A RECOMMENDER SYSTEM FOR GROCERIES CONTRACTOR

APPLIED DATA SCIENCE CAPSTONE

ABHISHEK KUMAR

INTRODUCTION

Part 1: Problem Description

- To provide a groceries contractor (in one of the boroughs of Toronto - Scarborough) recommendations for optimized warehouse location

- Supplier to various Restaurants, Bakery, Breakfast Spot, Brewery, Cafe (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.

- The contractor should build this warehouse where it is closest to its customers in order to minimize the cost of transportation in addition to the example above.

- Finding the right neighborhood is our mission and our recommender system will provide this contractor with a sorted list of neighborhoods in which the first element of the list will be the best suggested neighborhood.

Part 2: Data We Need

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.

- Geo-locational information (latitude and longitude)about that specific borough and the neighborhoods in that borough.

- Data about different venues in different neighborhoods of that specific borough.

- "Foursquare" locational information to be used. (basic and advanced information about that venue)

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

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

DATA WE NEED

`scarborough_venues.head()`

	Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Venue Category	Distance
0	M1W	Steeles West	43.799525	-79.318389	Mr Congee Chinese Cuisine 龍粥記	This spot is popular	Chinese Restaurant	72
1	M1W	Steeles West	43.799525	-79.318389	Agincourt Bakery	This spot is popular	Bakery	759
2	M1W	Steeles West	43.799525	-79.318389	Little Sheep Mongolian Hot Pot 小肥羊	This spot is popular	Hotpot Restaurant	972
3	M1W	Steeles West	43.799525	-79.318389	Phoenix Restaurant 金鳳餐廳	This spot is popular	Chinese Restaurant	147
4	M1W	Steeles West	43.799525	-79.318389	Price Chopper	This spot is popular	Grocery Store	16

• Part 3: 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"
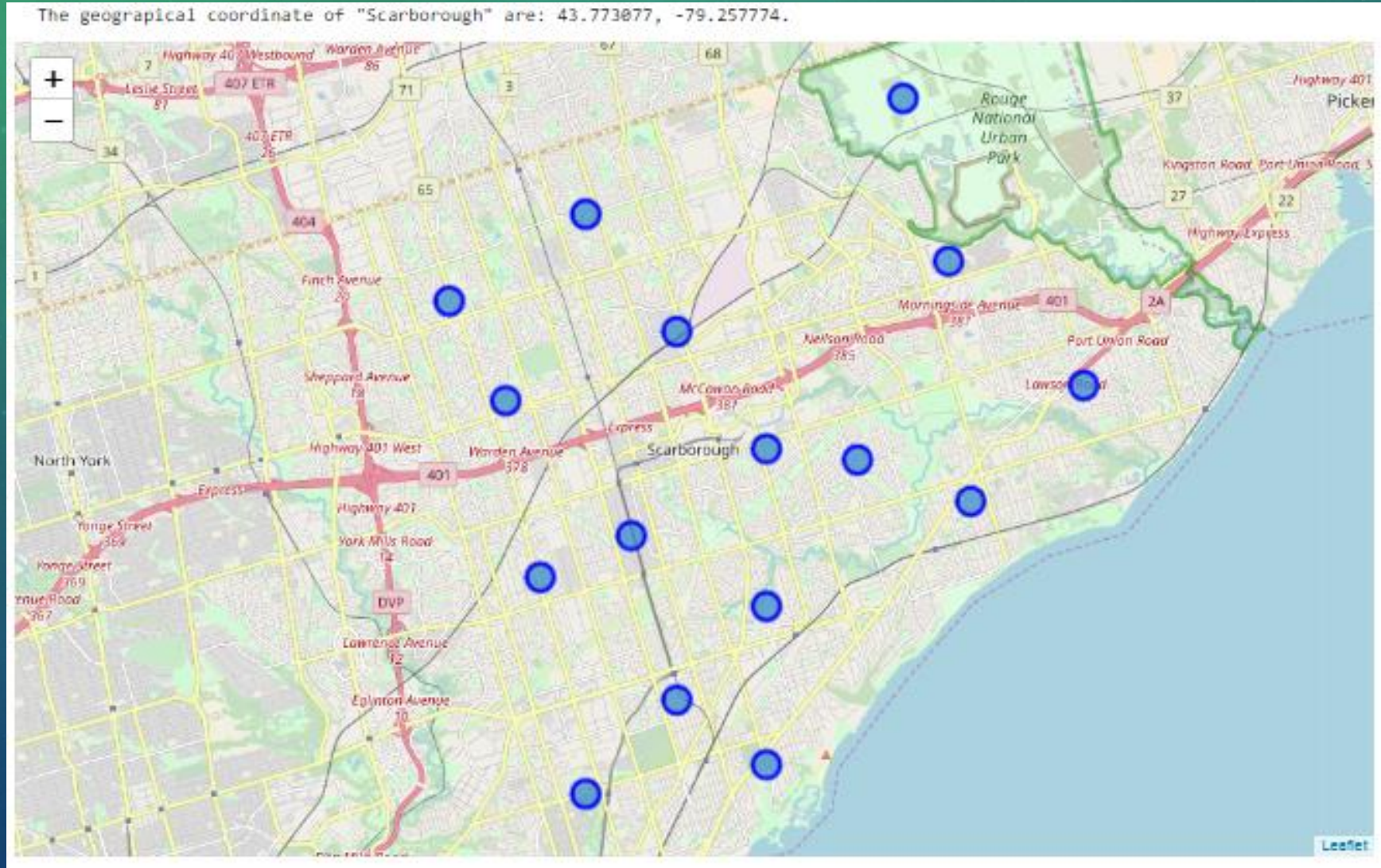
Part 4: Connecting to Foursquare and Retrieving
Locational Data for Each Venue in Every Neighborhood

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.

- 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.

PART 5: PROCESSING THE RETRIEVED DATA AND CREATING A DATAFROME 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" wil be One-hot encoded and different venues will have different feature-columns. After On-hot encoding we will integrate all restaurant columns to one column "Total Restaurants" and all food joint columns to "Total Joints" column.

```
scarborough_venues.head()
```

	Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Venue Category	Distance
0	M1W	Steeles West	43.799525	-79.318389	Mr Congee Chinese Cuisine 龍粥記	This spot is popular	Chinese Restaurant	72
1	M1W	Steeles West	43.799525	-79.318389	Agincourt Bakery	This spot is popular	Bakery	759
2	M1W	Steeles West	43.799525	-79.318389	Little Sheep Mongolian Hot Pot 小肥羊	This spot is popular	Hotpot Restaurant	972
3	M1W	Steeles West	43.799525	-79.318389	Phoenix Restaurant 金凰餐廳	This spot is popular	Chinese Restaurant	147
4	M1W	Steeles West	43.799525	-79.318389	Price Chopper	This spot is popular	Grocery Store	16

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

`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	1	0	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

PART 6: APPLYING ONE OF MACHINE LEARNING TECHNIQUES (K-MEANS CLUSTERING)

```python
# import k-means from clustering stage
from sklearn.cluster import KMeans

# run k-means clustering
kmeans = KMeans(n_clusters = 5, random_state = 0).fit(scarborough_onehot)
```

	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

DECISION MAKING AND REPORTING RESULTS

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.

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

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

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

DECISION MAKING AND REPORTING RESULTS

Best Neighborhood Is ...

```
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}
```