Introduction

Toronto, the most populous city in Canada, is an international center of business, finance, arts, and culture. Its economy is highly diversified with strengths, such as technology, financial services, education, art, and tourism. [1] In the city of Toronto, booksellers could enjoy being part of a community, encouraging the pleasure of reading in adults, and helping to make lifelong readers out of children. For someone who is looking to open a bookstore, it is vital to choose the neighborhood and retail location. The goal of this project is to figure out where a bookstore should be set up for success with data analysis.

"Logistics is generally the detailed organization and implementation of a complex operation. In a general business sense, logistics is the management of the flow of things between the point of origin and the point of consumption to meet the requirements of customers or corporations. The resources managed in logistics may include tangible goods such as materials, equipment, and supplies, as well as food and other consumable items."[i]

Logistics costs represent a big percentage of the products we acquire, independently of the type or nature of it. The logistics process involves several elements, which add costs to our products, these are:
- Storage, warehousing and materials handling.
- Packaging and unitization.
- Inventory.
- Transport.
- Information and control.

The project scope seeks to locate a strategic distribution center for space rental in the Metropolitan Area of Toronto, it aims to cover and influence several aspects of the logistic process for many of the stores located around a heavy populated area.
This opportunity seeks to improve our customers
- ✓ Improve Customer Satisfaction by right on time deliveries.
- ✓ Minimize Translation times.
- ✓ Centralize Procurement.
- ✓ Consolidate Shipments.
- ✓ Data analytics for Decision Making.
- ✓ Automate Warehouse and Logistics Processes through data analytics.
- ✓ Engage and develop a data based software for our customers based on a free trial.

Data Acquisition

1. Neighborhoods in Toronto — Wikipedia[2].
2. Using Folium to get geological location by address name.
3. Using Foursquare API to get the most common venues of given Borough of Toronto.
4. Using Foursquare API to get the venues' record of given venues of Toronto.

[1] https://en.wikipedia.org/wiki/Logistics
[1] https://www12.statcan.gc.ca/census-recensement/2016/as-sa/fogs-spg/Facts-can-eng.cfm?Lang=Eng&GK=CAN&GC=01&TOPIC=1

Methodology

Scrape the Wikipedia page [2] and transform it into a pandas data frame. Postal codes beginning with M are located within the city of Toronto. The data frame consists of three columns: 'Postal Code', 'Borough', and 'Neighborhood'. Since some cells are 'Not assigned', I drop the cells with a borough that is 'Not assigned', and change the cells' neighborhood to borough for cells having a 'Not assigned' neighborhood. I also combine Neighborhoods with the same postal code.

The data frame cleaned has 3 columns and 103 rows. The first five rows are shown below:

	Postal Code	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Acquire the data of latitude and the longitude coordinates in Toronto using Folium. In addition, combine the geological location of each neighborhood with the data frame above. Here I have the dataset that contains Postcode, Borough, Neighborhood, Latitude, and Longitude of each neighborhood. The table below shows the first rows of the new data frame. To brief the project, I choose only to analyze the boroughs that contain the word 'Toronto'.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	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

Once done, we want to include data to find out the population we might target along with the average income, we bring the census data from Canada[ii] 2016 census.

#		Postal Code	AfterTaxIncome2015	Population_2016_x	Bourough	Neighborhood_x	Latitude_x	Longitude_x
92	66	M2P	115237.0	7843.0	North York	York Mills West	43.752758	-79.400049
47	55	M5M	111821.0	25975.0	North York	Bedford Park, Lawrence Manor East	43.733283	-79.419750
78	61	M4N	109841.0	15330.0	Central Toronto	Lawrence Park	43.728020	-79.388790
46	74	M5R	108271.0	26496.0	Central Toronto	The Annex, North Midtown, Yorkville	43.672710	-79.405678
88	98	M8X	97210.0	10787.0	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944

Then after merging the information and using the foursquare app, we will collect and review all the venues in Toronto and check in which locations we have more.
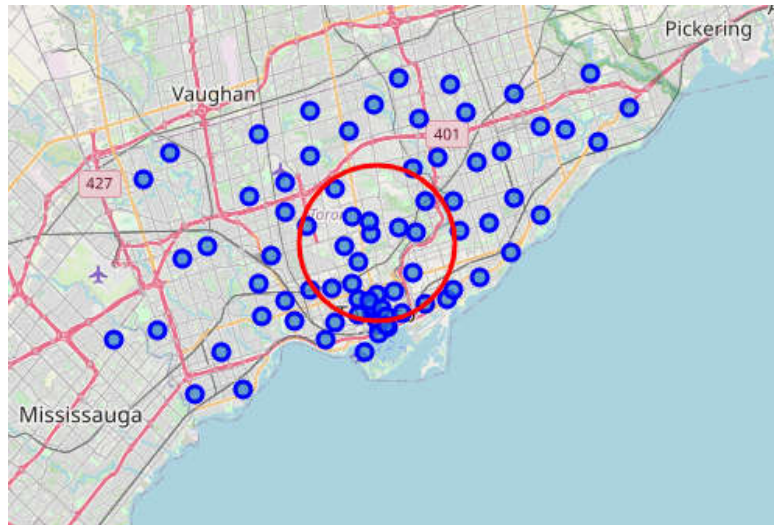
We want to locate our logistic center in a strategic place, which will cover more area and customers in a radius, following our cost effective premises shared before.

Cluster and segment the venues and rank them for most common in the area in Toronto.

Moreover, use folium to visualize the distribution and locate the best area to locate the distribution center which will be in an equal center to all in Toronto.

[1] https://en.wikipedia.org/wiki/Logistics
[1] https://www12.statcan.gc.ca/census-recensement/2016/as-sa/fogs-spg/Facts-can-eng.cfm?Lang=Eng&GK=CAN&GC=01&TOPIC=1

Explaining the results

As we built, our list of neighborhoods with all the venues we discovered came to advise that the greatest concentration is in Central Toronto and downtown Toronto. This is not ideal for a logistic center, but we can minimize the impact by locating our center in a broader spectrum and covering other areas.

This might seem obvious but it would also appear that these are some of the most affluent neighborhoods in Toronto so we could sense a be correlation.

By Locating our logistic center in the general vicinity we can minimize most of the costs incurred in time, logistics could be geographically centered in this cluster and poised to service a broader customer base with greatest efficiency.

When we built our K-Means dataset we used Silhouette analysis to tell us there was a lot of similarity between neighborhoods and knowing the most common venues will let you prepare the warehouses or slots according, along with a better customer approach on their needs. For our surprise, the analysis led to only two types of cluster or neighborhoods in greater Toronto. The vast majority of those were in 1 cluster. So Toronto venues are evenly located near the center of Toronto.

Of the 103 Toronto Neighborhoods gathered only 55.3% are above the median after-tax income. 37.8% are below the median after-tax income and 6.7% neighborhoods did not register, as it appears their populations are too low. We want to be close to or near a middle range, where all of our services will be of benefit for the stores and population in the area, however after a close look to our map, this information was disregarded as it won't affect due to the proximity on them. Special note: this could be assessed further with a deeper analysis.

Conclusion

The subject area covers all the needs and it gives us many options to choose, giving the flexibility to pick for a place with better roads and communication between areas and venues.

While simple, a small analysis can give further information regarding many subjects and I think that this was accomplished through all this course.

Thank you

---

ii https://www12.statcan.gc.ca/census-recensement/2016/as-sa/fogs-spg/Facts-can-eng.cfm?Lang=Eng&GK=CAN&GC=01&TOPIC=1

[1] https://en.wikipedia.org/wiki/Logistics
[1] https://www12.statcan.gc.ca/census-recensement/2016/as-sa/fogs-spg/Facts-can-eng.cfm?Lang=Eng&GK=CAN&GC=01&TOPIC=1