The Battle of Neighborhoods - Exploring Neighborhoods of Toronto

Sangita Mitra

December 21, 2020

Introduction

Toronto is Canada's largest city and the capital of the province of Ontario. It's a dynamic metropolis with a core of soaring skyscrapers, all dwarfed by the iconic, free-standing CN Tower. This capstone project explores neighborhoods in Toronto, including Central Toronto, East Toronto, West Toronto, and Downtown Toronto. Many people migrate to Canada every year from around the world. Toronto is one of the most desired cities for them. The aim is to help people to explore better facilities around the neighborhoods of Toronto. It will help newcomers access cafes, schools, supermarkets, medical shops, grocery shops, malls, theatres, hospitals, etc. Our findings will help immigrants make informed decisions and address any concerns they have, including the different kinds of cuisines, pubs, parks, provision stores, and what the city has to offer. It will help people make smart and efficient decisions on selecting great neighborhoods out of numbers of other Toronto neighborhoods.

In this project, I've also explored Thai restaurants in Toronto and tried to find a suitable location to open a new Thai restaurant in the neighborhoods.

Business Problem

This project explores Toronto's neighborhoods and finds the most suitable location for an entrepreneur to open a Thai restaurant in Toronto. There are many venues to explore, including different cuisines, movie theatres, parks, markets, etc. This project would use Four-square API as its prime data gathering source. It has a database of millions of places, especially their places API, which provides the ability to perform location search, location sharing, and details about a business. We will be using the KMeans

Clustering Machine learning algorithm to cluster similar neighborhoods together. The business problems that this project trying to solve are -

1. What are the most common venues for different neighborhoods of Toronto?
2. If an entrepreneur wants to open a new Thai restaurant, which place would be most suitable?

Target Audience

This project's target audience is the newcomers in Toronto neighborhoods and the people who want to open a new Thai restaurant in Toronto, Canada.

Data Description

We require geographical location data for Toronto, Canada. Using Postal codes, we can find out the neighborhoods, boroughs, longitude, and latitude of neighborhoods, venues, most popular venue categories, and the venue data related to Thai restaurants.

Data Extraction

Data Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

The Wikipedia page contains information about -

1. Postal Code
2. Borough
3. Neighborhoods
4. Latitude
5. Longitude

Foursquare API Data

We will need data about different venues in different neighborhoods of that specific Borough. To gain that information, we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus, and even photos. The foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API. After finding the neighborhoods' list, we connect to the Foursquare API to gather information about

venues inside every neighborhood. For each neighborhood, we have chosen the radius to be 100 meters.

The data retrieved from Foursquare contained venues' information within a specified distance of the postcodes' longitude and latitude. The information obtained per venue as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue
6. Venue Latitude
7. Venue Longitude
8. Venue Category

Methodology

We will be creating our model with Python's help, so we import all the required packages. Packages that are used in this project are the following.

Pandas: To collect and manipulate data in JSON and HTML and then data analysis

numpy: create an array object

requests: Handle HTTP requests

matplotlib: Detailing the generated maps

folium: Generating maps of Toronto

Sklearn: To import Kmeans, which is the machine learning model that we are using.

Step 1: Extracting the list of Toronto neighborhoods by web scrapping using pandas packages from Wikipedia page. (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

Step 2: Extracting coordinates with latitude and longitude of each neighborhood of Toronto using IBM's csv file. (http://cocl.us/Geospatial_data)

Step 3: As we want to explore Toronto's city, I created a new data frame with Borough that contains Toronto. This data frame has 39 boroughs, including central Toronto, east Toronto, west Toronto, and downtown Toronto. The data frame looks like this:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M4E	East Toronto	The Beaches	43.676357	-79.293031
1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
2	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
3	M4M	East Toronto	Studio District	43.659526	-79.340923
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

Step 4: After gathering all these coordinates, the map of Toronto is visualized using the Folium package to verify whether these are correct coordinates.
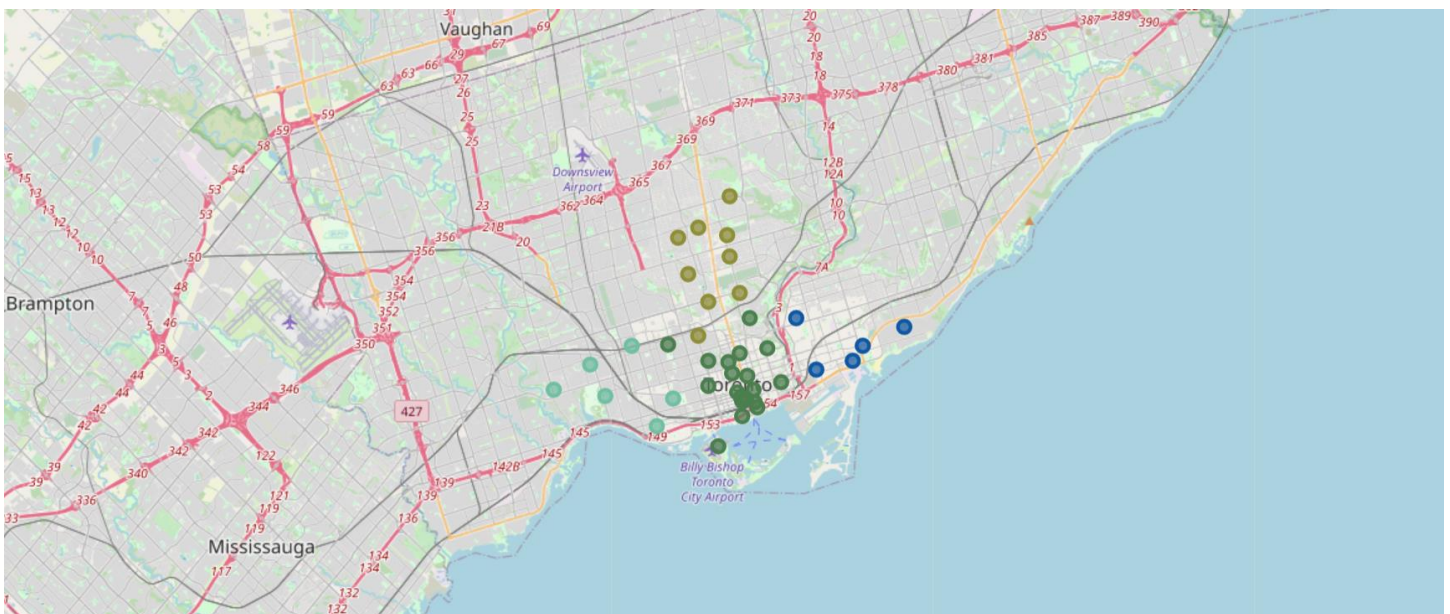

Step 5: Using Foursquare API, we define a function that collects information about each neighborhood, including the neighborhood, geo-coordinates, venue, and venue categories. With the help of Foursquare API, a list is created with the top 200 venues within 500 meters. I have pulled the data using Foursquare client ID, client secret, and version. From Foursquare, I can pull the names, categories, latitude, and longitude of the venues. With this data, I can also check how many unique categories I can get from these venues. Then, I analyze each neighborhood by grouping the rows by neighborhood and taking the mean on each venue category's frequency of occurrence. In my analysis, I got 234 unique venues in Toronto.

Step 6: Implemented One Hot Encoding to convert the categorical datatype of values categories into numerical values.

Step 7: Finding the top 10 most common venue categories in Toronto's neighborhood using a function. The data frame with the most common venues looks like this.

	Neighbourhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Berczy Park	Coffee Shop	Bakery	Cocktail Bar	Pharmacy	Beer Bar	Farmers Market	Cheese Shop	Seafood Restaurant	Restaurant	Irish Pub
1	Brockton, Parkdale Village, Exhibition Place	Café	Breakfast Spot	Coffee Shop	Gym / Fitness Center	Bakery	Stadium	Burrito Place	Restaurant	Climbing Gym	Pet Store
2	Business reply mail Processing Centre, South C...	Light Rail Station	Gym / Fitness Center	Auto Workshop	Park	Comic Shop	Pizza Place	Restaurant	Burrito Place	Brewery	Skate Park
3	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport Lounge	Airport Terminal	Sculpture Garden	Harbor / Marina	Rental Car Location	Plane	Coffee Shop	Boat or Ferry	Boutique
4	Central Bay Street	Coffee Shop	Sandwich Place	Italian Restaurant	Café	Salad Place	Bubble Tea Shop	Burger Joint	Thai Restaurant	Yoga Studio	Modern European Restaurant

Step 8: Using unsupervised K-means clustering, I've clustered the neighborhoods of Toronto into 3 clusters, and the map of clustered neighborhoods is visualized using Folium.
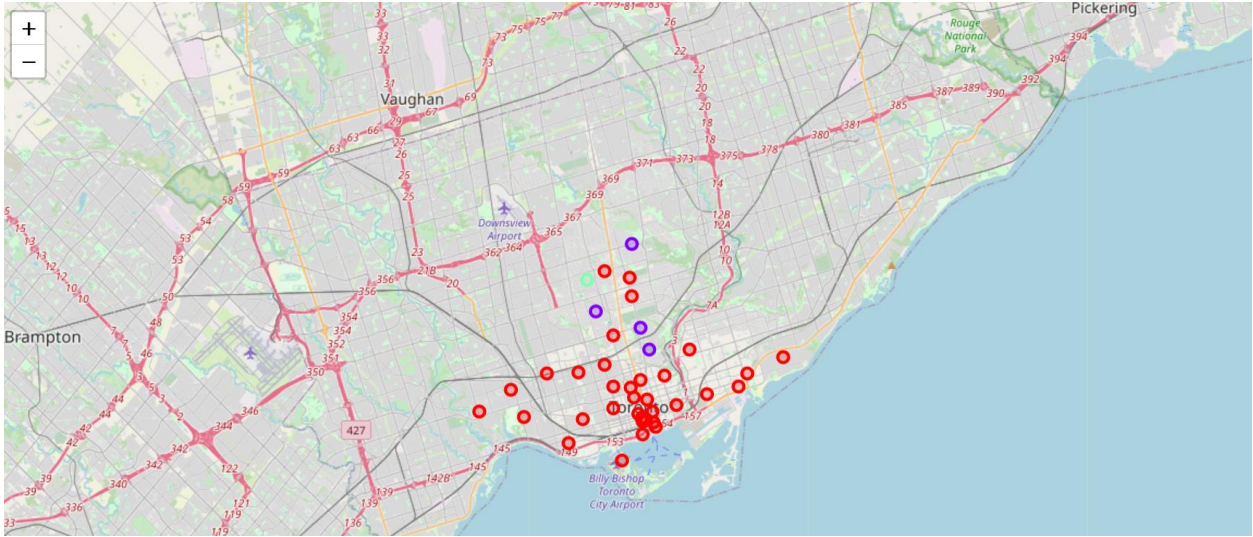

Step 9: Analyzed the most common venues for each cluster of Toronto.

Results

I have clustered the neighborhoods in Toronto into 3 clusters based on their frequency of occurrence for "Thai food" using K-means clustering and visualized the map using Folium. K-means clustering algorithm identifies k number of centroids and then allocates every data point to the nearest cluster

while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms, and it is highly suited for this project.
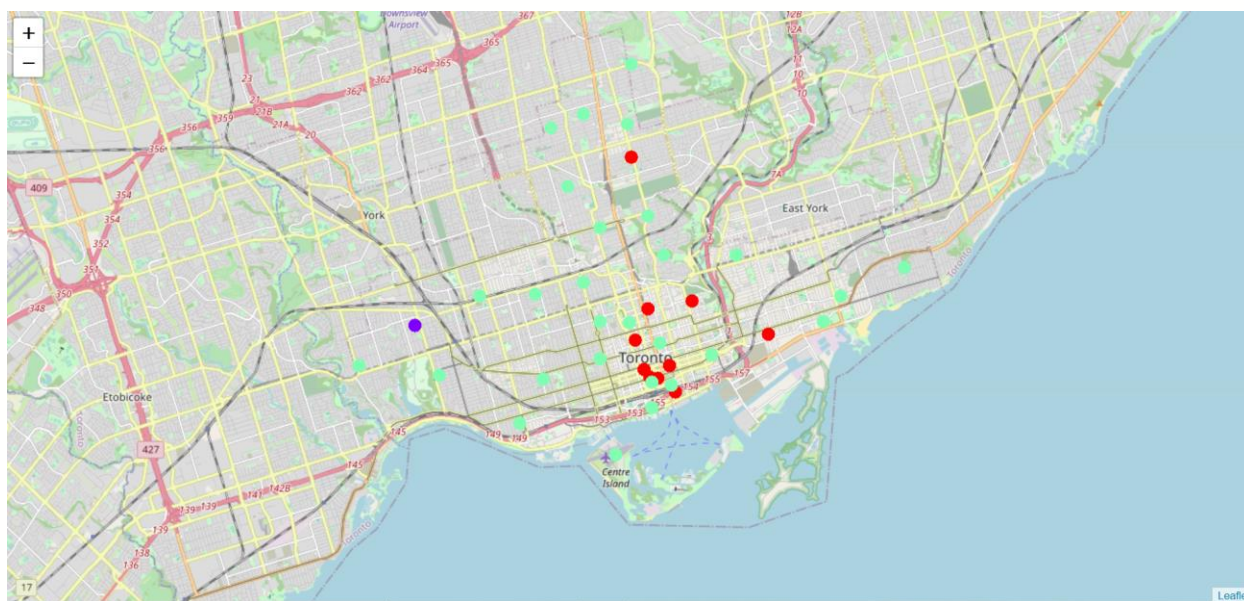
The results from k-means clustering show that we can categorize Toronto neighborhoods into 3 clusters based on how many Thai restaurants are in each neighborhood:

● Cluster 0: Neighborhoods with a low number of Thai restaurants

● Cluster 1: Neighborhoods with no Thai restaurants

● Cluster 2: Neighborhoods with a high number of Thai restaurants

The results are visualized in the above map with Cluster 0 in red color, Cluster 1 in purple color, and Cluster 2 in light green color.

Discussions

The neighborhoods of Toronto are very multicultural. There are many different cuisines, including Indian, Italian, Japanese, Chinese, Thai, Greek, etc. Toronto covers all citizens' necessities by having many restaurants, bars, juice bars, coffee shops, supermarkets, Fish and Chips shops, and Breakfast spots. It has many shopping options, including flower shops, toy store, boutique stores, and clothing stores. For leisure, the neighborhoods are set up to have lots of parks, art galleries, gyms, and historic sites. Overall, the city of Toronto offers a multicultural, diverse, and certainly entertaining experience.

Most Thai restaurants are in Cluster 2 (Green), which is around Regent Park, India Bazar, Queen's park, Toronto Island areas, and lowest (close to zero) in Cluster 1 (Purple) areas, which are High Park

areas. Therefore, there are good opportunities to open near High Park, Toronto, as the competition below. The number of Thai restaurants is lower in cluster 0 (Red) and around First Canadian Place, Richmond, and Victoria Hotel. Therefore, there are lots of places that will be suitable to open a new Thai restaurant. This project recommends the entrepreneur open an authentic Thai restaurant in these locations with little to no competition.

Conclusion

This project aimed to explore Toronto's city and find the most common venues in each neighborhood for potential tourists and migrants. We explored all Toronto neighborhoods based on their postal codes. We then extrapolated the common venues present in each neighborhood, concluding with clustering similar neighborhoods together using k-means clustering. We could see that each of the neighborhoods has a wide variety of experiences to offer, unique in its way. The cultural diversity is quite evident, which also gives the feeling of a sense of inclusion. This project also recommends the entrepreneur to open an authentic Thai restaurant near High Park as the competition will be low in that area.

References

1. [List of Neighborhoods of Toronto](#)
2. [Foursquare API](#)
3. [Foursquare Developer Documentation](#)