

The Battle of Neighborhoods – Capstone Project Final Report: Opening a Thai Restaurant in Toronto

by

Abu Sadat Nurullah

June 03, 2021

1. Introduction

1.1 Objective and Background:

The objective of this project is to explore the best neighborhood in Toronto to open a Thai restaurant. This is based on Foursquare location data as well as available geospatial and Wikipedia data on various neighborhoods in Toronto. Although a lot of restaurants have permanently closed across Canada due to COVID-19 pandemic since last year [1], there could be opportunity to open new restaurants as soon as the pandemic subsides. Thai cuisine is the 3rd most popular ethnic cuisines in Canada behind Italian and Chinese ones, as indicated by Google Trends data [2]. Similarly, the choice of location as Toronto is partly based on the fact that the majority of Thai populations reside in the province of Ontario, and Toronto being the capital city of that province.

Toronto is one of the most densely populated and multicultural cities in North America. This creates a great opportunity for entrepreneurs to start or grow their business ventures, such as opening a new Thai restaurant. However, it will be competitive to open a restaurant in Toronto because there are already many Thai restaurants in the city. The main issues to consider are the frequency of other existing restaurants as well as safety of neighborhoods (e.g., based on crime rates) to determine the most suitable spot.

1.2 Business Problem:

Given the increasing popularity of Thai cuisine in Canada, entrepreneurs and stakeholders may be interested in opening a Thai restaurant in Toronto that would generate profit. The aim of this project is to find a “good” location for opening a Thai restaurant in Toronto. The primary task is to figure out and narrow down an area

where Thai restaurant is not amongst the most common venues so as to weed out any potential competition. To accomplish this task, I'll analyze restaurant data in Toronto. Therefore, I'll attempt to solve the following business problems:

- Finding out areas where there are limited or lack of Thai restaurants
- Figuring out the most common venues/restaurants in the area
- Identifying areas with lower than average crime rate in the neighborhood
- Deciding on the appropriate neighborhood to open that restaurant

1.3 Target Audience:

The key target audience of this project includes entrepreneurs, business owners or stakeholders who are interested in opening a new Thai Restaurant or develop their existing business further. The stakeholders or investors will benefit from the analysis in making better decisions and drawing insights from this project in finding the most suitable neighborhood. In addition, people interested in enjoying Thai cuisine in Toronto will find this convenient. Finally, data scientists who want to explore Toronto neighborhoods using machine learning techniques will find this project useful.

2. Data Description:

There are four main sources of data used in this project. They are:

- a) Toronto Postal Codes, Borough, and Neighborhoods data collected from a Wikipedia page [3]. The Wikipedia page has all the basic information we need to explore and cluster the neighborhoods in Toronto.
- b) Toronto Postal Codes together with their corresponding latitudes and longitudes from Geospatial Coordinates data file [4].
- c) Gathering the number of restaurants within the certain radius of each neighborhood and then filtering out Thai restaurants using Foursquare API. This includes mapping the neighborhood clusters, which uses latitudes, longitudes, venue details etc. from Foursquare API and then use Folium to create maps.
- d) Toronto crime data by each neighborhood [5]. The data include information on 7 major crimes i.e., assault, auto theft, break & enter, robbery, theft over/stealing, shootings, and homicide rates per 100,000 people by each neighborhood, provided by Toronto Police Service.

3. Methodology and Analysis

3.1 Data processing, cleaning and preparation:

The data preparation from each source was done separately. First, I've scraped the neighborhood data from Wikipedia page to obtain the Postal Codes, Borough, and Neighborhood names within Toronto. This also included cleaning and wrangling the data, and putting it in a pandas dataframe. It was followed by merging the latitudes and longitudes information with that data based on the postal codes. The resulting dataframe contains Postal Codes, Borough, Neighborhoods, Latitudes and Longitudes of Toronto, as shown in Figure 1.

	PostalCode	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	Queen's Park	Queen's Park	43.662301	-79.389494

Figure 1: Toronto neighborhoods and their geospatial coordinates

Since my focus is on Toronto, I decided to work only with boroughs that contain the word Toronto in it.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
15	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
19	M4E	East Toronto	The Beaches	43.676357	-79.293031
20	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306

Figure 2: Selected boroughs that contain the word 'Toronto'

The selection resulted in 39 neighborhoods across 5 boroughs: Downtown Toronto (18), East Toronto (5), West Toronto (6), East York/East Toronto (1), and Central Toronto (9). Figure 3 shows the number of neighborhoods in each borough.

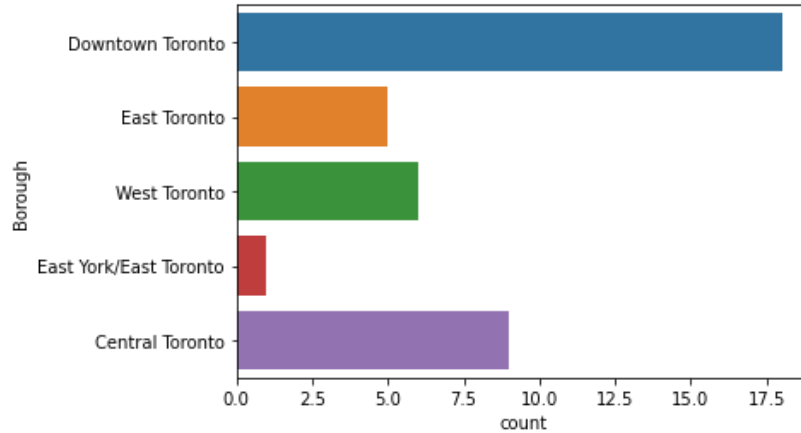


Figure 3: Number of neighborhoods in each borough

After the neighborhoods data were ready for further analysis, I used the Foursquare API to get information about the nearby venues within each neighborhood. Based on latitude and longitude components, the radius was set to 500 meters, and limit to top 100 venues for each neighborhood, such as coffee shops, different types of restaurants, park, bakery, gym etc. The first 5 venues are shown in Figure 4.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
4	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa

Figure 4: Venues information for each neighborhood

The exploration of Foursquare API was crucial to analyzing the number of Thai restaurants in Toronto. The result indicated that there are 230 unique categories of venues across the 39 neighborhoods. I also checked for the number of Thai restaurants there, and found out that there are 19 Thai restaurants already existing in those 39 neighborhoods. The venues list is shown in Figure 5.

```
print (toronto_venues['Venue Category'].value_counts())
```

Coffee Shop	152
Café	82
Restaurant	50
Italian Restaurant	41
Hotel	39
Park	36
Japanese Restaurant	32
Bakery	30
Pizza Place	27
Gym	24
Seafood Restaurant	24
Clothing Store	23
Bar	23
Sushi Restaurant	22
Sandwich Place	20
Thai Restaurant	19
Pub	19
Cocktail Bar	18
Gastropub	18

Figure 5: List of venues in different neighborhoods of Toronto

3.2 Exploratory Data Analysis:

In order to get the frequency of different types of venues or restaurants, I performed One-hot encoding of the data (or creating dummies of each type). This technique is used to transform data into binary category of 0 or 1 for machine learning algorithms. For each venue, it indicated the frequency at how many of those venues were located in each neighborhood. Subsequently, I decided to focus only on restaurants, so I narrowed the search string to the word 'Restaurant' and got the total number of restaurants per neighborhood. In addition, I prepared a dataframe with only Thai restaurants, and then merged those two to form a new one that shows the number of Thai restaurants vs. all other restaurants in each neighborhood (see Figure 6).

	Neighborhood	Thai Restaurant	Total Restaurants
25	Richmond, Adelaide, King	3	24
6	Commerce Court, Victoria Hotel	2	30
12	First Canadian Place, Underground city	2	30
3	Central Bay Street	2	18
16	High Park, The Junction South	2	7

Figure 6: Number of Thai restaurants vs. Total restaurants per neighborhood

Afterwards, I extracted data of the 10 most common restaurant types in each neighborhood. The first few rows are shown in the figure below.

	Neighborhood	1st Most Common Cuisine	2nd Most Common Cuisine	3rd Most Common Cuisine	4th Most Common Cuisine	5th Most Common Cuisine	6th Most Common Cuisine	7th Most Common Cuisine	8th Most Common Cuisine	9th Most Common Cuisine	10th Most Common Cuisine
0	Berczy Park	Total Restaurants	Seafood Restaurant	Restaurant	Comfort Food Restaurant	Eastern European Restaurant	French Restaurant	Greek Restaurant	Japanese Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant
1	Brockton, Parkdale Village, Exhibition Place	Total Restaurants	Italian Restaurant	Restaurant	Doner Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant
2	CN Tower, King and Spadina, Railway Lands, Har...	Total Restaurants	Doner Restaurant	Gluten-free Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant	Eastern European Restaurant
3	Central Bay Street	Total Restaurants	Italian Restaurant	Thai Restaurant	Japanese Restaurant	Portuguese Restaurant	Falafel Restaurant	French Restaurant	Korean Restaurant	Middle Eastern Restaurant	Modern European Restaurant
4	Christie	Total Restaurants	Italian Restaurant	Restaurant	Doner Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant

Figure 7: Top 10 most common restaurant types in each neighborhood

3.3 Using K-Means Clustering Approach:

I used an unsupervised Machine Learning technique called “K-means clustering” to segment the neighborhoods with similar clusters on the basis of each neighborhood data. The ‘Elbow Method’ was used to determine the optimal k number to divide neighborhoods into clusters. Using both Distortion and Inertia, the K-Elbow Visualizer demonstrated that the best value of k in our model was 3. Therefore, the number of neighborhood clusters was set to 3 (Figure 8).

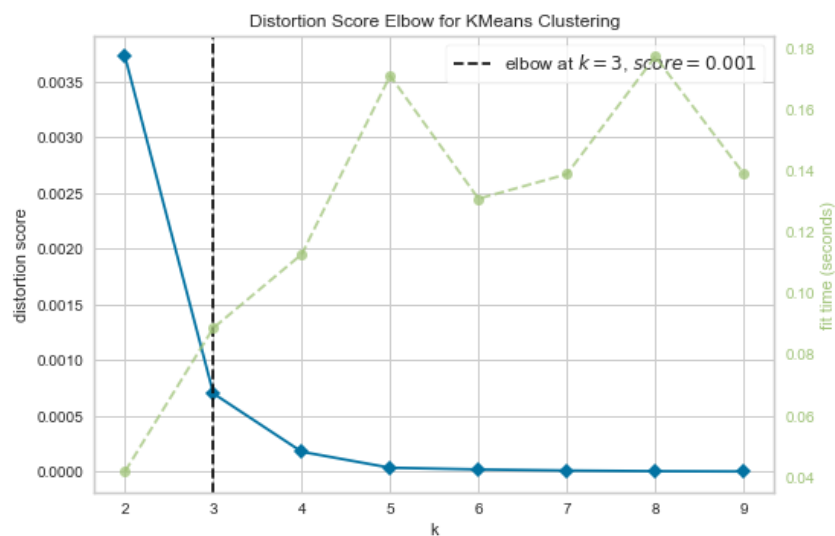


Figure 8: Finding the best k via the ‘Elbow Method’

For further analysis, I've merged the Toronto data with the top 10 most common restaurant types in each neighborhood along with cluster labels and latitude/longitude information. The resulting dataframe is depicted in Figure 9.

Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common cuisine	2nd Most Common cuisine	3rd Most Common cuisine	4th Most Common cuisine	5th Most Common cuisine	6th Most Common cuisine	7th Most Common cuisine	8th Most Common cuisine	9th Most Common cuisine	10th Most Common cuisine
Regent Park, Harbourfront	43.654260	-79.360636	0	Total Restaurants	Restaurant	Asian Restaurant	French Restaurant	Mexican Restaurant	Doner Restaurant	German Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant
Garden District, Ryerson	43.657162	-79.378937	0	Total Restaurants	Middle Eastern Restaurant	Italian Restaurant	Japanese Restaurant	Ramen Restaurant	Fast Food Restaurant	Mexican Restaurant	Modern European Restaurant	Chinese Restaurant	New American Restaurant
St. James Town	43.651494	-79.375418	0	Total Restaurants	Seafood Restaurant	Moroccan Restaurant	Restaurant	Italian Restaurant	American Restaurant	Thai Restaurant	Belgian Restaurant	Sushi Restaurant	Asian Restaurant
The Beaches	43.676357	-79.293031	2	Total Restaurants	Doner Restaurant	Gluten-free Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant	Eastern European Restaurant
Berczy Park	43.644771	-79.373306	0	Total Restaurants	Seafood Restaurant	Restaurant	Comfort Food Restaurant	Eastern European Restaurant	French Restaurant	Greek Restaurant	Japanese Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant

Figure 9: Top 10 most common restaurants with geospatial data and cluster labels

3.4 Mapping the Clusters:

In conjunction with the merged data, I've utilized Foursquare API to gather additional information about nearby venues and to narrow down the location of Thai restaurants in each neighborhood of Toronto. Furthermore, I've created maps of Toronto based upon restaurant data using Folium that will provide the stakeholders with insights as to which neighborhoods have the best potential to open their restaurant in. Folium makes it easy to visualize data in Python on an interactive leaflet map. It not only binds data to a map but also creates visualizations as markers on the map (using OpenStreetMap, Mapbox etc.).

The map in Figure 10 shows that the neighborhoods are grouped into three clusters. Each cluster is randomly color coded to distinguish one from the other. Each dot in the map represents Latitude, Longitude, Neighborhood name, and Cluster Label. At a first glance, it seems that the blue cluster contains the largest neighborhoods. I'll explore each individual cluster in detail in the results section.

Subsequently, I've created another dataframe to filter out only Thai restaurants as top 10 most common venues in each neighborhood. This was done to show which neighborhoods are unsuitable for opening a new Thai restaurant because of the existing ones. Then, I've re-plotted the map to visually illustrate those neighborhoods (Figure 11).

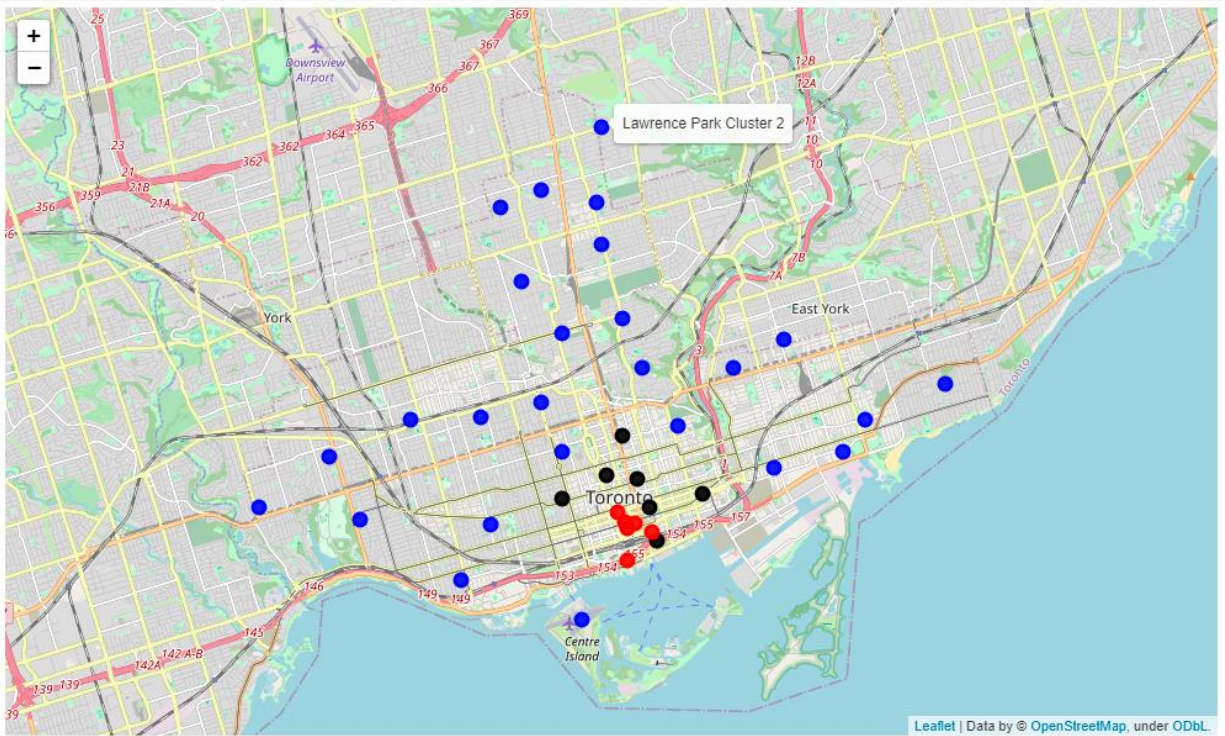


Figure 10: Folium map of Toronto for the clusters of different neighborhoods

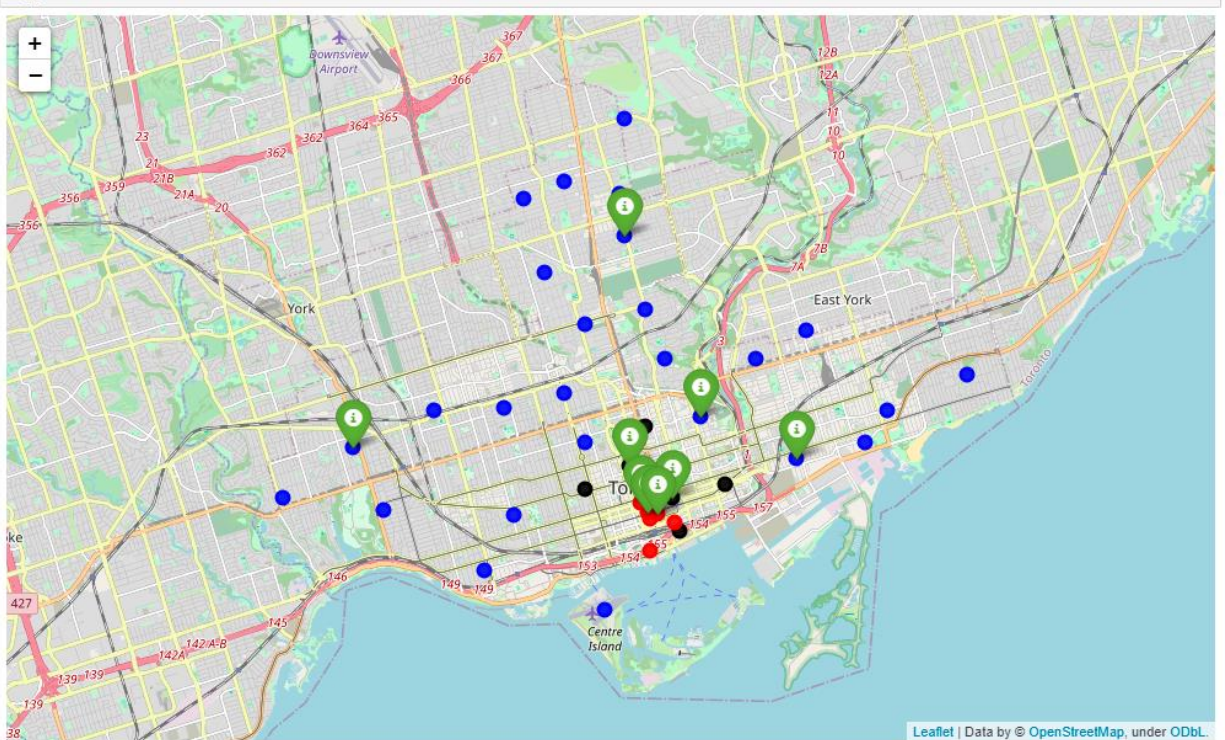


Figure 11: Toronto map with existing Thai restaurants in each neighborhood

4. Results

After running the K-means clustering we found out which Toronto neighborhoods were assigned to each of the three clusters. Let's explore each cluster one by one using Figures 12 and 13.

Looking into the neighborhoods in the first cluster, we find that 6 neighborhoods belong to this cluster. There are 6 already existing Thai restaurants in this cluster.

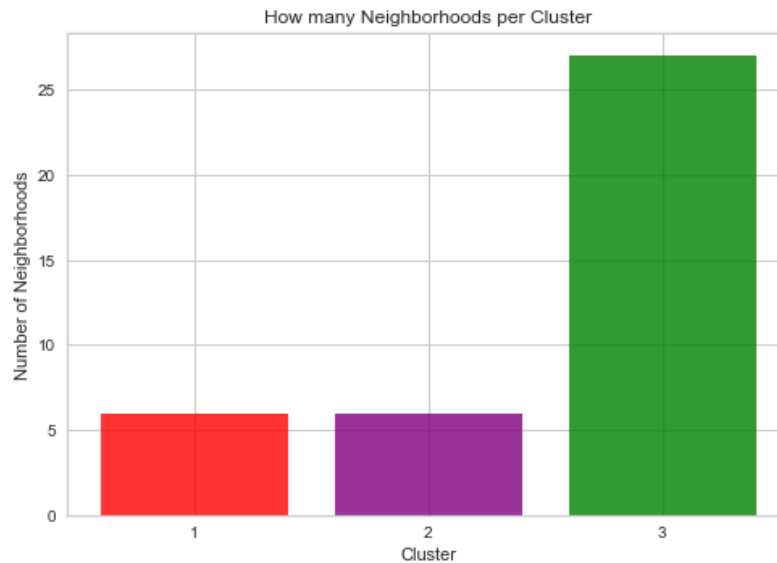


Figure 12: The number of neighborhoods in each cluster

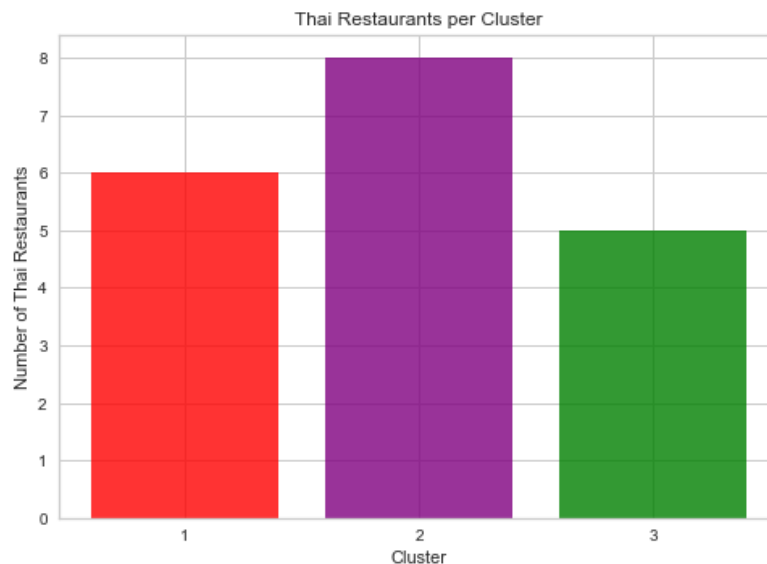


Figure 13: The number of Thai Restaurants in each cluster

The second cluster contains 6 Toronto neighborhoods as well. However, it has the most number of Thai restaurants at 8. Hence, it does not look like a suitable cluster to open a new Thai restaurant.

Finally, the third cluster comprises the most number of neighborhoods at 27. And yet, it has the least number of Thai restaurants at 5. Therefore, this cluster could be the ideal location to open a new Thai restaurant.

At this juncture, I wanted to examine the crime rates of different Toronto neighborhoods to find the most suitable location with lower crimes on average for a potential new Thai restaurant. The Toronto crime data include information on 7 major crimes i.e., assault, auto theft, break & enter, robbery, theft over/stealing, shootings, and homicide rates per 100,000 people by each neighborhood. Because the aim is to see recent crimes, I focused on analyzing the rates for the last three years (2018-2020) of available data. First, I calculated the average rate of each crime for the last 3 years, and then the Total Crime Rate by summing all 7 types of crime. Below is a sorted list of first few neighborhoods and their associated crime rates (Figure 14).

	Neighborhood	Crimes_Rate			
126	Steeles	491.993422			
0	Yonge, St.Clair	500.172143	70	High Park, Swansea	717.460890
36	Guildwood	563.262814	10	Humewood, Cedarvale	721.457310
127	Lawrence Park North	571.106238	97	Rouge	723.635242
21	Centennial Scarborough	582.303423	83	L'Amoreaux	723.951576
124	Maple Leaf	622.680181	8	Thorncliffe Park	725.995132
82	Eringate, Centennial, West Deane	625.138094	112	Woodbine, Lumsden	730.393257
64	Edenbridge, Humber Valley	627.720897	119	Lawrence Park South	734.898010
113	Bayview Village	634.751809	32	Etobicoke West Mall	735.410690
135	Pleasant View	646.067931	107	Parkwoods, Donalda	740.210424
137	Leaside, Bennington	661.711867	125	Markland Wood	743.445932
45	Bayview Woods, Steeles	662.532260	4	Stonegate, Queensway	753.930820
79	Agincourt North	669.637431	29	Broadview North	783.114284
100	Mount Pleasant East	679.351487	86	Alderwood	783.326559
92	Lambton Baby Point	680.050087	71	Highland Creek	784.285678
95	Willowdale East	681.214945	25	Flemington Park	800.499160
84	Banbury, Don Mills	682.906097	117	Don Valley Village	801.504084
58	Humber Heights, Westmount	683.215860	52	Kingsway South	802.100430
42	Westminster, Branson	686.173577	69	High Park North	814.287843
9	Danforth East York	690.259810	61	Hillcrest Village	816.651646
30	Princess, Rosethorn	702.441273	85	Bathurst Manor	817.005183
55	Henry Farm	708.230413	6	The Beaches	820.389906

Figure 14: Crime rate of Toronto neighborhoods

5. Discussion and Conclusion

Our primary objective of this project was to figure out the optimal neighborhood in Toronto to open a new Thai restaurant. During data analysis, the clustering method as well as mapping with Folium helped to highlight the optimal neighborhoods in Toronto and to consolidate information for making better decisions. When comparing among the clusters, the 3rd Cluster includes the most number of neighborhoods at 27, but has only 5 Thai restaurants. Thus, cluster 3 seems to be the optimal candidate for exploring a restaurant business venture.

Now, our criteria at the beginning were to find a place with limited number of existing Thai restaurants and safety/security of the neighborhood. When considering both of these criteria, here are the recommendations of this project:

- The optimum location to open a quality Thai Restaurant seems to be in the neighborhood of Lawrence Park.
- The second best neighborhood with great potential for growing a new Thai restaurant is The Beaches, and the third best is Rosedale.
- All three of these locations have no Thai restaurants in it, and they have low/ relatively lower crime rate per 100,000 people.

Therefore, investors and stakeholders can open a quality Thai restaurant in these locations because there would be little to no competition and they are comparatively safer than other neighborhoods.

However, it should be noted that the Foursquare API data can be a bit limiting. For a more realistic scenario, one would need real-time monitoring data of neighborhood locations and include other variables in the model, such as the relative cost of operating a restaurant in each neighborhood, number of tourist attractions, income level of residents etc. In addition, the analysis of this project can be extended further by looking into the rating and tips of existing restaurants for a better understanding of any competition that stakeholders may face. Nevertheless, the results of this project provide valuable insights into the best neighborhoods in Toronto to open a new Thai Restaurant.

References:

1. "10,000 restaurants have closed across Canada since the start of the pandemic". Accessed at: <https://dailyhive.com/vancouver/restaurants-canada-jobs-loss-10000>
2. "Top 10 Most Popular Ethnic Cuisines in Canada According to Google". Accessed at: <https://www.chefspencil.com/top-10-most-popular-ethnic-cuisines-in-canada/>
3. "List of postal codes of Canada: M". Accessed at: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
4. https://cocl.us/Geospatial_data
5. "Neighbourhood Crime Rates 2020". Accessed at: <https://data.torontopolice.on.ca/datasets/TorontoPS::neighbourhood-crime-rates-2020-1/about>