

Identifying Toronto Neighborhoods that are like Barrhaven in Ottawa

Temitope Akinwumi

May 2019

1. Introduction (Business Problem)

1.1 Background

Toronto is the commercial capital of Canada and the capital of the province of Ontario. It is a metropolitan city with booming economy that offers exciting and promising opportunities for both intra-Canadian migration and immigration into Canada from other parts of the world.

Migrating to Toronto is not an easy task, Toronto is a major commercial hub with lots of activities which make it difficult to find residential neighborhoods that are calm and without the hassles of living in a big city like Toronto. One of the major challenges faced by intending migrants to Toronto especially people moving from a calm city/neighborhood is finding good residential locations in Toronto that are similar to where they currently live.

This project focuses on migration from Ottawa to Toronto. The case study used is “migration from Barrhaven Ottawa to Toronto”. I attempt to apply machine learning to determine Toronto neighborhoods with similarities to Barrhaven in Ottawa – the model will also work perfectly well for any other cities of consideration in any part of the world.

This project aims to determine clusters of neighborhoods in Ottawa and compare them with clusters of neighborhoods in Toronto. We will determine neighborhoods of Ottawa that are in similar clusters as Toronto neighborhoods.

2. Data

There are two major data requirements for this project viz:

- a. List of neighborhoods in Toronto and Ottawa with corresponding longitude and latitude coordinates of each neighborhood. This data was not readily available, alternative source of getting the data was identified and utilized. List of neighborhoods in Ottawa and Toronto were scraped from the internet and corresponding coordinates (longitude & latitude) were obtained using the geopy geodecoder python library. This may not give a 100% accuracy but very good level of accuracy to achieve the desired objective of this project. Below is a sample of the neighborhood data after acquisition and wrangling:

	Neighbourhood	City	Latitude	Longitude
0	ByWard Market, Ottawa	Ottawa	45.427917	-75.692601
1	Centretown, Ottawa	Ottawa	45.416340	-75.694286
2	Centretown West, Ottawa	Ottawa	45.416340	-75.694286
3	Downtown, Ottawa	Ottawa	45.433464	-75.669203
4	The Glebe, Ottawa	Ottawa	45.402221	-75.691623

Fig: sample data set showing Ottawa neighborhoods and their coordinates

	Neighbourhood	City	Latitude	Longitude
0	Alexandra Park, Toronto	Toronto	43.650758	-79.404298
1	The Annex, Toronto	Toronto	43.670338	-79.407117
2	Baldwin Village, Toronto	Toronto	43.669818	-79.497172
3	Cabbagetown, Toronto	Toronto	43.664473	-79.366986
4	CityPlace, Toronto	Toronto	43.639248	-79.396387

Fig: sample data set showing Toronto neighborhoods and their coordinates

A total of 239 neighborhoods were eventually extracted for Ottawa and 191 neighborhoods were extracted for Toronto

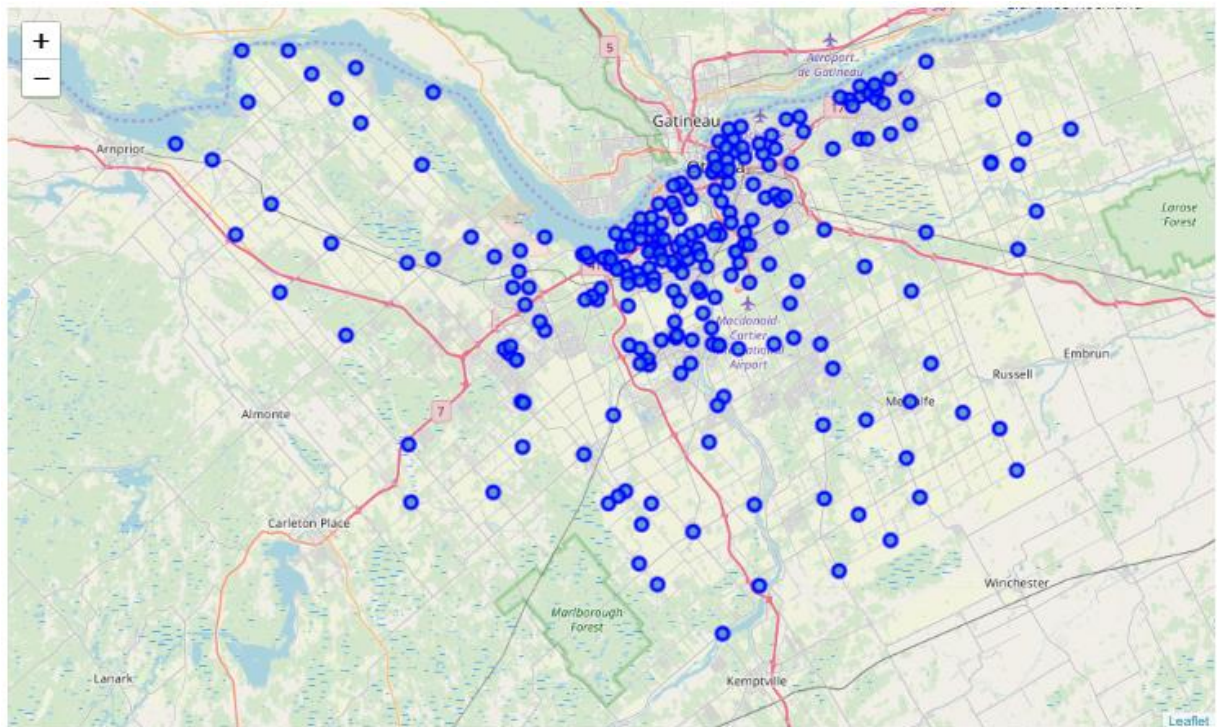


Fig: plot of the 239 extracted neighborhoods of Ottawa on the map of Ottawa

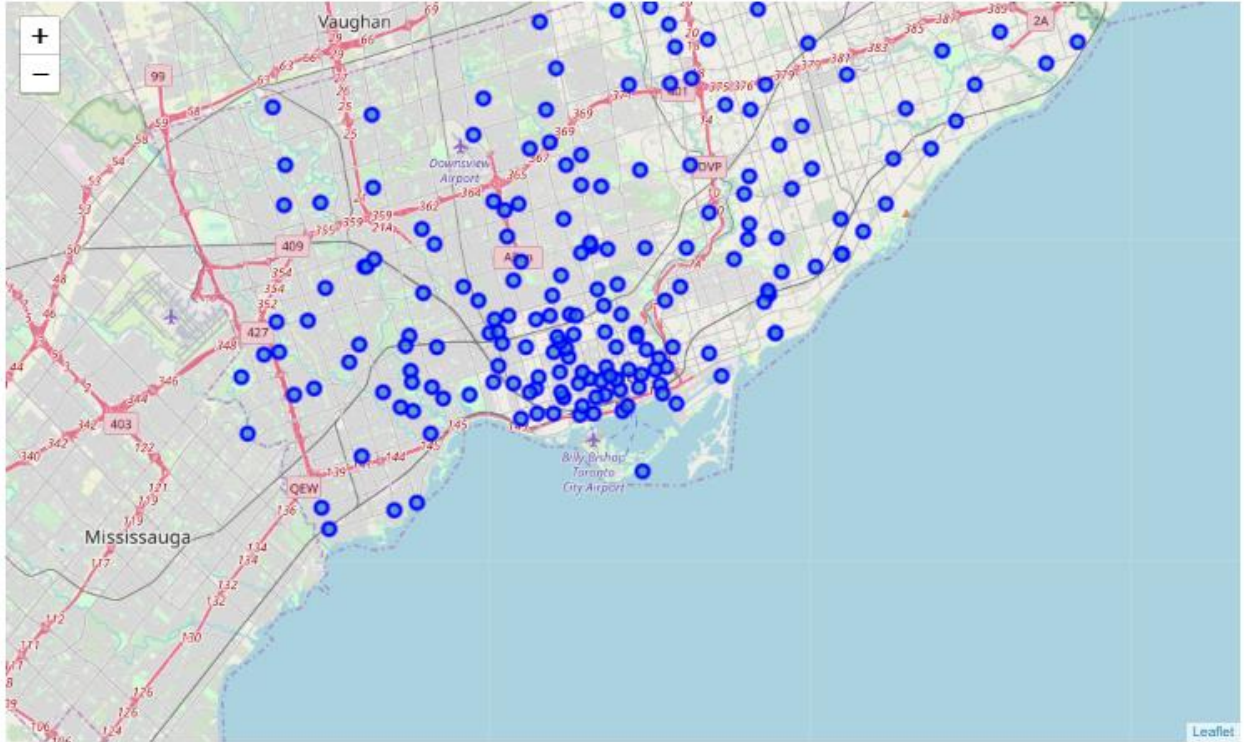


Fig: plot of the 191 extracted neighborhoods of Toronto on the map of Toronto

- b. List of top venues and amenities within 1000meters (1km) of each neighborhood in the two cities segmented by categories. This is necessary to understand the nature of each neighborhoods in the two cities. This data was obtained using the Foursquare location data API. Across the two cities, a total of 14,051 venues spread across 383 unique categories were returned from Foursquare API. These categories became the attributes for modeling and this was achieved using one hot encoding.

3. Methodology & Analysis

To identify neighborhoods in Toronto that are similar to Barrhaven in Ottawa, the key considerations from our data attributes are:

- a. Peculiarity of the environments in each neighborhood based on venues and amenities generated from Foursquare location data API
- b. Distance of candidate neighborhood from Downtown Toronto. We want to make sure that proximity from Downtown Toronto is a factor as well because most people migrating to Toronto would likely work in Downtown or would want to be as close as possible to Downtown while still living in a relaxed/calm neighborhood.

Based on this, the distance of each neighborhood from Downtown Toronto was calculated in kilometer using the formula below:

Let's say:

R (Radius of the earth) = 6373.0km
downtown_lat = latitude of Downtown Toronto
downtown_lon = longitude of Downtown Toronto
lat = latitude of candidate neighborhood
lon = longitude of candidate neighborhood

then:

dlat = radians(downtown_lat) – radians(lat)
dlon = radians(downtown_lon) – radians(lon)

$a = \sin(dlat/2)^2 + \cos(\text{radians}(\text{lat})) * \cos(\text{radians}(\text{downtown_lat})) * \sin(dlon/2)^2$
 $c = 2 * \text{aTan2}(\text{sqrt}(a), \text{sqrt}(1-a))$
distance (km) = R * c

This was applied to all neighborhoods in Toronto with an additional attribute created to show in one glance – the neighborhood name & it's distance to Downtown Toronto

	Neighbourhood	City	Latitude	Longitude	Dist_to_downtown_km	map_label
0	Alexandra Park, Toronto	Toronto	43.650758	-79.404298	2	Alexandra Park, Toronto, (Distance to Downtown...
1	The Annex, Toronto	Toronto	43.670338	-79.407117	3	The Annex, Toronto, (Distance to Downtown: 3km)
2	Baldwin Village, Toronto	Toronto	43.669818	-79.497172	10	Baldwin Village, Toronto, (Distance to Downtow...
3	Cabbagetown, Toronto	Toronto	43.664473	-79.366986	1	Cabbagetown, Toronto, (Distance to Downtown: 1km)
4	CityPlace, Toronto	Toronto	43.639248	-79.396387	2	CityPlace, Toronto, (Distance to Downtown: 2km)

Fig: sample output data showing Toronto neighborhoods and their distance to Downtown

3.1 Modeling (clustering)

K-Means algorithm was used to model the data to segment each neighborhood of Toronto and Ottawa into clusters. To achieve this, the neighborhood data of both Toronto and Ottawa were combined into a single data set and fed into the model as one unit. This resulted in the model classifying the combined data of neighborhoods into clusters

In order to select the optimal value of K for the actual segmentation, the elbow method was used by running the algorithm multiple times and plotting the value of k against cost for all the outputs generated. The algorithm was run 50 times and it was observed that cost continuously decreased with increase in k. However due to computation time & cost, 40 was selected as a value for k and the model was built to cluster the neighborhoods into 40 segments.

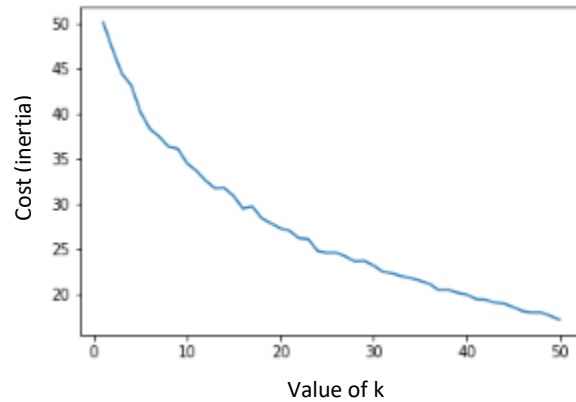


Fig: plot of k values vs cost to determine the optimal value of k for K-Means algorithm

4. Results & Discussion

It was observed that Barrhaven Ottawa belongs to cluster 25, there were 49 other neighborhoods in Ottawa that belong to this cluster as well.

On checking the neighborhoods in Toronto that belong to cluster 25, it was observed that 5 neighborhoods in Toronto belong to this cluster.

Port Lands, Toronto, (Distance to Downtown: 3km)
 Downsview, North York, Toronto, (Distance to Downtown: 12km)
 Milliken, Scarborough, Toronto, (Distance to Downtown: 20km)
 Brown's Corners (historical), Scarborough, Toronto, (Distance to Downtown: 21km)
 Highland Creek, Scarborough, Toronto, (Distance to Downtown: 22km)

Looking at the characteristics of these 5 neighborhoods in Toronto, we have:

map_label	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
Port Lands, Toronto, (Distance to Downtown: 3km)	Boat or Ferry	Circus	Athletics & Sports	Beach	Pool	Music Venue	Nightclub	Theater	Supermarket	Harbor / Marina
Downsview, North York, Toronto, (Distance to D...	Athletics & Sports	French Restaurant	Soccer Field	Gym / Fitness Center	Gym Pool	Fast Food Restaurant	Mediterranean Restaurant	Sandwich Place	Arts & Crafts Store	Basketball Court
Brown's Corners (historical), Scarborough, Tor...	Light Rail Station	Gym	Chinese Restaurant	Fast Food Restaurant	Sandwich Place	Automotive Shop	Vietnamese Restaurant	Train Station	Chocolate Shop	Indian Restaurant
Highland Creek, Scarborough, Toronto, (Distan...	Breakfast Spot	Italian Restaurant	Burger Joint	Neighborhood	Flower Shop	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant	Fish & Chips Shop
Milliken, Scarborough, Toronto, (Distance to D...	Chinese Restaurant	Japanese Restaurant	Bakery	Vegetarian / Vegan Restaurant	Asian Restaurant	Hong Kong Restaurant	Vietnamese Restaurant	Bubble Tea Shop	Noodle House	Bank

- a. These neighborhoods have a high concentration of sports & gym facilities (including swimming pools)
- b. There is also a variety of restaurants to pick from ranging from European, Asian and Middle East restaurants
- c. They also seem to have good transportation network considering the presence of train stations

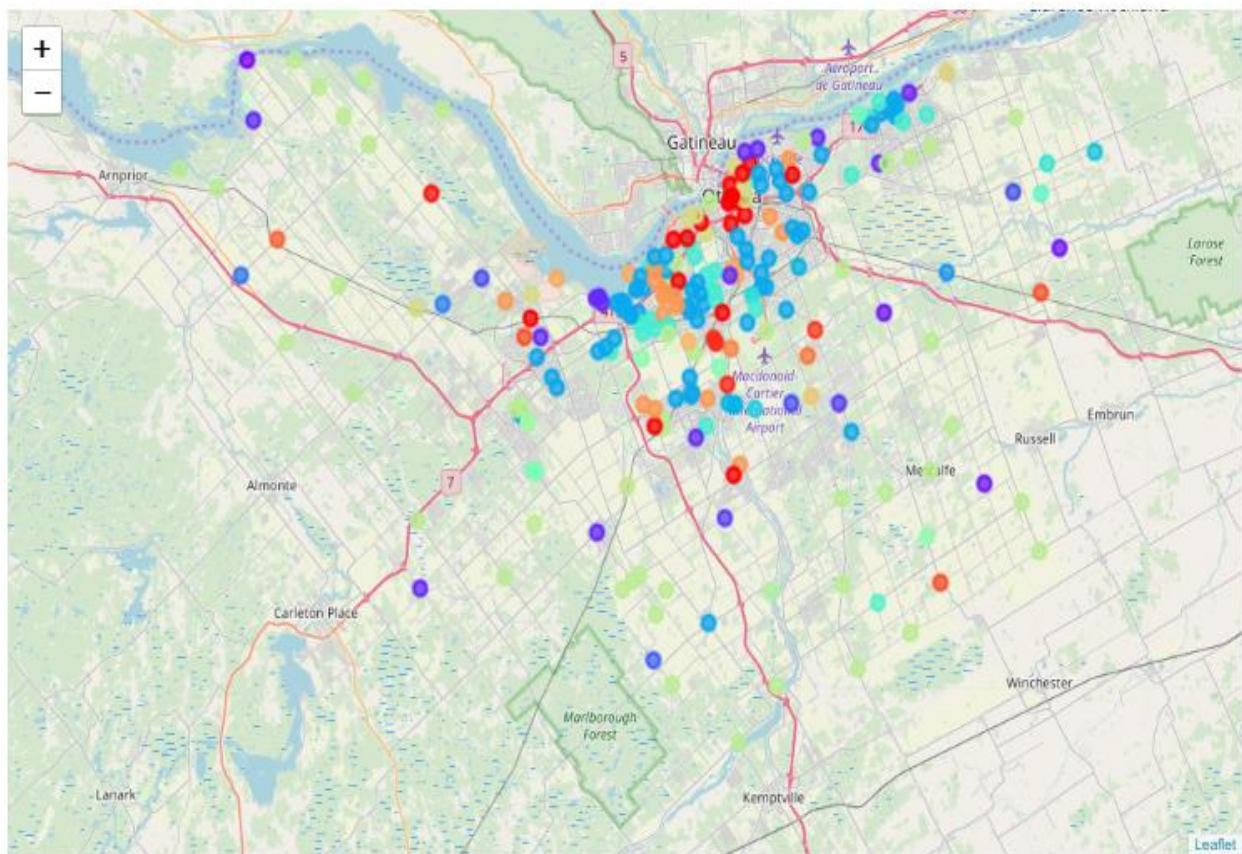


Fig: outlook of the clusters on Ottawa map

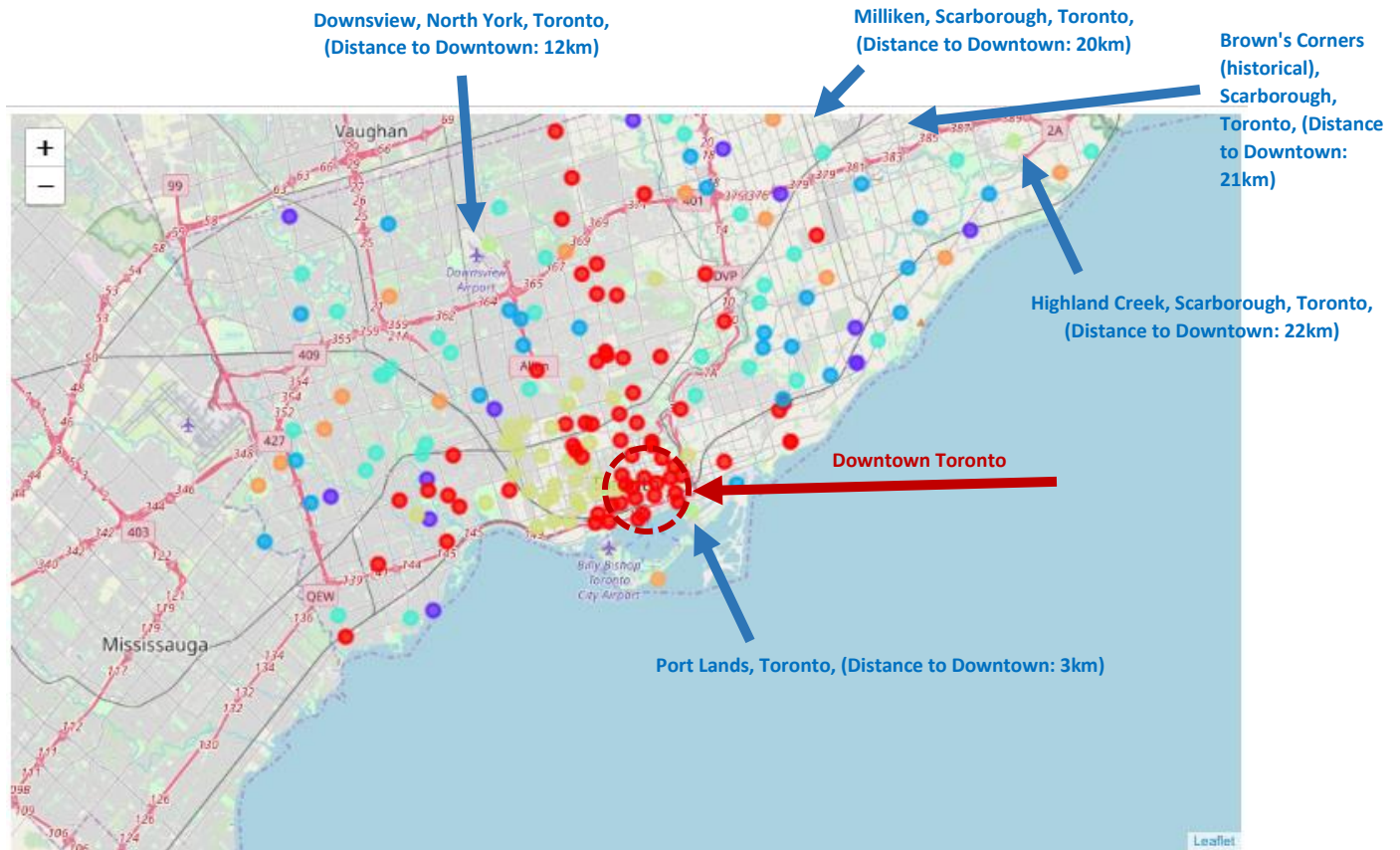


Fig: outlook of the clusters on Toronto map

5. Conclusion

Through this model/analysis, we have been able to identify 5 neighborhoods within Toronto that are like Barrhaven in Ottawa, they are:

- Port Lands, Toronto, (Distance to Downtown: 3km)
- Downsview, North York, Toronto, (Distance to Downtown: 12km)
- Milliken, Scarborough, Toronto, (Distance to Downtown: 20km)
- Brown's Corners (historical), Scarborough, Toronto, (Distance to Downtown: 21km)
- Highland Creek, Scarborough, Toronto, (Distance to Downtown: 22km)

Future works are encouraged to improve on the variables used for this analysis like: cost of accommodation in the neighborhoods, crime rate and security of the neighborhood, population per sq. meter etc. These can help to be more decisive in selection of a preferred location among the 5 candidate neighborhoods within the list.