Opening a Pharmacy in Toronto| Project Report

-Ishaasamyuktha. S

(www.github.com/Ishaasamyuktha)

Introduction

Toronto is considered as one of the safest countries in the world. Neighborhoods of Toronto are filled with nightlife, colourful restaurant scene, museum, and art galleries in addition, Toronto provides almost free health services to the people even if you have just moved to Canada. If you have been a resident of Ontario for three months, you are entitled to health care services paid for by the Ontario Health Insurance Plan (OHIP). This provides additional benefits in migrating to Toronto along with good housing prices and reputed schools for children.

This project aims in helping pharmacists in exploring neighborhoods who wants to open a new pharmacy in Toronto area. This project will aid in making smart and effective decision on choosing the better neighbourhood for business development.

Objective

The major aim of this project, is to suggest a better neighborhood in a new city for an entrepreneur who wants to start a business here. This helps the entrepreneur wanting to start a medical shop or pharmacy here. The competition of that business in different neighborhoods, and which neighbourhood needs a pharmacy in their locality. And the objective can be further extended to knowledge of people are shifting here about medical shops in here neighbourhood.

Target Audience

- Budding entrepreneurs who wants to start a pharmacy in a new city.
- Canadian immigrants who want to explore and find neighborhoods with better medical facilities.

Data Acquisition

For this project we require the following:

- List of neighborhoods in Canada.
- The location i.e. Latitude and Longitude of these neighborhoods.
- Number of pharmacy/Medical Shops in each neighborhood.

Data Extraction

- Using Wikipedia to scrap the data of Toronto city containing boroughs, neighborhoods and postal codes.

- Geocoder Package to obtain latitudes and longitudes of these neighborhoods.
- Use of Foursquare app to get the venue details regarding medical shops.

Methodology

1. Data Cleansing

The data acquired need to be sorted and cleaned before using for our problem solving. The data is extracted from wikipedia page (“https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M ”) and the data is scraped using beautiful soup to handle http requests. And XML is used to separate data from presentation and XML stores data in plain text format. Next the coordinates of the neighbourhood data is obtained using geocoder package.

The Toronto neighbourhood are visualised using folium maps to verify whether they are correct coordinates. Foursquare app is then utilised to get the list of venues near these neighborhoods. Foursquare API is used to pull the list of top 100 venues within 500 meters radius. To use this, we need to create a Foursquare developer account in order to obtain account ID and API key to pull the data. From Foursquare, it is possible to pull the names, categories, latitude and longitude of the venues. With this data, it is checked how many unique categories are there from these venues. Each neighbourhood is analysed by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category. This is to prepare clustering to be done later. Now as for our aim pharmacies are searched in each neighbourhood to analyse and find number of medical shops available in each neighbourhood.

2. Foursquare API

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. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

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 100 meter.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

- Neighborhood
- Neighborhood Latitude
- Neighborhood Longitude
- Venue
- Name of the venue
- Venue Latitude
- Venue Longitude
- Venue Category

3. Libraries Imported

Pandas: For creating and manipulating data frames.

Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Data.

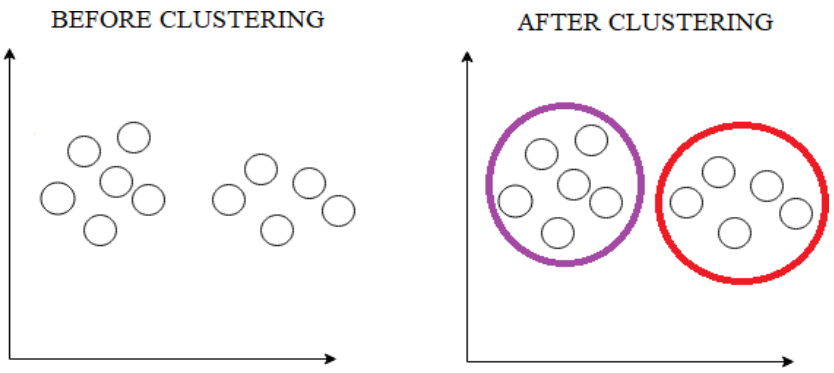
Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.

4. Clustering Approach

K-Means

K-Means is a form of non-parametric unsupervised learning. It works on the basis of forming clusters with the data available. Clustering groups objects that are similar to each other, but different from others. The clustering technique followed in K-Means is centroid-based. A centroid is a point lying at the centre of the data. Similarity between data points is established based on their proximity to the centroid.



The algorithm will assume a required number of centroids (representing a different cluster each), and for any input point, the distance to either of the centroids will be computed. The point belongs to the cluster to which the distance is least. For the project the number of clusters is set as 3 based on their frequency of occurrence for “Pharmacy”. Based on the concentration of clusters, it will be easy to determine and recommend the ideal location to open a medical shop. The working algorithm is seen in figure.

Results and Discussion

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M4E	East Toronto	The Beaches	43.67703	-79.29542
1	M4K	East Toronto	The Danforth West, Riverdale	43.68375	-79.35528
2	M4L	East Toronto	India Bazaar, The Beaches West	43.66797	-79.31468
3	M4M	East Toronto	Studio District	43.66091	-79.33503
4	M4N	Central Toronto	Lawrence Park	43.72898	-79.39173

Fig.1: Neighborhoods in Toronto with their Latitude and Longitude Coordinates using Geocoder

Fig.1 shows the how postal codes, boroughs and neighborhoods are scrapped from Wikipedia and made into data frames. Their geographical coordinated are extracted using geocoder.

	Neighborhood	Pharmacy	Cluster Labels
0	Berczy Park	0.030303	1
1	Brockton, Parkdale Village, Exhibition Place	0.011628	2
2	Business reply mail Processing Centre, South C...	0.010000	2
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0
4	Central Bay Street	0.000000	0
5	Christie	0.000000	0
6	Church and Wellesley	0.000000	0
7	Commerce Court, Victoria Hotel	0.000000	0
8	Davisville	0.000000	0
9	Davisville North	0.000000	0

Fig2: Neighborhood segregated into 3 clusters based on their distance from medical shops

The nearby venues are found using Foursquare app for these neighborhoods and classified using unsupervised learning algorithm K-Means to cluster these areas with respect to the distances from pharmacies represented in Fig.2.

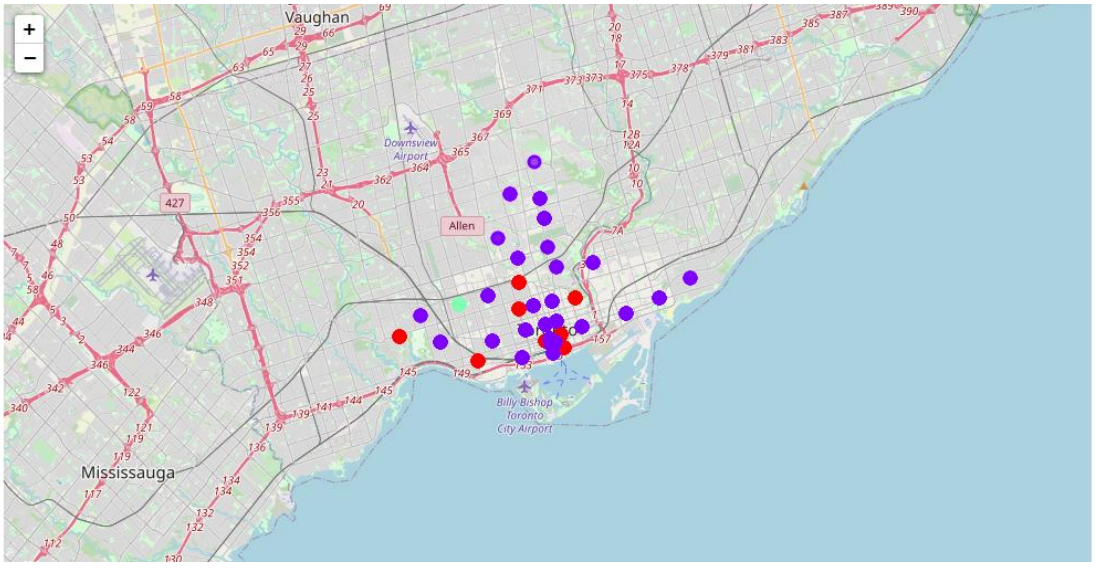


Fig3: Clusters viewed using Folium

The results from K- means clustering is visualized using Folium as seen in Fig.3. Toronto neighborhoods are classified into 3 clusters based on how many pharmacies are in each neighborhood:

- Cluster 0: Neighborhoods with no medical shops.
- Cluster 1: Neighborhoods with high number of medical shops.
- Cluster 2: Neighborhoods with little or no medical shops.

The results are visualized in the above map with Cluster 0 in green colour, Cluster 1 in blue colour and Cluster 2 in red colour.

6	Church and Wellesley	0.0	0	43.66659	-79.38130	Barbara Hall Park	43.666879	-79.381068	Park
6	Church and Wellesley	0.0	0	43.66659	-79.38130	Fabarnak	43.666377	-79.380964	Restaurant
25	Richmond, Adelaide, King	0.0	0	43.64970	-79.38258	The Cambridge Club	43.651663	-79.383075	Gym
6	Church and Wellesley	0.0	0	43.66659	-79.38130	Smith	43.666927	-79.381421	Breakfast Spot
5	Christie	0.0	0	43.66878	-79.42071	Marlenes Just Babies	43.671824	-79.420499	Baby Store
5	Christie	0.0	0	43.66878	-79.42071	Loblaws	43.671657	-79.421364	Grocery Store
5	Christie	0.0	0	43.66878	-79.42071	Stubbe Chocolates	43.671566	-79.421289	Candy Store
5	Christie	0.0	0	43.66878	-79.42071	Faema Caffe	43.671046	-79.419297	Café
5	Christie	0.0	0	43.66878	-79.42071	Vinny's Panini	43.670679	-79.426148	Italian Restaurant
5	Christie	0.0	0	43.66878	-79.42071	Vermont Square Park	43.670493	-79.415399	Playground
5	Christie	0.0	0	43.66878	-79.42071	Starbucks	43.671530	-79.421400	Coffee Shop
5	Christie	0.0	0	43.66878	-79.42071	Contra Cafe	43.669107	-79.426105	Café
5	Christie	0.0	0	43.66878	-79.42071	Scout and Cash Caffe	43.667360	-79.419938	Café
5	Christie	0.0	0	43.66878	-79.42071	Fiesta Farms	43.668471	-79.420485	Grocery Store
4	Central Bay Street	0.0	0	43.65609	-79.38493	Tim Hortons	43.655212	-79.380063	Coffee Shop
5	Christie	0.0	0	43.66878	-79.42071	Queens Club	43.672386	-79.418106	Athletics & Sports

Fig4: Localities in cluster 0

	Neighborhood	Pharmacy	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
33	The Annex, North Midtown, Yorkville	0.033333	1	43.67484	-79.40185	Ezra's Pound	43.675153	-79.405858	Café
33	The Annex, North Midtown, Yorkville	0.033333	1	43.67484	-79.40185	Playa Cabana	43.676112	-79.401279	Mexican Restaurant
33	The Annex, North Midtown, Yorkville	0.033333	1	43.67484	-79.40185	Big Crow	43.675896	-79.403680	BBQ Joint
33	The Annex, North Midtown, Yorkville	0.033333	1	43.67484	-79.40185	Le Paradis	43.675007	-79.400036	French Restaurant
33	The Annex, North Midtown, Yorkville	0.033333	1	43.67484	-79.40185	L'Unita Restaurant	43.674387	-79.396488	Italian Restaurant
33	The Annex, North Midtown, Yorkville	0.033333	1	43.67484	-79.40185	Glen Edyth Drive Parkette	43.678294	-79.404868	Park
33	The Annex, North Midtown, Yorkville	0.033333	1	43.67484	-79.40185	food plus market 24 hrs	43.674515	-79.407138	Grocery Store
33	The Annex, North Midtown, Yorkville	0.033333	1	43.67484	-79.40185	Nancy's Cheese	43.674851	-79.406753	Cheese Shop
33	The Annex, North Midtown, Yorkville	0.033333	1	43.67484	-79.40185	Annex Grill and Bar	43.674913	-79.406709	Asian Restaurant
33	The Annex, North Midtown, Yorkville	0.033333	1	43.67484	-79.40185	Shoppers Drug Mart	43.674959	-79.407986	Pharmacy

Fig5: Localities in Cluster 1

29	Cabbagetown	0.021277	2	43.66788	-79.36649	Bistro	43.666856	-79.368752	Café
29	St. James Town, Cabbagetown	0.021277	2	43.66788	-79.36649	Cabbagetown Brew	43.666923	-79.369289	Café
29	St. James Town, Cabbagetown	0.021277	2	43.66788	-79.36649	Absolute Bakery & Café	43.667469	-79.369277	Bakery
29	St. James Town, Cabbagetown	0.021277	2	43.66788	-79.36649	Butter Chicken Factory	43.667072	-79.369184	Indian Restaurant
29	St. James Town, Cabbagetown	0.021277	2	43.66788	-79.36649	Fair Trade Jewellery Co.	43.665348	-79.368362	Jewelry Store
29	St. James Town, Cabbagetown	0.021277	2	43.66788	-79.36649	Toronto Dance Theatre	43.666232	-79.367075	General Entertainment
29	St. James Town, Cabbagetown	0.021277	2	43.66788	-79.36649	Riverdale Farm	43.666850	-79.361376	Farm
29	St. James Town, Cabbagetown	0.021277	2	43.66788	-79.36649	Riverdale Park West	43.666048	-79.360941	Park
29	St. James Town, Cabbagetown	0.021277	2	43.66788	-79.36649	House on Parliament	43.663646	-79.367854	Gastropub
29	St. James Town, Cabbagetown	0.021277	2	43.66788	-79.36649	St. Jamestown Delicatessen	43.665811	-79.368648	Butcher
28	St. James Town	0.011494	2	43.65143	-79.37557	Tim Hortons	43.647955	-79.373833	Coffee Shop
37	University of Toronto, Harbord	0.017544	2	43.66311	-79.40180	Good For Her	43.661816	-79.407737	Adult Boutique

Fig6: Localities in Cluster 2

Fig.4, Fig.5, Fig.6 represents the neighborhoods which fall under the three clusters with no pharmacy, high and little number of pharmacies in Cluster 0,1,2 respectively.

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing the machine learning by utilizing k-means clustering and providing recommendation to the stakeholder.

Recommendation

Most of Pharmacies are present in Cluster 1 which is around Berczy park, Yorkville, and Dovercourt village areas and lowest (close to zero) in Cluster 0 areas which are Central Bay Street, Reyerson, Richmond, Adelaide, Church and Wellesley. Also, there are good opportunities to open near University of Toronto, St James town as the competition seems to be low. Looking at nearby venues, it seems Cluster 0 might be a good location as there are not a lot of medical shops in these areas. Therefore, this project recommends the entrepreneur to open a Pharmacy in these locations with little to no competition. Nonetheless, if the quality of the medicine, biomedical waste management, hygiene and service and hospitality is good, It is possible that pharmacy will have great following everywhere.

Future Work

In this project, only one consideration is taken i.e., existence of medical shops in each neighborhood. There are many factors that can be taken into consideration such as population density, income of residents, rent that could influence the decision to open a new pharmacy. Future research can take into consideration of these factors. In addition, future research can take into consideration of other variables such as existence of medical camps, hospitals and health services, accident prone areas which require immediate medical shops for first aid in each neighborhood.

References

1. List of neighborhoods in Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2. Foursquare Developer Documentation: https://developer.foursquare.com/docs