Capstone Project: The Battle of Neighbourhoods

Mojdeh Soltani

November 2020

1 Introduction

1.1 Background

Toronto, Canada's largest city with 2,731,571 population in 2016, is the capital of Ontario Province, the most populous city in Canada, and the fourth largest populous city in North America. This city is a world leader in such areas as business, finance, technology, entertainment and culture. Its large population of immigrants from all over the globe has also made Toronto one of the most multicultural cities in the world. Toronto encompasses a geographical area formerly administered by many separate municipalities. These municipalities have each developed a distinct history and identity over the years, and their names remain in common use among Torontonians. Former municipalities include East York, Etobicoke, Forest Hill, Mimico, North York, Parkdale, Scarborough, Swansea, Weston and York. Throughout the city there exist hundreds of small neighbourhoods and some larger neighbourhoods covering a few square kilometres which made Toronto a city of neighbourhoods.

1.2 Problem Definition

In this project, we want to explore the neighborhoods in Toronto and group them into similar and dissimilar clusters. There can be many factors to consider some regions similar, including the facilities, events, restaurants, parks, schools, etc. in each neighborhood.

1.3 Interest

This study can be interesting for those who want to live temporary or for a long period in Toronto including new residents, tourists, and people who want to change their neighborhood. Imagine that someone wants to live a new neighborhood (whether they are tourists or Toronto residents), it is important for them to know their new neighborhood and compare it to their previous or desired districts. Hence, this project will help them to know every area in Toronto and choose their new and favorite neighborhood.

2 Data Acquisition and Cleaning

2.1 Data Sources

The following Wikipedia page is used to get information about neighborhoods in Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. This defines the scope of this project, which is the city of Toronto in Canada.

Also, we use the following CSV file to extract the geographical coordinates of different postal codes (neighborhoods): http://cocl.us/Geospatial_data. This is required to get the venue data and plot the map.

Finally, we request the venue data for each neighborhood from the Foursquare API. This data is used to execute clustering on the neighborhoods.

2.2 Data Cleaning

We combine the data downloaded from multiple sources into one table. After transforming the data into the Pandas data frame, we ignore the rows with 'Not assigned' label in the Borough column. Then we merge the neighborhoods with the same postal code. Finally, if a neighborhood has 'Not assigned' name, we consider the name of their borough as their neighborhood's name.

2.3 Feature Selection

After all the merging and cleaning data that we mentioned above, we consider postal code, borough, neighborhood's name, latitude, and longitude of each neighborhood as shown in the following table (there are 103 rows and five columns). Note that in the methodology section, we will discuss how to consider and insert different events for each neighborhood as a new data frame.

	Postalcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

3 Methodology

3.1 Preparing the Primary Data

First, we use the BeautifoulSoup package to read the data about Toronto neighborhoods on the Wikipedia page, and then we transform it into the Pandas data frame as below.

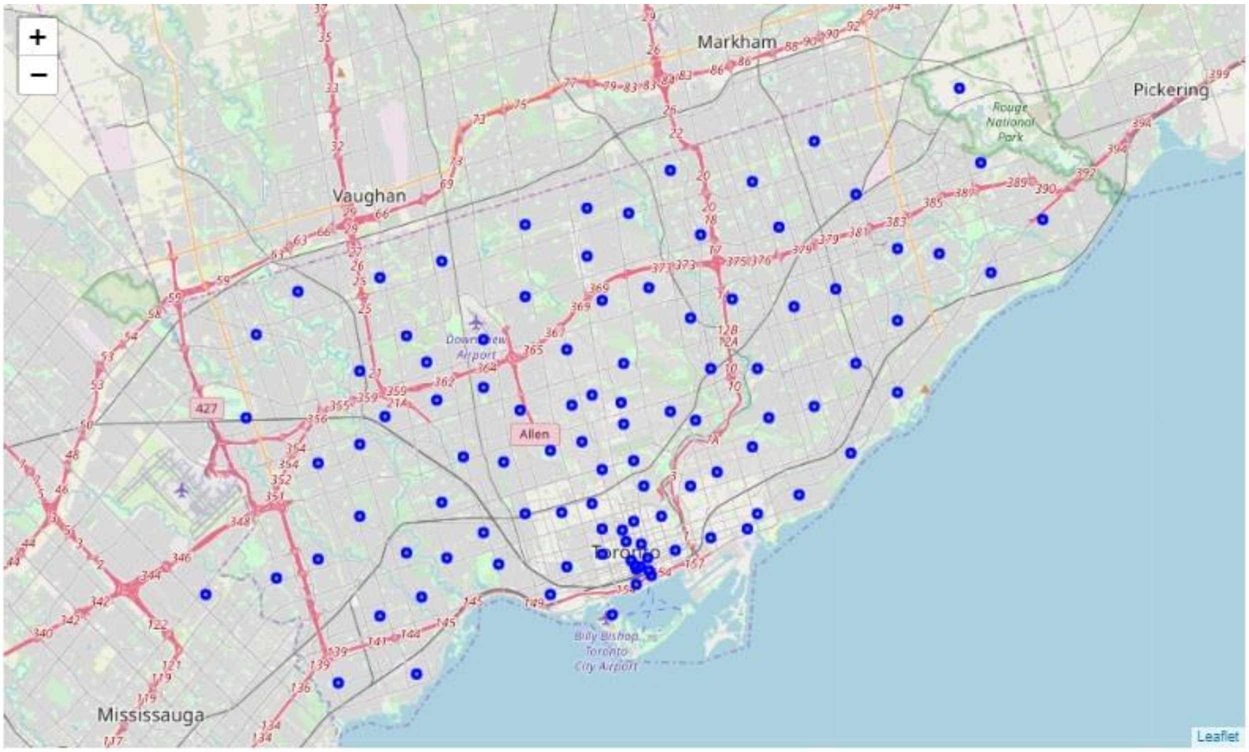
	Postalcode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront

Second, we use the CSV file to extract the geographical coordinates of different neighborhoods. Finally, after doing some data cleaning mentioned in section 2.2, we combine the data as follows.

	Postalcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

3.2 Showing the Toronto Neighborhoods on the Map

By using the latitude and longitude of each neighborhood, and the Python folium library, we generate the following map to visualize the data (Toronto neighborhoods).



3.3 Using Foursquare API to Explore each Neighborhood

By using the Foursquare API, we explore the neighborhoods to find out what venues exist in each neighborhood. We get the top 50 venues of each neighborhood within the radius of 600 meters of their geographical coordinates. Eventually, we create a new data frame as follows to display the ten most common venues of each neighborhood.

	Cluster Labels	Neighborhood	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	0	Adelaide, King, Richmond	Coffee Shop	Café	Steakhouse	Bar	Restaurant	Burger Joint	Bakery	Asian Restaurant	Thai Restaurant	Cosmetics Shop
1	0	Berczy Park	Coffee Shop	Cocktail Bar	Beer Bar	Farmers Market	Bakery	Seafood Restaurant	Steakhouse	Cheese Shop	Café	Greek Restaurant
2	0	Brockton, Exhibition Place, Parkdale Village	Breakfast Spot	Café	Nightclub	Coffee Shop	Yoga Studio	Pet Store	Stadium	Burrito Place	Restaurant	Climbing Gym
3	0	Business Reply Mail Processing Centre 969 Eastern	Skate Park	Auto Workshop	Brewery	Smoke Shop	Spa	Restaurant	Farmers Market	Fast Food Restaurant	Burrito Place	Recording Studio
4	0	CN Tower, Bathurst Quay, Island airport, Harbo...	Airport Service	Airport Lounge	Airport Terminal	Coffee Shop	Harbor / Marina	Rental Car Location	Sculpture Garden	Bar	Boat or Ferry	Airport

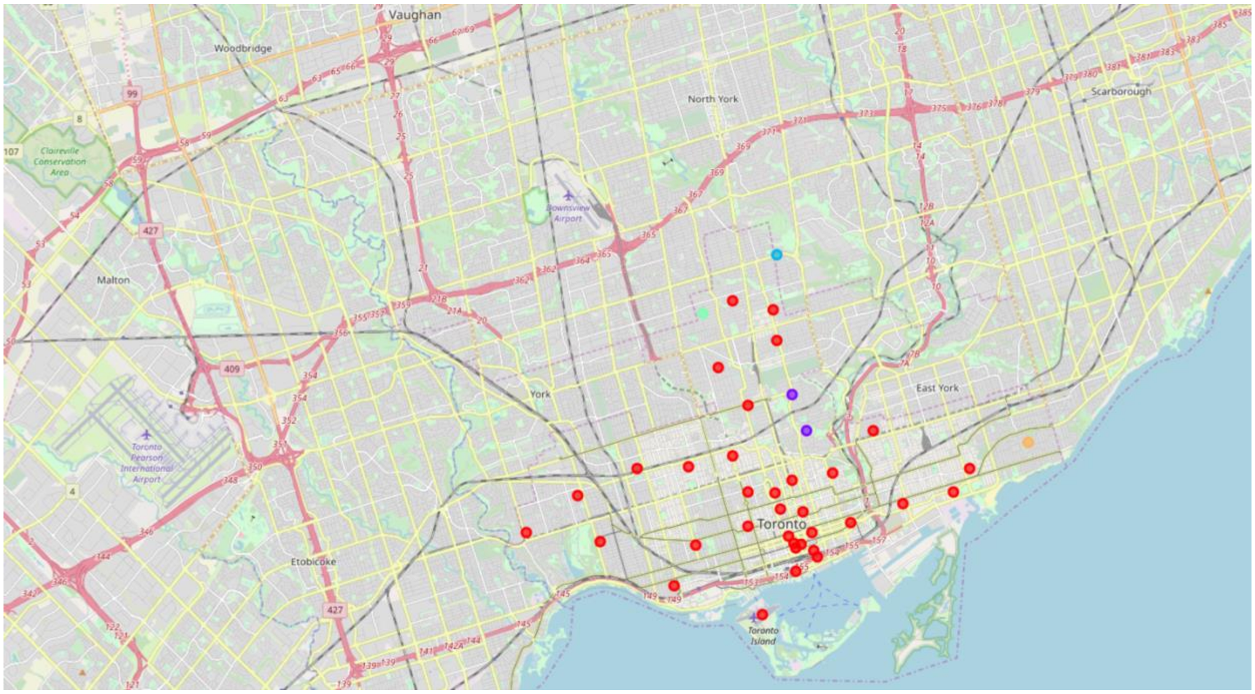
3.4 K-Means Clustering Algorithm

There are some common venues among neighborhoods. So, the K-means algorithm is a suitable way to group neighborhoods into different categories in which each category shows similar neighborhoods.

K-means algorithm is a popular unsupervised machine learning algorithm, which is used for clustering data. Note that we categorize neighborhoods into 6 clusters.

4 Results

The following map visualizes the cluster of each neighborhood in Toronto by using the Folium package and Matplotlib library.



5 Discussion

Now, we know about the most common venues in each neighborhood. Also, we categorized similar neighborhoods into 5 clusters. This helps those who want to live in a new place to choose the neighborhood which is similar to their previous or desired neighborhood.

We can analyze the clusters and see similar neighborhoods in each cluster. For example, the below table shows a part of the neighborhoods in cluster 1.

	Borough	Cluster Labels	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
41	East Toronto	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Furniture / Home Store	Restaurant	Bubble Tea Shop	Grocery Store	Pub	Pizza Place
42	East Toronto	0	Park	Board Shop	Sushi Restaurant	Sandwich Place	Brewery	Liquor Store	Burger Joint	Italian Restaurant	Burrito Place	Fast Food Restaurant
43	East Toronto	0	Café	Coffee Shop	Gastropub	Bakery	Italian Restaurant	Brewery	American Restaurant	Yoga Studio	Bookstore	Sandwich Place
45	Central Toronto	0	Park	Department Store	Breakfast Spot	Sandwich Place	Food & Drink Shop	Hotel	Gym	Comic Shop	Dim Sum Restaurant	Eastern European Restaurant
46	Central Toronto	0	Clothing Store	Coffee Shop	Sporting Goods Shop	Salon / Barbershop	Restaurant	Rental Car Location	Café	Chinese Restaurant	Park	Mexican Restaurant
47	Central Toronto	0	Dessert Shop	Sandwich Place	Coffee Shop	Sushi Restaurant	Gym	Café	Italian Restaurant	Pizza Place	Brewery	Restaurant
49	Central Toronto	0	Coffee Shop	Pub	Pizza Place	Sushi Restaurant	Sports Bar	Fried Chicken Joint	Restaurant	American Restaurant	Supermarket	Liquor Store
51	Downtown Toronto	0	Restaurant	Coffee Shop	Italian Restaurant	Pizza Place	Bakery	Pub	Café	Butcher	Sandwich Place	Breakfast Spot
52	Downtown Toronto	0	Coffee Shop	Japanese Restaurant	Gay Bar	Sushi Restaurant	Restaurant	Fast Food Restaurant	Men's Store	Mediterranean Restaurant	Hotel	Gym
53	Downtown Toronto	0	Coffee Shop	Park	Bakery	Café	Pub	Breakfast Spot	Restaurant	Mexican Restaurant	Farmers Market	Event Space
54	Downtown Toronto	0	Coffee Shop	Clothing Store	Café	Japanese Restaurant	Cosmetics Shop	Electronics Store	Tea Room	Ice Cream Shop	Bubble Tea Shop	Pizza Place
55	Downtown Toronto	0	Coffee Shop	Café	Restaurant	Cocktail Bar	Cosmetics Shop	Italian Restaurant	Beer Bar	Clothing Store	American Restaurant	Hotel
56	Downtown Toronto	0	Coffee Shop	Cocktail Bar	Beer Bar	Farmers Market	Bakery	Seafood Restaurant	Steakhouse	Cheese Shop	Café	Greek Restaurant
57	Downtown Toronto	0	Coffee Shop	Sandwich Place	Café	Italian Restaurant	Ice Cream Shop	Juice Bar	Japanese Restaurant	Burger Joint	Salad Place	Chinese Restaurant
58	Downtown Toronto	0	Coffee Shop	Café	Steakhouse	Bar	Restaurant	Burger Joint	Bakery	Asian Restaurant	Thai Restaurant	Cosmetics Shop
59	Downtown Toronto	0	Coffee Shop	Aquarium	Italian Restaurant	Hotel	Café	Sporting Goods Shop	Restaurant	Fried Chicken Joint	Scenic Lookout	Brewery
60	Downtown Toronto	0	Coffee Shop	Café	Hotel	Restaurant	Steakhouse	Gastropub	Seafood Restaurant	Bar	Deli / Bodega	Italian Restaurant
61	Downtown Toronto	0	Coffee Shop	Café	Hotel	Restaurant	Gym	Seafood Restaurant	Bakery	Italian Restaurant	Deli / Bodega	Gastropub

6 Conclusion and Future Directions

In this project, we explored the neighborhoods in Toronto through preparing data, categorize neighborhoods into six groups by performing K-means clustering algorithm (which is an unsupervised machine learning algorithm). Lastly, we developed recommendations to the people who want to live temporary or for a long period in Toronto including new residents, tourists, and people who want to change their neighborhood.

As new research, some can consider other algorithms to cluster neighborhoods and compare the results of different algorithms. Also, we can find a way to determine the optimal number of clusters (k) before performing the K-means algorithm.

7 References

- Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- CSV file for geographical data: http://cocl.us/Geospatial_data
- Foursquare API