

# A RECOMMENDER SYSTEM FOR GROCERIES BUSINESS



# Introduction

- **Part 1: Problem Description**

**There is a groceries contractor in one of the boroughs of Toronto (Downtown Toronto). This contractor provides places such as: Different types of Restaurants, Bakery, Breakfast Spot, and Café with fresh and high-quality groceries. The contractor wants to build a grocery markets it buys 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.**



# Introduction

## **Part 2: Data We Need**

**1. We will need geo-locational information about that specific borough and the neighborhoods in that borough. We assume it is "Downtown Toronto" in Toronto. This is easily provided for us by the contractor, because the contractor has already made up his mind about the borough.**

**2. We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. A typical request from Foursquare will provide us with the following information:**



# Main Part

## Part 1: Identifying Postal Codes (and then Neighborhoods) in “Downtown Toronto”

	Postcode	Borough	Neighborhood	Postal Code	Latitude	Longitude
0	M4W	Downtown Toronto	Rosedale	M4W	43.679563	-79.377529
1	M4X	Downtown Toronto	Cabbagetown, St. James Town	M4X	43.667967	-79.367675
2	M4Y	Downtown Toronto	Church and Wellesley	M4Y	43.665860	-79.383160
3	M5A	Downtown Toronto	Harbourfront, Regent Park	M5A	43.654260	-79.360636
4	M5B	Downtown Toronto	Ryerson, Garden District	M5B	43.657162	-79.378937
5	M5C	Downtown Toronto	St. James Town	M5C	43.651494	-79.375418
6	M5E	Downtown Toronto	Berczy Park	M5E	43.644771	-79.373306
7	M5G	Downtown Toronto	Central Bay Street	M5G	43.657952	-79.387383
8	M5H	Downtown Toronto	Adelaide, King, Richmond	M5H	43.650571	-79.384568
9	M5J	Downtown Toronto	Harbourfront East, Toronto Islands, Union Station	M5J	43.640816	-79.381752
10	M5K	Downtown Toronto	Design Exchange, Toronto Dominion Centre	M5K	43.647177	-79.381576
11	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	M5L	43.648198	-79.379817
12	M5S	Downtown Toronto	Harbord, University of Toronto	M5S	43.662696	-79.400049
13	M5T	Downtown Toronto	Chinatown, Grange Park, Kensington Market	M5T	43.653206	-79.400049
14	M5V	Downtown Toronto	CN Tower, Bathurst Quay, Island airport, Harbo...	M5V	43.628947	-79.394420
15	M5W	Downtown Toronto	Stn A PO Boxes 25 The Esplanade	M5W	43.646435	-79.374846
16	M5X	Downtown Toronto	First Canadian Place, Underground city	M5X	43.648429	-79.382280
17	M6G	Downtown Toronto	Christie	M6G	43.669542	-79.422564



# Main Part

## Part 1: Identifying Postal Codes (and then Neighborhoods) in “Downtown Toronto”





# Main Part

## **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.**



# Main Part

## **Part 3: Processing the Retrieved Data and Creating a DataFrame for All the Venues inside the Downtown Toronto**

**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.**



# Main Part

## Part 3: Processing the Retrieved Data and Creating a DataFrame for All the Venues inside the Downtown Toronto

	Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Venue Category	Distance
0	M4W	Rosedale	43.679563	-79.377529	Summerhill Market	This spot is popular	Grocery Store	764
1	M4W	Rosedale	43.679563	-79.377529	Black Camel	This spot is popular	BBQ Joint	994
2	M4W	Rosedale	43.679563	-79.377529	Toronto Lawn Tennis Club	This spot is popular	Athletics & Sports	896
3	M4W	Rosedale	43.679563	-79.377529	Tinuno	This spot is popular	Filipino Restaurant	945
4	M4W	Rosedale	43.679563	-79.377529	Pie Squared	This spot is popular	Pie Shop	826



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

	Grocery Store	Pie Shop	Breakfast Spot	Sandwich Place	Diner	Deli / Bodega	Steakhouse	Pizza Place	General Entertainment	Creperie	Supermarket	Bakery	Farmers Market	Food Truck	Burrito Place
Neighborhood															
Adelaide, King, Richmond	0	0	1	0	0	1	2	2	0	0	0	2	0	0	
Berczy Park	1	0	3	0	1	1	2	1	0	1	1	4	2	1	
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Cabbagetown, St. James Town	0	1	0	0	2	1	1	0	0	0	0	0	0	0	
Central Bay Street	1	0	2	1	1	0	1	2	0	1	1	1	0	0	



# K-means Clustering

	Grocery Store	Pie Shop	Breakfast Spot	Sandwich Place	Diner	Deli / Bodega	Steakhouse	Pizza Place	General Entertainment	Creperie	Supermarket	Bakery	Farmers Market	Food Truck	Burrito Place	Total Restaurants	Total Joints	Total Sum
G5	3.500	0.000000e+00	0.50	1.00	2.0	0.500	0.50	2.0	5.000000e-01	0.5	0.500	0.5	0.000	0.000	0.00	32.000	2.50	48.00
G3	0.500	0.000000e+00	0.50	0.50	0.0	0.000	0.00	1.5	0.000000e+00	0.5	1.000	5.0	1.000	0.000	1.00	28.000	2.00	46.50
G4	0.375	-2.775558e-17	1.50	0.25	1.0	0.875	1.75	1.5	-1.387779e-17	0.5	0.375	2.0	0.625	0.625	0.75	21.875	1.25	37.50
G1	0.500	2.500000e-01	1.25	0.00	1.0	1.250	1.50	1.0	0.000000e+00	0.5	0.500	2.0	1.000	0.500	0.00	14.250	1.75	29.25
G2	1.000	5.000000e-01	0.50	0.50	0.0	0.000	0.00	0.0	0.000000e+00	0.0	0.000	0.0	0.000	0.000	0.00	1.000	0.50	4.00



# Result

**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.} This algorithm although is pretty straightforward yet is strongly powerful.**



# Result

**Result:**

**Best Group is G5;**

**Second Best Group is G3;**

**Third Best Group is G4;**

	Neighborhood	Group
0	Rosedale	4
1	Cabbagetown, St. James Town	1
2	Church and Wellesley	2
3	Harbourfront, Regent Park	1
4	Ryerson, Garden District	4
5	St. James Town	3
6	Berczy Park	5
7	Central Bay Street	5
8	Adelaide, King, Richmond	4
9	Harbourfront East, Toronto Islands, Union Station	4
10	Design Exchange, Toronto Dominion Centre	4
11	Commerce Court, Victoria Hotel	3
12	Harbord, University of Toronto	1
13	Chinatown, Grange Park, Kensington Market	4
14	CN Tower, Bathurst Quay, Island airport, Harbo...	2
15	Stn A PO Boxes 25 The Esplanade	4
16	First Canadian Place, Underground city	4
17	Christie	1



# Result

## Best Neighborhood Are...

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

	Neighborhood	Group
6	Berczy Park	5
7	Central Bay Street	5

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

```
{'Postal Code': 'M5E',  
 'Neighborhood': 'Berczy Park',  
 'Neighborhood Latitude': 43.644770799999996,  
 'Neighborhood Longitude': -79.3733064}
```

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

```
{'Postal Code': 'M5G',  
 'Neighborhood': 'Central Bay Street',  
 'Neighborhood Latitude': 43.6579524,  
 'Neighborhood Longitude': -79.3873826}
```