

IBM Coursera Data Science Professional Certificate
Capstone Project: The Battle of the Neighborhoods
March 5, 2020
Marcelo Guimarães



**Comparing Neighbourhoods in New York
City and Toronto**

Introduction

When you have to move from your home, it is always difficult to find the right neighborhood to live. Throughout my life I have moved among different neighborhoods in the same city, from one city to another inside the same state, from one state to another, and even from one county to another, including countries in different continents. Everytime I was moving from one place to another, the same question arises: where in this new city will I find the right place to live? This problem could be minimized if we were able to compare the neighborhoods in different cities and make a list of the best candidates, or at least the neighborhoods that are similar to the one we like.

What if we could create a recommendation system for neighbourhoods? We will try to create such system by gathering information about the neighbourhoods using the Foursquare API. The recommendation system will be based on our preferred venues and their ratings, its output will be a list of possible candidates. It is not a complete solution, but it is a start.

In this project we will consider a client that lives in Toronto, specifically in the neighbourhood called Little Portugal. The client will move to New York City and would like to know which neighbourhoods would be similar to the current one.

All the details of the process are well documented and described in the Jupyter Notebook associated with this final report.

Data

We will collect data from different sources in order to understand the distribution of venues in New York City and Toronto, and start to search for good areas to live.

For New York City, we will collect information about each neighbourhood and borough from the website: https://cocl.us/new_york_dataset

It returns a JSON file, which will be open using Pandas and read into a Dataframe. The first 5 rows for this dataframe is displayed in Table 1.

Table 1: New York City Neighbourhood Information.

Borough	Neighbourhood	Latitude	Longitude
Bronx	Wakefield	40.89	-73.85
Bronx	Co-op City	40.87	-73.83
Bronx	Eastchester	40.89	-73.83
Bronx	Fieldston	40.9	-73.91
Bronx	Riverdale	40.89	-73.91

For Toronto we will use the Wikipedia page containing information about the postcodes and neighbourhoods of the city. We will use the Wikipedia library for Python in order to extract the table with the information important to us. The latitude and longitude for each neighbourhood will be extracted from the CSV file: `Geospatial_Coordinates.csv`. We can see the first 5 rows of the final dataframe below.

Table 2: Toronto Neighbourhood Information

Borough	Neighbourhood	Latitude	Longitude
North York	Parkwoods	43.75	-79.33
North York	Victoria Village	43.73	-79.32
Downtown Toronto	Harbourfront	43.65	-79.36
North York	Lawrence Heights, Lawrence Manor	43.72	-79.46
Downtown Toronto	Queen's Park	43.66	-79.39

We will use the Foursquare API to retrieve relevant data for New York City and Toronto and organize it into pandas Dataframes.

We limited our search in 100 venues/neighbourhood and a search radius of 500 meters, centred in the latitude and longitude of the neighbourhood.

For New York City we have 306 neighbourhoods, distributed in 5 boroughs and our query using the Foursquare API returned 10278 venues, with 429 unique categories.

The first 5 rows of the dataframe with the venues data are shown in Table 3.

Table 3: Example of venues returned by the Foursquare API. New York Venues Dataframe.

Neighbourhood	Neigh. Latitude	Neigh. Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Wakefield	40.89	-73.85	Lollipops Gelato	40.89	-73.85	Dessert Shop
Wakefield	40.89	-73.85	Rite Aid	40.9	-73.84	Pharmacy
Wakefield	40.89	-73.85	Carvel Ice Cream	40.89	-73.85	Ice Cream Shop
Wakefield	40.89	-73.85	Walgreens	40.9	-73.84	Pharmacy
Wakefield	40.89	-73.85	Dunkin'	40.89	-73.85	Donut Shop

For Toronto we have 103 neighbourhoods, distributed in 10 boroughs and a total of 2228 venues, as shown in the table below. These 2228 venues are distributed in 267 unique categories. An example of the Dataframe is shown in Table 4.

Table 4: Example of venues returned by the Foursquare API. Toronto Venues Dataframe.

Neighbourhood	Neigh. Latitude	Neigh. Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Parkwoods	43.75	-79.33	Brookbanks Park	43.75	-79.33	Park
Parkwoods	43.75	-79.33	Variety Store	43.75	-79.33	Food & Drink Shop
Victoria Village	43.73	-79.32	Victoria Village Arena	43.72	-79.32	Hockey Arena
Victoria Village	43.73	-79.32	Tim Hortons	43.73	-79.31	Coffee Shop
Victoria Village	43.73	-79.32	Portugril	43.73	-79.31	Portuguese Restaurant