Context

This report is the final project from the IBM Data Science Specialization. The goal of this project is to use the tools and skills acquired through the different courses to develop and idea to leverage the Foursquare location data to explore or compare neighborhoods or cities of our choice in order to come up with a problem that you can be solved using the Foursquare location data

Introduction/Business Problem

QB is an international company focused on providing consulting services to people who intends to move their business to another city. QB is well-known to always use the top trend tecnologies like machine learning, Blockchain and quantum computing in order to bring better solutions for their clients.

Recently, Michael Rain has asked for QB services. Michael Rain wants to move his warehouse for groceries from Central Toronto (Roselawn) to a neighbourhood in Scarborough. He want to keep their lifestyle and also have the most near groceries as posible to improve the quality of his warehouse.

Data

- Neighbourhood information: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Geo locational information about the neighbourhoods in Scarborough
- Foursquare information on venue categories, top tips, location data, ratings, etc.

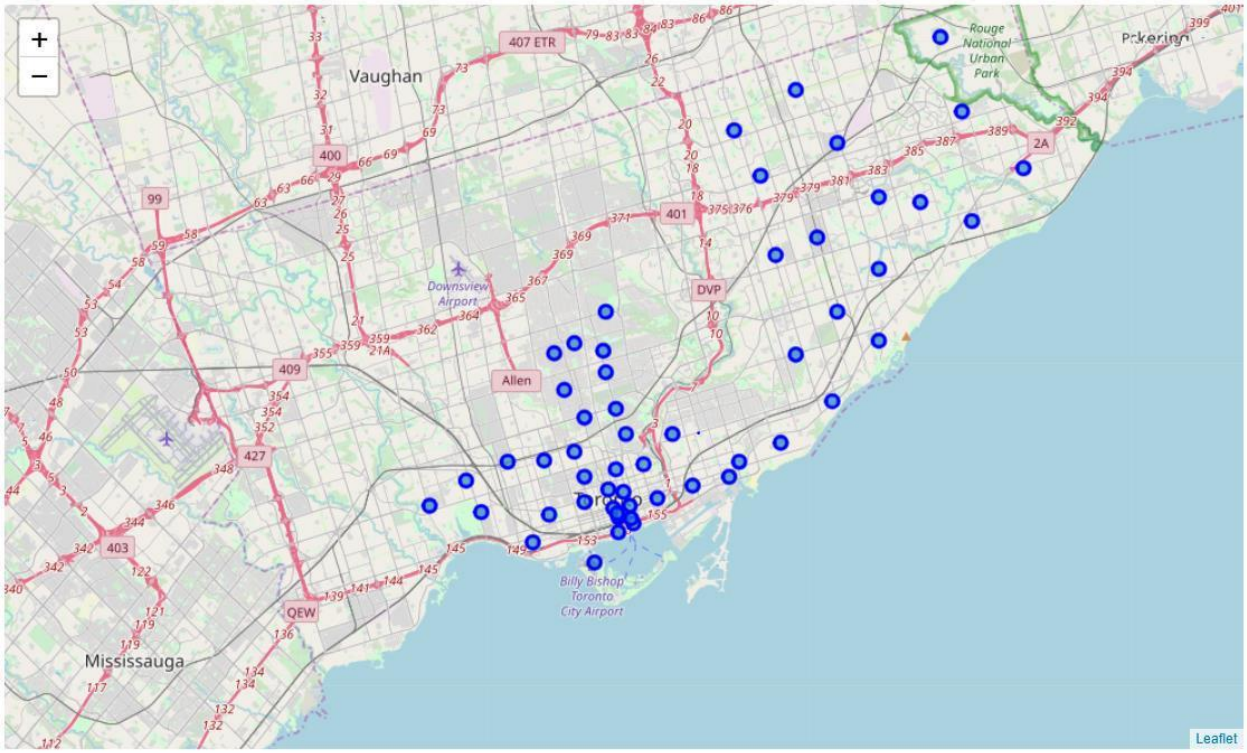
Methodology

1. We import data from week 3 with the information from Toronto's Neighborhood (Postal Code, Borough, Neighborhood, Latitude and Longitude). This information was created by scraping the web page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and adding the longitude and latitude from a csv file

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

2. For the scope of our problem, we filter the data by the boroughs we need to analyze (Downtown Toronto, East Toronto, West Toronto, Central Toronto, Scarborough)

In the picture below, we can see all the initial candidates for Michael's grocery warehouse (55 Neighborhoods)



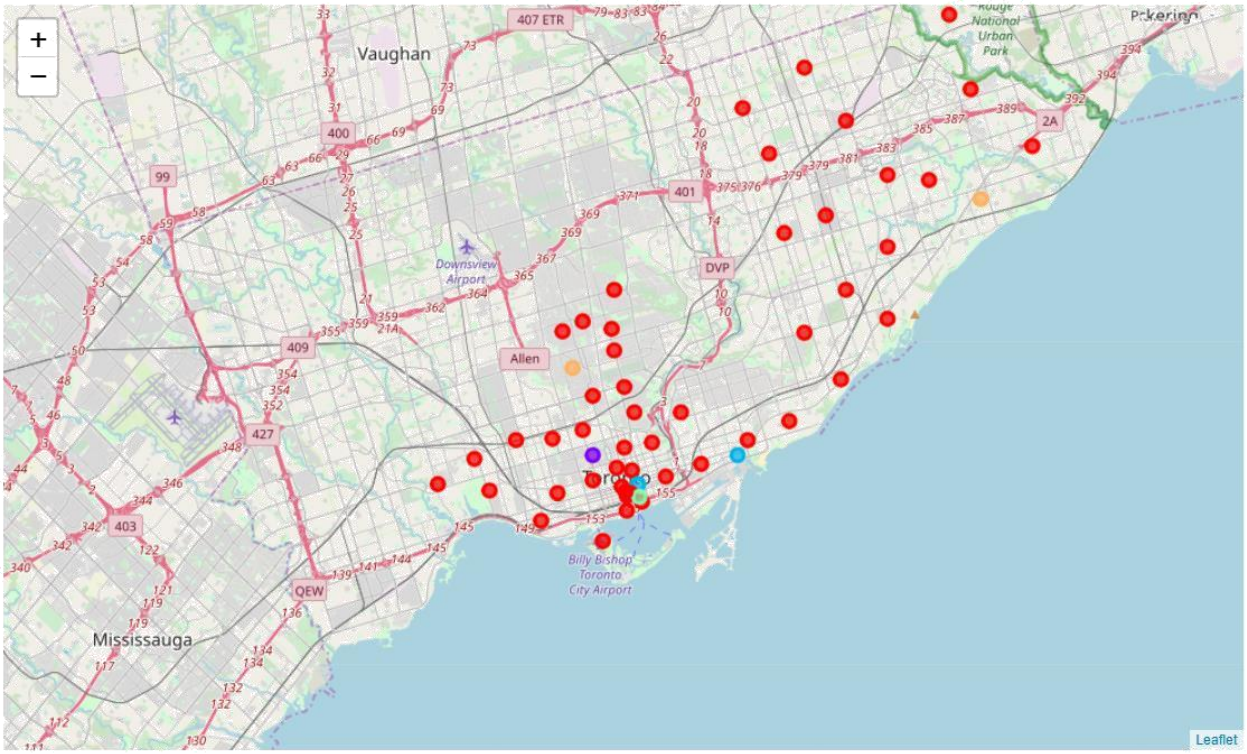
3. We access the Foursquare API using our credentials (CLIENT_ID, CLIENT_SECRET) and the version 20180605 and define some functions to get the nearby venues in Toronto. We limit the call to 100 and radius to 500. After that, we encode the results using one-hot encoding by the Venue Category. Having this, we group the venues by neighborhood and get the mean in order to get the proportion of a Venue Category in each neighborhood.

	Neighborhood	Yoga Studio	Adult Boutique	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	...	Thrift / Vintage Store	Toy / Game Store	Trail	Train Station
0	Adelaide, King, Richmond	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.00000	0.000000
1	Agincourt	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.00000	0.000000
2	Agincourt North, L'Amoreaux East, Milliken, St...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.00000	0.000000
3	Berczy Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.00000	0.000000
4	Birch Cliff, Cliffside West	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.00000	0.000000
5	Brockton, Exhibition Place, Parkdale Village	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.00000	0.000000
6	Business reply mail Processing Centre969 Eastern	0.052632	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.00000	0.000000
7	CN Tower, Bathurst Quay, Island airport, Harbo...	0.000000	0.000000	0.000000	0.071429	0.071429	0.071429	0.142857	0.142857	0.142857	...	0.000000	0.000000	0.00000	0.000000
8	Cabbagetown, St. James Town	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.00000	0.000000

4. We extract the top 10 venues in each Neighborhood ordering by the proportion calculated in step 3.

	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	Adelaide, King, Richmond	Coffee Shop	Café	Thai Restaurant	Steakhouse	American Restaurant	Asian Restaurant	Hotel	Bar	Restaurant	Bakery
1	Agincourt	Lounge	Clothing Store	Skating Rink	Breakfast Spot	Women's Store	Dim Sum Restaurant	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant
2	Agincourt North, L'Amoreaux East, Milliken, St...	Park	Playground	Women's Store	Department Store	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop
3	Berczy Park	Coffee Shop	Cocktail Bar	Restaurant	Farmers Market	Steakhouse	Seafood Restaurant	Cheese Shop	Café	Italian Restaurant	Bakery
4	Birch Cliff, Cliffside West	College Stadium	Skating Rink	General Entertainment	Café	Dessert Shop	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant

5. We apply K-mean Cluster to segment all 55 neighborhoods into 5 clusters. This clusters were similar in most common venue and we use this as an indicator of a similar lifestyle in the neighborhoods for each clusters.



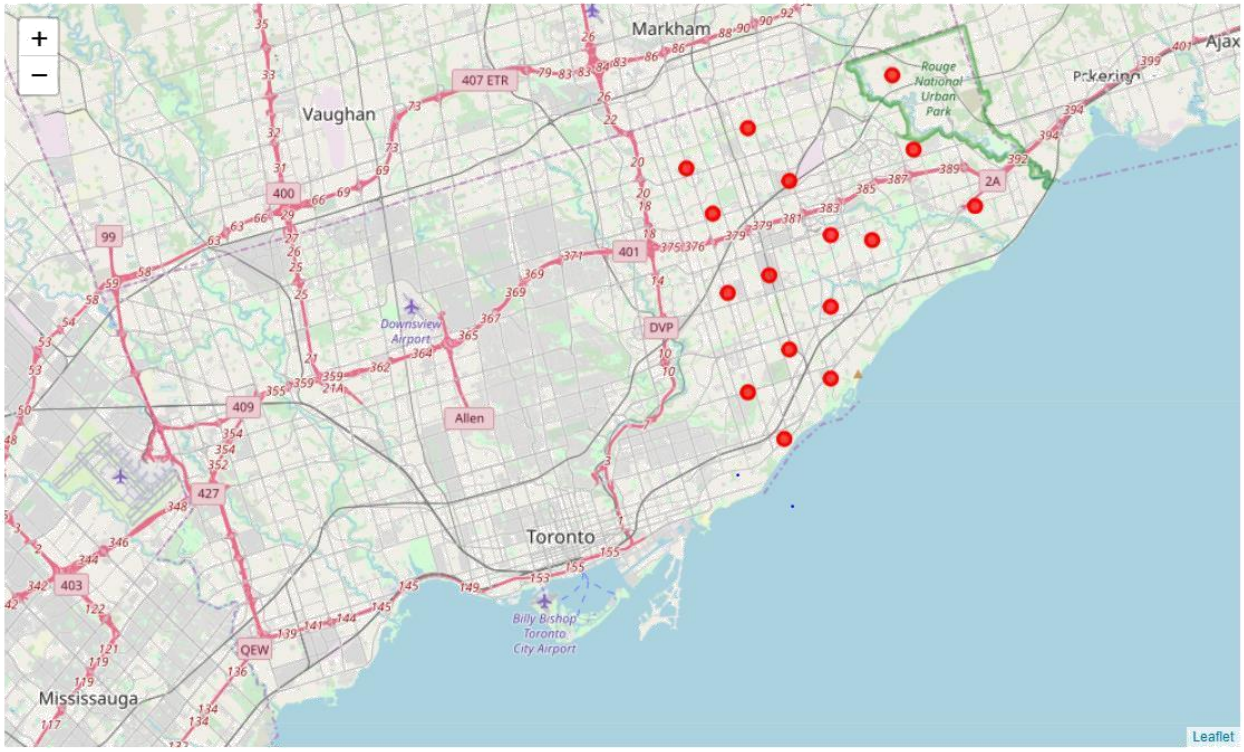
6. To find the best candidate, we need to find the neighborhood who satisfies the requirements of the client. First of all, the client wants to move from Roselawn. According to our analysis, Roselawn is part of cluster 0. So, the new neighborhood has also to be part of the cluster 0.

	PostalCode	Borough	Neighborhood	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 Most Common Venue
39	M5N	Central Toronto	Roselawn	43.711695	-79.416936	0	Garden	Pool	Home Service	Women's Store	Doner Restaurant	Dim Sum Restaurant	Diner

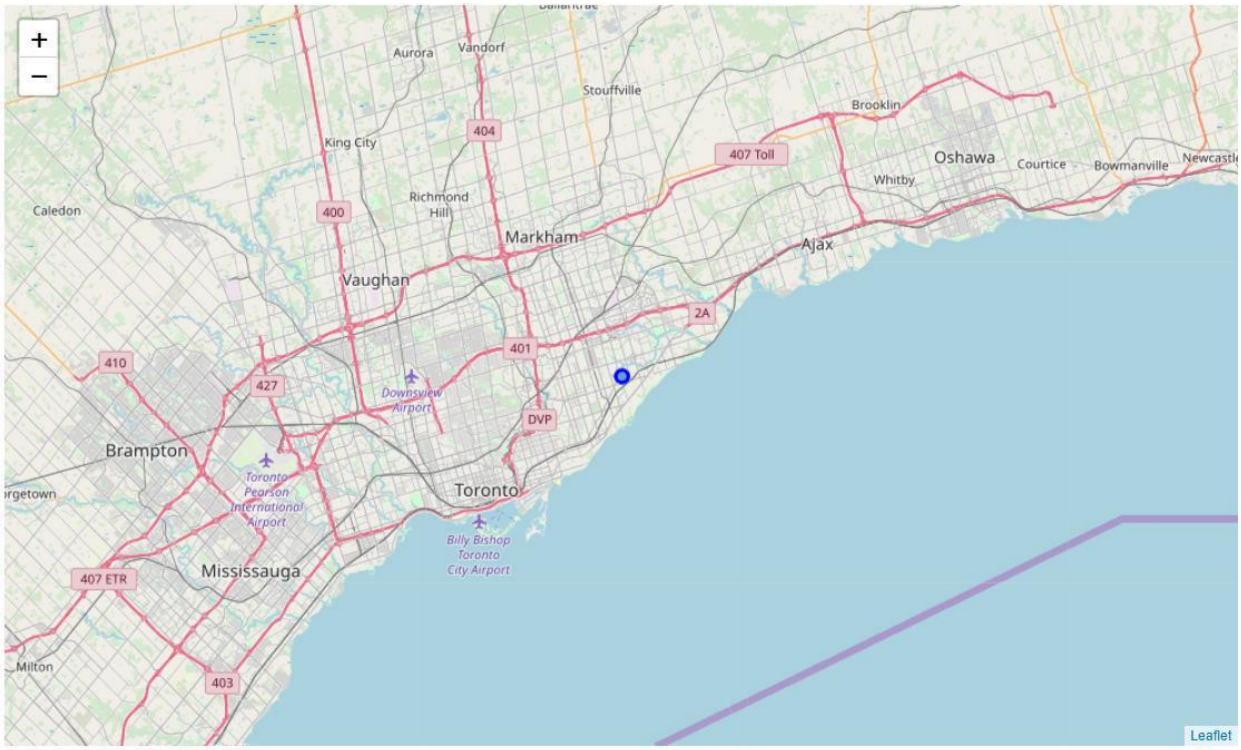
The other requirement of the client was that the borough should be Scarborough.

	PostalCode	Borough	Neighborhood	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 Most Common Venue
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353	0	Fast Food Restaurant	Women's Store	Dessert Shop	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	0	Bar	Women's Store	Fast Food Restaurant	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store
3	M1G	Scarborough	Woburn	43.770992	-79.216917	0	Coffee Shop	Korean Restaurant	Soccer Field	Women's Store	Dim Sum Restaurant	Falafel Restaurant	Event Space
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	0	Bank	Athletics & Sports	Caribbean Restaurant	Thai Restaurant	Bakery	Fried Chicken Joint	Hakka Restaurant
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476	0	Grocery Store	Spa	Playground	Women's Store	Department Store	Ethiopian Restaurant	Electronics Store

We get 11 solutions.



However, one of the most important requirements, was the need of groceries around the neighborhood chosen to maintain the quality of the grocery warehouse. The only candidate who satisfies the 3 requirements was candidate 5 (Scarborough Village).



Finally, we make a deeper analysis about the ratings of the groceries store in a radius of 1000 from Scarborough Village. Half of the groceries store has not been rated yet, and the other were around 7.

Results

By exploring the requirement of the clients, we only find one neighborhood (Scarborough Village) that match with this requirement (similar lifestyle, Scarborough, groceries store near)

Discussion

The result is somehow limited for the assumptions we have made. First, being in a cluster does not mean having a similar lifestyle at all with the other neighborhoods in the clusters. The clusters were made by using just the top 10 venues, so we do not analyze completely the neighborhoods with more venues but with less frequency. Furthermore, we get the top 10 venues in a radius of 500, so we should need to analyze if the change of the parameter radius has an impact on the results

Conclusion

According to our analysis, the best neighborhood is Scarborough Village. However, we must take into account the assumptions we have made that may have an impact on the results.