SL – Capstone Project – Battle of Neighborhoods:

Vegan Franchising

By Sohail Ladha

June 2020

1. Introduction / Business Problem

It is widely understood that cities have their own individual culture and vibe. New York has a different culture to Paris or London or Toronto or Tokyo. However, it is important to note that individual neighborhoods have their own sub-cultures. A neighborhood in New York, for example, may have more in common with a neighborhood in Amsterdam than with other neighborhoods in New York. The specific attributes of the neighborhood may have more of an impact on the success of a business than the impact of the city itself. This is one of the key themes of this paper.

Somerville

Somerville is an example of a unique neighborhood in New York. New York Times recently published an article highlighting the benefits and status of the neighborhood, which has been under the radar for a long time.

https://www.nytimes.com/2019/08/14/realestate/somerville-nj-a-walkable-suburban-alternative.html

Fresh Tiki, a vegan franchise

Mr. Xi operates a famous vegan restaurant in Somerville, Fresh Tiki. This restaurant has received excellent reviews from food critics and the company is planning to open in numerous locations globally.

Recently, Mr. Xi was approached by Ms.Varga, a prominent Toronto businesswoman, to open five locations of Fresh Tiki in Toronto. Mr. Xi is amenable to opening such locations on a franchise basis, but wants to provide his consent regarding locations.

The Problem

Mr. Xi is a firm believer in Somerville and he believes that the eclectic nature of the neighborhood, with its high walkability and urban / suburban feel, have been strong contributors to the success of Fresh Tiki. Therefore, one of his requirements is that Fresh Tiki be opened in similar locations in Toronto.

Mr. Xi wants to shortlist up to 10 locations to serve as the initial neighborhoods for opening Fresh Tiki in Toronto. The requirements for this shortlist are that the selected locations 'must be similar to Summerville'. He has engaged our data science firm to assist with this real world problem.

Target Audience

1. The initial target audience is Mr. Xi, who is seeking to find similar neighborhoods to Somerville (New York neighborhood) in Toronto in order to open is chain of vegan restaurants.

2. The same solution can be useful to any business or household seeking to shift from Somerville (New York) to Toronto, while maintaining a similar feel.

Other variations of the analysis can be used for companies deciding to move from one neighborhood in New York to Toronto or vice versa (or expand from one city to another).

2. The Data

In order to solve this problem, I will be using three possible sources of data:

1. Data regarding Toronto neighborhoods. This data is available from the following Wikipedia page:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

This includes data regarding postal codes, boroughs, neighborhoods. In this data-set we cover 103 neighborhoods in Toronto.

2. Data regarding New York neighborhoods. The data is available from the following JSON file:

https://cocl.us/new_york_dataset

This includes data regarding boroughs, neighborhoods. In this data-set we cover 306 neighborhoods in New York.

3. Data in the Foursquare API. We will make calls to the Foursquare API to explore individual neighborhoods. The data collected from Foursquare will be the types of locations located within a 500 meter radius of the neighborhood.

In order to work with the data, we will have to clean the data in New York and Toronto to remove unassigned neighborhoods. There are 409 combined neighborhoods; once we remove certain duplicate items, there are 395 unique neighborhoods.

We will also need to extract locational data (longitude and latitude data) so that we can use it to extract interesting locations in or around the neighborhood from Foursquare API. This data should be sufficient for us to populate the locations with key characteristics of surroundings of the neighborhood.

This forms our final data set for further analysis.

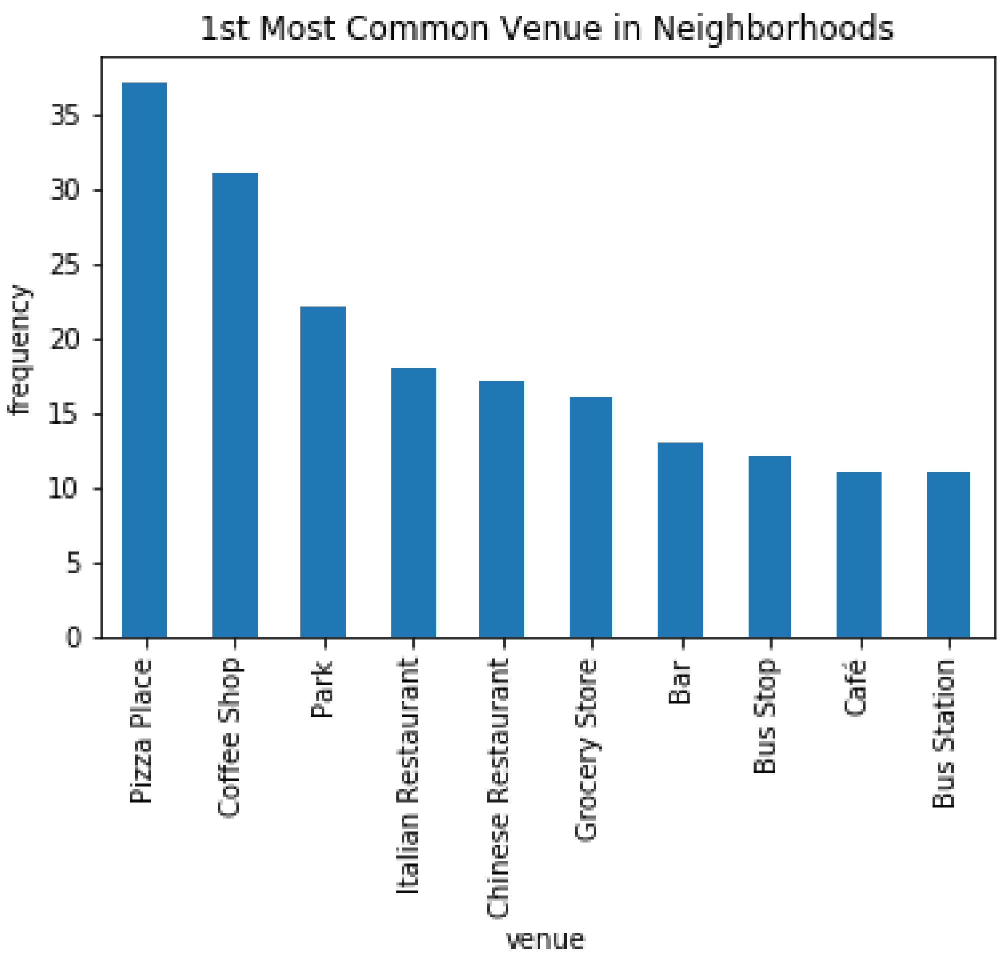
3. Methodology

Preliminary data analysis

Once our data set is ready for further analysis, we completed some preliminary data analysis through running some basic statistics regarding features of the data. This included assessing the most frequent location types in the neighborhoods as well as the most concentrated neighborhoods.

- Our data set consisted of 12,089 different venues in 395 individual neighborhoods.
- There are, in fact, 463 unique categories of venues within those neighborhoods.

- We are able to group and sort data so that we can assess the most frequent venues in the individual neighborhoods.
- The most frequent venues in Neighborhoods are shown in the chart below.



Model implementation

In order to solve our problem, we will use the K-means Clustering Algorithm. This is an unsupervised machine learning approach, especially useful when we have unlabeled data. The idea is to cluster our neighborhoods into various groups, without providing any guidance related to what is similar about these neighborhoods. A key input into this clustering approach is the need to specify the number of cetroids, i.e. K. We used K=5, i.e. to cluster neighborhoods in New York and Toronto in five clusters.

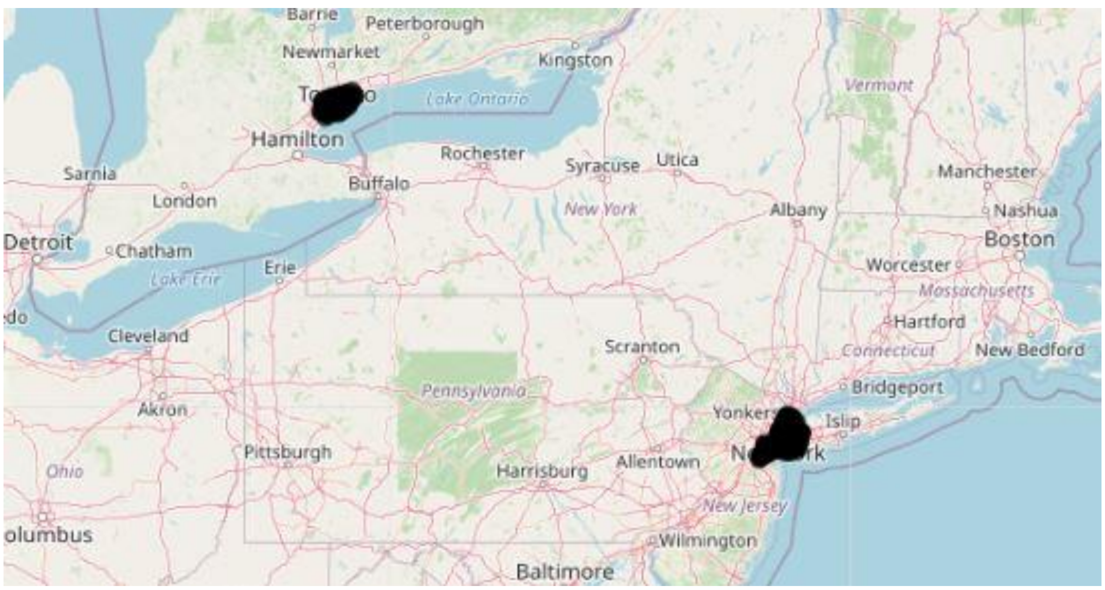
Once the clusters were properly formed, we:

- Assessed the individual clusters to assess composition and similarities.
- Compared neighborhoods within clusters as well as across clusters.

The focus of the analysis was to assess the cluster consisting of Somerville, a neighborhood in New York to find similar neighborhoods in Toronto. We were successful in finding several neighborhoods in Toronto, which were in the same cluster.

4. Results

- The 395 neighborhoods fell within five clusters.



- Some clusters were far more concentrated than other clusters.
- One cluster was purely focused on Toronto.
- Other clusters were purely New York clusters.
- Some of our important results included the following:

	Number of Neighborhoods	Key features
Cluster 1	12	Closeness to parks, fitness studios, farms, farmers' markets, women's stores and ethnic restaurants
Cluster 2	300	Emphasis on cafes, fast food restaurants, bars, sushi restaurants, etc.
Cluster 3	2	Baseball field centric clusters in Toronto
Cluster 4	86	Clusters with a strong emphasis on pizza places, pubs and cafes.
Cluster 5	4	Beach / pier centric neighborhoods in New York.

Our cluster of interest is Cluster 1, the cluster of Somerville.

		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
Toronto	0	Parkwoods	Food & Drink Shop	Park	Fireworks Store	Falafel Restaurant	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service
	21	Caledonia-Fairbanks	Park	Women's Store	Pool	Farm	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service
	35	East Toronto, Broadview North (Old East York)	Convenience Store	Metro Station	Park	Intersection	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service
	61	Lawrence Park	Bus Line	Swim School	Park	Flower Shop	Flea Market	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant
	64	Weston	Park	Convenience Store	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant
	66	York Mills West	Convenience Store	Park	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant
	83	Moore Park, Summerhill East	Gym	Park	Women's Store	Falafel Restaurant	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service
	85	Milliken, Agincourt North, Steeles East, L'Amo...	Park	Playground	Women's Store	Farm	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service
	98	The Kingsway, Montgomery Road, Old Mill North	Park	River	Women's Store	Farm	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service
New York	27	Clason Point	Park	Convenience Store	Pool	South American Restaurant	Bus Stop	Boat or Ferry	Grocery Store	Egyptian Restaurant	Eastern European Restaurant	Electronics Store
	192	Somerville	Park	Women's Store	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant
	203	Todt Hill	Park	Trail	Women's Store	Falafel Restaurant	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service

5. Discussion

This cluster of interest, the "Somerville" cluster (Cluster 1) seems to validate Mr. Xi's assertion that Somerville is a fairly unique neighborhood in New York. There are only three other neighborhoods in New York with similar characteristics / vibe.

This cluster can be characterized as:

- Being close to parks, gyms, women's stores.
- Having a high concentration of Mediterranean, Eastern European and empanada restaurants.
- Open spaces.
- Unique areas, e.g. flea markets, flowers shops, farms.

There are 9 locations similar to Somerville neighborhood in Toronto.

As Fresh Tiki considers opening various locations in Toronto, these are the 9 possible neighborhoods that provide the same feel. These neighborhoods include:

- Parkwoods
- Caledonia-Fairbanks
- East Toronto, Broadview North
- Lawrence Park
- Weston
- York Mills West
- Moore Park, Summerhill East
- Mililikwn, Agincourt North, Steeles East
- The Kingsway, Montgomery Road, Old Mill North

6. Conclusion

Mr. Xi's assertion was that the Somerville neighborhood was unique in New York. He further stated that the feel of this neighborhood and contributed to the success of his restaurant, Fresh Tiki. Even though we cannot comment on how the neighborhood helped the location succeed, we can confirm through our analysis that there are only two other neighborhoods which fall within the same cluster. We were able to assess this through our machine learning technique, K-means clustering.

Through the analysis, we were also able to successfully identify several neighborhoods in Toronto, which fall within the same cluster. Therefore, these could provide a short-list for Mr. Xi as he considers franchising several locations in Toronto.

This analysis could be helpful to other people and businesses considering moving from their preferred neighborhood in Toronto to New York, or vice versa.

Further areas of analysis for this might include adding other metrics to the analysis, such as crime data, education data and other socio-economic parameters. These may provide a better understanding of the individual neighborhoods.