

Characteristic of Toronto city related with opening a business

Maria Triagni

Jan 26,2020

1. Introduction

1.1 Background

Toronto is provincial capital of Ontario and the most populous city in Canada, with population of 2,731,571 as of 2016. It is also known as an international center of business , finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

There are 140 neighborhoods officially recognized by the city of Toronto with some unofficial neighborhoods within the city's boundary.

The “inner ring” suburbs of York and East York are older, predominantly middle-income area , and ethnically diverse. Many of neighborhood in these area were built up as streetcar suburbs and contain many dense and mixed use some of which are one-way.

The “outer ring” suburbs of Etobicoke , Scarborough, and North York are much suburbanin nature although these borough are developing urban centers of their own such as North York City Centre around Mel Lastman Square

1.2 Problem

For someone who is interested to open a business in Toronto or anywhere else, it is important to understand the characteristic of the location before decide which location that good for that business. To decide which location we want to open the business, we have to know what characteristic of the city that we want target. The characteristic of the city can be seen from the top 10 venue that the city has.

If we want to open a restaurant, the characteristic of the right target location is neighborhood that has restaurant as the top 10 venue. If we want to open construction, the characteristic of the right target location is neighborhood that has construction as the top 10 venue.

This project will explain about the characteristic of neighborhood in Toronto city to help choose which location is good to open a certain business in Toronto city

2. Data acquisition and cleaning

2.1 Data source

This project will use these 3 major data :

- List of Toronto's postal code, borough, neighborhoods which I scrapped from Toronto wikipedia :
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Coordinate (Latitude and Longitude) of each neighborhood in Toronto which I will get from geocoders
- List of venue from each neighborhood which I will get from Foursquare

2.2 Data cleaning

When I downloaded and scrapped the data from Toronto Wikipedia source and find the coordinate of each neighborhood, there are some missing values. Below is the attribute that has missing value and how I deal with this :

- There are many missing borough for some postal code. For this case, I delete the record
- There is a missing neighborhood from one borough. For this case, I copy the borough name to the missing neighborhood
- There are some neighborhood that doesn't found it's coordinate. For this case, I delete the record

After the data cleaning, there were 103 neighborhood which has coordinate

2.3 Feature selection

After the data cleaned, I got the list of Toronto city which has attribute postal code, borough, neighborhood, latitude, longitude. For each neighborhood, I download the venue that listed in that neighborhood from Foursquare. From that venue from Foursquare we will see the characteristic of Toronto city.

3. Methodology

For repository, I used GitHub to store the downloaded data, notebook, also the report.

To scrap data from wikipedia Toronto, I used BeautifulSoup from bs4 . And to get the coordinate of each neighborhood, I used Nominatim from geopy.geocoders.

Below is the top 5 result of scrapped data with it's coordinate :

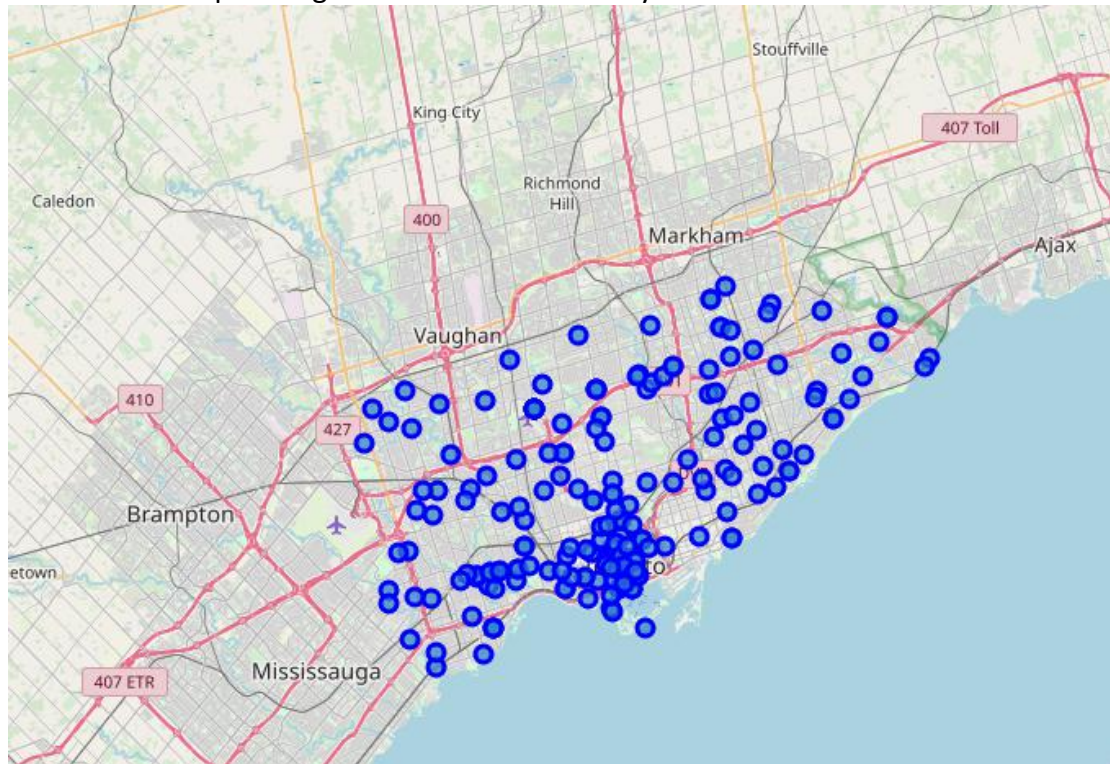
	Postcode	Borough	Neighbourhood	latitude	longitude
0	M5R	Central Toronto	Yorkville	43.671386	-79.390168
1	M3J	North York	York University	43.779242	-79.483559
2	M2P	North York	York Mills West	43.744039	-79.406657
3	M2L	North York	York Mills	43.744039	-79.406657
4	M4C	East York	Woodbine Heights	43.699930	-79.319132

For each neighborhood above, I will apply K-means as the model to cluster the neighborhood. I define 5 cluster for Toronto city. I also use mean of the number of venue as the value that will be applied with K-means algorithm

4. Result

4.1 Map of Toronto city

Below is the map of neighborhood in Toronto city :



Those dots are is the point of the neighborhood of Toronto city.

4.2 Top 10 venues for each neighborhood

Below is the top 10 venues for each neighborhood :

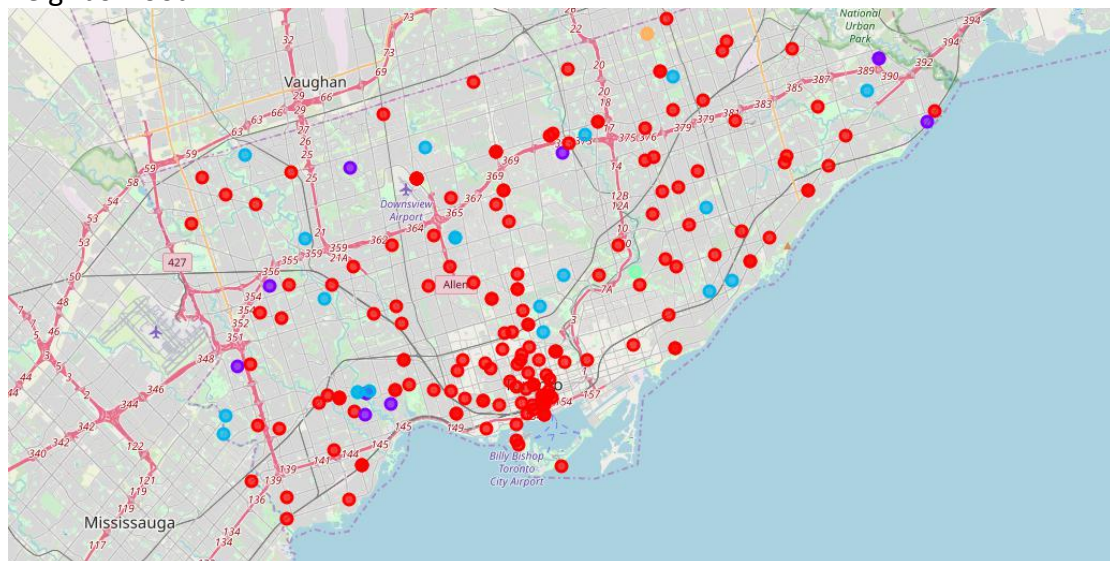
	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	0	Adelaide	Coffee Shop	Restaurant	Café	Gastropub	Cosmetics Shop	Breakfast Spot	American Restaurant	Hotel	Bookstore	BBQ Joint
1	0	Agincourt	Chinese Restaurant	Shopping Mall	Cantonese Restaurant	Restaurant	Asian Restaurant	Coffee Shop	Train Station	Korean Restaurant	Vietnamese Restaurant	Hong Kong Restaurant
2	0	Agincourt North	Chinese Restaurant	Bakery	Ice Cream Shop	Movie Theater	Taco Place	Beer Store	Liquor Store	Sporting Goods Shop	Frozen Yogurt Shop	Fried Chicken Joint
3	0	Albion Gardens	Grocery Store	Video Store	Sandwich Place	Liquor Store	Hardware Store	Coffee Shop	Fast Food Restaurant	Beer Store	Pharmacy	Fried Chicken Joint
4	0	Alderwood	Pizza Place	Sandwich Place	Coffee Shop	Gym	Pub	Pharmacy	Skating Rink	Electronics Store	Empanada Restaurant	Ethiopian Restaurant
5	2	Bathurst Manor	Playground	Convenience Store	Baseball Field	Park	Yoga Studio	Fast Food Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant
6	0	Bathurst Quay	Coffee Shop	Café	Park	Grocery Store	Pub	Sushi Restaurant	Bank	Sculpture Garden	Garden	Harbor / Marina
7	0	Bayview Village	Metro Station	Outdoor Supply Store	Sporting Goods Shop	Bubble Tea Shop	Burger Joint	Fish Market	Gas Station	Fast Food Restaurant	Bank	Breakfast Spot

I calculate the top 10 venues for each neighborhood in Toronto by calculate the mean of the frequency of the occurrence of the venue in each neighborhood. After that , for each neighborhood, it will show the top 10 that has the most frequency of the occurrence venue .

After all the neighborhood has it's top 10 most common venue, I apply the K-means algorithm with 5 cluster and get the cluster label for each neighborhood.

4.3 Map of Toronto segmentation

After apply the K-means algorithm, below is the visualization of the cluster of the neighborhood:



Red : cluster 0

Orange : cluster 1

Light green : cluster 2

Blue : cluster 3

Purple : cluster 4

From the picture, we can see that the majority of the neighborhood are in cluster 0. After that, most of the neighborhoods are in cluster 3, and then cluster 4. The least neighborhoods are in cluster 2.

4.4 Segmentation of the neighborhood

Cluster 0

Borough	Neighborhood	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
Central Toronto	Yorkville	0.0	Café	Restaurant	Italian Restaurant	Boutique	Coffee Shop	Hotel	Spa	Women's Store	Clothing Store	French Restaurant
North York	York University	0.0	Bakery	Gas Station	Martial Arts Dojo	Business Service	Yoga Studio	Filipino Restaurant	Exhibit	Falafel Restaurant	Farm	Farmers Market
North York	York Mills West	0.0	Coffee Shop	Restaurant	Japanese Restaurant	Pub	Thai Restaurant	Gym	Gym / Fitness Center	Business Service	Sandwich Place	Burrito Place
North York	York Mills	0.0	Coffee Shop	Restaurant	Japanese Restaurant	Pub	Thai Restaurant	Gym	Gym / Fitness Center	Business Service	Sandwich Place	Burrito Place
East York	Woodbine Heights	0.0	Skating Rink	Cosmetics Shop	Dance Studio	Athletics & Sports	Pharmacy	Bus Stop	Park	Event Space	Exhibit	Falafel Restaurant
East York	Woodbine Gardens	0.0	Coffee Shop	Bakery	Park	Empanada Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant	Farm	Farmers Market

Cluster 0 has 162 neighborhood . 83 % of the neighborhoods are in this cluster. Most of the neighborhood in cluster 0 has restaurant, cafe, bakery shop as it's main venue.

Cluster 1

Borough	Neighborhood	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
Scarborough	Upper Rouge	1.0	Park	Fast Food Restaurant	Yoga Studio	Festival	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant	Farm	Farmers Market
West Toronto	Swansea	1.0	Park	Bus Line	Pilates Studio	Yoga Studio	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant	Farm	Farmers Market
North York	Silver Hills	1.0	Middle Eastern Restaurant	Park	Festival	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant
Scarborough	Rouge	1.0	Park	Fast Food Restaurant	Yoga Studio	Festival	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant	Farm	Farmers Market
Scarborough	Port Union	1.0	Park	Yoga Studio	Festival	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant
Etobicoke	Old Mill South	1.0	Park	Metro Station	Spa	American Restaurant	River	Yoga Studio	Fast Food Restaurant	Ethiopian Restaurant	Event Space	Exhibit
Etobicoke	Old Mill North	1.0	Park	Metro Station	Spa	American Restaurant	River	Yoga Studio	Fast Food Restaurant	Ethiopian Restaurant	Event Space	Exhibit
North York	Northwood Park	1.0	Park	Baseball Field	Yoga Studio	Festival	Event Space	Exhibit	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant
Etobicoke	Kingsview Village	1.0	Park	Yoga Studio	Festival	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant

Cluster 1 has 11 neighborhood. 6% of the neighborhoods are in this cluster.

Most of the neighborhood in cluster 1 has park, fast food restaurant, yoga studio. festival and exhibit as it's main venue

Cluster 2

Borough	Neighborhood	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
Etobicoke	Westmount	2.0	Pizza Place	Park	Gas Station	Yoga Studio	Fast Food Restaurant	Empanada Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant
Downtown Toronto	Rosedale	2.0	Food Truck	Playground	Bike Trail	Park	Japanese Restaurant	Festival	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant
Etobicoke	Old Burnhamthorpe	2.0	Dog Run	Flower Shop	Gas Station	Park	Yoga Studio	Fast Food Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant
Scarborough	Oakridge	2.0	Ice Cream Shop	Convenience Store	Park	Dessert Shop	Bus Stop	Restaurant	Fast Food Restaurant	Ethiopian Restaurant	Event Space	Exhibit
Central Toronto	Moore Park	2.0	Gym	Convenience Store	Playground	Tennis Court	Farmers Market	Electronics Store	Empanada Restaurant	Ethiopian Restaurant	Event Space	Exhibit
Etobicoke	Markland Wood	2.0	Golf Course	Baseball Field	Park	Plano Bar	Yoga Studio	Fast Food Restaurant	Event Space	Exhibit	Falafel Restaurant	Farm
East York	Leaside	2.0	Sandwich Place	Japanese Restaurant	Convenience Store	Park	Business Service	Empanada Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant
North York	Lawrence Manor East	2.0	Electronics Store	Bank	Kids Store	Park	Doctor's Office	Fast Food Restaurant	Event Space	Exhibit	Falafel Restaurant	Farm
North York	Lawrence Manor	2.0	Electronics Store	Bank	Kids Store	Park	Doctor's Office	Fast Food Restaurant	Event Space	Exhibit	Falafel Restaurant	Farm
Etobicoke	Kingsway Park South West	2.0	Bus Line	Metro Station	Park	Bus Stop	Festival	Event Space	Exhibit	Falafel Restaurant	Farm	Farmers Market

Cluster 2 has 20 neighborhood. 20 % of the neighborhoods are in this cluster. Most neighborhood in cluster 2 has pizza place or sandwich place or ice cream shop, electronic shop, bank, and convenience store as it's main venue

Cluster 3

Borough	Neighborhood	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
East York	Parkview Hill	3.0	Construction & Landscaping	Yoga Studio	Filipino Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant

Cluster 3 has 1 neighborhood. 1 % of the neighborhoods are in this cluster Neighborhood in cluster 3 has construction & landscaping, yoga studio as it's main venue

Cluster 4

Borough	Neighborhood	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
Scarborough	Steeles East	4.0	Playground	Health & Beauty Service	Yoga Studio	Fast Food Restaurant	Empanada Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant	Farm
Etobicoke	South Steeles	4.0	Playground	Health & Beauty Service	Yoga Studio	Fast Food Restaurant	Empanada Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant	Farm

Cluster 4 has 2 neighborhood. 1 % of the neighborhoods are in this cluster

Neighborhood in cluster 4 has playground , health & beauty services, and yoga studio as it's main venue

5. Discussion

When I use K-mean clustering, I use 5 as the number of cluster. It is good that if the number of cluster is based on the elbow method.

6. Conclusion

Based on the cluster and segmentation for Toronto city above, I can recommend to the business owner if they want to open a new business / store, they can use the cluster as the reference to choose the location. For example :

- For the business owner that want to open a restaurant, it is recommended to open in neighborhood in cluster 0, because restaurant is one of the main venue in neighborhood cluster 0.
- For the business owner that want to open a fast food restaurant, it is recommended to open in neighborhood in cluster 1, because fast food restaurant is one of the main venue in neighborhood cluster 1.
- For the business owner that want to open a ice cream shop / pizza / sandwich place, it is recommended to open in neighborhood in cluster 2, because ice cream shop / pizza / sandwich is one of the main venue in neighborhood cluster 2.
- For the business owner that want to open a construction & landscaping business, it is recommended to open in neighborhood in cluster 3, because construction & landscaping business is one of the main venue in neighborhood cluster 3.
- For the business owner that want to open a health & beauty services, it is recommended to open in neighborhood in cluster 4, because health & beauty services business is one of the main venue in neighborhood cluster 4.

So, for the business owner that want to open a business can have a look which location that has main venue related with the business they want to open

7. Reference

<https://en.wikipedia.org/wiki/Toronto#Cityscape>

https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Toronto