

Applied Data Science Capstone Project

The Battle of the Neighborhoods in Toronto, Canada

By : Gökberk Yücel KAYA

Date : May 22 , 2021

Table of Contents

1. Introduction: Business Problem	3
2. Data	3
2.1. Data Collection	4
2.1.1. Neighborhood Data	4
2.1.2. Geospatial Coordinates Data	4
2.1.3. Explore the neighborhoods in Toronto	5
2.1.4. Foursquare Data	6
3. Methodology	7
4. Analysis	7
5. Results and Discussion	10
6. Conclusion	11

1. Introduction: Business Problem

There are many factors to consider when starting a food business. The main factors can be listed as financing, location, theme and menu. To begin with, it can be easily said that location is one of the most important of these. In this project we will try to help find the best location for a food business.

The factors to be considered to find the best location can be listed as follows: demography, labor costs and minimum wage, access, proximity to suppliers, competition, visibility, future growth, health regulations and zoning, security / crime rates.

This report will focus specifically on the competitive factor. Being too close to established competition can be a double-edged sword. To put it simply, being close to your competitors can help commercial marketing, but if you are too close to the competition, you may have a hard time gaining a foothold in the community.

Finally, this report will specifically target stakeholders who want to open a food venue in Toronto, Canada. Since there are many food venues in Toronto, we will do our work based on the neighborhoods to make it more understandable. Our aim here is to reveal the most or the least amount of food venues in the neighborhoods. And also to provide alternative options to stakeholders by grouping similar neighborhoods.

2. Data

Based on definition of our problem, factors that will influence our decision are:

- number of existing food venues in the neighborhood (any type of food venue)
- The most common food venues in the neighborhood

Following data sources will be needed to extract/generate the required information:

- Information such as the postal code, borough, and neighborhood of the city of Toronto will be collected by scraping the [wikipedia](#) page.
- Toronto Neighborhood coordinates will be collected using the prepared geospatial data csv file.
- The latitude and longitude values of the city of Toronto will be collected using the geopy library.

- The number, type, location, postal code and distance of the food venues in each neighborhood will be obtained using the Foursquare API.

2.1. Data Collection

2.1.1. Neighborhood Data

This [wikipedia](#) page contains an html table containing data of the Toronto, Canada postal codes. Using this table, a data frame consisting of the Postal Code, Borough, and Neighborhood data of the city of Toronto is obtained.

	Postal Code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Ontario Provincial Government

Figure 1 - Neighborhood Data

2.1.2. Geospatial Coordinates Data

Using the provided [Geospatial Coordinates](#) csv file we get the latitude and longitude values of each neighborhood.

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Figure 2 - Geospatial Data

2.1.3. Explore the neighborhoods in Toronto

The combination of the obtained neighborhood data and geospatial coordinate data is as follows.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Ontario Provincial Government	43.662301	-79.389494

Figure 3 - Merged Data

The visualized version of the data set created on the map for better understanding is as shown below.

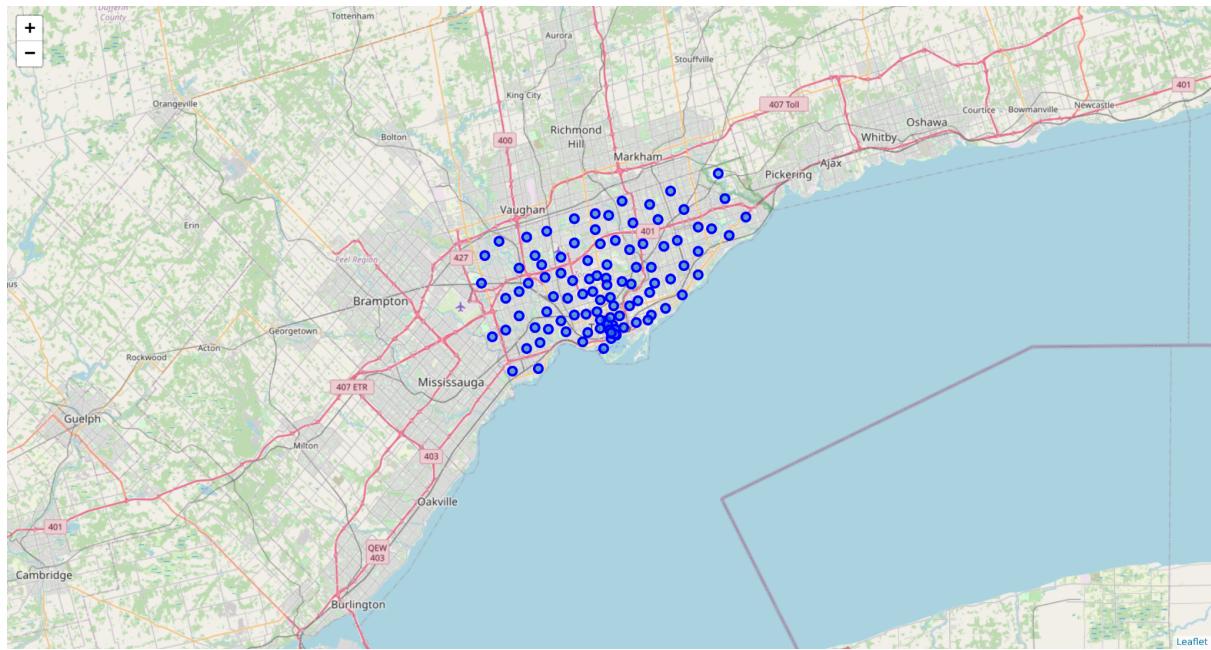


Figure 4 - Map of Neighborhoods

2.1.4. Foursquare Data

This project deals with 'food' category venues such as coffee shops, pizzerias, patisseries etc. Therefore, only food category venues are included in the list. Food venues are collected according to the neighborhoods. The radius is taken as 2000 so as not to miss any food venue. Of course, this will have some side effects. For neighborhoods that are close to each other, some spaces will overlap. This problem will be solved in the data preprocessing part. Finally, to put it briefly, the name of the food venue, latitude and longitude, postal code and distance are extracted from the json file obtained using the foursquare api. This data is then collected in a dataframe.

	Neighborhood Postal Code	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Postal Code	Venue Distance	Venue Category
0	M3A	North York	Parkwoods	43.753259	-79.329656	Allwyn's Bakery	43.759840	-79.324719	M3A 1Z5	833	Caribbean Restaurant
1	M3A	North York	Parkwoods	43.753259	-79.329656	Tim Hortons	43.760668	-79.326368	M3A 1Z5	866	Café
2	M3A	North York	Parkwoods	43.753259	-79.329656	Darband Restaurant	43.755194	-79.348498	M3B 1Y5	1530	Middle Eastern Restaurant
3	M3A	North York	Parkwoods	43.753259	-79.329656	Island Foods	43.745866	-79.346035		1553	Caribbean Restaurant
4	M3A	North York	Parkwoods	43.753259	-79.329656	Me Va Me Kitchen Express	43.754957	-79.351894	M3B 1Y4	1798	Mediterranean Restaurant

Figure 5 - Venues Data

3. Methodology

In this project, studies are being directed to determine the density of food venues in the neighborhoods of Toronto and the type of food venues that are most common in the neighborhoods.

In the first step, the necessary data were collected: name, location, category, postal code and distance of every food venue in Toronto city.

The second step of the analysis will be to prepare the data set created using the Foursquare api for analysis. This will be done by arranging overlapping or unclear spaces to which neighborhood they belong to. Then, necessary measures will be taken to find the 'food venue density' and 'frequency of similar food venues'.

In the third and final step, a heat map will be presented to show the density of neighborhoods in terms of food venues and focus on what implications these can have for stakeholders. In addition, basic information will be provided to stakeholders, taking into account the most common types of food venues in these neighborhoods. This will allow stakeholders to choose a neighborhood-level starting point. Finally, clustering (using k-mean clustering) will be established to offer stakeholders different alternatives in terms of neighborhood.

4. Analysis

The first thing to notice after the Venue dataframe was created was that the same venues were more than once. The reason for this was that when using the foursquare api, the radius value was set as 2000. Because in neighborhoods close to each other, venues were overlapping. The postal code was used to fix this. Venues that did not match the Neighborhood postal code and overlapped were removed. However, some venues did not have a postal code. To solve this, distance

value was used. Venues that do not have postal code are assigned to the nearest neighborhood.

	Neighborhood Postal Code	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Postal Code	Venue Distance	Venue Category
0	M3A	North York	Parkwoods	43.753259	-79.329656	Allwyn's Bakery	43.759840	-79.324719	M3A	833	Caribbean Restaurant
1	M3A	North York	Parkwoods	43.753259	-79.329656	Tim Hortons	43.760668	-79.326368	M3A	866	Café
2	M3A	North York	Parkwoods	43.753259	-79.329656	A&W	43.760643	-79.326865	M3A	852	Fast Food Restaurant
3	M3A	North York	Parkwoods	43.753259	-79.329656	High Street Fish & Chips	43.745260	-79.324949	M3A	967	Fish & Chips Shop
4	M3A	North York	Parkwoods	43.753259	-79.329656	Pizza Pizza	43.760231	-79.325666	M3A	839	Pizza Place

Figure 6 - Processed Venues Data

Afterwards, a map showing the heat map / intensity of food venues was created to extract meaningful information v. As you can see, the map is very hot, especially the neighborhoods of Downtown Toronto, Central Toronto, West Toronto and East Toronto are on fire. As can be understood from here, these areas are very valuable places for food venues.

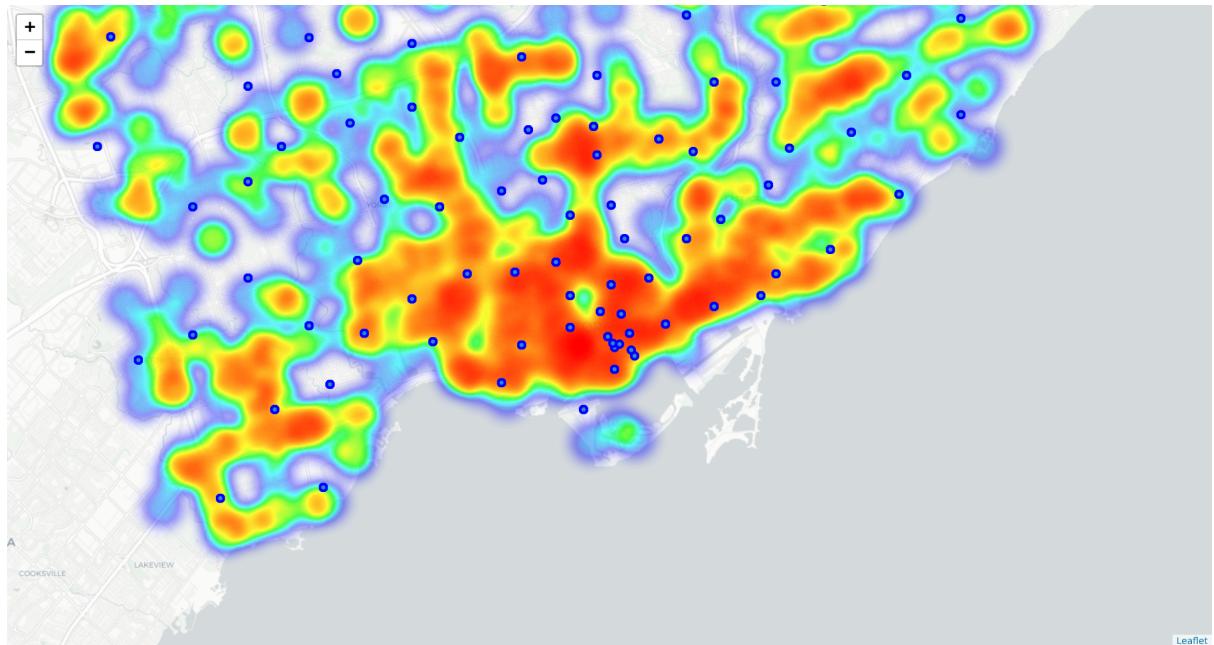


Figure 7 - Heat Map of Venues

The frequencies of venues in the neighborhoods were calculated and ranked using the Venue dataframe. As an example, you can examine the chart below.

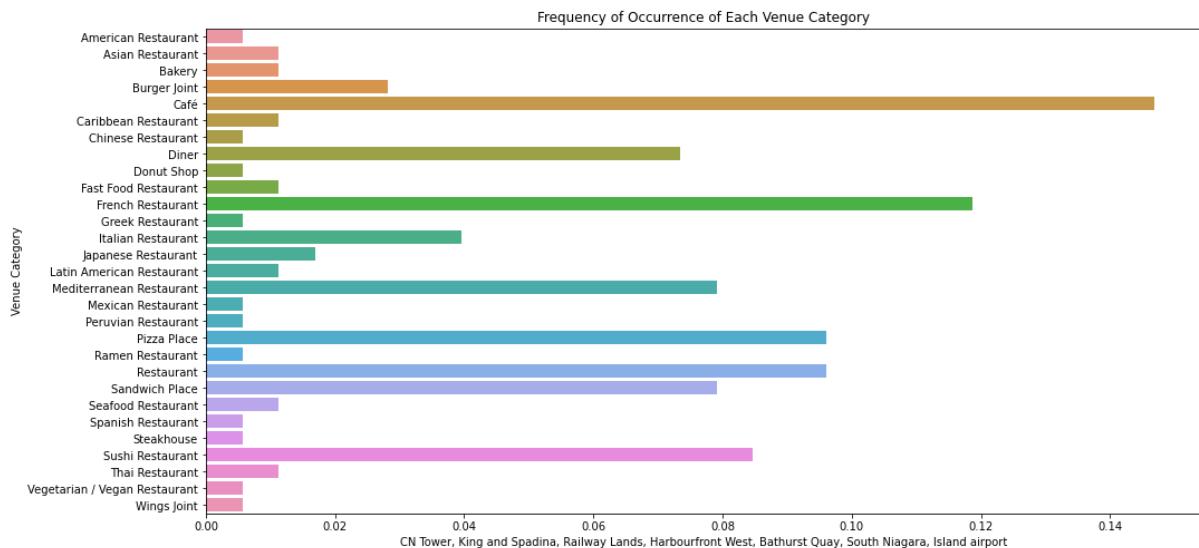


Figure 8 - Frequency of venues

The CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport neighborhoods are the first area with the most venues. As seen in the graph, the most common types of food venues in this region are Cafes, French Restaurants, gastropubs, Pizza places and Restaurants. Although this region is very valuable for food venues, it may not be a good idea for a new business to open one of these types of venues, which are already in large numbers here.

Finally, neighborhoods were clustered using kmeans clustering to see similar neighborhoods. This will enable different neighborhoods to be offered to stakeholders as an alternative. The results achieved are shown in the image below.

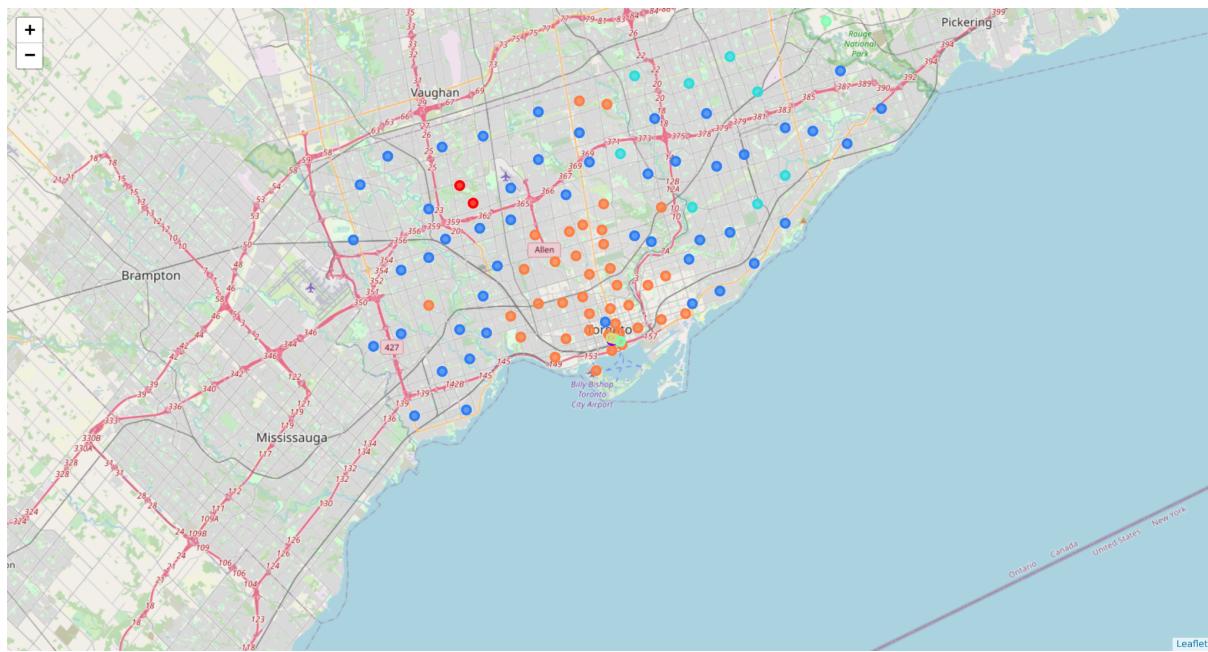


Figure 9 - Clustering of Neighborhoods

5. Results and Discussion

The analysis shows that the busiest area in the city of Toronto in terms of food venues is the neighborhoods connected to the Downtown Toronto borough. The main ones are CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport neighborhoods. Only from this analysis it is understood that this area is a valuable area for food venues. However, there are many things to consider when choosing a place for the food venue. Although this area is a valuable place, it may not be a good choice for some food venues.

Of course, there are many factors to consider when choosing a food venue location, but here we focus only on the 'competition' factor. In order to evaluate this factor, we presented the neighborhood-based venues we collected and how often they were in which neighborhood as a result of our analysis. For example, although the area consisting of CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport neighborhoods seems like an attractive location, it may not be a good idea for venues like Cafe or French

Restaurant here. If we are looking for a venue in this style, we can also consider the following areas.

Finally, the fact that this study is not inclusive may reveal some drawbacks. One of them is that the information we provide is far from offering pinpoint options for stakeholders. The reason for this is that our analyzes are far from the real field. Therefore, it should not be forgotten that the information we provide will only help stakeholders to choose a starting point in finding a new location for their business.

6. Conclusion

The aim of this project was to assist stakeholders trying to find venues for a food venue in the city of Toronto. There are many factors that affect the location of a food venue. In this study, analyzes were made considering only the competition factor. First of all, all venues in the food category were collected on the basis of Neighborhood using the Foursquare api. As a result of the analysis made afterwards, the density of food venues in the city was revealed. Thus, the competition factor at the general level was examined. Then, going a little more specific, the frequency of food venues in the neighborhoods was revealed as a result of analysis. Thus, stakeholders will be able to easily select a starting point by taking this information into consideration during the location selection phase.

Finally, it should be noted that the stakeholders should pay attention to all factors when choosing the location for the food venue. We can list these factors as follows: demography, labor costs and minimum wage, access, proximity to suppliers, competition, visibility, future growth, health regulations and zoning, security / crime rates. When the stakeholders consider all these factors, they will be able to choose the best location for their food venues.