Predicting the Best Location to Open Up a Coffee Shop in Toronto

Seung Youn Lee

June 14, 2019

1. Introduction

These days, every city is overflowing with coffee shops, and you can see multiple coffee shops on every block. Yet, no one can deny a coffee shop is an attractive business – a possibility of high profit with comparably low investment. I mean… who doesn't love coffee?

If someone were to open up a coffee shop in Toronto, what would be an ideal location? For obvious reasons, one would want to open up a coffee shop where there is not much competition but much demand. Office workers are one of the most valuable customers to coffee shops, in my opinion. They usually order take outs and thus do not sit for long hours with a cup of coffee, which is usually the case for students. Furthermore, they tend to be regular customers. Human is a creature of habits, so unless the taste of coffee is really bad or the service is terrible, most will continue to go to the nearby coffee shop to get caffeinated when they are tired in the morning or after lunch. This will especially be the case when there are not many other options nearby.

Based on this reasoning, for those who are thinking about opening up a coffee shop in Toronto, an ideal location may be where there are many offices but not many coffee shops nearby. Let's try to come up with an ideal location for a new coffee shop in Toronto!

2. Data

Location data for coffee shops and offices can be found on Foursquare. However, Foursquare data lacks the postal code, so I scraped the list of postal codes of Canada from Wikipedia. Once I read it in as a pandas data frame, I had three columns – postal code, borough, and neighborhood. Only cells that have an assigned borough should be worked with, so by filtering out cells with 'Borough' column not assigned, I was able to ignore with any cell with unassigned borough. Since I was working on postal code level, I combined all rows with same postal codes but different neighborhoods. For example, under postal code M5A, there would be two neighborhoods – Harbourfront and Regent Park. Also, some cells had a borough but did not have an assigned neighborhood. In this case, I replaced them with the name of the borough.

Now that we have all the postal codes and neighborhood data cleaned, we can move on to coordinate data. I read in a csv file with all the geospatial coordinates and was able to match them with the corresponding postal code/neighborhood by doing an outer join. After all the data cleaning and processing, I ended up with 11 boroughs and 103 neighborhoods in total.

3. Methodology - Exploratory Data Analysis

Using the data frame created from the step above, neighborhoods in Toronto can be explored on. For example, there is a neighborhood called Victoria Village, and one can find out the latitude of the village is 43.725882299999995 and the longitude of the village is -79.31557159999998. Using the Foursquare data, I looked at the nearby venues for each

neighborhood. For example, in the neighborhood Parkwoods, there is a park called Brookbanks Park and a fast food restaurant called KFC, as well as a food and drink shop called Variety Store.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	KFC	43.754387	-79.333021	Fast Food Restaurant
2	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
3	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
4	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
5	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
6	Victoria Village	43.725882	-79.315572	Eglinton Ave E & Sloane Ave/Bermondsey Rd	43.726086	-79.313620	Intersection
7	Victoria Village	43.725882	-79.315572	Pizza Nova	43.725824	-79.312860	Pizza Place

Furthermore, I looked at the number of venues for each neighborhood. For example, neighborhood named Agincourt has 5 reported venues according to Foursquare. I also looked at the statistics on venue category and found out there are 278 unique categories of venue category in Toronto.

```
'Park', 'Fast Food Restaurant', 'Food & Drink Shop',
'Hockey Arena', 'Coffee Shop', 'Portuguese Restaurant',
'Intersection', 'Bakery', 'Gym / Fitness Center', 'Spa',
'Breakfast Spot', 'Restaurant', 'Pub', 'Historic Site',
'Chocolate Shop', 'Farmers Market', 'Dessert Shop',
'Performing Arts Venue', 'Mexican Restaurant', 'Café', 'Theater',
'French Restaurant', 'Italian Restaurant', 'Event Space',
'Yoga Studio', 'Cosmetics Shop', 'Shoe Store', 'Art Gallery',
'Brewery', 'Electronics Store', 'Bank', 'Beer Store', 'Hotel',
'Health Food Store', 'Antique Shop', 'Boutique',
'Furniture / Home Store', 'Vietnamese Restaurant',
'Accessories Store', 'Clothing Store', 'Fraternity House',
"Women's Store", 'Miscellaneous Shop', 'Gym', 'Sushi Restaurant',
'Creperie', 'Hobby Shop', 'Arts & Crafts Store', 'Burrito Place',
'Persian Restaurant', 'Diner', 'Japanese Restaurant',
'Wings Joint', 'Burger Joint', 'Seafood Restaurant', 'Nightclub',
'Smoothie Shop', 'Sandwich Place', 'College Auditorium', 'Bar',
'Vegetarian / Vegan Restaurant', 'Print Shop',
'Caribbean Restaurant', 'Baseball Field', 'Gastropub', 'Pharmacy'
'Pizza Place', 'Pet Store', 'Athletics & Sports', 'Comic Shop',
'Plaza', 'Ramen Restaurant', 'Tea Room', 'Thai Restaurant',
'Taco Place', 'Music Venue', 'Movie Theater', 'Beer Bar',
'Steakhouse', 'American Restaurant', 'Sporting Goods Shop',
'Food Court', 'Middle Eastern Restaurant', 'Tanning Salon',
'Bookstore', 'Modern European Restaurant', 'College Rec Center',
'Shopping Mall', 'Department Store', 'Chinese Restaurant',
'Lounge', 'Toy / Game Store', 'Juice Bar', 'Hookah Bar',
'Bubble Tea Shop', 'Wine Bar', 'Video Game Store', 'Smoke Shop',
```

For the purpose of my project, I am only interested in venue categories related to coffee shops and offices. Going through the entire list of venue categories, I found out that coffee shops exist under different names such as ‘Dessert Shop’ and ‘Café’, and offices exist number different names such as ‘Tech Startup’ and ‘Coworking Space’. Thus, I filtered the data frame to only

look at rows with those venue categories. Then, I replaced all the different names into 'Coffee Shop' and 'Office' for consistency. Then, I ended up with 310 rows of data – coffee shops and offices. Out of 310, 305 of them were coffee shops, while only 5 of them were offices. Thus, there may be inaccuracy or bias in Foursquare data. I also did a quick analysis to see how many coffee shops or offices there were in each neighborhood.

Central Bay Street	**Coffee Shop**	19
	Office	1
Chinatown, Grange Park, Kensington Market	**Coffee Shop**	14
Christie	**Coffee Shop**	4
Church and Wellesley	**Coffee Shop**	8
Commerce Court, Victoria Hotel	**Coffee Shop**	18
Davisville	**Coffee Shop**	7
Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West	**Coffee Shop**	2

As you can see, there are 19 coffee shops in Central Bay Street, while there is just one office in that neighborhood.

4. Results

Clustering was used to cluster neighborhoods and look at the trend of coffee shops and offices. K-means clustering was used with kclusters of 5.

Neighborhood	Neighborhood Latitude_x	Neighborhood Longitude_x	Venue	Venue Latitude	Venue Longitude	Venue Category	Cluster Labels	Neighborhood Latitude_y	Neighborhood Longitude_y
Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop	1	43.725882	-79.315572
Harbourfront, Regent Park	43.654260	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop	0	43.654260	-79.360636
Harbourfront, Regent Park	43.654260	-79.360636	Starbucks	43.651327	-79.364329	Coffee Shop	0	43.654260	-79.360636
Harbourfront, Regent Park	43.654260	-79.360636	Sumach Espresso	43.658135	-79.359515	Coffee Shop	0	43.654260	-79.360636
Harbourfront, Regent Park	43.654260	-79.360636	Rooster Coffee	43.651900	-79.365609	Coffee Shop	0	43.654260	-79.360636

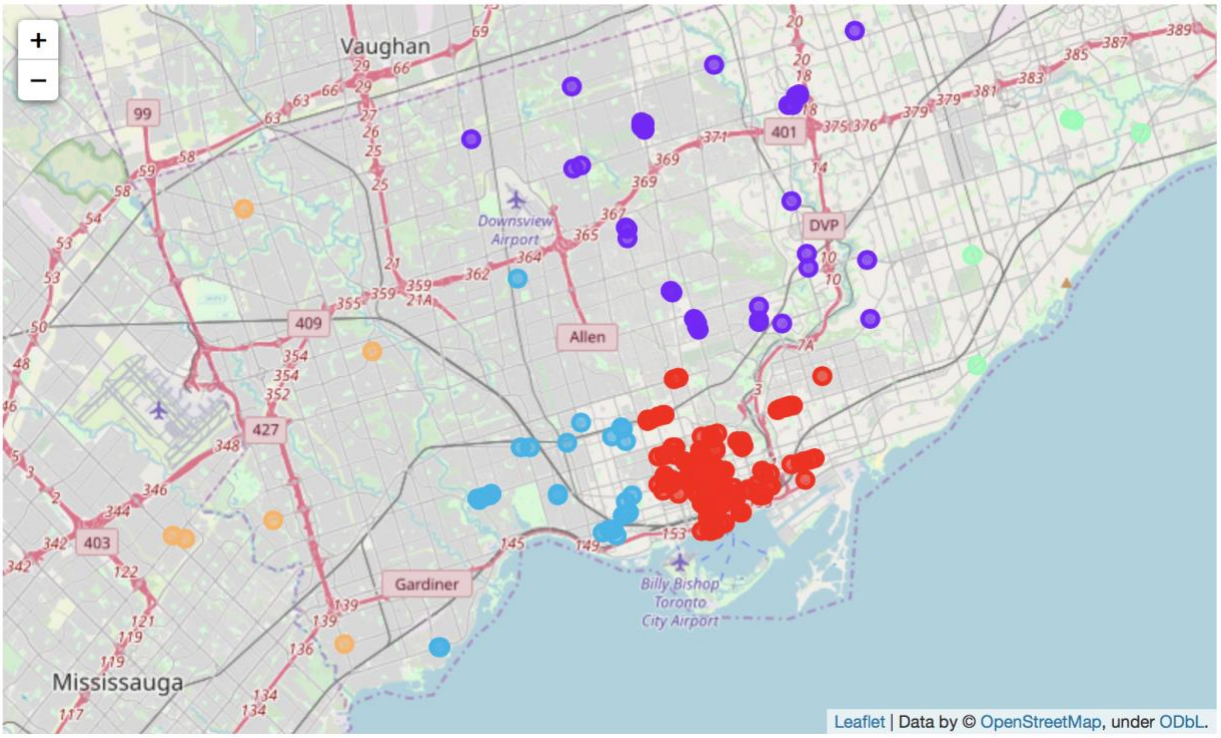
Let's take a look at how many coffee shops there are for each cluster.

Cluster Labels	
0	508
1	40
2	29
3	4
4	6

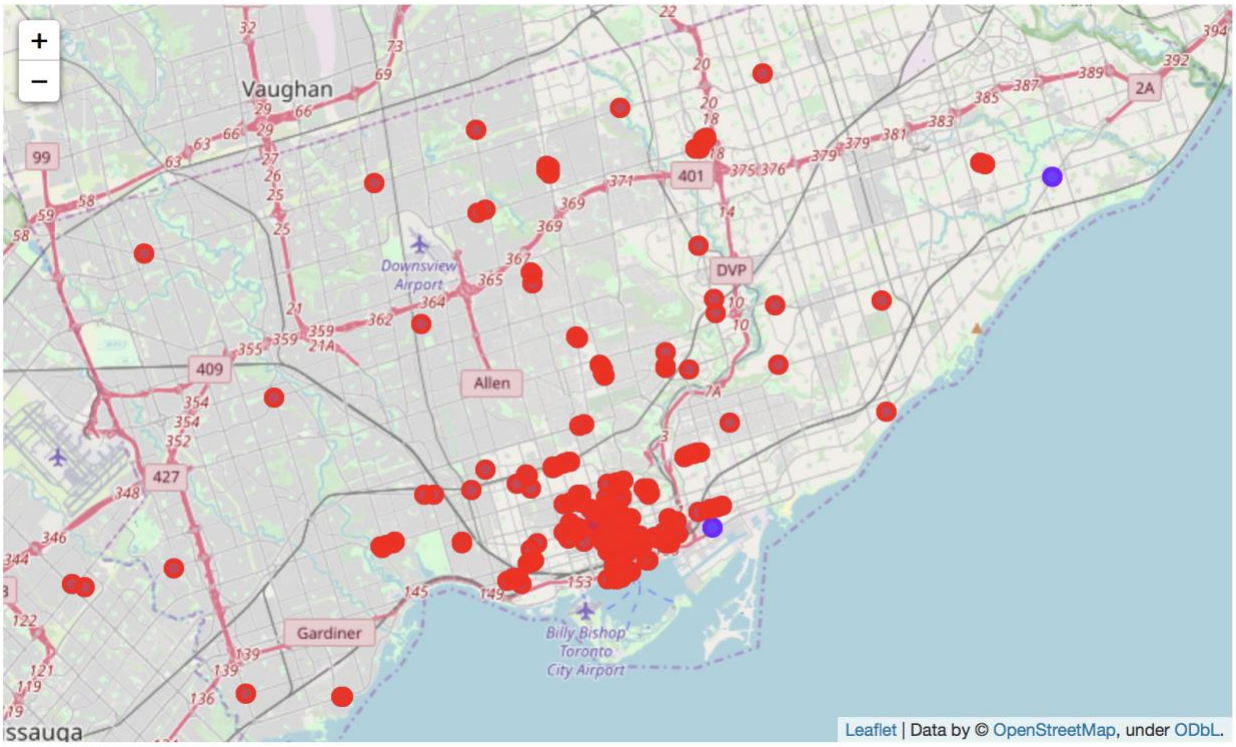
Let’s take a look at how many offices there are for each cluster.

Cluster Labels	
0	6
3	1

Now, let’s take a look at the clusters on a map and see how coffee shops and offices are clustered around.



As expected, you can see a lot in downtown. Now let’s get back the original goal of the project and take a look at the distribution of coffee shops and offices on the map – visualize it, and determine where would be ideal for opening up a new coffee shop.



The dots in red represent coffee shops, and the dots in purple represent offices. You can see there is an office in downtown, but it has way too many coffee shops nearby. In fact, there are too may coffee shops and not enough offices. If you look on the right, there is an office that is surrounded by only 2 coffee shops which are not even that close by. Clicking on the purple dot indicates it is an office called chatr Mobile. Looking it up on the data frame, I was able to find out it was in the neighborhood – Guildwood, Morningside, and West Hill.

5. Discussion/ Conclusion

I wanted to decide on the best location to open up a coffee shop when I am interested in opening it up where there are not many coffee shops nearby but there is an office nearby. Satisfying this condition was the neighborhoods – Guildwood, Morningside, and West Hill. I would recommend the coffee shop to be opened up as close as possible to the office – chatr Mobile. The latitude is 43.765917, and the longitude is -79.191672 if anyone wanted to look up the address. We have to keep in mind though that Foursquare data was not perfect or completely fair because there were way more coffee shops data than offices data. More complete data of coffee shops and offices would give a better recommended location for the new coffee shop.