Comparing New York and Toronto's Neighbourhoods

Erik Saphores Zaldivar

1. Introduction

Location is everything. There have been loads of studies on what made Starbucks successful, and a lot of experts claim that their key asset is their coffee shops locations.

Imagine you own a restaurant or a coffee house in a certain city and you're doing very well. The idea of opening a second restaurant or coffee house wonders your head but, where should you put it? If you're doing well with the one that you already have, maybe you just want to replicate that in the other side of town. If your current one is not doing well, maybe you should move a different neighbourhood. Data science is a good way to solve this question.

It's an interesting case study because it's valid for other situations. If you like your actual neighbourhood but you must move to another city, you can use the same method to determine which neighbourhood of another city is a good choice for you. This should be done with more than just location information (crime rates, average life span, contamination figures, etc.). This analysis will stay with just the Foursquare data for Toronto and New York, but in this case, we'll include the whole New York data (not only Manhattan).

2. Data

For this analysis, the New York Data from the previous labs (https://cocl.us/new_york_dataset) will be used. This data contains the names of the neighbourhoods from New York, their Boroughs, latitudes and longitudes.

For Toronto, there's two sources of data. First, there's the Wikipedia postal codes table (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). It contains the names, postal code and borough of every neighbourhood in Toronto. In second place, a csv file that has the latitude and longitude for every postal code in Toronto (https://cocl.us/Geospatial_data) was used.

And finally, we use the Foursquare API to get information about the most popular venues around within a 500 meters radius of every neighbourhood's coordinates. With this we can have a notion of what kind of neighbourhood is every one of them.

3. Methodology:

For the New York data, we must load the JSON file and extract the important features with a *for loop*.

The first set of Toronto's data was imported it by using the pandas library method *pd.read_html()*. Before using the data, it had to be cleaned. This because it has some rows with *NaN* values in it.

Then both dataframes were merged and end up having the same information that we had from New York of Toronto.

Then, the Foursquare API was used to retrieve the most popular venues in every neighbourhood. This is done by providing the coordinates of the neighbourhood and the Foursquare Developer Account's credentials in the API query.

With all the venues from the neighbourhoods, the next step is to use one hot encoding with the venues categories and use a machine learning algorithm called K-Means for being able to cluster the neighbourhoods in both Toronto and New York City. For that, the Python's library *Sci-Kit Learn* was used.

4. Results:

As a result, five clusters were obtained. Cluster 1 neighbourhoods were displayed in red, cluster 2 neighbourhoods were displayed in purple, cluster 3 neighbourhoods were displayed in blue, cluster 4 neighbourhoods were displayed in cyan/light-green and cluster 5 neighbourhoods were displayed in orange.

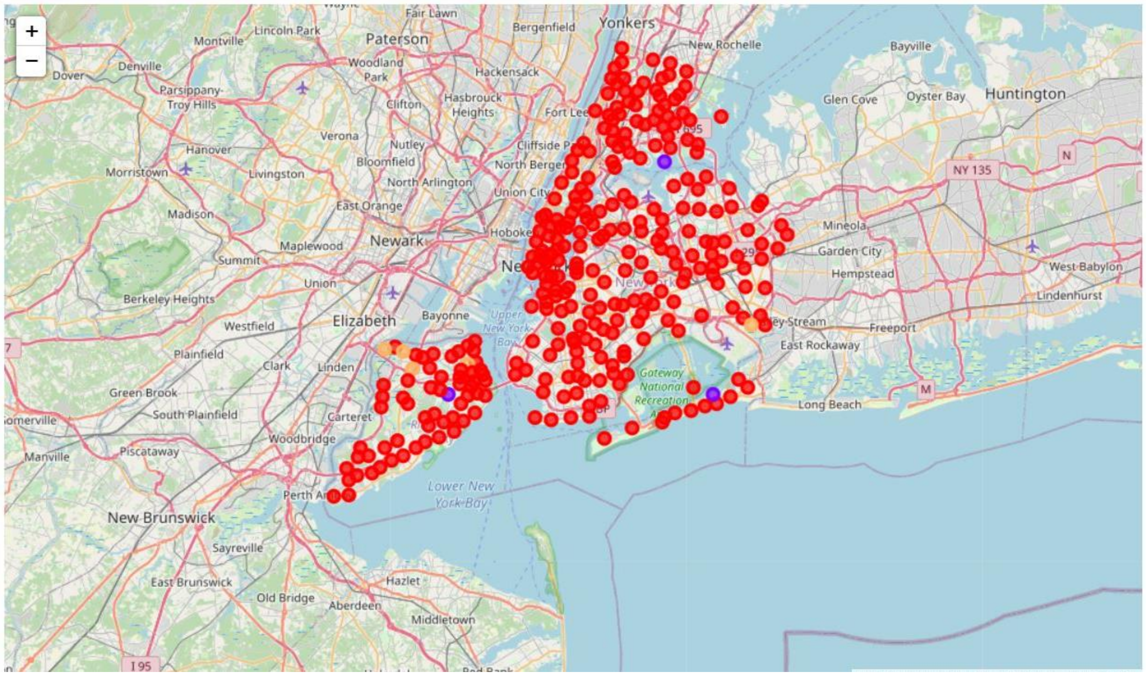

Figure 1: Neighbourhoods in New York.

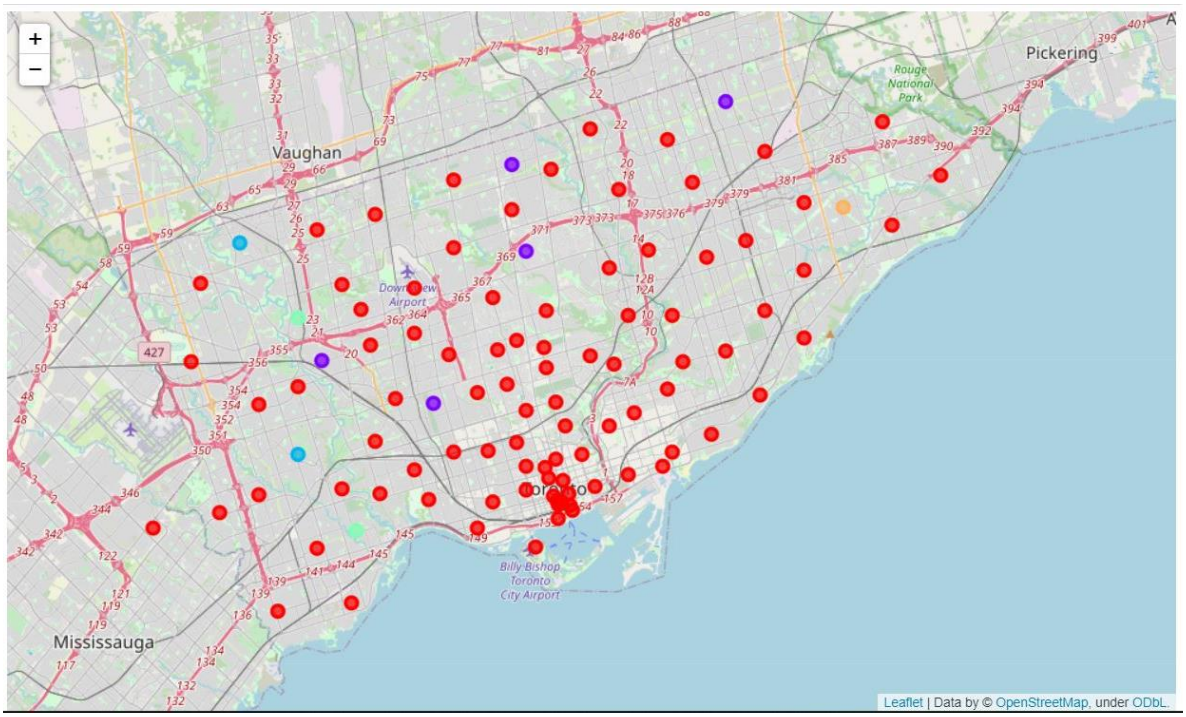

Figure 2: Neighbourhoods in Toronto.

	index	Neighbourhood	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	City
0	27	Clason Point	Park	Convenience Store	Bus Stop	Boat or Ferry	Grocery Store	Pool	South American Restaurant	Ethiopian Restaurant	Event Space	Farmers Market	New York
1	192	Somerville	Park	Yoga Studio	Egyptian Restaurant	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	New York
2	203	Todt Hill	Park	Yoga Studio	Egyptian Restaurant	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	New York
3	327	Caledonia-Fairbanks	Park	Women's Store	Pool	Fast Food Restaurant	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Toronto
4	358	Willowdale, Newtonbrook	Park	Yoga Studio	Egyptian Restaurant	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Toronto
5	370	Weston	Convenience Store	Park	Yoga Studio	Fast Food Restaurant	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Toronto
6	372	York Mills West	Park	Convenience Store	Yoga Studio	Fast Food Restaurant	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Toronto
7	391	Milliken, Agincourt North, Steeles East, L'Amo...	Park	Playground	Yoga Studio	Fast Food Restaurant	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Toronto

Figure 3: Cluster number 2.

	index	Neighbourhood	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	City
0	311	Islington Avenue, Humber Valley Village	Pizza Place	Yoga Studio	Farmers Market	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Toronto
1	356	Humber Summit	Pizza Place	Yoga Studio	Farmers Market	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Toronto

Figure 4: Cluster number 3.

	index	Neighbourhood	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	City
0	363	Humberlea, Emery	Baseball Field	Yoga Studio	Electronics Store	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Factory	Toronto
1	407	Old Mill South, King's Mill Park, Sunnylea, Hu...	Baseball Field	Business Service	Yoga Studio	Field	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Toronto

Figure 5: Cluster number 4.

	index	Neighbourhood	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	City
0	193	Brookville	Recording Studio	Deli / Bodega	Fast Food Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	New York
1	202	Grymes Hill	Deli / Bodega	Dog Run	Yoga Studio	Field	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	New York
2	226	Graniteville	Boat or Ferry	Grocery Store	Yoga Studio	Field	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	New York
3	227	Arlington	Deli / Bodega	Bus Stop	Coffee Shop	Construction & Landscaping	Boat or Ferry	Eye Doctor	Farmers Market	Farm	Falafel Restaurant	Factory	New York
4	257	Howland Hook	Boat or Ferry	Yoga Studio	Field	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Factory	New York
5	328	Woburn	Coffee Shop	Korean Restaurant	Fast Food Restaurant	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Toronto

Figure 6: Cluster number 5.

Cluster number 1 wasn't displayed in detail because it's too large. We still can see it in the maps from Figures 1 and 2 as the neighbourhoods in red.

5. Discussion:

Based on the results, some conclusions can be made:

- Clason Point, Somerville and Todt Hill from New York are similar to Caledonia-Fairbanks, Willowdale, Newtonbrook, Weston, York Mills West, Milliken, Agincourt North, Steeles East and L'Amoreaux East from Toronto.

- Islington Avenue, Humber Valley Village and Humber Summit from Toronto are similar.

- Humberlea, Emery, Old Mill South, King's Mill Park, Sunnylea Humber Bay, Mimico NE, The Queensway East, Royal York South East and Kingsway Park South East from Toronto are similar.

- Brookville, Grymes Hill, Graniteville, Arlington and Howland Hook from New York are similar to Woburn from Toronto.

- And the rest of the neighbourhoods from New York are similar to the rest of the neighbourhoods from Toronto.

Based on the most common venues of each cluster, some cluster characteristics can be drawn:

- Cluster 1 neighbourhoods have banks, bus stations, pizza places and pharmacies, among others.
- Cluster 2 neighbourhoods have parks, yoga studios, convenience stores, event spaces and restaurants.
- Cluster 3 neighbourhoods have pizza places, yoga studios, farmer's markets, electronics stores, restaurants, entertainment services, event spaces and exhibits.
- Cluster 4 neighbourhoods have baseball fields, yoga studios, restaurants, eye doctors and exhibits.
- Cluster 5 neighbourhoods have deli/bodegas, restaurants, entertainment services and boats or ferries.

6. Conclusion:

It can be concluded that there are a lot of neighbourhoods from both cities that are similar. And also, there's a lot of neighbourhoods that are very similar, even if they are not close from each other. This is a demonstration of how powerful and useful machine learning algorithms can be and how universally they can be used.

It's important to clarify that the similarity that's been found it's based on the main venues categories from each neighbourhood. It would be interesting to conduct the same study but adding new features like population density, crime rates, human development index, men/women ratio, among others.