Capstone Project - The Battle of Neighborhoods

1. Introduction

1.1 Background

Toronto is a city with scores of neighborhoods, some of which have identities that have been imposed upon them by realtors, whereas others are of much longer standing and have a more distinctive character. The main premise of our research will center on investigating and determining the ideal locale in Toronto, Canada, for the establishment of a new dining experience, a restaurant targeting people whose taste palette is unconventional and outside the box. In this study, I endeavor to identify the right location by finding the right cluster of people and amenities, where income, age, education, household size combined with the right competitive mixture to create an enabling environment to support business and those who depend on it.

1.2 Business Problem

Opening a successful restaurant is all about location; physical brick-and-mortar venues matter even in today's world of virtual reality. Location can make or break a restaurant. Toronto is Canada's business and financial capital. The city is the second largest financial services center in North America and has one of the highest concentrations of financial services company headquarters in the Americas. With its reputation for safety, soundness and stability, Toronto is fast becoming a global location destination for financial services. The object of this project is to identify the ideal location to start a new restaurant in Toronto, Canada.

1.3 Target Audience

Our target audience are prospective investors and entrepreneurs who consider to establish a restaurant in Toronto. One such critical aspect of an exercise of this nature, is selecting a location, that is, often what type of community, city, state or even country to locate their business in. This capstone project hopes to make this process a little less burdensome and scientific, by providing a robust and grounded analytical framework through which one can better assess the costs and benefits of a location relative to each other.

2. Data Acquisition and Cleaning

Neighborhoods in Toronto is provided in Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). I will use Python Requests and Beautiful-soup packages to scrape and extract the data table from the Wikipedia page. Then I used Toronto_Geo list (https://cocl.us/Geospatial_data) to get latitude and longitude into the Toronto neighborhood dataset based on postal code. Finally, I will connect Python with Foursquere API to get venue data for Toronto neighborhoods.

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	Riverdale, The Danforth West	43.679557	-79.352188	MenEssentials	43.677820	-79.351265	Cosmetics Shop
5	Riverdale, The Danforth West	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant
6	Riverdale, The Danforth West	43.679557	-79.352188	Cafe Fiorentina	43.677743	-79.350115	Italian Restaurant
7	Riverdale, The Danforth West	43.679557	-79.352188	Dolce Gelato	43.677773	-79.351187	Ice Cream Shop
8	Riverdale, The Danforth West	43.679557	-79.352188	La Diperie	43.677530	-79.352295	Ice Cream Shop
9	Riverdale, The Danforth West	43.679557	-79.352188	Moksha Yoga Danforth	43.677622	-79.352116	Yoga Studio
10	Riverdale, The Danforth West	43.679557	-79.352188	Mezes	43.677962	-79.350196	Greek Restaurant
11	Riverdale, The Danforth West	43.679557	-79.352188	Louis Cifer Brew Works	43.677663	-79.351313	Brewery
12	Riverdale, The Danforth West	43.679557	-79.352188	Alexandros	43.678304	-79.349486	Greek Restaurant
13	Riverdale, The Danforth West	43.679557	-79.352188	Valley Farm Produce	43.677999	-79.349969	Fruit & Vegetable Store
14	Riverdale, The Danforth West	43.679557	-79.352188	7 Numbers	43.677062	-79.353934	Italian Restaurant

(Dataframe: toronto_venues)

I removed “Not Assigned” cells and only kept those boroughs which have "Toronto" written in the names in order to clean the data. As a result, there are 103 postal codes, 10 boroughs, and 98 neighborhoods left in the dataset.

	Postal code	Borough	Neighborhood
count	103	103	103
unique	103	10	98
top	M4G	North York	Downsview
freq	1	24	4

3. Methodology

An unsupervised machine learning algorithm – K-means clustering will be conducted to help determine where is the best place to run what kind of business. For this project the focus will be on looking for the cluster of neighborhood which has more restaurants than the others.

3.1 Frequency of Venue Category by Neighborhood

First, I converted Venue Category into dummy variables with values in 1 or 0.

Neighbourhood	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium
The Beaches	0	0	0	0	0	0	0	0	0
The Beaches	0	0	0	0	0	0	0	0	0
The Beaches	0	0	0	0	0	0	0	0	0
The Beaches	0	0	0	0	0	0	0	0	0
Riverdale, The Danforth West	0	0	0	0	0	0	0	0	0

(Dataframe: toronto_onehot)

Then I got the frequency of each venue category for neighborhoods.

Neighbourhood	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop
Adelaide, King, Richmond	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.000000
Bathurst Quay, CN Tower, Harbourfront West, Is...	0.058824	0.058824	0.058824	0.117647	0.176471	0.117647	0.000000	0.000000
Berczy Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Brockton, Exhibition Place, Parkdale Village	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

(Dataframe: toronto_grouped)

Sorted the frequencies in descending order and kept 5 most common venue categories. Here is how the result looks:

Neighbourhood	1st	2nd	3rd	4th	5th
Adelaide, King, Richmond	Coffee Shop	Café	Seafood Restaurant	Lounge	Smoke Shop
Bathurst Quay, CN Tower, Harbourfront West, Is...	Airport Service	Airport Lounge	Airport Terminal	Airport	Harbor / Marina
Berczy Park	Coffee Shop	Cocktail Bar	Beer Bar	Seafood Restaurant	Park
Brockton, Exhibition Place, Parkdale Village	Café	Breakfast Spot	Nightclub	Coffee Shop	Pet Store
Business Reply Mail Processing Centre 969 Eastern	Yoga Studio	Spa	Garden	Gym / Fitness Center	Fast Food Restaurant

(Data frame: toronto_neighborhood)

3.2 Clustering Neighborhoods

I conducted K-means clustering (k=10) with data frame toronto_grouped which has frequencies of venue category for each neighborhood. Once I got Cluster Labels I added Latitude and Longitude by merging data frame toronto_venues on Neighborhoods to complete the final data frame for visualization on the map.

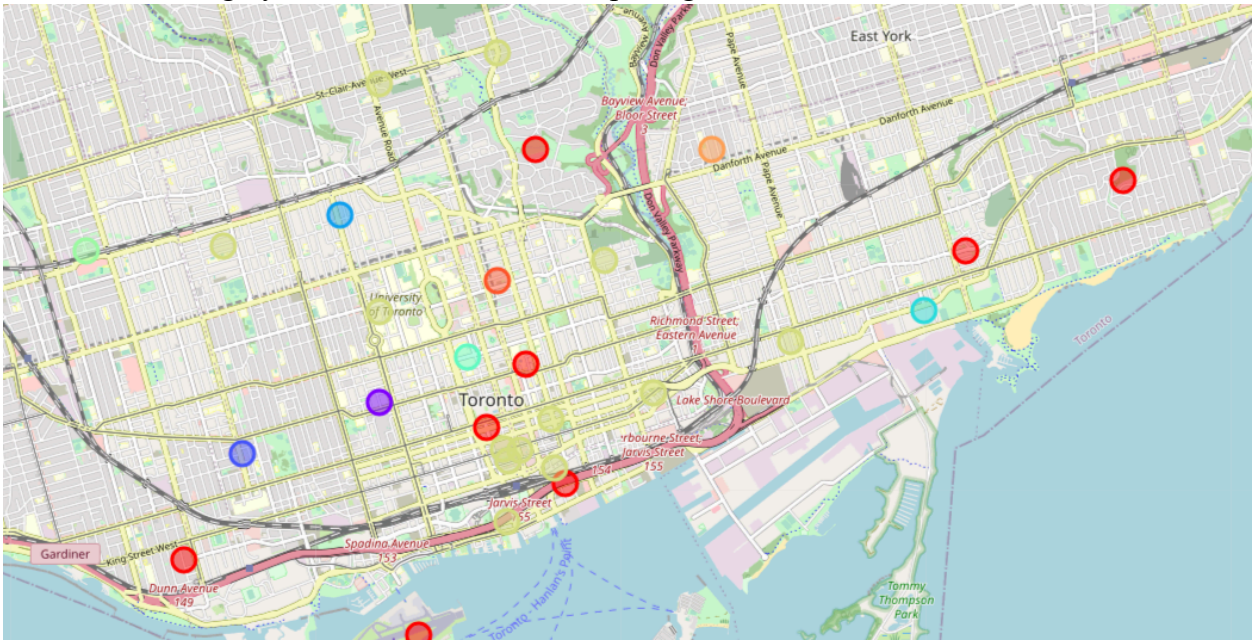
Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st	2nd	3rd	4th	5th
East Toronto	The Beaches	43.676357	-79.293031	0	Neighborhood	Trail	Pub	Health Food Store	Yoga Studio
East Toronto	Riverdale, The Danforth West	43.679557	-79.352188	8	Greek Restaurant	Italian Restaurant	Ice Cream Shop	Yoga Studio	Pizza Place
East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572	0	Sandwich Place	Fast Food Restaurant	Gym	Liquor Store	Ice Cream Shop
East Toronto	Studio District	43.659526	-79.340923	7	Café	Coffee Shop	Bakery	Neighborhood	Middle Eastern Restaurant
Central Toronto	Lawrence Park	43.728020	-79.388790	7	Park	Bus Line	Construction & Landscaping	Swim School	Donut Shop

(Data frame: toronto_Nei_Lable)

4. Results

Let us use Folium to populate the clusters on the map. The orange cluster represented Cluster 8

which has more restaurants than the other clusters. Cluster 8 is located around East Toronto area which I would highly recommend to consider opening new restaurants here.



5. Discussion

Density of a certain kind of business is an important factor to consider when looking for a location. Higher density means better popularity but also means more competition in this region. To make a restaurant outstanding among the competition, investors and entrepreneurs also need to continue analyzing more details such as types of surrounding restaurants and local culture in order to determine target customers and marketing strategy.

6. Conclusion

Machine learning provides useful algorithms to explore a greater volume and variety of data quickly. Due to the size and complexity of these data sets, machine learning can help unlock value from all this data in a way that humans cannot. Decision makers can leverage technologies to drive efficiency. However, decision-making steps still need humans involved to make sure the results support the initial purposes of the project.