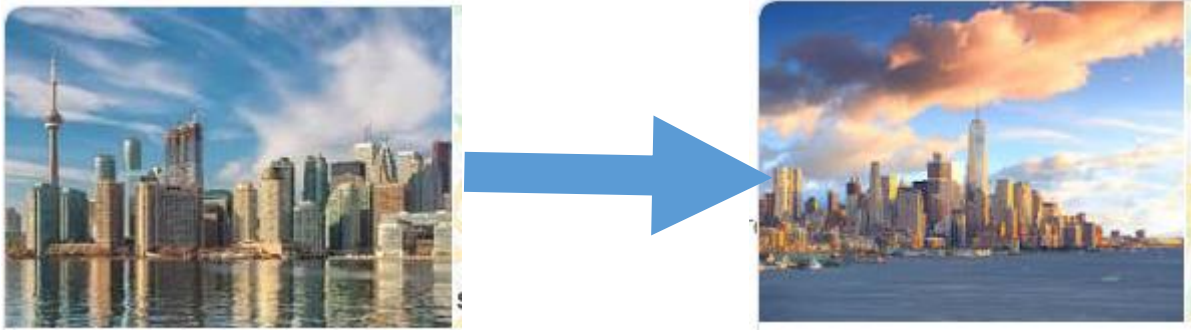# Coursera Capstone Project – The Battle of Neighborhoods

IBM Data Science Professional Certificate

Migrating from Toronto to New York

By: Sameer Mahajan

March 2020

Introduction

When you decide to live in a neighborhood, it is mainly because of your personal choices. These choices are for amenities or venues that are in your neighborhood. These venues could be like gourmet restaurants, pharmacies, parks, schools and so on.

Business problem

Let's say you live on the west side of city of Toronto in Canada say Weston neighborhood in York Borough with postal code of M9N to be more specific. You love your neighborhood based on above mentioned factors. You receive a job offer from a great company in city of New York with great career

prospect. Because of location change you will have to move if you decide to accept the offer. Wouldn't it be great if you are able to decide the neighborhoods in New York that are exactly like your current neighborhood in Toronto? You can then pick a neighborhood that is exactly like your current neighborhood and also very close to your new workplace to move to.

Target audience of this project

This project is particularly useful for anyone moving from Toronto to New York City. It can also be used by recruiting agencies to motivate, guide and / or assist employees moving across these cities. The same technique can be applied for move between any two cities by leveraging similar data from those cities.

Data

We need

- List of neighborhoods in these two cities
- Geographical coordinates of latitude and longitude information of these neighborhoods and
- Information about venues (categories like restaurant, park, hospital etc.) around these neighborhoods

For list of neighborhoods we will either get readymade data like https://ibm.box.com/shared/static/fbpwbovar7lf8p5sgddm06cgipa2rxpe.json for New York City or extract it from Wikipedia pages like https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M for Toronto. Here is some sample data for Toronto:

	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Rouge,Malvern,
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union,
2	M1E	Scarborough	Guildwood,Morningside,West Hill,
3	M1G	Scarborough	Woburn,
4	M1H	Scarborough	Cedarbrae,

We will get the geographical coordinates of these neighborhoods using python's geocoder package. Here is some sample data for Toronto:

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711

	Postal Code	Latitude	Longitude
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

We will use Foursquare API (https://developer.foursquare.com/) to get information about venues around these neighborhoods. It has data of over 105 million places. It provides many categories of the venue data. Here is some sample data for Toronto:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rouge,Malvern,	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge,Malvern,	43.806686	-79.194353	Interprovincial Group	43.805630	-79.200378	Print Shop
2	Highland Creek,Rouge Hill,Port Union,	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
3	Guildwood,Morningside,West Hill,	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
4	Guildwood,Morningside,West Hill,	43.763573	-79.188711	Marina Spa	43.766000	-79.191000	Spa

We will leverage category of venues to identify similar neighborhoods.

You can refer to notebook https://github.com/sameermahajan/Coursera_Capstone/blob/master/Neighborhoods%20in%20Toronto.ipynb for details on how it can be done for Toronto.

Processing

This project applies various data science techniques like

- Web scraping to gather data from Wikipedia page
- Working with APIs (Foursquare)
- Data cleaning

- Data wrangling
- Machine learning (k means clustering)
- Map visualization (folium)

To arrive at the solution. The next section will elaborate on methodology applied, steps taken and how we arrived at the solution.