

# Using clustering method to find the most convenient place for a friend to live in Toronto

Zuyu Yang 20210309

## Introduction

Toronto is the capital city of the Canadian province of Ontario. It is the most populous city in Canada and the fourth most populous city in North America. In the decades after World War II, Chinese job-seekers started to arrive Toronto, contributing to its population growing. In 2016, the three most commonly reported ethnic origins overall were Chinese (332,830 or 12.5 per cent), English (331,890 or 12.3 per cent) and Canadian (323,175 or 12.0 per cent)(<https://en.wikipedia.org/wiki/Toronto>).

One of my friends named Lin who will relocate from Beijing to Toronto. She is asking my suggestion where is the most convenient for her to live. Lin likes doing exercise and enjoys going out for dinner. So the regions that she prefers to live need to be close to gyms, parks, as well as restaurant.

## Methods

I used the library BeautifulSoup to scrape the list of neighbourhoods of Toronto Wikipedia. Each neighbourhood's coordinates were retrieved using Geocoder. Then I retrieve the list of venues for each neighbourhood from FoursquareAPI. I used a K-means algorithm to perform the clustering analysis after dataset combined and cleaned. The neighbourhoods were displayed with different colors on a map using Folium. With this clustering method, we can see neighbourhoods that are similar using their most common types of venues. Then look detail of each cluster and find the regions that match Lin's requirement for setting her new home.

## Analysis and results

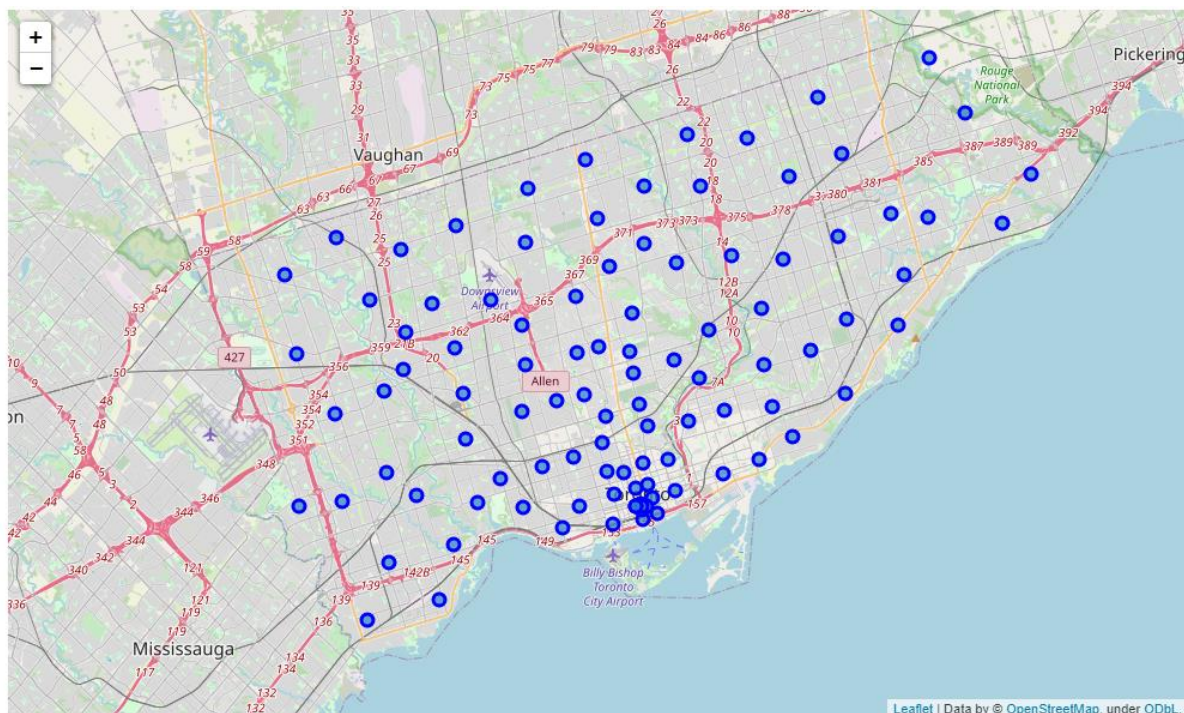
**(1) Scrape the list of neighbourhoods:** The Wikipedia page contains a table of postal codes followed in Toronto, along with the boroughs and neighborhoods in Toronto city. I used the library BeautifulSoup to scrape the list of neighbourhoods of Toronto Wikipedia. Use the head() function, the data in the form of table looks something like this before it enters the pre-processing phase of the project, followed by further step of filtering all the unassigned neighbourhood.

	PostalCode	Borough	Neighborhood
0	None	None	None
1	M1A	Not assigned	Not assigned
2	M2A	Not assigned	Not assigned
3	M3A	North York	Parkwoods
4	M4A	North York	Victoria Village

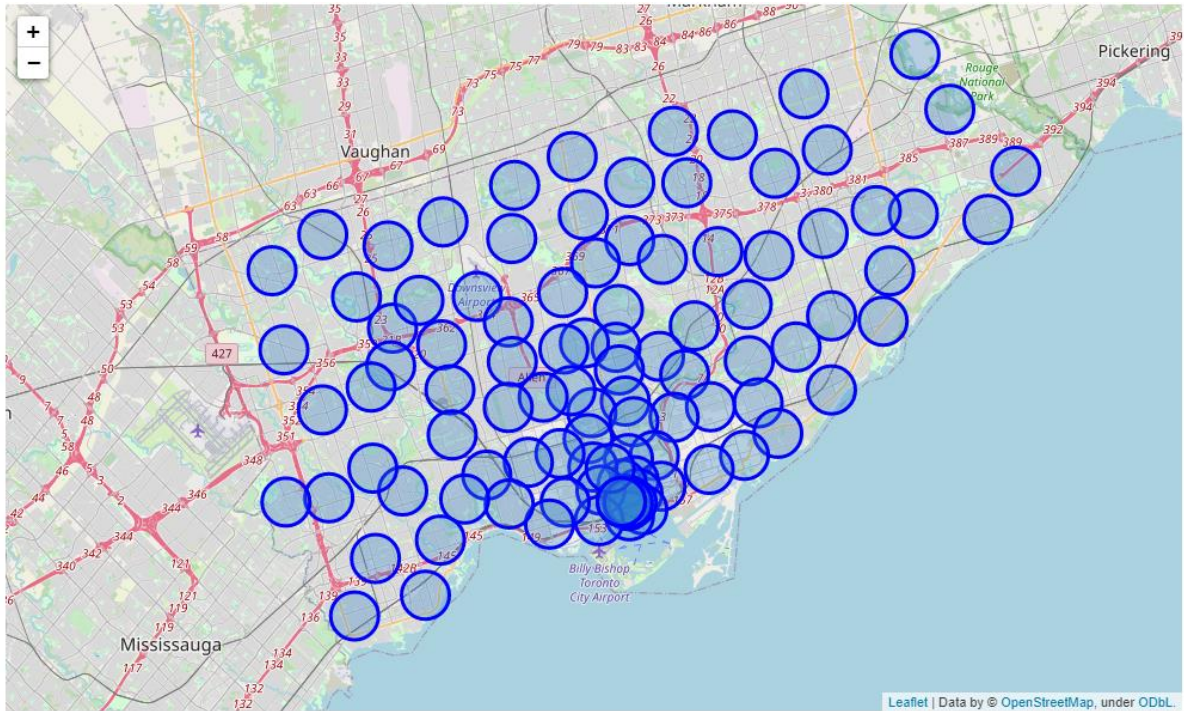
**(2) Combine the latitude and the longitude coordinates of each neighbourhood:** Each neighbourhood's coordinates were retrieved using Geocoder. Merge the coordinates dataframe with the original neighbourhoods dataframe. With the head() function, the data in the form of table now looks like this:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.81139	-79.19662
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78574	-79.15875
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.76575	-79.17470
3	M1G	Scarborough	Woburn	43.76812	-79.21761
4	M1H	Scarborough	Cedarbrae	43.76944	-79.23892

**(3) Draw the city of Toronto with the neighbourhoods from the dataframe:** create map of Toronto using latitude and longitude values and add markers to map:



Redraw the same map, but with each neighbourhood represented by a circle having a radius equals to the radius size `CONST_venuesRadiusScan`



**(4) Get the venues in Toronto, using FoursquareAPI:** Retrieve the list of venues for each neighbourhood from FoursquareAPI, the follow table shows its first 5 lines:

	PostalCode	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	M1B	Scarborough	Malvern, Rouge	43.81139	-79.19662	Images Salon & Spa	43.802283	-79.198565	Spa
1	M1B	Scarborough	Malvern, Rouge	43.81139	-79.19662	Canadiana exhibit	43.817962	-79.193374	Zoo Exhibit
2	M1B	Scarborough	Malvern, Rouge	43.81139	-79.19662	Wendy's	43.807448	-79.199056	Fast Food Restaurant
3	M1B	Scarborough	Malvern, Rouge	43.81139	-79.19662	Wendy's	43.802008	-79.198080	Fast Food Restaurant
4	M1B	Scarborough	Malvern, Rouge	43.81139	-79.19662	Tim Hortons	43.802000	-79.198169	Coffee Shop

**(5) Analyse each neighborhood :** As we are going to perform a K-means clustering on the dataframe each type of venue will be one hot encoded.

	PostalCode	Borough	Neighborhood	ATM	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	American Restaurant	...	Video Store	Vietnamese Restaurant	War
0	M1B	Scarborough	Malvern, Rouge	0	0	0	0	0	0	0	...	0	0	0
1	M1B	Scarborough	Malvern, Rouge	0	0	0	0	0	0	0	...	0	0	0
2	M1B	Scarborough	Malvern, Rouge	0	0	0	0	0	0	0	...	0	0	0
3	M1B	Scarborough	Malvern, Rouge	0	0	0	0	0	0	0	...	0	0	0
4	M1B	Scarborough	Malvern, Rouge	0	0	0	0	0	0	0	...	0	0	0

5 rows × 339 columns

**(6) Group rows by postal code, borough and neighbourhood, using means**

	PostalCode	Borough	Neighborhood	ATM	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	American Restaurant	...	Video Store	Vietnamese Restaurant	Wa
0	M1B	Scarborough	Malvern, Rouge	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0
2	M1E	Scarborough	Guildwood, Morningside, West Hill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0
3	M1G	Scarborough	Woburn	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.047619	0.0
4	M1H	Scarborough	Cedarbrae	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0

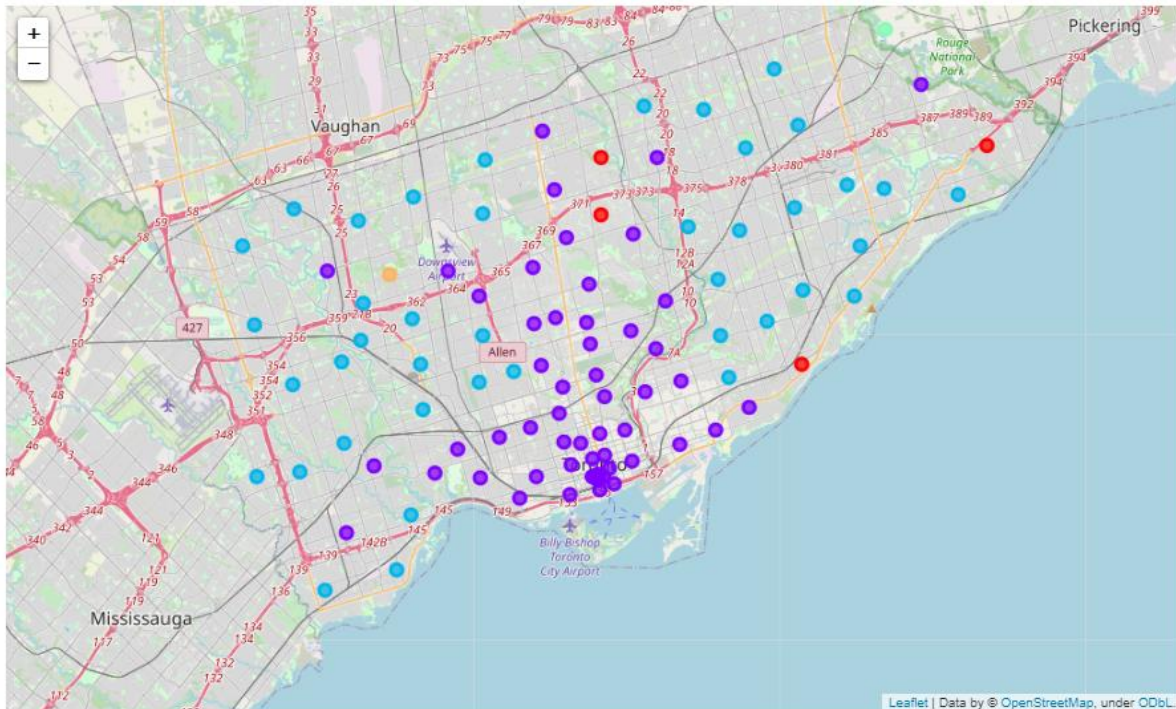
5 rows × 339 columns

**(7) Create the new dataframe and display the top 10 venues for each neighborhood.**

	PostalCode	Borough	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
0	M1B	Scarborough	Malvern, Rouge	Zoo Exhibit	Coffee Shop	Fast Food Restaurant	Other Great Outdoors	Spa	Trail	Hobby Shop	Other Repair Shop	Organic Grocer
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	Park	Burger Joint	Breakfast Spot	Italian Restaurant	Playground	Gym / Fitness Center	Music Venue	Opera House	Movie Theatre
2	M1E	Scarborough	Guildwood, Morningside, West Hill	Park	Convenience Store	Pizza Place	Breakfast Spot	Bank	Grocery Store	Gym / Fitness Center	Fast Food Restaurant	Gymnastics Gym
3	M1G	Scarborough	Woburn	Indian Restaurant	Park	Fast Food Restaurant	Coffee Shop	Pizza Place	Supermarket	Bank	Bakery	Thrift / Vintage Store
4	M1H	Scarborough	Cedarbrae	Indian Restaurant	Coffee Shop	Gas Station	Grocery Store	Bakery	Athletics & Sports	Bus Line	Music Store	Caribbean Restaurant



**(8) Cluster neighbourhoods:** set number of clusters as 5, and run k-means clustering. visualise the clusters using folium.



## (9) Examine clusters:

**Cluster 0** (red) focuses on Parks, Gyms, and places to practice sports and/or chillout outside. This cluster is the places that you want to go if you want to do exercise and outdoor activities.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	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
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78574	-79.15875	0	Park	Burger Joint	Breakfast Spot	Italian Restaurant	Playground	Gym / Fitness Center
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.69510	-79.26466	0	Park	Gym Pool	Restaurant	Skating Rink	General Entertainment	Café
19	M2K	North York	Bayview Village	43.78112	-79.38060	0	Park	Chinese Restaurant	Bank	Intersection	Trail	Japanese Restaurant
20	M2L	North York	York Mills, Silver Hills	43.75698	-79.38060	0	Gym	Park	Intersection	Gym / Fitness Center	Shopping Mall	Furniture / Home Store

**Cluster 1 (purple)** focuses on the diversity of coffee shops, Restaurant, and bars. Its 1st common Venue is Zoo Exhibit, We can see that the downtown of Toronto is highly categorised in this cluster.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	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
0	M1B	Scarborough	Malvern, Rouge	43.81139	-79.19662	1	Zoo Exhibit	Coffee Shop	Fast Food Restaurant	Other Great Outdoors	Spa	Trail
18	M2J	North York	Fairview, Henry Farm, Oriole	43.78097	-79.34781	1	Clothing Store	Coffee Shop	Restaurant	Japanese Restaurant	Sandwich Place	Bakery
21	M2M	North York	Willowdale, Newtonbrook	43.79182	-79.41372	1	Coffee Shop	Korean Restaurant	Café	Middle Eastern Restaurant	Restaurant	Grocery Store
22	M2N	North York	Willowdale, Willowdale East	43.76774	-79.40728	1	Coffee Shop	Japanese Restaurant	Sandwich Place	Ramen Restaurant	Korean Restaurant	Grocery Store
23	M2P	North York	York Mills West	43.74778	-79.40033	1	Coffee Shop	Park	Restaurant	Food Court	Bowling Alley	Burrito Place
26	M3B	North York	Don Mills	43.74923	-79.36186	1	Coffee Shop	Pizza Place	Electronics Store	Japanese Restaurant	Supermarket	Restaurant
27	M3C	North York	Don Mills	43.72168	-79.34352	1	Coffee Shop	Gym	Japanese Restaurant	Beer Store	Sandwich Place	Park
30	M3K	North York	Downsview	43.73384	-79.46828	1	Park	Turkish Restaurant	Coffee Shop	Vietnamese Restaurant	Latin American Restaurant	Thrift / Vintage Store
37	M4E	East Toronto	The Beaches	43.67709	-79.29547	1	Pub	Coffee Shop	Beach	Breakfast Spot	Japanese Restaurant	Pizza Place
38	M4G	East York	Leaside	43.70902	-79.36349	1	Coffee Shop	Bakery	Grocery Store	Indian Restaurant	Bank	Sporting Goods Shop

**Cluster 2 (blue)** focuses on the diversity of Restaurant, Grocery storesm, Convenience store. This cluster is predominant outside of the downtown.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	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
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.76575	-79.17470	2	Park	Convenience Store	Pizza Place	Breakfast Spot	Bank	Grocer
3	M1G	Scarborough	Woburn	43.76812	-79.21761	2	Indian Restaurant	Park	Fast Food Restaurant	Coffee Shop	Pizza Place	Superr
4	M1H	Scarborough	Cedarbrae	43.76944	-79.23892	2	Indian Restaurant	Coffee Shop	Gas Station	Grocery Store	Bakery	Athletic Sports
5	M1J	Scarborough	Scarborough Village	43.74446	-79.23117	2	Fast Food Restaurant	Sandwich Place	Ice Cream Shop	Convenience Store	Restaurant	Chines Restau
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	43.72582	-79.26461	2	Chinese Restaurant	Grocery Store	Fast Food Restaurant	Coffee Shop	Discount Store	Sandw Place
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.71289	-79.28506	2	Intersection	Grocery Store	Bakery	Sandwich Place	Bus Station	Soccer
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	43.72360	-79.23496	2	Ice Cream Shop	Pharmacy	Sporting Goods Shop	Pizza Place	Park	Gift Sh
10	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...	43.75998	-79.26940	2	Indian Restaurant	Coffee Shop	Pizza Place	Furniture / Home Store	Chinese Restaurant	Bank
11	M1R	Scarborough	Wexford, Maryvale	43.75075	-79.30054	2	Middle Eastern Restaurant	Pizza Place	Asian Restaurant	Grocery Store	Intersection	Furnitu Home
12	M1S	Scarborough	Agincourt	43.79452	-79.26708	2	Chinese Restaurant	Shopping Mall	Cantonese Restaurant	Pizza Place	Department Store	Medite Restau

**Cluster 3 (green)** is an outlier, its 1st common venue is Ski Area.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	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
16	M1X	Scarborough	Upper Rouge	43.83412	-79.21668	3	Ski Area	ATM	Nightclub	Other Repair Shop	Other Great Outdoors	Organic Grocery	Optical Shop

**Cluster 4 (orange)** focuses on Food Truck, Park and Baseball field and restaurants.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	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
32	M3M	North York	Downsview	43.73224	-79.50178	4	Food Truck	Park	Baseball Field	Vietnamese Restaurant	ATM	Nightclub	Other Great Outdoors

## Discussion

Relocating from a city where you lived for many years to a completely new city could be a challenge. My friend Lin, she lives in a community near Beijing Sport University, she likes doing exercise such as jogging and playing tennis. The community she stayed in Beijing provides all the necessary places that suit her needs. So the regions in Toronto that she prefers to live in need to be close to gyms, parks, as well as restaurants. By the clustering analysis of top 10 venues for each neighborhood in Toronto, I found 5 clusters. Cluster 0 focuses on Parks, Gyms, and places to practice sports and/or chill out outside. This cluster is the places that you want to go if you want to do exercise and outdoor activities. Cluster 1 focuses on the diversity of coffee shops, Restaurants, and bars. Its 1st common Venue is Zoo Exhibit. We can see that the downtown of Toronto is highly categorized in this cluster. Cluster 2 (blue) focuses on the diversity of Restaurants, Grocery stores, Convenience stores. This cluster is predominant outside of the downtown. Cluster 3 is an outlier, its 1st common venue is Ski Area. Cluster 4 (orange) focuses on Food Truck, Park and Baseball field and restaurants. Since Lin prefers to live close to gyms, parks, as well as restaurants, the North York region of near both cluster 0 and cluster 1 could be a good choice for her. But this analysis has limitations, without considering the house price or rental cost, convenience of transport et. al. Integrating more will be needed to answer more specific questions.

