# Recommender System for
# Supermarket Dealer to build Warehouse
### (Data Science Capstone–IBM Professional Certification)

## Anandhavalli Muniasamy
## May 15, 2019

## 1. Introduction

Recommender systems are ending up being an essential business tool in e-commerce, as more and more companies are implementing this function into their website. Recommender systems usually are kinds of collective filtering that include predictive designs, heuristic search, data collection, user interaction and design upkeep. The system generally has to be upgraded regularly with recently added ratings, products and users. In shorts a recommender system is an information filtering technology created to figure out choices that are most likely to the customer's tastes. After the best items have been determined they are recommended to the user. Recommender systems connect with users on their choices and form a profile of each customer typically based upon scores of items. The various profiles are compared with each other with aid of an algorithm and are utilized to quote and forecast the items that are most likely to the user's tastes. Simply put recommender systems are a kind of heuristic search that uses gathered and stored information of users and or choices to predicts and suggest what items users will like.

### 1.1 Background

This project mainly focusses on the supermarket dealer in one of the boroughs of Toronto (Scarborough). This dealer provides places such as: Different types of Restaurants, Bakery, Breakfast Spot, Brewery and Café with fresh and high-quality supermarket products. The dealer wants to build a warehouse for the products which are buy from villagers and farmers inside the borough, so that they will support more customers and also bring better "Quality of Service" to the old customers.

### 1.2 Problem

If the warehouse is close to those old and famous restaurants, then the vegetables and other products 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 problem here is how to find the best place for building the warehouse neighbourhood? This project aims to help the dealers to find the nearest place of the warehouse to its customers in order to minimize the cost of transportation, which neighbourhood (particularly Scarborough) would be a better choice for the dealer to build the warehouse in that neighbourhood.

## 1.3 Interest

Obviously, supermarket dealers would be very interested in accurate finding the suitable place to build their warehouse in borough, for competitive advantage and business values. Others who are interested in this recommender system are customers.

# 2. Data acquisition and cleaning

## 2.1 Data Source & Description

Data that might contribute to building recommender system for supermarket dealer include :
- Data set 1: Postal Codes of different regions inside Scarborough to find the list of neighborhoods. The dataset will be consider from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
- Data set 2: The data about different venues in different neighborhoods of that specific borough will be collected from "Foursquare" locational information (https://foursquare.com/). Foursquare is a local search-and-discovery service mobile app which provides search results for its users (Wikipedia). 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)]*

## 2.2 Data cleaning

Data pre-processing starts with fetching data from the internet url (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) and extracting the raw table inside that webpage.

In the Neighborhood column the information is ending with \n this needs to be removed. All the rows in the Borough column containing Not assigned must be deleted. If the Neighborhood is Not assigned then the Borough is assigned as the neighbourhood itself.

Each Postal Code can contain multiple neighborhoods, so combining all the neighborhoods into single line separated by comma ( ,). So there is no duplicity in the Postal Code i.e there is only one occurrence of each postal code in the dataset.

Converting the received html content into a soup object from "bs4" package for easier extraction of necessary information. Finding the table containing the postal codes and neighborhood of Toronto.

By iterating through all the rows of the table, the content can be accessed and stored as a list. Pre-processing ends with a creation of a DataFrame using the list containing Postal Codes, Borough and Neighborhood of Toronto.

## 2.3 Feature selection

Geo-locational information about that specific borough and the neighbourhoods in that borough which comprises the latitude and longitude numbers of that borough. In this project,, it is assumed that it is "Scarborough" in Toronto. This is easily provided for us by the dealer because the dealer has already made up his mind about the borough. The Postal Codes that fall into that borough (Scarborough) would also be sufficient for this analysis. In fact, we will first find neighborhoods inside Scarborough by their corresponding Postal Codes.
Data about different venues in different neighborhoods of that specific borough. In order to gain that information, "Foursquare" locational information will be useful. By locational information for each venue, it means basic and advanced information about that venue. As basic information, we can obtain its precise latitude and longitude and also its distance from the centre of the neighborhood. 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

# 3 Exploratory Data Analysis

## 3.1 Methodology - Work flow

HTTP requests would be made to this Foursquare API server using zip codes of the Seattle city neighborhoods to pull the location information (Latitude and Longitude).

Foursquare API search feature is enabled to collect the nearby places of the neighborhoods. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 700.

Folium- Python visualization library is used to visualize the neighborhoods cluster distribution of Seattle city over an interactive leaflet map.

Extensive comparative analysis of two randomly picked neighborhoods is carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.

Unsupervised machine learning algorithm K-mean clustering is applied to form the clusters of different categories of places residing in and around the neighborhoods. These clusters from each of those two chosen neighborhoods would be analyzed individually collectively and comparatively to derive the conclusions.

### 3.2 Identifying Neighborhoods inside "Scarborough"

Postal Codes of different regions inside Scarborough to find the list of neighborhoods are identified from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and then retrieved data in the table from this site for processing
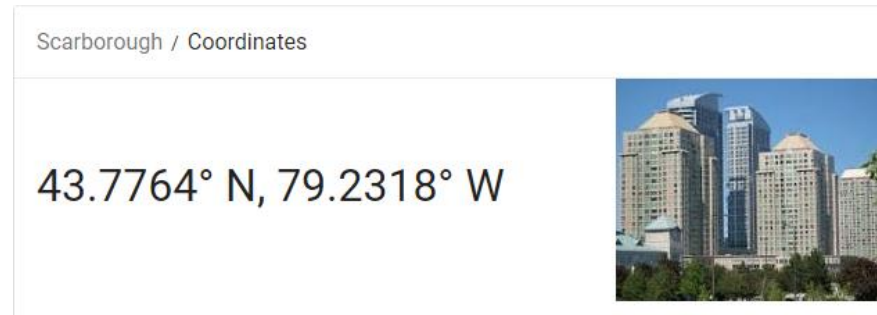


image source: google.com

### 3.2 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. Here the distance is measured by latitude and longitude of venues and neighborhoods, and it is not the walking distance for venues.

### 3.3 Processing the Retrieved Data and Creating a Data Frame for All the Venues inside the Scarborough

When the data is completely gathered, we 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" is One-hot encoded and different venues will have different feature-columns. After On-hot encoding we 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.

### 3.4 Applying one of Machine Learning Techniques (K-Means Clustering)

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

# 4. Results & Discussion

- **Venues** in for each neighborhood in Scarborough are listed below:

Steeles West, Scarborough Village, Woburn, Highland Creek, Rouge Hill, Port Union, Birch Cliff, Maryvale, Wexford, Agincourt North, Milliken, Cedarbrae, Tam O'Shanter, Cliffcrest, Cliffside, Morningside, West Hill, Rouge, Malvern, Agincourt, Ionview, Kennedy Park, Dorset Park, Scarborough Town Centre, Wexford Heights, Upper Rouge, Clairlea, Golden Mile, Oakridge.

- Identifying Postal Codes (and then Neighborhoods) in "Scarborough"

scarborough_data

|    | Postcode | Borough | Neighbourhood | Latitude | Longitude |
|----|----------|---------|---------------|----------|-----------|
| 0  | M1W | Scarborough | Steeles West | 43.799525 | -79.318389 |
| 1  | M1J | Scarborough | Scarborough Village | 43.744734 | -79.239476 |
| 2  | M1G | Scarborough | Woburn | 43.770992 | -79.216917 |
| 3  | M1C | Scarborough | Highland Creek, Rouge Hill, Port Union | 43.784535 | -79.160497 |
| 4  | M1N | Scarborough | Birch Cliff | 43.692657 | -79.264848 |
| 5  | M1R | Scarborough | Maryvale, Wexford | 43.750072 | -79.295849 |
| 6  | M1V | Scarborough | Agincourt North, Milliken | 43.815252 | -79.284577 |
| 7  | M1H | Scarborough | Cedarbrae | 43.773136 | -79.239476 |
| 8  | M1T | Scarborough | Tam O'Shanter | 43.781638 | -79.304302 |
| 9  | M1M | Scarborough | Cliffcrest, Cliffside | 43.716316 | -79.239476 |
| 10 | M1E | Scarborough | Morningside, West Hill | 43.763573 | -79.188711 |
| 11 | M1B | Scarborough | Rouge, Malvern | 43.806686 | -79.194353 |
| 12 | M1S | Scarborough | Agincourt | 43.794200 | -79.262029 |
| 13 | M1K | Scarborough | Ionview, Kennedy Park | 43.727929 | -79.262029 |
| 14 | M1P | Scarborough | Dorset Park, Scarborough Town Centre, Wexford ... | 43.757410 | -79.273304 |
| 15 | M1X | Scarborough | Upper Rouge | 43.836125 | -79.205636 |
| 16 | M1L | Scarborough | Clairlea, Golden Mile, Oakridge | 43.711112 | -79.284577 |

- Identifying Postal Codes (and then Neighborhoods) in "Scarborough"

The geograpical coordinate of "Scarborough" are: 43.773077, -79.257774.



- Categories are nearly 56 unique categories in Scarborough.

```
scarborough_venues.groupby('Neighborhood').count()
```

| Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|
| Agincourt | 4 | 4 | 4 | 4 | 4 | 4 |
| Agincourt North, Milliken | 2 | 2 | 2 | 2 | 2 | 2 |
| Birch Cliff | 4 | 4 | 4 | 4 | 4 | 4 |
| Cedarbrae | 7 | 7 | 7 | 7 | 7 | 7 |
| Clairlea, Golden Mile, Oakridge | 10 | 10 | 10 | 10 | 10 | 10 |
| Cliffcrest, Cliffside | 2 | 2 | 2 | 2 | 2 | 2 |
| Dorset Park, Scarborough Town Centre, Wexford Heights | 8 | 8 | 8 | 8 | 8 | 8 |

- Grouping the neighbourhood is based on one-hot encoding method.

```
scarb_onehot.head()
```

| | Neighborhood | American Restaurant | Athletics & Sports | Auto Garage | BBQ Joint | Bakery | Bank | Bar | Breakfast Spot | Brewery | Bus Line | Bus Station | Café | Caribbean Restaurant | Chinese Restaurant | Coffee Shop | Colle Stadiu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Steeles West | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | |
| 1 | Steeles West | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | |
| 2 | Steeles West | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 3 | Steeles West | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 4 | Steeles West | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |

```
scarb_grouped.head(7)
```

| | Neighborhood | American Restaurant | Athletics & Sports | Auto Garage | BBQ Joint | Bakery | Bank | Bar | Breakfast Spot | Brewery | Bus Line | Bus Station | Café | Caribbean Restaurant | Chinese Restaurant | Cc S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Agincourt | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.000000 | 0.000000 | 0.0 | 0.25 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.000000 | |
| 1 | Agincourt North, Milliken | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.000000 | 0.000000 | 0.0 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.000000 | |
| 2 | Birch Cliff | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.000000 | 0.000000 | 0.0 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.25 | 0.000000 | 0.000000 | |
| 3 | Cedarbrae | 0.000000 | 0.142857 | 0.0 | 0.0 | 0.142857 | 0.142857 | 0.0 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.142857 | 0.000000 | |
| 4 | Clairlea, Golden Mile, Oakridge | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.111111 | 0.000000 | 0.0 | 0.00 | 0.000000 | 0.222222 | 0.111111 | 0.00 | 0.000000 | 0.000000 | |
| 5 | Cliffcrest, Cliffside | 0.333333 | 0.000000 | 0.0 | 0.0 | 0.000000 | 0.000000 | 0.0 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.000000 | |
| 6 | Dorset Park, Scarborough Town Centre, Wexford ... | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.000000 | 0.000000 | 0.0 | 0.00 | 0.111111 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.111111 | |

- Also, find the top 10 venues per neighborhood.

```
neighborhoods_venues_sorted
```

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Agincourt | Skating Rink | Breakfast Spot | Lounge | Clothing Store | Vietnamese Restaurant | Coffee Shop | Grocery Store | General Entertainment | Fried Chicken Joint | Fast Food Restaurant |
| 1 | Agincourt North, Milliken | Park | Playground | Chinese Restaurant | Grocery Store | General Entertainment | Fried Chicken Joint | Fast Food Restaurant | Electronics Store | Discount Store | Department Store |
| 2 | Birch Cliff | General Entertainment | Skating Rink | Café | College Stadium | Vietnamese Restaurant | Clothing Store | Grocery Store | Fried Chicken Joint | Fast Food Restaurant | Electronics Store |
| 3 | Cedarbrae | Hakka Restaurant | Athletics & Sports | Thai Restaurant | Bakery | Bank | Fried Chicken Joint | Caribbean Restaurant | College Stadium | Grocery Store | General Entertainment |
| 4 | Clairlea, Golden Mile, Oakridge | Bakery | Bus Line | Ice Cream Shop | Bus Station | Park | Soccer Field | Intersection | Metro Station | Construction & Landscaping | Cosmetics Shop |

- Scarborough Data frame covers the following labels & data

```
scarb_merged
```

| | Postcode | Borough | Neighbourhood | Latitude | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Mc Comm Ven |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | M5R | Central Toronto | The Annex, Yorkville | 43.672710 | -79.405678 | 1 | Café | Coffee Shop | Sandwich Place | Pizza Place | Jewish Restaurant | Indian Restaurant | F |
| 1 | M4Y | Downtown Toronto | Church and Wellesley | 43.665860 | -79.383160 | 1 | Japanese Restaurant | Coffee Shop | Gay Bar | Sushi Restaurant | Restaurant | Gastropub | Nightcl |
| 2 | M5V | Downtown Toronto | CN Tower, King and Spadina, Railway Lands, Sou... | 43.628947 | -79.394420 | 1 | Airport Service | Airport Lounge | Airport Terminal | Harbor / Marina | Boutique | Airport | Airp Fc Co |
| 3 | M6K | West Toronto | Exhibition Place, Parkdale Village | 43.636847 | -79.428191 | 1 | Coffee Shop | Café | Breakfast Spot | Bar | Grocery Store | Furniture / Home Store | Ital Restaur |

- **Applying one of Machine Learning Techniques (K-Means Clustering)**

The complete dataset for applying k-means clustering analysis purpose are as follows:

scarborough_onehot

| Neighborhood | Bakery | Breakfast Spot | Diner | Fish Market | Food & Drink Shop | Fruit & Vegetable Store | Grocery Store | Noodle House | Pizza Place | Sandwich Place | Total Restaurants | Total Joints |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Agincourt | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 21 | 0 |
| Agincourt North, Milliken | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 13 | 2 |
| Birch Cliff | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |
| Cedarbrae | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 7 | 3 |
| Clairlea, Golden Mile, Oakridge | 2 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Cliffcrest, Cliffside | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 1 | 1 |
| Dorset Park, Scarborough Town Centre, Wexford Heights | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 14 | 3 |
| Highland Creek, Rouge Hill, Port Union | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Ionview, Kennedy Park | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 1 | 6 | 1 |
| Maryvale, Wexford | 0 | 1 | 0 | 1 | 0 | 0 | 3 | 0 | 3 | 0 | 8 | 1 |
| Morningside, West Hill | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 3 | 3 |
| Rouge, Malvern | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 7 | 0 |
| Scarborough Village | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 3 | 0 |
| Steeles West | 2 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 7 | 1 |
| Tam O'Shanter | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 13 | 1 |
| Woburn | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 |

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 dealer. {Note: Total Sum = Total Restaurants + Total Joints + Other Venues.} This algorithm although is pretty straightforward yet is strongly powerful.

| | Bakery | Breakfast Spot | Diner | Fish Market | Food & Drink Shop | Fruit & Vegetable Store | Grocery Store | Noodle House | Pizza Place | Sandwich Place | Total Restaurants | Total Joints | Total Sum |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G5 | 2.000000 | 1.000000 | 0.000000 | 0.0 | 0.000000 | 0.00 | 0.000000 | 1.000000 | 1.000000 | 2.000000 | 21.000000 | 0.000000 | 28.000000 |
| G1 | 1.333333 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0.00 | 0.333333 | 0.666667 | 1.666667 | 1.000000 | 13.333333 | 2.000000 | 20.333333 |
| G4 | 0.000000 | 1.000000 | 0.000000 | 1.0 | 0.000000 | 0.00 | 3.000000 | 0.000000 | 3.000000 | 0.000000 | 8.000000 | 1.000000 | 17.000000 |
| G3 | 1.500000 | 0.250000 | 0.000000 | 0.0 | 0.000000 | 0.25 | 1.000000 | 0.000000 | 0.750000 | 0.750000 | 6.750000 | 1.250000 | 12.500000 |
| G2 | 0.285714 | 0.142857 | 0.285714 | 0.0 | 0.142857 | 0.00 | 0.142857 | 0.000000 | 0.857143 | 0.428571 | 2.000000 | 0.714286 | 5.000000 |

Visualize the clusters after applying k-means algorithm:

The geograpical coordinate of "Scarborough" are: 43.773077, -79.257774.
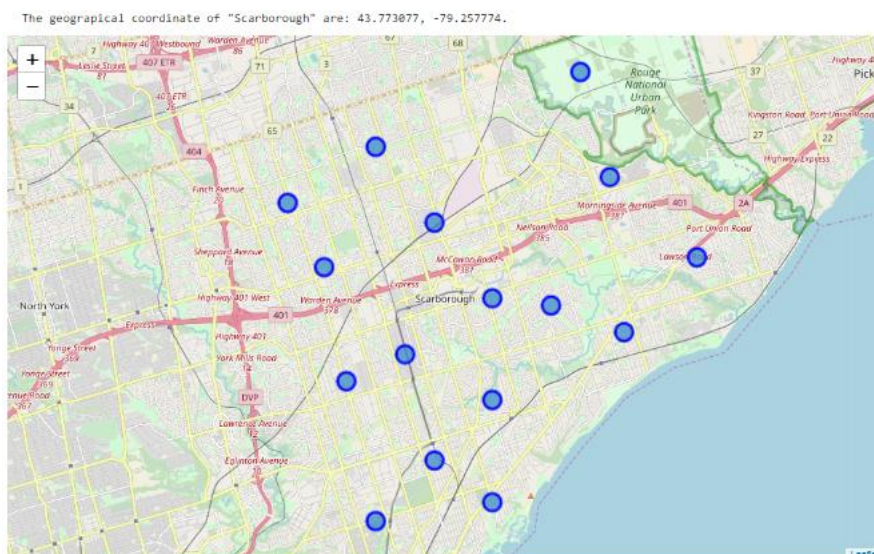
Examine the five clusters created for finding the best neighbourhood.

### Examine each of the five clusters

```
scarb_merged.loc[scarb_merged['Cluster Labels'] == 0, scarb_merged.columns[[1] + list(range(5, scarb_merged.shape[1]))]]
```

| | Postcode | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Co |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | M1W | -79.318389 | 0 | Chinese Restaurant | Coffee Shop | Fast Food Restaurant | Cosmetics Shop | Pharmacy | Pizza Place | Breakfast Spot | Grocery Store | Japanese Restaurant | Elec |
| 2 | M1G | -79.216917 | 0 | Coffee Shop | Korean Restaurant | Vietnamese Restaurant | Health & Beauty Service | Grocery Store | General Entertainment | Fried Chicken Joint | Fast Food Restaurant | Electronics Store | Di |
| 3 | M1C | -79.160497 | 0 | Bar | Vietnamese Restaurant | Coffee Shop | Hakka Restaurant | Grocery Store | General Entertainment | Fried Chicken Joint | Fast Food Restaurant | Electronics Store | Di |
| 4 | M1N | -79.264848 | 0 | General Entertainment | Skating Rink | Café | College Stadium | Vietnamese Restaurant | Clothing Store | Grocery Store | Fried Chicken Joint | Fast Food Restaurant | Elec |
| 5 | M1R | -79.295849 | 0 | Sandwich Place | Auto Garage | Smoke Shop | Shopping Mall | Breakfast Spot | Discount Store | College Stadium | Construction & Landscaping | Cosmetics Shop | Depa |

```
scarb_merged.loc[scarb_merged['Cluster Labels'] == 1, scarb_merged.columns[[1] + list(range(5, scarb_merged.shape[1]))]]
```

| | Postcode | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | M1X | -79.205636 | 1 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | Na |

```
scarb_merged.loc[scarb_merged['Cluster Labels'] == 2, scarb_merged.columns[[1] + list(range(5, scarb_merged.shape[1]))]]
```

| | Postcode | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | M1B | -79.194353 | 2 | Fast Food Restaurant | Vietnamese Restaurant | Train Station | Hakka Restaurant | Grocery Store | General Entertainment | Fried Chicken Joint | Electronics Store | Discount Store | Departmen Stor |

```
scarb_merged.loc[scarb_merged['Cluster Labels'] == 3, scarb_merged.columns[[1] + list(range(5, scarb_merged.shape[1]))]]
```

| | Postcode | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | M1H | -79.239476 | 3 | Hakka Restaurant | Athletics & Sports | Thai Restaurant | Bakery | Bank | Fried Chicken Joint | Caribbean Restaurant | College Stadium | Grocery Store | Genera Entertainme |

```
scarb_merged.loc[scarb_merged['Cluster Labels'] == 4, scarb_merged.columns[[1] + list(range(5, scarb_merged.shape[1]))]]
```

| | Postcode | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | M1J | -79.239476 | 4 | Playground | Construction & Landscaping | Health & Beauty Service | Vietnamese Restaurant | Clothing Store | Grocery Store | General Entertainment | Fried Chicken Joint | Fast Food Restaurant | Electronic Stor |
| 12 | M1S | -79.262029 | 4 | Skating Rink | Breakfast Spot | Lounge | Clothing Store | Vietnamese Restaurant | Coffee Shop | Grocery Store | General Entertainment | Fried Chicken Joint | Fast Foo Restaurar |

Decision Making and Reporting Results

| | Neighborhood | Group |
|---|---|---|
| 0 | Agincourt | 5 |
| 1 | Agincourt North, Milliken | 1 |
| 2 | Birch Cliff | 2 |
| 3 | Cedarbrae | 3 |
| 4 | Clairlea, Golden Mile, Oakridge | 2 |
| 5 | Cliffcrest, Cliffside | 2 |
| 6 | Dorset Park, Scarborough Town Centre, Wexford ... | 1 |
| 7 | Highland Creek, Rouge Hill, Port Union | 2 |
| 8 | Ionview, Kennedy Park | 3 |
| 9 | Maryvale, Wexford | 4 |
| 10 | Morningside, West Hill | 2 |
| 11 | Rouge, Malvern | 3 |
| 12 | Scarborough Village | 2 |
| 13 | Steeles West | 3 |
| 14 | Tam O'Shanter | 1 |
| 15 | Woburn | 2 |

- Best Neighbourhood is G5 group.

```
neigh_summary[neigh_summary['Group'] == 5]
```

| | Neighborhood | Group |
|---|---|---|
| 0 | Agincourt | 5 |

```
name_of_neigh = list(neigh_summary[neigh_summary['Group'] == 5]['Neighborhood'])[0]
scarborough_venues[scarborough_venues['Neighborhood'] == name_of_neigh].iloc[0,1:5].to_dict()
```

```
{'Postal Code': 'M1S',
 'Neighborhood': 'Agincourt',
 'Neighborhood Latitude': 43.7942003,
 'Neighborhood Longitude': -79.26202940000002}
```

# 5. Conclusions

In this study, the analysis can be concluded as the best group is G5. Second best is G1 and third best group is G4. These models can be very useful in helping supermarket dealer to find best neighbourhood in Scarborough.

# 6. Future directions

Models in this study mainly focused on k-means clustering method. In future I am interested to apply other clustering methods to carry out comparative analysis to find the best model for this neighbourhood analysis problem.