

NY vs. Toronto:

1. Introduction:

1.1. Background

Different cities in the world have many different venues that define the culture of the cities. Despite the dissimilarities, it is somewhat possible to group together the similar kind of neighborhoods in different cities. You can segment the different venues in the neighborhood according to venue category, and then group neighborhoods together that incorporate similar kind of neighborhoods.

1.2. Problem

Since people make changes in life and sometimes have to move from there city to another it is very hard to decide where to move to. Having the similar neighborhoods grouped together it can help make decision when people consider moving out of a city to another.

1.3. Interest

I think many people will gain from this problem being solved. Since when they will be promoted from work for example and will have to move to a different city they will be able to make there move easier since they will move to the most similar city.

2. Data:

We will use 2 datasets.

The first consists of New York's different neighborhoods and their respective geometric coordinates.

https://cocl.us/new_york_dataset

The second dataset consists of Toronto's different borough and their respective postcodes.

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

Clean it and make the data frame comfortable to read.

The first data set after cleaning will look like:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

The second data set will look like:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Heights, Lawrence Manor	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park	43.662301	-79.389494

Next we will use Foursquare for each data frame to find the venues in the neighborhoods. We will make a new df for each city with the venues clustered into groups.

1.

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentin Restaurant
0	Allerton	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	Annadale	0.000000	0.00	0.000000	0.000000	0.076923	0.000000	0.000000	0.000000	0.000000
2	Arden Heights	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	Arlington	0.000000	0.00	0.000000	0.000000	0.142857	0.000000	0.000000	0.000000	0.000000
4	Arrochar	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

2.

	Neighborhood	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery
0	Adelaide, King, Richmond	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.02	0.0	0.0	0
1	Agincourt	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0
2	Agincourt North, L'Amoreaux East, Milliken, S...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0
3	Albion Gardens, Beaumont Heights, Humbergate,...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0
4	Alderswood, Long Branch	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0

Last step will to merge both data frames and cluster the same venues.

	Unnamed: 0	Neighborhood	Accessories Store	Afghan Restaurant	American Restaurant	Antique Shop	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint
0	0	Allerton	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0
1	1	Annadale	0.0	0.0	0.076923	0.0	0.0	0.0	0.0	0.000000	0.0
2	2	Arden Heights	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0
3	3	Arlington	0.0	0.0	0.142857	0.0	0.0	0.0	0.0	0.000000	0.0
4	4	Arrochar	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.045455	0.0