A RECOMMENDER SYSTEM FOR CAFE CONTRACTOR

APPLIED DATA SCIENCE CAPSTONE

IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

BRAHIM AMRHAR (BASY15@GMAIL.COM)

MAY 2019

A RECOMMENDER SYSTEM FOR CONTRACTOR

Image Source : http://www.deeptenger.com/init/static/images/conversionswords.jpg

CONTEXT AND PROBLEM DESCRIPTION

- A new brand of Café contractor wants to spring-up in Toronto. Globally, restaurants don't have a variety of cafés for their clients and therefore, client get done with one café without any particularities in terms of flavor, taste, *aromas, etc. Moreover, Café culture declines a variety of coffee types by seasons and by origins.*
- *The main ide*a behind the business is to promote a good quality with divers tastes, all over the seasons, to restaurants so that the "Café" moment become more a deep pleasure than just appetizers' ends. It's matter of renew with the grand café values.
- This will benefit to restaurants so they will create clients' loyalty and retain their actual ones. This will also benefit to the contractor as he will have broad range of customers to supply and make his trademark more known. Consequently, it's win-win business relationship with more income in both parts.
- Then the contractor is going to work on this field of the market and find potential restaurants to hold this idea. The geographical scope of the study is Toronto and particularly North York borough where the contractor want to develop first this business.

Data acquisition and cleaning

- We use Postal Codes of different regions inside North York to find the list of neighborhoods. We essentially obtain our information from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and then process the table inside this site.
- Given that the Geocoder package can be very unreliable, we used a link to a csv file that has the geographical coordinates of each postal code from : http://cocl.us/Geospatial_data.
- We will need data about different venues in different neighborhoods of that specific borough from : https://foursquare.com/

LATITUDE AND LONGITUDE COORDINATES ARE: 43.761539, -79.411079.

NORTH YORK IS A DISTRICT OF TORONTO, ONTARIO, CANADA, WITH THE POPULATION OF ABOUT 640,000 PEOPLE. IT IS A FORMER TOWNSHIP WHICH HAS MERGED WITH THE DEVELOPING TORONTO. NORTH YORK IS ALSO KNOWN AS "THE CITY WITH HEART". IT IS A LARGE AND DOMINANTLY YOUNG COMMUNITY, WITH VERY VIVID AND ACTIVE LIFESTYLE, MANY PARKS, CINEMA THEATERS, PERFORMING ARTS THEATERS, AND OTHER ENTERTAINMENT ESTABLISHMENTS.

DATA WE NEED

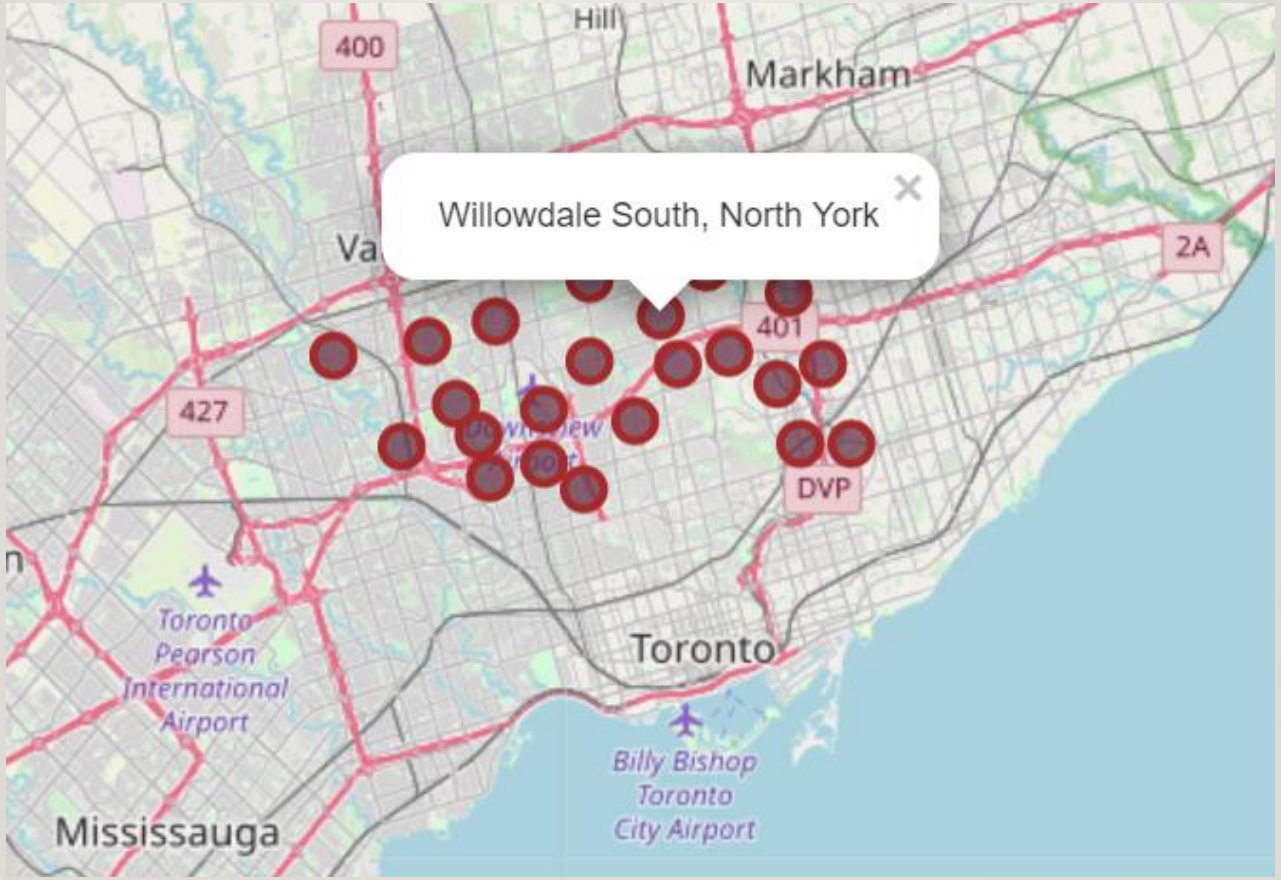
We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. A typical request from Foursquare will provide us with the following information:

	Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Venue Category	Distance
0	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

Identifying Postal Codes in “North York"

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
3	M6A	North York	Lawrence Heights, Lawrence Manor	43.718518	-79.464763
7	M3B	North York	Don Mills North	43.745906	-79.352188
10	M6B	North York	Glencairn	43.709577	-79.445073

Identifying Postal Codes in “North York"



Retrieve 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.

Processing Retrieved Data for all venues inside North York

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.

Processing Retrieved Data for all venues inside North York

	Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Venue Category	Distance
0	M3A	Parkwoods	43.753259	-79.329656	Allwyn's Bakery	This spot is popular	Caribbean Restaurant	833
1	M3A	Parkwoods	43.753259	-79.329656	Brookbanks Park	This spot is popular	Park	245
2	M3A	Parkwoods	43.753259	-79.329656	Tim Hortons	This spot is popular	Café	866
3	M3A	Parkwoods	43.753259	-79.329656	A&W Canada	This spot is popular	Fast Food Restaurant	852
4	M3A	Parkwoods	43.753259	-79.329656	Food Basics	This spot is popular	Supermarket	895
5	M3A	Parkwoods	43.753259	-79.329656	Bruno's valu-mart	This spot is popular	Grocery Store	882
6	M3A	Parkwoods	43.753259	-79.329656	Shoppers Drug Mart	This spot is popular	Pharmacy	953
7	M3A	Parkwoods	43.753259	-79.329656	High Street Fish & Chips	This spot is popular	Fish & Chips Shop	967
8	M3A	Parkwoods	43.753259	-79.329656	Variety Store	This spot is popular	Food & Drink Shop	312
9	M3A	Parkwoods	43.753259	-79.329656	Shoppers Drug Mart	This spot is popular	Pharmacy	926

Now, the dataset is fully ready to be used for machine learning purposes

Neighborhood	American Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Eastern European Restaurant	Empanada Restaurant	Falafel Restaurant	French Restaurant	Greek Restaurant	Indian Restaurant	Indonesian Restaurant
Bathurst Manor, Downsview North, Wilson Heights	0	0	0	0	0	0	0	0	0	0	0	0	0
Bayview Village	0	0	0	1	0	0	0	0	0	0	0	0	0
Bedford Park, Lawrence Manor East	1	0	0	0	1	0	0	0	0	0	1	1	0
CFB Toronto, Downsview East	0	0	0	0	0	0	0	0	0	0	0	0	0

Display the top 10 venues for each neighborhood

	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	Bathurst Manor, Downsview North, Wilson Heights	Sushi Restaurant	Restaurant	Middle Eastern Restaurant	Mediterranean Restaurant	Vietnamese Restaurant	Greek Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant
1	Bayview Village	Japanese Restaurant	Chinese Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Caribbean Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Eastern European Restaurant	Empanada Restaurant
2	Bedford Park, Lawrence Manor East	Italian Restaurant	American Restaurant	Thai Restaurant	Sushi Restaurant	Restaurant	Comfort Food Restaurant	Indian Restaurant	Greek Restaurant	French Restaurant	Asian Restaurant
3	CFB Toronto, Downsview East	Turkish Restaurant	Vietnamese Restaurant	Middle Eastern Restaurant	Latin American Restaurant	Italian Restaurant	Greek Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant
4	Don Mills North	Japanese Restaurant	Greek Restaurant	Thai Restaurant	Asian Restaurant	Caribbean Restaurant	Restaurant	Vietnamese Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant

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

```
# set number of clusters (k)
k = 3

NorthYork_grouped_clustering = NorthYork_grouped.drop('Neighborhood', 1)

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

# check cluster labels generated for each row in the dataframe
kmeans.labels_
```

```
array([0, 2, 0, 0, 2, 0, 0, 0, 0, 0, 2, 2, 0, 0, 0, 0, 2, 2, 0, 0, 0, 1,
       0, 0])
```

	Total Sum	American Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Eastern European Restaurant	Empanada Restaurant	Falafel Restaurant	French Restaurant	Greek Restaurant	Inc Restaur
G2	29.000000	1.000000	0.000000e+00	0.000000	2.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
G3	7.666667	0.500000	6.666667e-01	0.333333	0.500000	0.000000	0.166667	0.000000	0.000000	0.166667	0.000000	0.166667	0.166
G1	3.352941	0.117647	2.775558e-17	0.117647	0.235294	0.058824	0.058824	0.058824	0.058824	0.058824	0.058824	0.176471	0.058

DECISION MAKING AND REPORTING RESULTS

Now, we focus on the centers of clusters and compare them for their fist common venues. The group which its center has the highest number of Asian restaurants will be our best recommendation to the contractor. Neighborhood Willowdale South is a good candidate to spring up the business of the new coffee brand.

DECISION MAKING AND REPORTING RESULTS

- Based on this analysis, the best recommended neighborhood will be:
- Neighborhood: Willowdale South,
- Postal Code: M2N,
- Neighborhood Latitude: 43.770120,
- Neighborhood Longitude: -79.408493
- The Best Group is G2
- The Second Group is G3

DECISION MAKING AND REPORTING RESULTS

Best Neighborhoods Are...

```
neigh_summary[neigh_summary['Group'] == 2]
```

	Neighborhood	Group
21	21	2

```
NorthYork_grouped.loc[21:21]
```

	Neighborhood	American Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Eastern European Restaurant	Empanada Restaurant	Falafel Restaurant	French Restaurant	Greek Restaurant	Ir Restau
21	Willowdale South	1	0	0	2	1	0	0	0	0	0	0	

Image is from: https://images.unsplash.com/photo-1499744937866-d7e566a20a61?ixlib=rb-1.2.1&auto=format&fit=crop&w=1000&q=80