Predicting what neighbourhoods are similar the one you live

Fernando Grillo

August, 05

1 Introduction

This project will be an applied “Battle of the Neighbourhoods” proposed by the Coursera Capstone Project for a person moving to a very different country.

1.1 Background

Globalization has been affecting people’s jobs in plenty of ways. Most people will not always live in the city they were born and are likely to move to another town because of job offers, for example, or better quality of life in general. This involves a variety of characteristics, for example: cost of living, climate, safety, political similarities, and others.

1.2 The Problem

Although moving to other towns or countries is a becoming quite common because of globalization, people might not be prepared to live in a different culture.

This may cause a variety of adaptation problems, leading, on some cases, to return to the previous town. A problem that, not only affects the economy in general, but the quality of life of residents and immigrants.

Using Data Science, we may help him to solve this problem. One way is to find the most similar neighbourhoods in the new town to the one the person currently lives. There are many variables that might be used to find these. In this project, the variables chosen are the frequency of venue types in the neighbourhoods, and will be presented as the following example story:

Let´s pretend someone received a quite good job offer in another country, a very different one compared to where he lives, and is considering moving.

Since the person lives in Brazil, and Toronto is on the other side of the world, it would be easier to adapt to a neighbourhood that is similar to the one he lives (Praia da Costa).

1.3 Interest

Everyone who might use this method to find the easiest neighbourhood to adapt on the new city.

Companies that hire employees from other parts of the world, and others.

2 Data Used

The used data is from Wikipedia’s page containing the list of postal codes in Canada.

“https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M”

Some tools were used to scrape the data and clean it and transform it into a data frame.

After, only the neighbourhoods in Toronto were selected. And lastly, a csv file containing the latitude and longitude of these neighbourhoods, obtained through Coursera’s course, was used to add this information to the frame.

	Neighbourhood	Latitude	Longitude
37	The Beaches	43.676357	-79.293031
41	The Danforth West, Riverdale	43.679557	-79.352188
42	The Beaches West, India Bazaar	43.668999	-79.315572
43	Studio District	43.659526	-79.340923
44	Lawrence Park	43.728020	-79.388790
45	Davisville North	43.712751	-79.390197
46	North Toronto West	43.715383	-79.405678
47	Davisville	43.704324	-79.388790
48	Moore Park, Summerhill East	43.689574	-79.383160

Part of the frame was shown above.

Also, the Foursquare database was used to find the 40 most common venues in each neighbourhood. For example let's pretend the most frequent venue of my neighbourhood is a coffee shop. Neighbourhoods in Toronto with this characteristic may be in the same cluster as the original neighbourhood.

3 Methodology

The methodology used consists in:

1) obtaining the name and coordinates of the neighbourhoods in Toronto (using Wikipedia) and the ones of my neighbourhood.

2) Using the Foursquare data to obtain the 40 most common venues in each neighbourhood

3) Using the K-nearest mean method to cluster the neighbourhoods based on the similarity of the venues’ frequency

4) Identifying which cluster Praia da Costa (the current neighbourhood) is in and list the similar neighbourhoods of the new town (Toronto).

The first part was done on the “Data Used” section. Later, I appended Praia da Costa’s latitude and longitude to the data frame.

In the second part, the Foursquare data was used. So, a data frame was made containing all the 40 venues in each neighbourhood. Part of this data frame is shown as it follows:

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Danforth West, Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant
5	The Danforth West, Riverdale	43.679557	-79.352188	Dolce Gelato	43.677773	-79.351187	Ice Cream Shop
6	The Danforth West, Riverdale	43.679557	-79.352188	MenEssentials	43.677820	-79.351265	Cosmetics Shop
7	The Danforth West, Riverdale	43.679557	-79.352188	Cafe Fiorentina	43.677743	-79.350115	Italian Restaurant
8	The Danforth West, Riverdale	43.679557	-79.352188	La Diperie	43.677530	-79.352295	Ice Cream Shop
9	The Danforth West, Riverdale	43.679557	-79.352188	Messini Authentic Gyros	43.677827	-79.350569	Greek Restaurant
10	The Danforth West, Riverdale	43.679557	-79.352188	Louis Cifer Brew Works	43.677663	-79.351313	Brewery

Later, each type of venue was counted and added to another data frame. That was used to find the frequency of each type in each neighbourhood.

	Neighbourhood	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...
0	Adelaide, King, Richmond	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.030000	0.000000	...
1	Berczy Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...
2	Brockton, Exhibition Place, Parkdale Village	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...
3	Business Reply Mail Processing Centre 969 Eastern	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...
4	CN Tower, Bathurst Quay, Island airport, Harbo...	0.000000	0.058824	0.058824	0.058824	0.117647	0.176471	0.117647	0.000000	0.000000	...
5	Cabbagetown, St. James Town	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.021277	0.000000	...

Then the most common venue was found for each neighbourhood:

```
----Adelaide, King, Richmond---
                venue  freq
0         Coffee Shop  0.07
1                Café  0.05
2          Steakhouse  0.04
3                 Bar  0.04
4      Thai Restaurant  0.04
5                 Gym  0.03
6       Cosmetics Shop  0.03
7      Breakfast Spot  0.03
8     Asian Restaurant  0.03
9           Restaurant  0.03
10  American Restaurant  0.03
11              Hotel  0.03
12        Pizza Place  0.02
13       Burger Joint  0.02
14       Concert Hall  0.02
15    Sushi Restaurant  0.02
16           Gastropub  0.02
17              Bakery  0.02
18        Salad Place  0.02
19              Plaza  0.01
20      Ice Cream Shop  0.01
21         Poke Place  0.01
22    Department Store  0.01
23       Deli / Bodega  0.01
24  Gym / Fitness Center  0.01
25        Record Shop  0.01
26          Jazz Club  0.01
27   Italian Restaurant  0.01
28   Salon / Barbershop  0.01
29  Colombian Restaurant  0.01
```

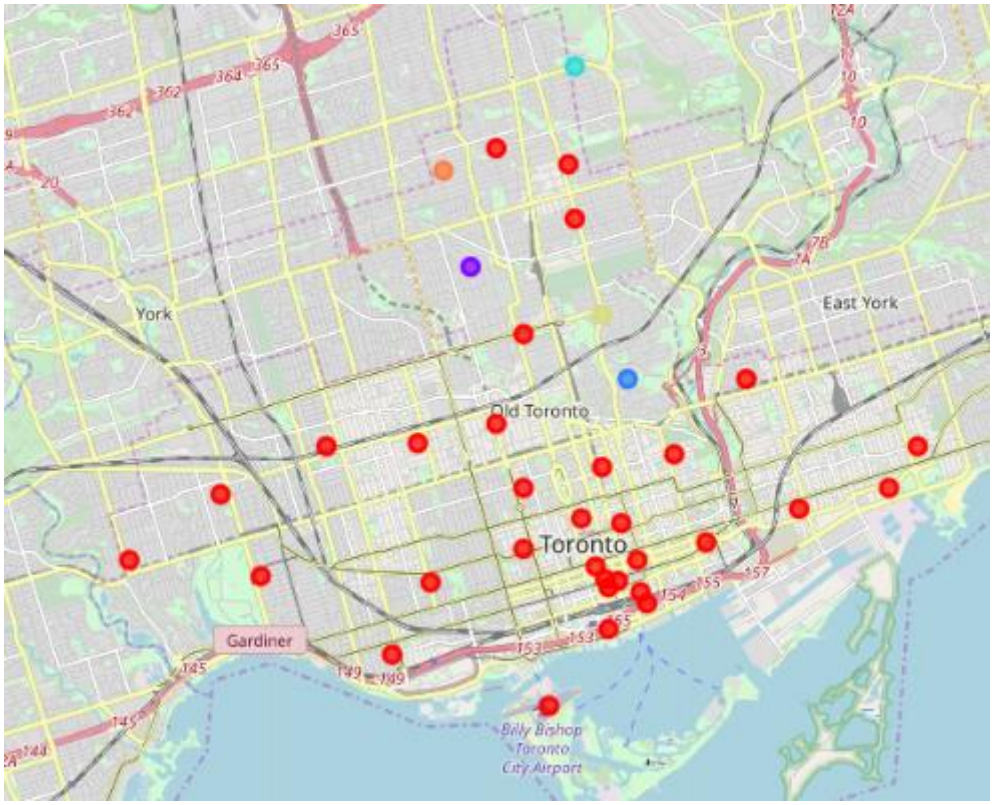
A data frame of the 40 most common venues in frequency was created, and the k-nearest mean was easily applied to cluster the neighbourhoods.

7 clusters were used.

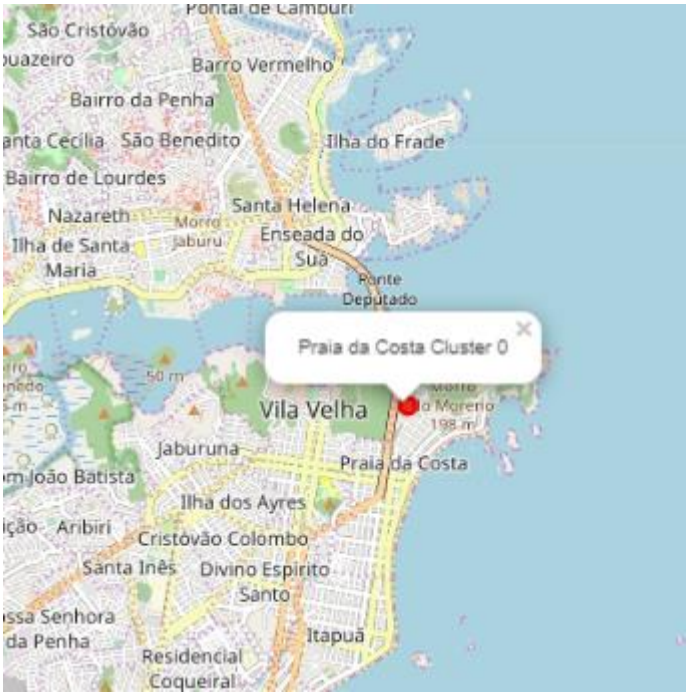
See an example of the frame:

	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
37	The Beaches	43.676357	-79.293031	4	Trail	Health Food Store
41	The Danforth West, Riverdale	43.679557	-79.352188	0	Greek Restaurant	Coffee Shop
42	The Beaches West, India Bazaar	43.668999	-79.315572	0	Park	Sandwich Place
43	Studio District	43.659526	-79.340923	0	Café	Coffee Shop

A map was generated, then:



The Brazilian neighbourhood was identified in cluster number 0:



This way, it is finally possible to obtain the similar neighbourhoods, that are:

```
The similar neighbourhoods are:
 0                          The Danforth West, Riverdale
1                       The Beaches West, India Bazaar
2                                      Studio District
3                                    Davisville North
4                                  North Toronto West
5                                           Davisville
6     Deer Park, Forest Hill SE, Rathnelly, South Hi...
7                           Cabbagetown, St. James Town
8                                  Church and Wellesley
9                             Harbourfront, Regent Park
10                             Ryerson, Garden District
11                                     St. James Town
12                                        Berczy Park
13                                 Central Bay Street
14                           Adelaide, King, Richmond
15    Harbourfront East, Toronto Islands, Union Station
16            Design Exchange, Toronto Dominion Centre
17                      Commerce Court, Victoria Hotel
18                 The Annex, North Midtown, Yorkville
19                      Harbord, University of Toronto
20            Chinatown, Grange Park, Kensington Market
21    CN Tower, Bathurst Quay, Island airport, Harbo...
22                    Stn A PO Boxes 25 The Esplanade
23              First Canadian Place, Underground city
24                                             Christie
25                        Dovercourt Village, Dufferin
26                           Little Portugal, Trinity
27         Brockton, Exhibition Place, Parkdale Village
28                       High Park, The Junction South
29                               Parkdale, Roncesvalles
30                                  Runnymede, Swansea
31    Business Reply Mail Processing Centre 969 Eastern
```

4 Results

There were 31 similar neighbourhoods in Toronto. A quite expressive amount considering two countries that are so far apart and have very different cultures.

This result is probably caused by the fact that Praia da Costa is one of the major neighbourhoods in the Brazilian town Vila Velha, close to the State Capital, Vitória.

People in bigger towns of Brazil have been more adapted to globalization, making the lifestyle similar to the one in global cities as Toronto. This reflects on the venues of the neighbourhoods, that were used to compare the similarity of them.

5 Discussions

The obtained results go toward the fact that the globalization is really making lifestyle in countries that have very few contacts. This reflects on the neighbourhoods' venues and, although adaptation problems exist, they tend to reduce in time, since the growing similarity of towns and cultures.

6 Conclusion

The project was useful to determine the similar neighbourhoods of two different countries, with different cultures and far apart from each other.

This method may be useful for people facing immigration situations and shows that, although there are differences between these two countries, the number of similar neighbourhoods are big enough to promote easy adaptation. This is probably mainly due to the globalization phenomenon.