

# *Data Science Capstone Project*

## *Best Neighborhood for groceries warehouse in Scarborough area of GTA*

*By*

Nirav Patel

*For*

IBM Data Science Professional  
Certificate, Coursera

## **BACKGROUND**

Loopmark is a Toronto based company focusing on providing consulting services for the commercial real estate. Here we will consider, we are working for the Logomark. Let us assume that our client wants to move their business to another city. Our job is to use the top trend technologies like machine learning and Data analytics to bring better solutions for our clients.

## **BUSINESS PROBLEM**

We have been assigned to serve our recent client name Mr. Jon. Mr. Jon has asked consulting from our company. He wants to move his warehouse for his business from Ottawa to Toronto. To be specific he is looking for a neighborhood in Scarborough area of the Greater Toronto Area. He is fine with the metro city lifestyle, but he wants to be in the area close to groceries and retail market, as his business defines those needs.

## **STAKEHOLDERS**

In this project, our main stakeholder is our client himself, Mr Jon. However, we can utilize this analysis for the future clients and provide the better service according to their need.

So, in general we can consider following stakeholders for this project.

- Loopmark pvt. Ltd
- Mr, Jon
- Future clients

## **DATA**

- [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada)
- Foursquare API information on venue categories, location data, ratings and more
- Geo locational information about the neighborhoods in Scarborough

The data collected from all these files will be cleaned and converted with the use of pre-defined libraries from the Python. Further, we will create some clusters and find the one name of the borough which satisfies the requirements for our client.

## METHODOLOGY

- 1) We have already performed the lab session in earlier project to find out the clustering of Toronto city and finding the different boroughs for specific location with area codes. In this project, we will use results from that project and start with the neighborhood only related to the Scarborough area from the Greater Toronto Area. Data sheet has Toronto's Neighborhood (Postal Code, Borough, Neighborhood, Latitude and Longitude). As we know, this data was created by scraping the web page [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) and adding the longitude and latitude from a csv file of the Toronto data.

	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

Fig.1 Scarborough neighborhood with location parameters

- 2) In the next step, we will filter the data according to boroughs we need to analyze for this project. (Downtown Toronto, East Toronto, West Toronto, Central Toronto, Scarborough)

In the picture below, we can see all the initial candidates for Mr. Jon's grocery warehouse which includes all neighborhoods.

Figure on the next page shows potential neighborhoods on the map of the Toronto.



Fig. 2 All possible venues in Toronto Area

- 3) Now, we will use the Foursquare API with my credentials (CLIENT\_ID, CLIENT\_SECRET) and the version 20190815 and define functions to get the nearby venues in Toronto. We are limiting the call to around 100 and radius of 500 meter only. After that, we will encode the results using one-hot encoding by the Venue Category. After that point, we will group the venues by neighborhood and get the mean to find out the proportion of a Venue Category for each neighborhood.
- 4) We will extract the top 10 venues in each Neighborhood ordering by the proportion calculated from earlier steps.

	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

Fig. 3 Top ten venues in Neighborhood

- 5) We will use the algorithm of K-mean Cluster to segment all neighborhoods into 5 clusters. These clusters are similar in most common venue and we use this as a parameter of similar lifestyle in the neighborhoods for each cluster.

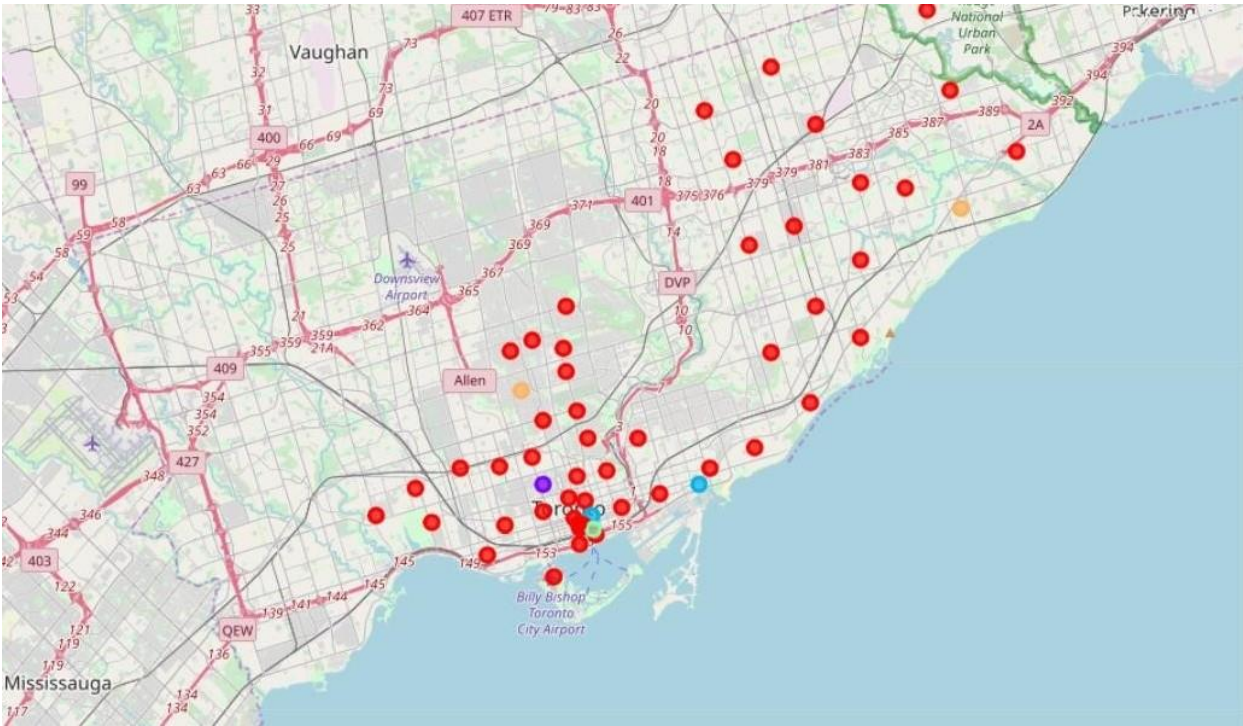


Fig 4 Five clusters with neighborhood

- 6) To find the best location out of available choices, we need to find the neighborhood which satisfies the requirements of the client in this case. Now we know that client prefers similar neighborhood for his new location. From history, we know he is moving from the Ottawa, according to him, it has similar locality and environment to the cluster 0 in our analysis here. So, we will choose the location available from the cluster 0.

From our analysis so far with the cluster 0 and available neighborhood is Roselawn with Central Toronto borough.

	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

Fig 5 Choosing the Neighborhood



- 7) Now, our main requirement for the client is, he wants location from the Scarborough. With this condition, we get 11 locations as shown in the map below.

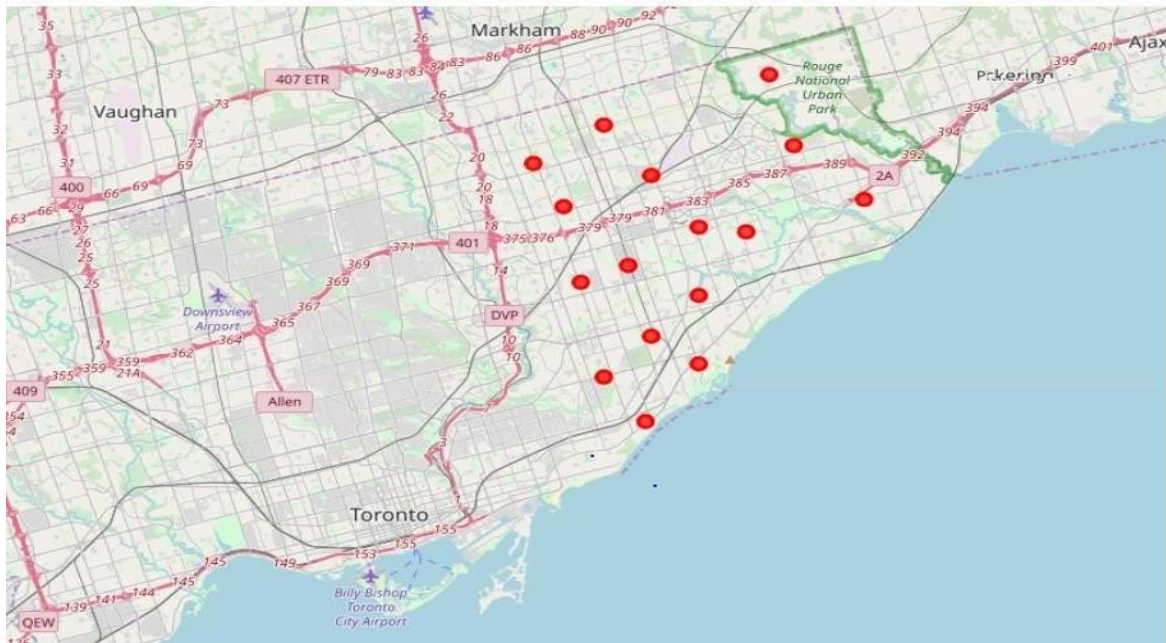


Fig 6 Possible locations from Scarborough area

Now, the last parameter to consider here is, our warehouse is supposed to be used for the groceries products, so client needs the location closest to the grocery store. To keep the location closest to the grocery store, we will choose one out of possible three locations.

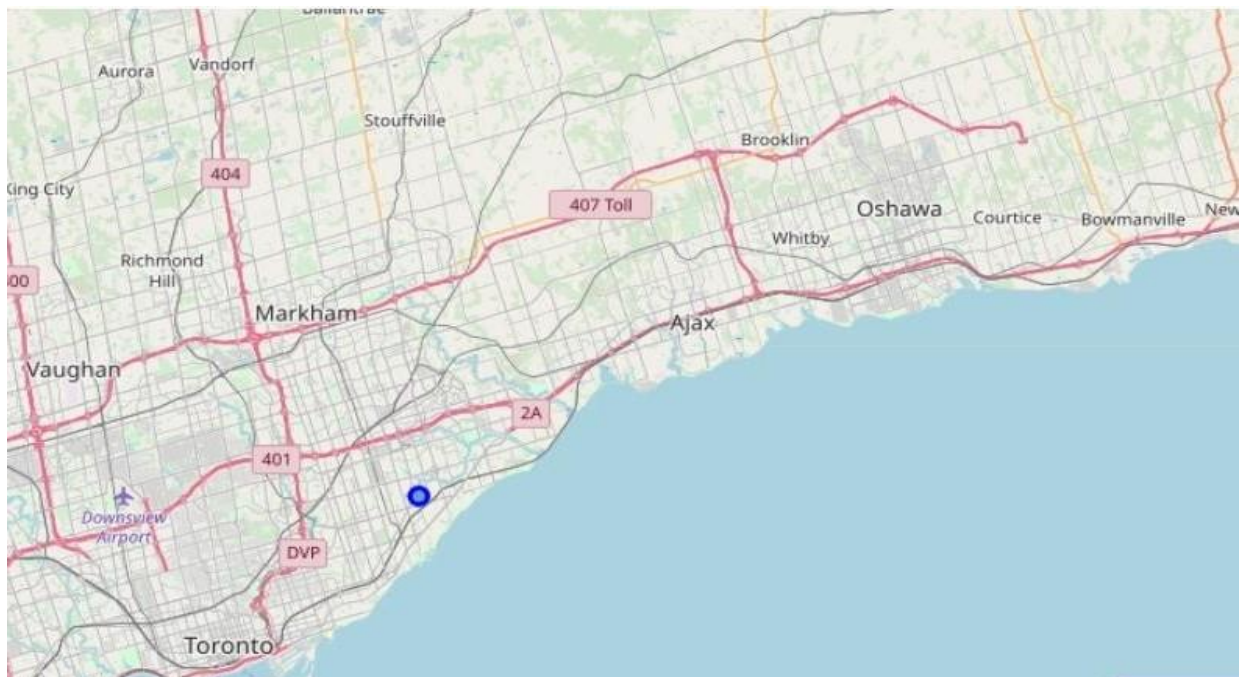


Fig 7 Nearest location to the grocery store in Scarborough village

One last step before coming to conclusion, we want to make sure the rating of the grocery stores is good. Considering the radius of one km through the village, half of the stores do not have ratings, and the others we are considering in our analysis have around 7 ratings which is not bad.

## **RESULTS**

By satisfying each of the requirement step by step as explained, we only find one neighborhood (Scarborough Village) which has the well rated grocery stores too.

## **DISCUSSION**

The result from this project has limitations. According to the scope of the project and available resources (Time and Skills in this case) we had to make some assumptions here. First, we considered that our client will get similar lifestyle and locality in the new locations like his current location. Further, the clusters were created by using the top ten venues, so we do not analyze here the total neighborhoods with all venues which could go hundreds in large scale. Next, we assume the radius of 500 meters and at last for the rating, we considered it secondary parameter, means we are not looking after the ratings of the stores as our primary criteria for the analysis.

## **CONCLUSION**

According to our analysis, the best neighborhood is Scarborough Village to choose for opening the warehouse relating to the groceries.