Analyzing the restaurant scenario in Downtown Toronto

Mauricio Tomé

April 5, 2021

1 Introduction and problem definition

We conducted a research about the businesses present in Toronto Central area, to look for a location to build a new restaurant in that region.

Restaurants benefit from being near central areas, where most people work and want to have lunch nearby, but, at the same time, direct competition (for example, restaurants that offer similar food) are not desirable. Restaurants that offer different food have less impact in the business of each other.

The requirements for the restaurant are:

- The restaurant has to be located between the central area and the lakeside, at a maximum distance of 1.5km from the city center;
- The type of food served in the restaurant has to be one of the popular cuisines in the region, but different from the offerings from restaurants in the vicinity.

So our task is to define the region where the restaurant might be located, while looking for neighborhoods that don't have too many of them, while also checking the most common restaurants in the region that aren't present in that neighborhood.

2 Data acquisition

We need a reference to locate the possible places of interest. To achieve that goal, we obtained the data from the Postal codes of Toronto, to look for the neighborhoods' locations.

This data was obtained from Wikipedia[1], and processed using Python's `BeaurifulSoup` library. Once we got the neighborhood's names, we used the `geocoder` library to fetch the latitude and longitude of the neighborhoods.

Having the geolocation data, we defined the region of interest and selected the neighborhoods matching the desired characteristics.

With the neighborhood list in hand, we proceeded to download information from the venues in the region using Foursquare API[2].

After collecting the venue data, we selected *only* the restaurants within that list, as well as their location and category. We then categorized the neighborhoods by amount of restaurants nearby using *k-means*. After that, we selected neighborhoods with low presence of restaurants and among them, which kinds of restaurants aren't present in the neighborhood, but are common elsewhere.

3 Exploring the data

3.1 Geolocation data from the neighborhoods

First, we overlay the geolocation data of the neighborhoods on top of Toronto city map, using the data scraped from Wikipedia together with Python's `folium` graphic library. We also marked the city center for reference, as we can see in **Figure 1**.

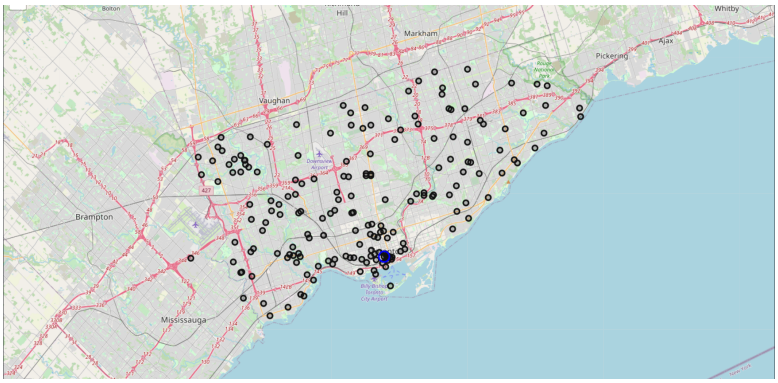

Figure 1: Neighborhoods of Toronto based on the Postal code list.

[1] `https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M`
[2] `https://developer.foursquare.com/`

We then define the region of interest, which is the area between the city center and the lakeside, within 1.5km of distance. To achieve that, we transformed the latitude and longitude coordinates in horizontal and vertical distances from the city center (marked in blue). The points inside the region of interest are marked in green in **Figure 2**.

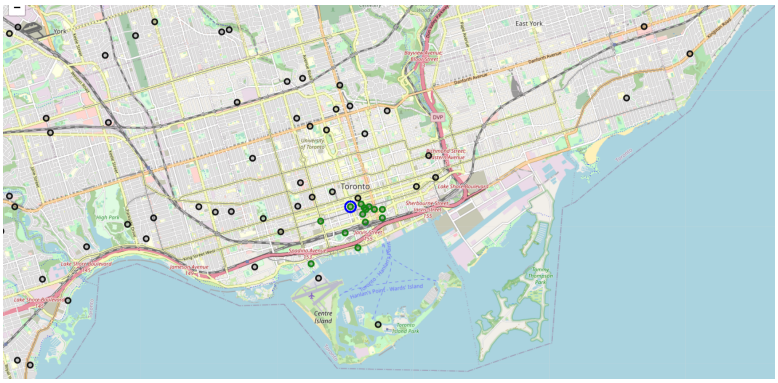

Figure 2: Detail of the region of interest.

3.2 Venue information from Foursquare API

After defining the region, we used the names and coordinates from the neighborhoods with the Foursquare API. The data collected has the neighborhood of reference, the venue names, location and category. A sample of the data can be seen in **Table 1**.

	Neighborhood			Venue			
	Name	Lat.	Long.	Name	Lat.	Long.	Category
0	Harbourfront	43.6392	-79.3830	Harbourfront Centre	43.6385	-79.3831	Performing Arts Venue
1	Harbourfront	43.6392	-79.3830	Harbourfront	43.6395	-79.3806	Neighborhood
2	Harbourfront	43.6392	-79.3830	Natrel Pond/Rink	43.6384	-79.3825	Skating Rink
3	Harbourfront	43.6392	-79.3830	Lick It Gelato	43.6392	-79.3846	Ice Cream Shop
4	Harbourfront	43.6392	-79.3830	Lake Ontario	43.6389	-79.3796	Lake

Table 1: Sample from the Foursquare data.

We can see several different categories for the venues, but we're interested only in restaurants, so we will filter the data. A sample of the filtered data can be seen in **Table 2**.

There are several categories from restaurant, and some similar ones, such as *Sushi* and *Japanese*, and so on. We grouped this similar categories under the same name to get more accurate results. Unfortunately, several restaurants are marked simply as *Restaurant*, so we are not able to discriminate what kind of food they offer.

	Neighborhood			Venue			
	Name	Lat.	Long.	Name	Lat.	Long.	Category
0	Underground city	43.6493	-79.3821	Cactus Club Cafe	43.6495	-79.3816	American
1	Garden District	43.6486	-79.3854	Bosk at Shangri-La	43.6490	-79.3858	Asian
2	Toronto Dominion C	43.6469	-79.3814	Canoe	43.6474	-79.3813	Uncategorized
3	Underground city	43.6493	-79.3821	Ruby Thai	43.6490	-79.3816	Thai
4	First Canadian Place	43.6484	-79.3809	Kupfert & Kim	43.6485	-79.3816	Gluten-free

Table 2: Sample from the Foursquare data, restaurants only.

The resulting dataset is composed of 327 lines, 29 unique food categories and 14 neighborhoods. The actual number of unique venues in the dataset is 118, and some results are duplicated because we selected neighborhoods that are within a small distance from each other, so several of them are within 500 meters of some restaurants.

3.3 Clustering of the neighborhoods using k-means

We now proceed to classify the neighborhoods by number of restaurants. We used k-means from the `scikit-learn` library to cluster the neighborhoods in 4 categories with regards to the amount of restaurants, namely *Low,*

Neighborhood	Restaurants nearby	Classification
Bathurst Quay	4	Low
Berczy Park	28	High
CN Tower	7	Low
Commerce Court	35	Very High
Design Exchange	33	Very High
First Canadian Place	29	High
Garden District	29	High
Harbourfront	15	Medium
King and Spadina	23	High
The Esplanade Enclave of M5E	24	High
Toronto Dominion Centre	27	High
Underground city	32	Very High
Union Station	18	Medium
Victoria Hotel	33	Very High

Table 3: Restaurants around neighborhoods

Medium, High and *Very High*. Results are shown in **Table 3**. We can see that many of the neighborhoods have several options of restaurants within a 500 meter range.

Figure 3 shows the location and classification of each one of the neighborhoods. We are interested in neighborhoods that have low number of restaurants nearby, but are close to other neighborhoods so that opening an restaurant there could attract people from other places. There are only two neighborhoods with low amount of restaurants: Bathurst Quay and CN Tower. We will choose CN Tower as a prospective place for the new restaurant, since there are more neighborthoods with several restaurants around.

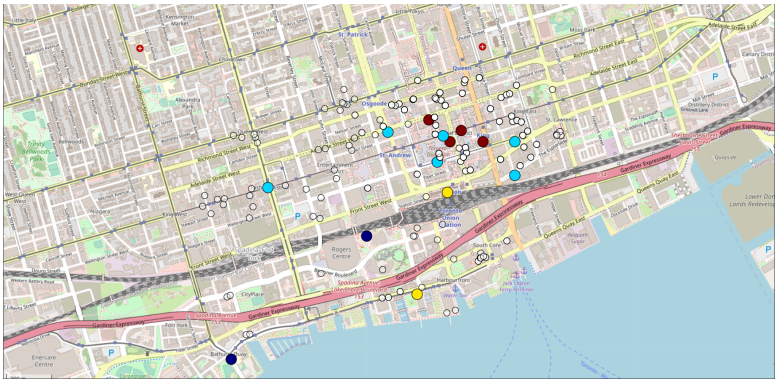

Figure 3: Classification of the neighborhoods by the amount restaurants nearby. Dark blue: Low, yellow: Medium, light blue: High, dark blue: Very High. Small white circles indicate the location of the restaurants.

Figure 4 shows the detaill of CN Tower neighborhood. It is possible to see that there are 7 restaurants in total less than 500 meters away from the neighborhood. The restaurants are Japanese(2), Greek(1), Thai(1), Indian(1), Italian(1) and Middle Eastern(1).

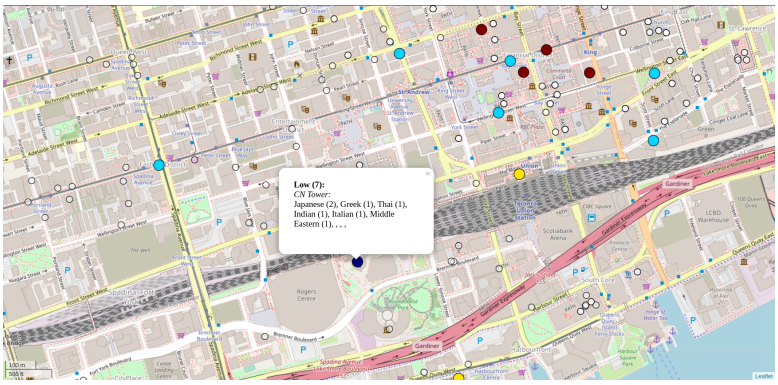

Figure 4: Detail of the CN tower neighborhood and category of the restaurants nearby.

3.4 Most popular restaurants in the region

Table 4 shows the top 10 unique counts of restaurants in the sleected neighborhoods. We can see that Japanese Restaurants are pretty popular in the city center, and even CN Tower already have two of them, so probably it's a bad choice to build another one in there. Same goes for Italian and Thai restaurants, both of which are also present.

Neighborhoods	Number of restaurants
Japanese Restaurant	19
Uncategorized Restaurant	16
Italian Restaurant	14
Seafood Restaurant	8
Thai Restaurant	6
Vegetarian / Vegan Restaurant	5
Fast Food Restaurant	5
Indian Restaurant	5
Chinese Restaurant	4
Asian Restaurant	4

Table 4: Top 10 most common restaurants in the region.

The first popular option to be missing from the neighborhood (and hence, a possible option for us) would be *seafood*, followed by *vegetarian, fast food, Chinese* and *Asian*.

4 Conclusion

In this study, we used data available from the city of Toronto, as well as the venues available on the Foursquare service, to determine a suitable neighborhood to open a new restaurant in downtown, as well as what kinds of restaurants we should consider to avoid having direct competitors within the same area (that is, offering the same kind of food). We identified the amount of restaurants in the region, as well as the most popular ones within each area. In this way, we believe to have provided a data-based approach which could help a potential businessperson to decide where to place his restaurant and which cuisine to offer.