Comparing the cities of Toronto and New York regarding the quantity of Health Centers

Alejandro Somarriba

December 5, 2019

1. Introduction/Business Problem

1.1 Background

New York City, located in the United States, and the city of Toronto, located in Canada, are their respective countries' financial capitals. Both countries are very developed and are among the world leaders. However, despite how similar they may seem, they have some key differences. For one, healthcare in Canada is socialized, while healthcare in the US relies on private health insurance. Given that Toronto and New York are their countries' financial capitals, it would be interesting to compare them using the proximity of their hospitals to each of their neighborhoods.

1.2 Problem

The objective of this project was to visually compare the way neighborhoods are spread throughout both cities, as well as determining which city had more hospitals closer to its neighborhoods. I chose to find out how many hospitals were located within 1 kilometer of the cities' neighborhoods. Then, based on the data I found, I would be able to determine which city is more prepared to provide healthcare services. Furthermore, the project also has the intention of categorizing the neighborhoods based on the most common hospital-related venue by category, such as Hospital, Medical Center, Hospital Ward, Emergency Room, and Urgent Care Center.

1.3 Audience

This project may be of interest to someone who lives on either city or intends to move to one of them, as they may want to see which healthcare options are closest to where they are. It may also be of interest to healthcare providers who may want to set up their own services close to neighborhoods where other venues are too far away.

2. Data

2.1 Data Sources

Parts of the data used in this project are lists of neighborhoods for Toronto and New York. The list used for the Toronto neighborhoods was scraped from a table from

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, which lists all of the neighborhoods in Toronto sorted by their postal codes. This list was cleaned to eliminate rows where values were not assigned, and then complemented by getting the coordinates of the rest of the neighborhoods. The list used for the New York neighborhoods was obtained from https://cocl.us/new_york_dataset, which is a JSON file that contains a list of New York neighborhoods, as well as their coordinates. The rest of the data used is mostly location data provided by the FourSquare API in order to find the closest hospital-related venues to the neighborhoods.

All this data will help me make markers on maps using the Folium Python library, which will allow me to see the distribution of venues for both cities. It will also help me when I use a clustering algorithm to separate the neighborhoods into distinct classes.

2.2 Data Cleaning

In order to clean the data, I first got rid of the useless rows from the Toronto neighborhoods list. Then, I used the GeoPy Python library to obtain the coordinates for all the neighborhoods in Toronto; this required having to eliminate some rows which yielded no results, as well as renaming other rows to get them. I also got rid of the 'Postal Code' and 'Borough' columns, as I only required the 'Neighborhood' column. Then, I cleaned the New York neighborhoods list by getting rid of the 'Borough' column, as I would only need the 'Neighborhood', 'Latitude', and 'Longitude' columns.

More data cleaning occurred further along in the process, once I had found the nearby venues I was looking for.

2.3 Data Examples

For example, a row from the Toronto neighborhoods table would have the neighborhood's name, latitude, and longitude, e.g.:

Neighborhood	Latitude	Longitude
Parkwoods	43.7588	-79.3202

A row from the New York neighborhoods table would be similar to the one from the Toronto neighborhoods table, with the neighborhood's name, latitude, and longitude, e.g.:

Neighborhood	Latitude	Longitude
Wakefield	40.894705	-73.847201

An example of the rows obtained from the FourSquare API would have the neighborhood's name, neighborhood's latitude, neighborhood's longitude, venue's name, venue's latitude, venue's longitude, and venue's category, e.g.:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
South of Bloor	43.667662	-79.394698	CAMH Russell Site	43.659131	-79.399194	Medical Center