Groceries Warehouse Recommender System

Nader Asadi

Problem Description

There is a groceries contractor in one of the boroughs of Toronto (Scarborough). This contractor provides places such as: Different types of Restaurants, Bakery, Breakfast Spot, Brewery and Café with fresh and high-quality groceries. The contractor wants to build a warehouse for the groceries 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.

Data

Part 1

We will need geo-locational information about that specific borough and the neighborhoods in that borough. We assume 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.

Data

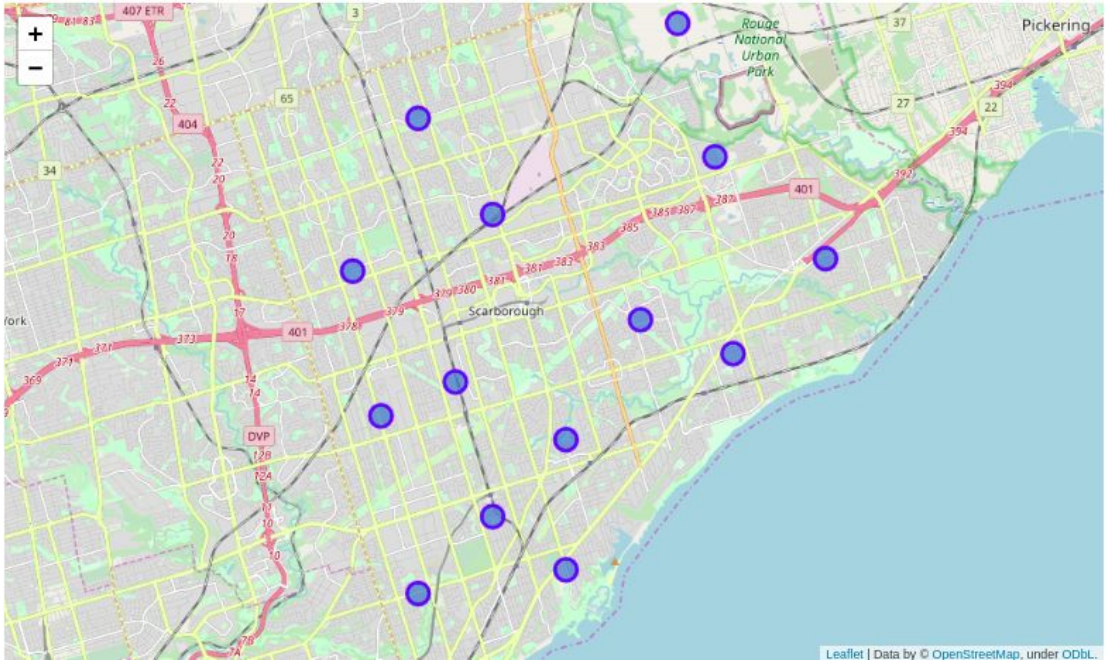
Part 2

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:

	Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Venue Category	Distance
0	M1V	Agincourt North, Milliken	43.815252	-79.284577	Fahmee Bakery & Jamaican Foods	This spot is popular	Caribbean Restaurant	669
1	M1V	Agincourt North, Milliken	43.815252	-79.284577	Jim Chai Kee Wonton Noodle 沾仔記	This spot is popular	Noodle House	689
2	M1V	Agincourt North, Milliken	43.815252	-79.284577	Lotus Pond Vegetarian Restaurant 蓮花素食	This spot is popular	Vegetarian / Vegan Restaurant	934
3	M1V	Agincourt North, Milliken	43.815252	-79.284577	DaanGo Cake Lab	This spot is popular	Bakery	809
4	M1V	Agincourt North, Milliken	43.815252	-79.284577	The Brighton Convention & Event Centre	This spot is popular	Event Space	890

Methodology

Part 1 : Identifying Postal Codes



Methodology

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.

Methodology

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.

Methodology

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

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	1	1	2	2	21	1
Agincourt North, Milliken	2	0	0	0	0	0	0	1	2	0	9	0
Birch Cliff	0	0	1	0	0	0	0	0	0	0	4	0
Clairlea, Golden Mile, Oakridge	2	0	1	0	0	0	1	0	1	1	4	0
Cliffcrest, Cliffside	0	0	0	0	0	0	0	0	3	0	3	2
Dorset Park, Scarborough Town Centre, Wexford Heights	2	0	0	0	0	0	1	0	1	1	13	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	2	1	5	1
Maryvale, Wexford	1	1	0	1	0	0	3	0	3	0	8	2
Morningside, West Hill	0	0	0	0	1	0	1	0	4	1	3	2
Rouge, Malvern	0	0	0	0	0	1	0	0	0	1	6	0
Scarborough Village	0	0	0	0	0	0	0	0	1	1	5	0
Tam O'Shanter	0	0	0	0	0	0	1	1	2	2	12	1
Woburn	0	0	0	0	0	0	0	0	0	0	3	0

Methodology

Part 4 : Applying K-means Clustering

	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
G3	2.000000	1.000000	0.000000	0.000000e+00	0.000000e+00	0.000000	1.000000	1.0	2.000000	2.000000	21.000000	1.000000	31.000000
G4	1.000000	0.000000	0.000000	0.000000e+00	0.000000e+00	0.000000	1.000000	0.5	1.500000	1.500000	12.500000	2.000000	20.000000
G1	1.500000	0.500000	0.000000	5.000000e-01	0.000000e+00	0.000000	1.500000	0.5	2.500000	0.000000	8.500000	1.000000	16.500000
G5	0.000000	0.000000	0.000000	0.000000e+00	3.333333e-01	0.000000	1.000000	0.0	3.000000	0.666667	3.666667	1.666667	10.333333
G2	0.333333	0.166667	0.333333	1.387779e-17	1.387779e-17	0.166667	0.166667	0.0	0.333333	0.500000	3.833333	0.166667	6.000000