Clustering Neighborhoods in Ottawa and Toronto, Canada

Valentín Irazabal

June 21, 2021

1. Introduction

1.1 Background

This project will intend to approach the resolution of a common problem for many Canadians: not knowing which neighborhood would suit them better when moving or travelling from Ottawa (the country's administrative capital) to Toronto (Ontario´s state capital as well as the country´s financial centre and most populated city). Since these two cities could be considered to be the country´s "top two", and given their proximity, Canadians living in one or the other may have to often visit the other city either for financial or administrative purposes. The aim of the project is to group similar neighborhoods in both cities into clusters in order to allow anyone to recognize which neighborhoods are more similar to their own in the other city.

1.2 Problem

Ottawa and Toronto are two large multicultural cities that can have very different neighborhoods, still, they are vastly interconnected since they are one the country´s administrative capital and the other the economic capita. The problem to solve will be in which Toronto neighborhood would someone coming from any Ottawa neighborhood fell more at home.

2. Data Acquisition and cleaning

2.1 Sources

I will use two main sources of data: first the Wikipedia pages on postal codes from Toronro and Ottawa, from where I will use BeautifulSoup to obtain Postal Codes of the different boroughs. This will then be very useful to use the geocoder API in order to obtain their geospatial coordinates. This information will be then useful to obtain the neighborhoods' venues by using the Foursquare API. Finally, once I obtain the venues I will be able to group the naighborhoods from both cities into clusters, in order to recognize which of them are similar to one another.

2.2 Data cleaning

As stated before, I will obtain all the data that I will need for the analysis in 3 steps. First, I will obtain the Boroughs and Neighborhoods' Postal Codes by scraping the Wikipedia page for both cities. After this, I will use the geocoder API to obtain their coordinates. Finally, Foursquare will allow me to obtain each neighborhoods nearby venues, which I will analyse in the next section.

From the whole Wikipedia page only postal codes, boroughs and neighborhood names were extracted into a clean dataframe. This was achieved thanks to the html scraper

BeautifulSoup. This was done twice, since both cities have two separate Wikipedia entries for this information (Ottawa: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_K , and Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M).

Coordinates for each neighborhood were obtained with the geocoder API and then a new dataframe was constructed by combining this information with the one previously obtained, this is to say, Postal Code, borough and neighborhood.

Finally, the venues obtained with Foursquare were grouped for each neighborhood and then only the 5 more frequent were selected to carry out the clustering. After completing this, Ottawa and Toronto dataframes containing the venues were merged into a single dataframe with all the neighborhoods to be clustered.

3. Methodology

3.1. K-means algorithm

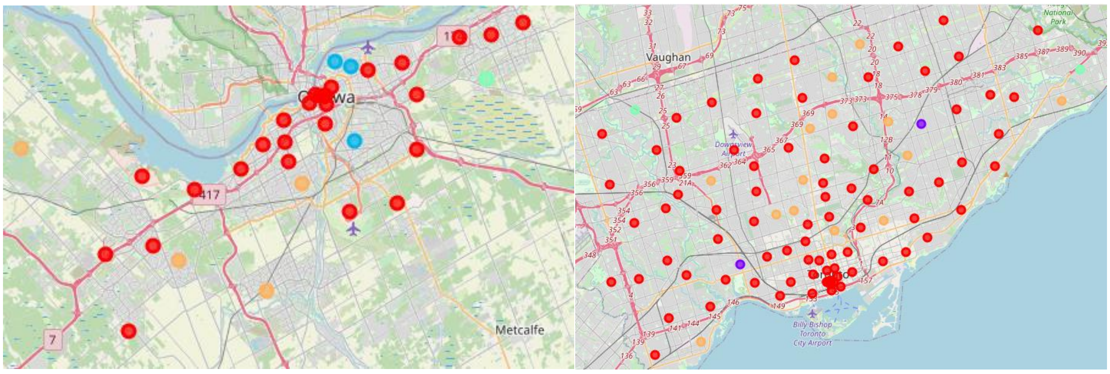
After the venues were grouped and the top 5 selected, the k-means algorithm was applied in order to group the different neighborhoods into the 5 different clusters. This is an unsupervised machine learning algorithm that groups the observations into groups that are the most similar within, and the most different when compared to other groups.

In the context of this application, the purpose of the algorithm is to use the most common venues from each neighborhood from one or the other city. This allows us to know which neighborhoods, even from different cities, are more alike; and which are different in terms of venues.

4. Results

4.1. Data visualization with Folium

After the clusters were obtained, the data was parsed into a Folium map using color code that allows any user to identify which color corresponds to their own neighborhood, and afterwards look to the neighborhoods in the other city which are alike theirs, with the corresponding color.



4.2. Viewing each cluster individually

Another output of the analysis is viewing each cluster and observing the most common venues in them. These dataframes are shown next.

First cluster:

Out[93]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Government of CanadaOttawa and Gatineau offices	0.0	Coffee Shop	Restaurant	Sandwich Place	Hotel	Juice Bar
1	Ottawa	0.0	Sporting Goods Shop	Fast Food Restaurant	Electronics Store	Gym	Gym / Fitness Center
2	Ottawa	0.0	Construction & Landscaping	Sandwich Place	Spa	Restaurant	Shopping Mall
3	Hawkesbury	0.0	Coffee Shop	Restaurant	Hotel	Sandwich Place	Clothing Store
5	PembrokeCentral and northern subdivisions	0.0	Coffee Shop	Restaurant	Hotel	Sandwich Place	Clothing Store
...	...	...	...	...	...	...	...
98	Etobicoke	0.0	Pool	Women's Store	Field	Ethiopian Restaurant	Event Space
99	Downtown Toronto	0.0	Coffee Shop	Sushi Restaurant	Japanese Restaurant	Restaurant	Gay Bar
100	East Toronto Business	0.0	Coffee Shop	Café	Asian Restaurant	Hotel	Restaurant
101	Etobicoke	0.0	Bank	Italian Restaurant	Sushi Restaurant	Coffee Shop	Fast Food Restaurant
102	Etobicoke	0.0	Bank	Eastern European Restaurant	Miscellaneous Shop	Food & Drink Shop	Mattress Store

137 rows × 7 columns

Seccond cluster:

Out[94]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
69	West Toronto	1.0	Convenience Store	Fish & Chips Shop	Event Space	Falafel Restaurant	Farm
71	Scarborough	1.0	Convenience Store	Auto Garage	Women's Store	Fish & Chips Shop	Event Space

Third cluster:

Out[95]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
20	Ottawa	2.0	Playground	Construction & Landscaping	Field	Ethiopian Restaurant	Event Space
36	Ottawa	2.0	Playground	Dog Run	Escape Room	Event Space	Falafel Restaurant
40	Ottawa	2.0	Playground	Fish & Chips Shop	Event Space	Falafel Restaurant	Farm

Fourth cluster:

Out[96]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
69	Ottawa	3.0	Home Service	Women's Store	Field	Ethiopian Restaurant	Event Space
12	Scarborough	3.0	Bar	Home Service	Fish & Chips Shop	Event Space	Falafel Restaurant
50	North York	3.0	Auto Garage	Home Service	Women's Store	Fish & Chips Shop	Event Space

Fifth cluster:

Out[97]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
16	Ottawa	4.0	Yoga Studio	Restaurant	Construction & Landscaping	Park	Chinese Restaurant
27	Ottawa	4.0	Convenience Store	Pizza Place	Bus Station	Park	Women's Store
41	Ottawa	4.0	Park	Pizza Place	American Restaurant	Restaurant	Gym
70	Ottawa	4.0	Construction & Landscaping	Women's Store	Fish & Chips Shop	Event Space	Falafel Restaurant
0	North York	4.0	Construction & Landscaping	Food & Drink Shop	Park	Fast Food Restaurant	Women's Store
1	North York	4.0	Park	German Restaurant	Nail Salon	Grocery Store	Women's Store
16	York	4.0	Trail	Park	Hockey Arena	Field	Grocery Store
18	Scarborough	4.0	Gym / Fitness Center	Construction & Landscaping	Park	Women's Store	Field
27	North York	4.0	Residential Building (Apartment / Condo)	Park	Entertainment Service	Ethiopian Restaurant	Event Space
35	East York/East Toronto	4.0	Intersection	Park	Home Service	Women's Store	Field
39	North York	4.0	Trail	Construction & Landscaping	Park	Women's Store	Event Space
45	North York	4.0	Park	Music Venue	Women's Store	Field	Event Space
49	North York	4.0	Bakery	Basketball Court	Park	Ethiopian Restaurant	Falafel Restaurant
66	North York	4.0	Convenience Store	Speakeasy	Coffee Shop	Park	Fast Food Restaurant
68	Central Toronto	4.0	Park	Women's Store	Escape Room	Event Space	Falafel Restaurant
73	Central Toronto	4.0	Playground	Gym Pool	Park	Entertainment Service	Ethiopian Restaurant
88	Etobicoke	4.0	Park	Yoga Studio	Skating Rink	Grocery Store	Convenience Store
91	Downtown Toronto	4.0	Park	Playground	Tennis Court	Gym / Fitness Center	Women's Store

5. Discussion

This can be a useful approach for travellers and people who move within these two cities. On top of this, the same methodology may be applied to any other city of the world, using the same Foursquare API. This has the potential to allow any traveller to feel “at home” when visiting any spot in the world by simply choosing wisely the neighborhood.

Of course this methodology carries some limitations, for example, the most amount of clusters were used until clusters with only one observation appeared, still, some of the clusters contain up to a hundred observations which is, from my point of view, too generic. Furthermore, there is a cluster unique to each city, meaning that people living in the neighborhoods contained in this clusters would be unable to find their corresponding in the other city.

6. Conclusion

In this study I obtained information from two different cities in Canada and was able to group them into clusters using their venues as criteria. This successfully fulfilled the purpose of obtaining similar and dissimilar neighborhoods in both cities.

Although this analysis may have some limitations, I consider this approach to be interesting to extrapolate to every city in the world, allowing anyone to know which neighborhood will they be the most familiar with when travelling or moving.