The Coffee Shop Problem

Where in Toronto Should you open a coffee shop?

William Brewer

July 10th, 2021

1 – Introduction

One of the most common establishments that can be found within any neighborhood of the city of Toronto is coffee shops. Although this shows there is a great market for this type of establishment, it also shows that competition would be fierce for a new business.

The problem is as follows; you are approached by an entrepreneur interested in starting their first small coffee shop. They are asking you to leverage available data to consult them on the most reasonable location within the city of Toronto to start their business. Key factors to consider are population, and the presence of competition. Assuming cost of operation due to the location is not a concern, where should the entrepreneur open their coffee shop?

The intended audience for this report is entrepreneurs' hoping to make the same decision as the hypothetical one mentioned above, using the analysis provided to help their decision. Secondary to this, other data scientists and the curious may also find this report useful as an insight into the business demographics of neighborhoods within their city.

2 – Data

This analysis will be performed with the list of postal codes of Toronto provided by Wikipedia here: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

The 2016 Canadian census data on forward sortation areas provided here (population by postal code): https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/comprehensive.cfm

The foursquare API will also be used to makes sense of the postal codes provided by Wikipedia by accessing nearby venues within each neighborhood in order to identify competition in the area.

The above data and API will be used to gather to create a map visualizing viable neighborhoods in Toronto and their potential value as a location for a new coffee shop.

3 – Methodology

In this report we will focus on finding prime neighborhood locations within Toronto featuring both high populations, and low frequency of coffee shops, maximizing demand and minimizing current supply. First, we will collect necessary data through Wikipedia and Stat Canada to create a master dataframe containing a comprehensive list of all neighborhoods in Toronto, along with their respective coordinates for later use, population, and name/borough. Second we will do some visualization work with our preliminary data using folium in order to visualize the population of each neighborhood on a single map. Third and finally we will start working with the foursquare API to augment our master dataframe to include the top 3 most common types of categories in each neighborhood and whether a coffee shop dominance exists.

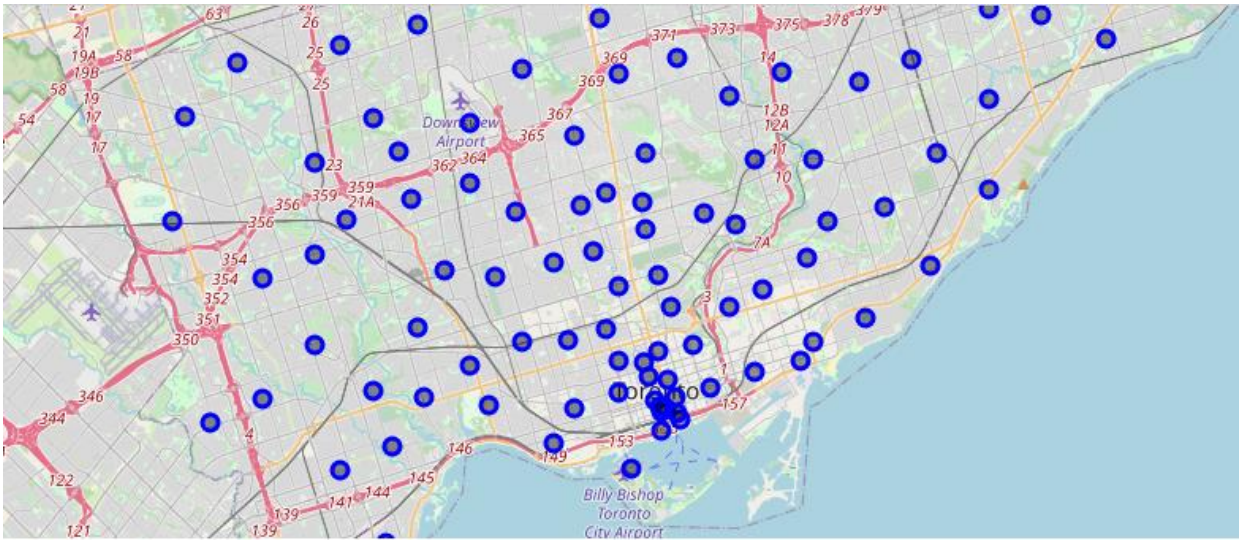
3.1 – Master Dataframe

After some collection and cleaning we find end up with our master data frame featuring all neighborhoods in Toronto, their population, and coordinates.

	Postal Code	Population	Borough	Neighborhood	Latitude	Longitude
0	M1B	66108.0	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	35626.0	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	46943.0	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	29690.0	Scarborough	Woburn	43.770992	-79.216917
4	M1H	24383.0	Scarborough	Cedarbrae	43.773136	-79.239476

3.2 – A simple mapping of the current dataframe

Just to test the folium package and demonstrate a visualization of our current dataframe, map all coordinates stored on a folium map.



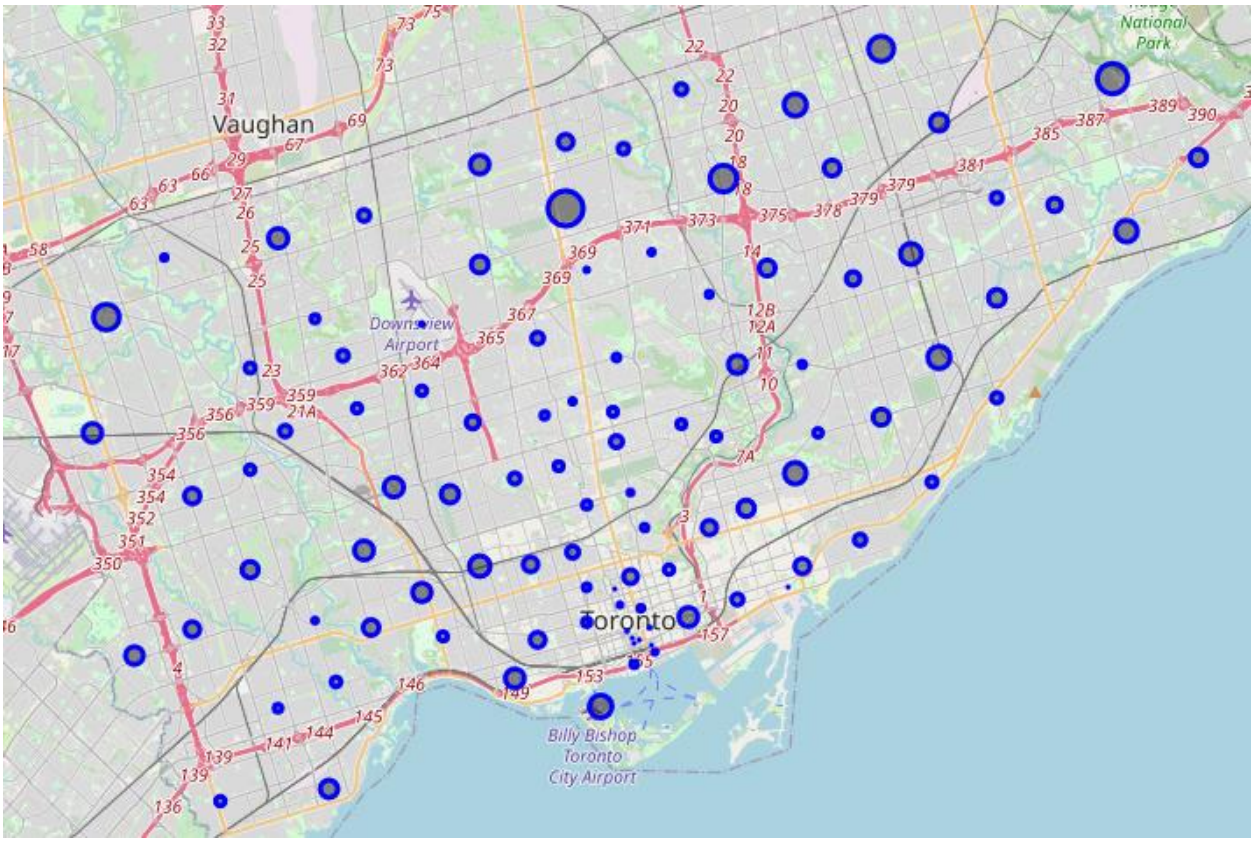
3.3 – Mapping based on population

We take it one step further by computing the radius of each marker to be proportional to their population, this will help visualize the more ideal neighborhoods to potentially open a coffee shop based on customer volume.

Now map markers with a larger normalized radius represent areas of higher population and a potentially more lucrative consumer base.

We calculate radius with:

$$radius = 10(\frac{population - population_{min}}{population_{max} - population_{min}})$$



3.4 – All venues within each Toronto Neighborhood

Using a new dataframe with foursquare API, we associate all found venues with their respective postal code, category, and coordinates.

	Postal Code	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	M1B	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	M1C	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	M1E	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank
3	M1E	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
4	M1E	43.763573	-79.188711	Sail Sushi	43.765951	-79.191275	Restaurant

Next, we group venues and find the top 3 most common categories per each neighborhood.

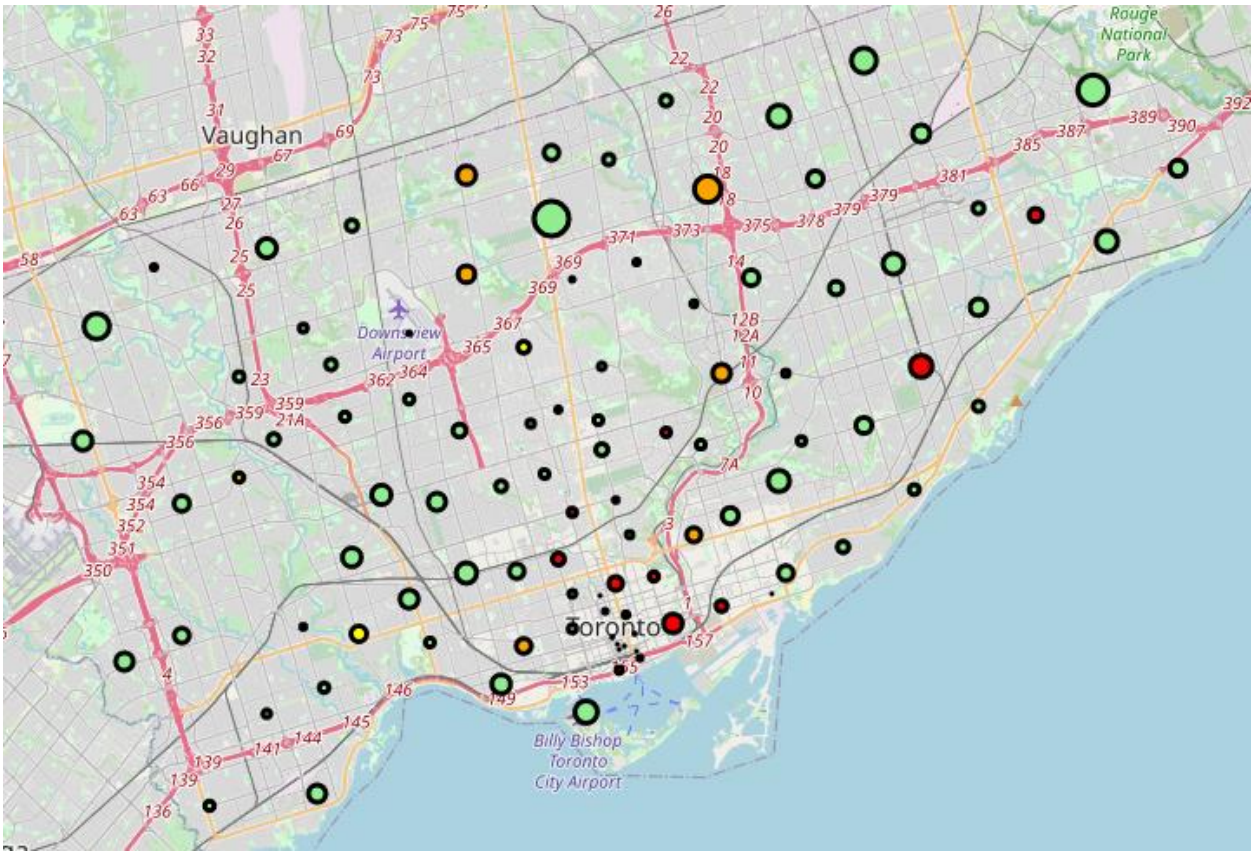
	Postal Code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	M1B	Fast Food Restaurant	Accessories Store	Museum
1	M1C	Bar	Accessories Store	Modern European Restaurant
2	M1E	Medical Center	Restaurant	Breakfast Spot
3	M1G	Coffee Shop	Pharmacy	Korean BBQ Restaurant
4	M1H	Bakery	Fried Chicken Joint	Gas Station

Finally, we merge this new dataframe to the previous master frame. Below is the final version of the dataframe necessary for our analysis. With data on population and the 3 most common venue types, we can make an educated decision on the best neighborhood in Toronto to open a new coffee shop.

	Postal Code	Population	Borough	Neighborhood	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	M1B	66108.0	Scarborough	Malvern, Rouge	43.806686	-79.194353	Fast Food Restaurant	Accessories Store	Museum
1	M1C	35626.0	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Bar	Accessories Store	Modern European Restaurant
2	M1E	46943.0	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	Medical Center	Restaurant	Breakfast Spot
3	M1G	29690.0	Scarborough	Woburn	43.770992	-79.216917	Coffee Shop	Pharmacy	Korean BBQ Restaurant
4	M1H	24383.0	Scarborough	Cedarbrae	43.773136	-79.239476	Bakery	Fried Chicken Joint	Gas Station

3.5 – One final visualization

Create one final map to visualize our fully prepped data. Now we will color markers based on their most common venue type, on top of computing their radius based on population. Markers in green indicate coffee shops are not in the top 3 most common venue types in that neighborhood, yellow means they are the 3rd most common, orange means 2nd, and red means coffee shops are the most common venue type there.

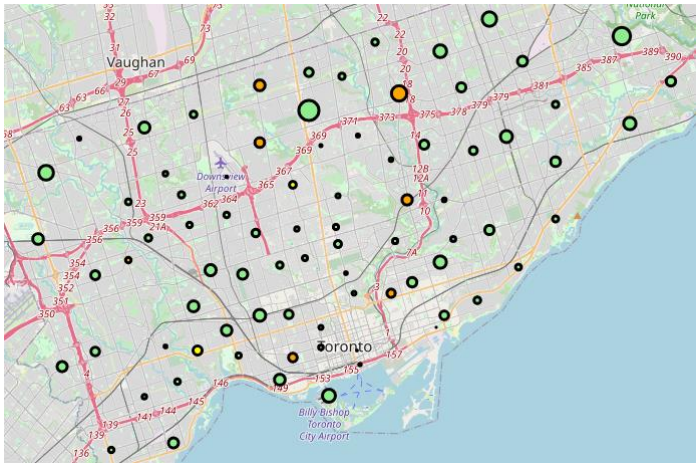


4 – Analysis

First, we will perform some exploration using our final master frame, along with some visualizations.

4.1 – Analysis A

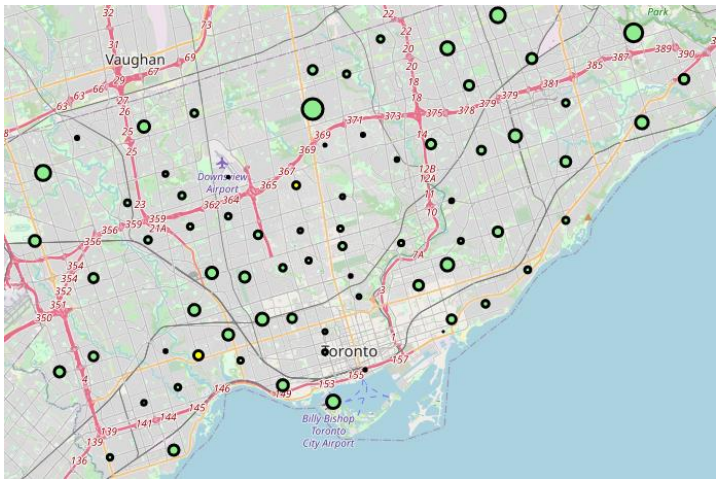
The first obvious step is to drop any neighborhood already dominated by coffee shops from our options. Remove all neighborhoods with coffee shops as their number one venue category.



As you can see, we've lost a number of neighborhoods already, lowering our options from 100 potential neighborhoods down to 81, as demonstrated above. However, it's worth noting that we still have a great number of neighborhoods with populations on the larger side left to choose from.

4.2 – Analysis b

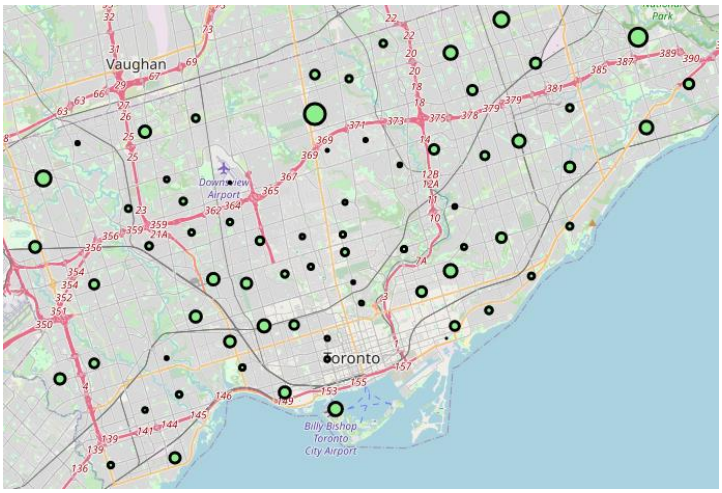
Next, we will drop neighborhoods with coffee shops as their second most common venue.



More neighborhoods cut out again, lowering our options from 81 potential neighborhoods down to 73.

4.3 – Analysis C

Finally, remove neighborhoods with coffee shops as their third most common venue.



More neighborhoods cut out again, lowering our options from 73 potential neighborhoods down to 70.

4.4 – Decision Basis

Let's reorder our newly trimmed dataframe by population to find our best neighborhood candidates.

	Postal Code	Population	Borough	Neighborhood	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	M2N	75897.0	North York	Willowdale South	43.770120	-79.408493	Ramen Restaurant	Pizza Place	Café
1	M1B	66108.0	Scarborough	Malvern, Rouge	43.806686	-79.194353	Fast Food Restaurant	Accessories Store	Museum
2	M9V	55959.0	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...	43.739416	-79.588437	Grocery Store	Pharmacy	Video Store
3	M1V	54680.0	Scarborough	Milliken, Agincourt North, Steeles East, L'Amo...	43.815252	-79.284577	Intersection	Playground	Park
4	M5V	49195.0	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.628947	-79.394420	Airport Service	Airport Lounge	Airport Terminal

So above are the best choices available to us. The top 5 ideal neighborhoods in toronto to open a new coffee shop in order of interest are:

1. South Willowdale
2. Malvern, Rouge
3. South Steeles / Silverstone / Humbergate / Jamestown / Mount Olive /
    a. Beaumond Heights / Thistletown / Albion Gardens
4. Milliken / Agincourt North / Steeles East / L'Amoreaux East
5. CN Tower / King and Spadina / Railway Lands / Harbourfront West /
    a. Bathurst Quay / South Niagara / YTZ

All featuring the highest potential customer base (population) and lowest competition (coffee shops do not occupy and of the top 3 venue categories).

5 - Results & Discussion

Thanks to our analysis, it is clear now the absolute best location to open a new coffee shop based on population and competition is Willowdale South. With a considerable population and very little competition, this neighborhood offers a great development opportunity for a new coffee shop.

It is worth noting, however, that cafes do however occupy the space as the 3rd most common venue type in **South Willowdale**. Although a different type of business, cafes are dangerously similar, so other strong alternative locations include Malvern Rouge, and the areas M9V, M1V, and M5V.

Although a good start we could greatly improve this analysis in a number of ways. First, as you noticed throughout the methodology section, we had to drop several neighborhoods due to uneven data. Ideally, we would avoid this either by finding stronger data to start with, or manually injecting the information we lost over the data treatment process. Furthermore, the accuracy of our decision could also have been improved greatly by considering more factors in our analysis. For example, rather than only concerning ourselves with population and competition, we could also factor in cost of operation, traffic areas, etc.

6 – Conclusion

The purpose of this analysis was to identify neighborhoods in Toronto that are prime locations to open a new coffee shop considering two factors: potential customer base, and local competition levels. Using census data, Wikipedia information, and the foursquare API, we concluded that **the best choice is Willowdale South**.

Next steps include improving our data collection and handling to avoid cutting any neighborhoods out of our analysis and increasing the number of factors we consider for our conclusion.