

The battle of the neighborhoods – Relocation between Toronto and New York City

by Ricardo F Reategui

Introduction

Historically, Toronto has attracted immigrants from all over the world. Today, there are about 2.7 million inhabitants within 200 square miles. More than 50 percent of residents belong to a visible minority population group, and over 200 distinct ethnic origins are represented among its inhabitants. While most Torontonians speak English as their primary language, over 160 languages are spoken in the city.

New York City, with about 8.5 million people within 300 square miles is more populated and denser than Toronto. It has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world.

Skilled workers, businessmen, science and technology professionals, entrepreneurs, artists, move around the world seeking better opportunities for themselves and contribute to the communities they settled in. Two of those communities are Toronto and New York City.

When market changes and new opportunities in Toronto arise for one of these individuals and their families who had settled in a neighborhood from New York City for a while, they probably will look for a place in Toronto with similar characteristics, in terms of housing, recreation, diversity, transportation, that form part of their way of life. The same scenario could be applied to a family moving from Toronto to New York City.

This project intends to leverage the use of Foursquare location data to explore the cities of Toronto and New York City for relocation purposes. It is possible to demonstrate that Foursquare, together with segmentation of the data generated, can provide the necessary information to individuals and families who want to relocate to a neighborhood with similar characteristics as their original place.

Data

The Foursquare API and the k-means clustering package will be used to segment and cluster the neighborhoods in the city of New York and Toronto.

In order to segment the neighborhoods of Toronto and New York City and explore them, we will use datasets that contains boroughs, the neighborhoods that exist in each borough, as well as the coordinates of each neighborhood.

The dataset for New York City is available on the web (https://geo.nyu.edu/catalog/nyu_2451_34572). This dataset was already uploaded to the server and used during the module “Segmenting and Clustering Neighborhoods in New York City”. The dataset was retrieved as a JSON file and transformed into a dataframe. Table 1 showed only 5 neighborhoods from The Bronx borough. A total of 5 boroughs and 306 neighborhoods were found in the dataframe.

Table 1. First 5 rows of a pandas dataframe with the coordinates of New York City neighborhoods

	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 dataset for Toronto was obtained by reading a Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) directly into a pandas dataframe. The Wikipedia page contained the postal codes and the boroughs of the Toronto neighborhoods. Table 2 displays only 5 neighborhoods with their postal codes.

Table 2. Example of 5 Toronto neighborhoods with their postal codes

	Postal code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park / Harbourfront
3	M6A	North York	Lawrence Manor / Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government

A Geospatial data of the Toronto postal codes and their coordinates were already available as a CSV file (http://cocl.us/Geospatial_data). This file was read into a pandas dataframe. Table 3 shows a portion of the data with the coordinates of 5 Toronto postal codes.

Table 3. Example with the coordinates of 5 Toronto postal codes

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Joining the dataframes (Tables 2 and Table 3) generate the Toronto dataframe that will be used for this project. Table 4 shows the first 5 rows of this dataframe.

Table 4. First 5 rows of a pandas dataframe with the coordinates of Toronto neighborhoods

	Postal code	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	Regent Park / Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor / Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government	43.662301	-79.389494

The Foursquare API will be used to get the most common venue categories in each neighborhood of New York City and Toronto. The neighborhoods will be grouped into clusters using the *k*-means clustering algorithm, and the clusters will be visualized in the maps of New York City and Toronto using the Folium library.

The clusters of New York City and Toronto will be compared, and the similarities of the most common venues will be discussed.