Location for Your Second Coffee Shop in Toronto

A capstone project for IBM Data Science Professional Certificate

Elly Zhao

6/18/2020

Executive summary: a coffee shop owner in Toronto decided to open a second coffee shop in the same city. This project is to analyze the neighborhoods and find a suitable location for the client to replicate their success.

Table of Contents

1.0 Introduction

A client of mine owns a coffee shop in Toronto and their business is doing very well that they now want to open a second coffee shop in the city. Imagine my client's coffee shop is in **Little Portugal** neighborhood of Toronto and with an average rating of 8.0 (out of 10.0). My task is to analyze the neighborhoods of Toronto to determine where would be the best location for my client to open their second shop.

To help my client succeed in their business, besides their management, the location of the shop is also critical. I tend to pick an environment that is very similar to where they have their first coffee shop. In such way, their second coffee shop will likely to attract similar customers, and they can expect similar business performance, similar traffic or even similar competitors. In conclusion I want to find which neighborhood is the most similar to my client's current neighborhood, **Little Portugal,** to open the second coffee shop.

2.0 Data

Data required for this project includes:

- Postal code, borough, and neighborhood of Canada from Wikipedia page, from where the Toronto neighborhood data will be filtered out.
- Latitude and longitude info of each neighborhood from here.
- Nearby venue data extracted from Foursquare database,
    - Nearby venues of each neighborhood,
    - Nearby coffee shops or coffee related venues of each neighborhood,
    - Details of each coffee shop: locations, ratings, price range, number of likes or dislikes, etc.

3.0 Methodology

The goal is to find a neighborhood that is similar to Little Portugal for my client to open their second coffee shop. To answer this question, first I will analyze the venue information in each neighborhood using the Foursquare location data, focusing on those drink and food venues. Next, build a machine learning model to segment neighborhoods in Toronto. The clustering technique, k-means, will be used for this task. I will examine each cluster and further segment the clusters if necessary. Then I will analyze the number of coffee shops in each neighborhood to see how competitive it is in these neighborhoods. Finally I will pick the most suitable neighborhood for my client's second coffee shop, and further analyze the quality of the existing coffee shops in the area, including their ratings, price range, number of likes or dislikes, etc.

4.0 Results

4.1 Overview of Toronto Neighborhoods

Canada postal code, borough and neighborhood data were cleaned up and 39 neighborhoods were found in Toronto region, as shown in Figure **1**. They belong to 4 boroughs: Downtown Toronto, East Toronto, West Toronto, and Central Toronto.

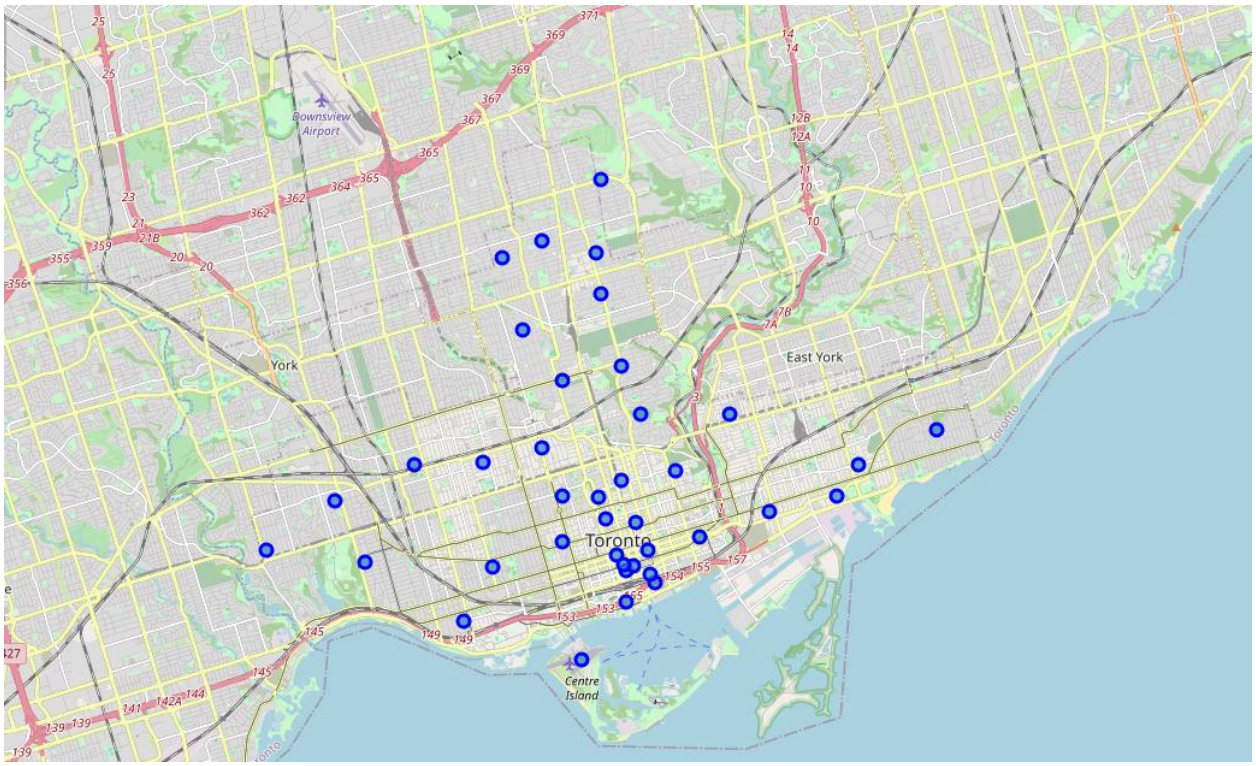

Figure 1: Toronto neighborhoods

4.2 Neighborhood Segmentation

Foursquare API was used to extract 100 nearby venues within 500 m of the center of each neighborhood in Toronto. A total of 233 venues with their name, location and category were obtained. Top 10 most common venues in each neighborhood were sorted and summarized. An example of the summarized table is shown in Figure **2** (a complete table can be found in the Jupyter Notebook).

	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
0	Berczy Park	Coffee Shop	Cocktail Bar	Café	Seafood Restaurant	Beer Bar	Restaurant	Cheese Shop	Bakery	Clothing Store	Basketball Stadium
1	Brockton, Parkdale Village, Exhibition Place	Café	Performing Arts Venue	Coffee Shop	Breakfast Spot	Yoga Studio	Bakery	Convenience Store	Pet Store	Climbing Gym	Restaurant
2	Business reply mail Processing Centre, South C...	Yoga Studio	Auto Workshop	Garden Center	Gym / Fitness Center	Fast Food Restaurant	Farmers Market	Light Rail Station	Comic Shop	Pizza Place	Recording Studio
3	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport Terminal	Sculpture Garden	Harbor / Marina	Rental Car Location	Plane	Coffee Shop	Boat or Ferry	Bar	Airport Lounge
4	Central Bay Street	Coffee Shop	Sandwich Place	Italian Restaurant	Japanese Restaurant	Café	Burger Joint	Department Store	Salad Place	Thai Restaurant	Bubble Tea Shop

Figure 2: Top 10 Common venues in each neighborhood

The venue categories were one-hot encoded in order to transform to numerical features to be used in clustering model. K-means was chosen to be the clustering technique for this problem. Initially the neighborhoods were clustered into 4 groups based on their venue types, as shown in Figure **3**. The four clusters are marked as 1-red, 2-blue, 3-purple, and 4-yellow.

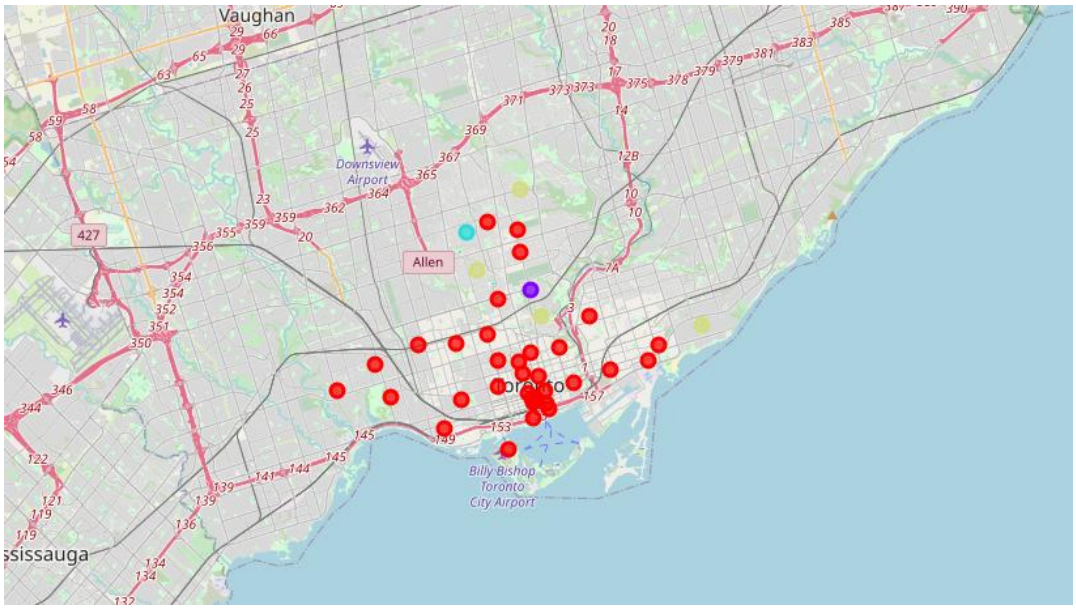

Figure 3: Clustered Toronto neighborhoods

Cluster 1 contains the most neighborhoods, a total of 33 neighborhoods including Little Portugal where my client has their first coffee shop, see details in Figure **4**. Based on their common venue types, we can picture what a typical neighborhood in Toronto would look like: there are plenty choices of coffee shops and cafes, restaurants, bars and many other food and drink venues are also easily to be found.

Cluster 2, 3, and 4 are quite different than cluster 1. They contain much less food and drink venues, but more parks, trails, yoga and dance studios. See details in Figure **5**, Figure **6**, and Figure **7**.

	Neighborhood	Cluster Labels	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
0	Regent Park, Harbourfront	0	Coffee Shop	Park	Pub	Bakery	Theater	Breakfast Spot	Café	Restaurant	Hotel	Spa
1	Queen's Park, Ontario Provincial Government	0	Coffee Shop	Sushi Restaurant	Yoga Studio	Bank	Beer Bar	Smoothie Shop	Sandwich Place	Burrito Place	Restaurant	Café
2	Garden District, Ryerson	0	Clothing Store	Coffee Shop	Cosmetics Shop	Café	Japanese Restaurant	Italian Restaurant	Bubble Tea Shop	Middle Eastern Restaurant	Bookstore	Bakery
3	St. James Town	0	Café	Coffee Shop	Cocktail Bar	American Restaurant	Gastropub	Hotel	Gym	Restaurant	Clothing Store	Italian Restaurant
5	Berczy Park	0	Coffee Shop	Cocktail Bar	Café	Seafood Restaurant	Beer Bar	Restaurant	Cheese Shop	Bakery	Clothing Store	Basketball Stadium
6	Central Bay Street	0	Coffee Shop	Sandwich Place	Italian Restaurant	Japanese Restaurant	Café	Burger Joint	Department Store	Salad Place	Thai Restaurant	Bubble Tea Shop
7	Christie	0	Grocery Store	Café	Park	Diner	Baby Store	Candy Store	Nightclub	Coffee Shop	Athletics & Sports	Restaurant
8	Richmond, Adelaide, King	0	Coffee Shop	Café	Restaurant	Deli / Bodega	Hotel	Gym	Thai Restaurant	Bookstore	Sushi Restaurant	Cosmetics Shop
9	Dufferin, Dovercourt Village	0	Bakery	Pharmacy	Park	Middle Eastern Restaurant	Café	Bar	Bank	Supermarket	Recording Studio	Brewery
10	Harbourfront East, Union Station, Toronto Islands	0	Coffee Shop	Aquarium	Café	Hotel	Brewery	Fried Chicken Joint	Scenic Lookout	Restaurant	Sporting Goods Shop	Pizza Place
11	Little Portugal, Trinity	0	Bar	Café	Coffee Shop	Asian Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Men's Store	Juice Bar	Korean Restaurant	Malay Restaurant

Figure 4: Common venues in cluster 1

	Neighborhood	Cluster Labels	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
29	Moore Park, Summerhill East	1	Gym	Trail	Yoga Studio	Deli / Bodega	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run

Figure 5: Common venues in cluster 2

	Neighborhood	Cluster Labels	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
19	Roselawn	2	Garden	Yoga Studio	Deli / Bodega	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run	Distribution Center

Figure 6: Common venues in cluster 3

	Neighborhood	Cluster Labels	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
4	The Beaches	3	Health Food Store	Asian Restaurant	Pizza Place	Pub	Trail	Distribution Center	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store
18	Lawrence Park	3	Park	Bus Line	Swim School	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run
21	Forest Hill North & West, Forest Hill Road Park	3	Park	Jewelry Store	Trail	Sushi Restaurant	Yoga Studio	Department Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant
33	Rosedale	3	Park	Playground	Trail	Dance Studio	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run	Distribution Center

Figure 7: Common venues in cluster 4

As the purpose of this project is to find a neighborhood similar to Little Portugal, the cluster 1 will be further segmented.

4.3 Further Segmentation

As discovered in previous section, cluster 1 contains Little Portugal, where my client has their current coffee shop, also contains most of the neighborhoods in Toronto. This cluster was further segmented into 4 groups to get more insights about these neighborhoods. Again, K-Means method was used for this task.

After further segmentation, 7 neighborhoods and Little Portugal were grouped into a smaller cluster, are shown in Figure **8**. That means these neighborhoods are more similar to Little Portugal. This list can be used as the list of the potential locations for my client's second coffee shop. These neighborhoods will be further examined in the next section.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M6G	Downtown Toronto	Christie	43.669542	-79.422564
1	M6H	West Toronto	Dufferin, Dovercourt Village	43.669005	-79.442259
2	M6J	West Toronto	Little Portugal, Trinity	43.647927	-79.419750
3	M4M	East Toronto	Studio District	43.659526	-79.340923
4	M6P	West Toronto	High Park, The Junction South	43.661608	-79.464763
5	M4R	Central Toronto	North Toronto West, Lawrence Park	43.715383	-79.405678
6	M5S	Downtown Toronto	University of Toronto, Harbord	43.662696	-79.400049
7	M5T	Downtown Toronto	Kensington Market, Chinatown, Grange Park	43.653206	-79.400049

Figure 8: Shortlist of the potential coffee shop locations

4.4 Analyze Coffee Shops in the Cluster

Using Foursquare API, a particular type of venues can be searched in an area. In this section, "coffee" was used as a keyword and search results returned 130 unique coffee shops in these 8 neighborhoods. The number of coffee shops in each neighborhood was calculated in Figure **9**.

	Neighborhood	count
0	Kensington Market, Chinatown, Grange Park	50
1	University of Toronto, Harbord	35
2	Little Portugal, Trinity	28
3	Christie	14
4	Dufferin, Dovercourt Village	11
5	Studio District	10
6	High Park, The Junction South	5
7	North Toronto West, Lawrence Park	3

Figure 9: Coffee shop count for each neighborhood

5.0 Discussion

By looking at the number of coffee shops in each neighborhood, we can tell that there is a lot of competition in Kensington Market, Chinatown, Grange Park areas, so these areas might not be the best choice. Meantime, North Toronto West, Lawrence Park, High Park, The Junction South have much less competition, however this might indicate a low demand. Instead, University of Toronto and Harbord might be a good choice because their coffee shop count is similar to Little Portugal. For this reason, University of Toronto and Harbord is chosen to be the most suitable location.

6.0 Conclusion

In this project, Toronto neighborhood segmentation was done based on the common venue types in each neighborhood. There are 39 neighborhoods in Toronto in which 33 neighborhoods are typical neighborhoods where you can easily find coffee shops, restaurants and bars. Then the 33 neighborhoods were further segmented and 7 neighborhoods were found to be more similar to our target neighborhood Little Portugal. Among the 7 finalist neighborhoods, the number of coffee shops was compared to get an idea of what kind of competitions my client will be facing.

Finally **University of Toronto & Harbord** area was chosen to be the most suitable location. This area is very similar to Little Portugal in terms of types of common venues in the area. This area also has a similar coffee shop count with Little Portugal, not the most competitive area, but also demonstrates a huge demand.

An extra step was to analyze the details of these coffee shops including average ratings, price range, number of likes or dislikes, etc., and then compare with my client's current coffee shop. Unfortunately I have reached my premium calls quota so the results cannot be displayed!