Determining best neighbourhood location for opening a new warehouse for a Groceries Supplier XYZ in Etobicoke, Toronto, ON, Canada

Problem Description

Background - One of the key aspects and challenges for the food chain outlets including restaurants and fast food corners is to procure fresh ingredients which can be speedily delivered. Distribution centres portrayed by warehouses play a critical role in supplying groceries especially perishable items promptly. Typically, warehouses are optimally located to cater to various delivery points which are spread across its operating regions and it take care of all food joints within its area of operation. In this specific case, we have considered a groceries supplier which have been providing fresh and high-quality groceries to most of the food joints in **Etobicoke (Toronto, ON, Canada)** and in the adjoining areas. However, in the recent times to due to increase demand of groceries in Etobicoke, there has been an additional stress on the existing supply chain. This increased demand is due to increase in population residing or migrated to Etobicoke due to increased economic activities and hence further potential of new food joints which will be coming up. This has impacted the existing supplies in terms of on time and quality of delivery especially perishable groceries due to nearest warehouse being in other adjoining Boroughs which are far from the delivery centres of Etobicoke. This has also led to increased transportation time and cost for its customer in Etobicoke.

Problem - Considering this scenario with increased demand of groceries, supplier has decided to open new warehouse at optimal location in Etobicoke to resolve this. The supplier plan to build new warehouse which shall be closest to its customers in order to minimize the cost of transportation for a foreseeable future to cater to increased demand.

Problem Statement – To determine best neighbourhood location for opening a new Distribution Centre for a Groceries Supplier in Etobicoke

Data We Need

To answer this business problem, we have leveraged the following data from various data sources:

1- We required the geo-locational information of Etobicoke with its corresponding neighbourhoods and its postal codes. We leveraged latitude and longitude numbers from the following link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2- We also required data about different venues especially of all the categories related to food in different neighbourhoods of Etobicoke. We leveraged "Foursquare" locational information with basic information about all venue especially all the food joints in Etobicoke with respective neighbourhood. We extracted precise latitude and longitude and also its distance from the centre of the neighbourhood. https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&limit={}'.format(
3- We have further leveraged demographics data including population density, change in population as per the last census and average income. Web scraping techniques was used to get Etobicoke's population density and demographics data from the following wiki link: https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods

Methodology

Determining the warehouses' locations can be seen as finding centroids of clusters of the corresponding served branches. Therefore, this is an excellent use case of K-means clustering. We have already known in the first place that the supplier will open new warehouse in Etobicoke where it is closest to the maximum food joints. We can, therefore, address the problem by first identifying the neighbourhood locations, which we can reasonably approximate with the longitude-latitude — comparable to (x,y) — of neighbourhood in Etobicoke and the total number of restaurants in each neighbourhood. Using these data points of city centres, we can actually run the standard K-means clustering to solve the problem already. Yet, this approach would fail to take into account the fact that some cities are more densely populated and have higher average income group than others which implies having a higher volume demand for groceries to be supplied.

In light of this, we also include demographic data like population density, percentage change and average income population data for becoming proxy to groceries demands. Below is the step by step approach which we have followed to solve this:

- Step 1 – First, we pulled using web scarping techniques Postal Codes of different regions inside Etobicoke to find the list of neighbourhoods from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
- Step 2 – We then pulled out demographic information of Toronto from https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods. We then cleaned and prepare the demographics data postal code and neighbourhood wise to match as per neighbourhood in step 1.
- Step 3 – We then connected to Foursquare and Retrieving Locational Data for Each Venue in every Neighbourhood of Toronto. After finding the list of neighbourhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighbourhood. For each neighbourhood, 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 centre of the neighbourhood.
- Step 4: We then Processed the Retrieved Data and Created a DataFrame. When the data is completely gathered, we performed 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 On-hot encoding we will integrate all restaurant columns to one column "Total Food Joints.
- Step 5: We further combined the rest of data for creating final set with defined features i.e. Location, total no. of restaurants, population density and average income of each neighbourhood.

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

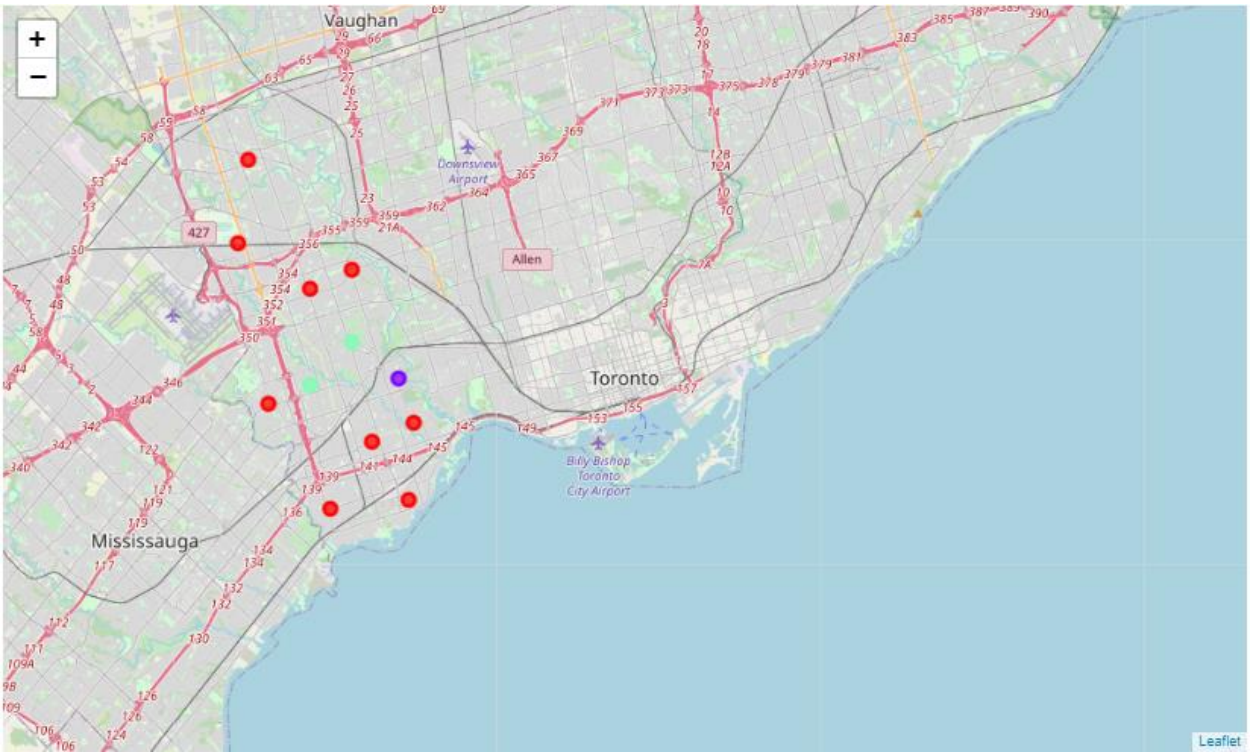
	Neighbourhood	Total Food joints	Density(people/km2)	% change in population since 2001	Average Income
0	New Toronto, Mimico South, Humber Bay Shores	20	7588	-5.7	39186
1	Alderwood, Long Branch	18	2360	-4.0	35239
2	The Kingsway, Montgomery Road, Old Mill North	17	3403	7.9	110944
3	Old Mill South, King's Mill Park, Sunnylea, Hu...	19	3366	-1.1	51398
4	Mimico NW, The Queensway West, South of Bloor,...	19	2629	15.4	47011
5	Islington Avenue, Humber Valley Village	19	2652	-0.1	80618
6	West Deane Park, Princess Gardens, Martin Grov...	20	2249	1.0	80607
7	Eringate, Bloordale Gardens, Old Burnhamthorpe...	15	3507	1.0	51695
8	Westmount	20	5932	5.0	35183
9	Kingsview Village, St. Phillips, Martin Grove ...	19	4013	-6.2	32004
10	South Steeles, Silverstone, Humbergate, Jamest...	23	4229	1.4	28955
11	Northwest, West Humber - Clairville	19	1268	-3.3	26610

Result

Below is the result of final cluster and cluster map:

	Cluster Labels	Neighbourhood	Total Food joints	Density(people/km2)	% change in population since 2001	Average Income	Postal Code	Latitude	Longitude
0	0	New Toronto, Mimico South, Humber Bay Shores	20	7588	-5.7	39186	M8V	43.605647	-79.501321
1	0	Alderwood, Long Branch	18	2360	-4.0	35239	M8W	43.602414	-79.543484
2	1	The Kingsway, Montgomery Road, Old Mill North	17	3403	7.9	110944	M8X	43.653654	-79.506944
3	0	Old Mill South, King's Mill Park, Sunnylea, Hu...	19	3366	-1.1	51398	M8Y	43.636258	-79.498509
4	0	Mimico NW, The Queensway West, South of Bloor,...	19	2629	15.4	47011	M8Z	43.628841	-79.520999
5	2	Islington Avenue, Humber Valley Village	19	2652	-0.1	80618	M9A	43.667856	-79.532242
6	2	West Deane Park, Princess Gardens, Martin Grov...	20	2249	1.0	80607	M9B	43.650943	-79.554724
7	0	Eringate, Bloordale Gardens, Old Burnhamthorpe...	15	3507	1.0	51695	M9C	43.643515	-79.577201
8	0	Westmount	20	5932	5.0	35183	M9P	43.696319	-79.532242
9	0	Kingsview Village, St. Phillips, Martin Grove ...	19	4013	-6.2	32004	M9R	43.688905	-79.554724
10	0	South Steeles, Silverstone, Humbergate, Jamest...	23	4229	1.4	28955	M9V	43.739416	-79.588437
11	0	Northwest, West Humber - Clairville	19	1268	-3.3	26610	M9W	43.706748	-79.594054

Cluster map



Clusters

Cluster 0 (red marker)	**This cluster has high population density with low population growth and low average income**
Cluster 1 (blue marker)	**This cluster has medium population density with high population growth and highest average income**
Cluster 2 (green marker)	**This cluster has low population density, low population growth and medium average income**

Result Discussion

From the results of the clustering algorithm, we recommend that neighbourhoods corresponding to cluster 1 will be the best choice for the location of opening warehouse.

	Cluster Labels	Neighbourhood	Total Food joints	Density(people/km2)	% change in population since 2001	Average Income	Postal Code	Latitude	Longitude
2	1	The Kingsway, Montgomery Road, Old Mill North	17	3403	7.9	110944	M8X	43.653654	-79.506944

Our recommendation is based on the following features of this cluster:

1. It has the highest average income as compared to other clusters and neighbourhood in the cluster has the highest average income
2. This cluster has the second highest cluster in terms of population growth. This hold the potential of further increased demand in near future.
3. Existing no. of restaurants are comparable to no. of restaurants in other clusters.
4. As per the cluster map, its quite evident that it is optimally located and almost a centre point for the entire Etobicoke area for the maximum coverage across other neighbourhoods.

Conclusion

Opening a warehouse in the right location is a complex and very critical task that impact the overall supply chain cost. Thus, extensive research about the area would further increase the chances to have the best located warehouse. From the project above, with the limited data, we have demonstrated the workflow necessary for data-based decision making to determine the best location.