns. After One-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 restaurants 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 neighbourhood 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 neighbourhood.

### **Decision Making and Reporting Results**

Now, we focus on the centres of clusters and compare them for their "Total Restaurants" and their "Total Joints". 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.} This algorithm although is pretty straightforward yet is strongly powerful.

#### **Results:**

Based on this analysis, the best recommended neighbourhood will be:

{'Neighbourhood': 'Agincourt',

'Postal Code': 'M1S',

'Neighbourhood Latitude': 43.7942003,

'Neighbourhood Longitude': -79.26202940000002}

**REPORT MADE BY : AMAN SHRIVASTAVA**

**E-MAIL : aman.shrivastava1497@gmail.com**

**Thank you!**

