



# Coursera Capstone Project

## The Battle of the Neighborhoods (Week 1)

### FOURTH ASSIGNMENT

Anna Fabiola Cedeno Maselli | Data Science Professional Certificate | 02/02/2021

## 1. A description of the problem and a discussion of the background.

How is it known Toronto, the capital of the province of Ontario, is a major Canadian city along Lake Ontario's northwestern shore. It's a dynamic metropolis with a core of skyscrapers, all dwarfed by the iconic, free-standing CN Tower.

On the other hand New York City comprises 5 boroughs sitting where the Hudson River meets the Atlantic Ocean. At its core is Manhattan, a densely populated borough that's among the world's major commercial, financial and cultural centers. Its iconic sites include skyscrapers such as the Empire State Building and sprawling Central Park.

Let's say that an Italian firm located in Texas City, United States, decides to move its headquarters to New York City or Toronto, Canada. They don't know which city is the best for them. The firm wants to know venues and neighborhoods to locate the company. The project will analyze the neighborhoods between New York City and Toronto, understand the differences and similarities, group the neighborhoods, visualize these groups on a map, and provide the best decision.

Also for this project the target audience are investors interested in moving their headquarters to the best city and may need an objective advice to choose the location for the company and its employees.

## 2. A description of the data and how it will be used to solve the problem.

The dataset used for this project and analyse the information are:

**a.** The websites that collect the information about Toronto and New York borough and their locations. The pages are: [https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572) and [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M).

**b.** The Foursquare API, that collect venues and their categories for each location within a radius 700 meters.

**c.** The Geopy and Folium libraries to get the coordinates of every location of Toronto and New York City. The page is [https://cocl.us/Geospatial\\_data](https://cocl.us/Geospatial_data)

**d.** Cluster venues of each neighborhood using k-means algorithm and analyze the top 10 most common venue in each cluste.

**e.** Visualize clusters on the map, thus showing the best locations.

**f.** The pandas library used for data manipulation and analysis.

**g.** The Numpy library used to work with arrays.

- h.** The Requests used to send HTTP/1.1 requests.
- i.** Matplotlib library used to create static, animated, and interactive visualizations in Python.
- j.** The json used to transfer data as text that can be sent over a network.
- k.** The Urllib used to fetch URLs (Uniform Resource Locators).
- m.** Bs4 library used to pulling data out of HTML and XML files.