**Capstone Project - The Battle of the Neighborhoods**

**Applied Data Science Capstone by IBM/Coursera**

**Introduction: Business Problem**

In this project I will try to explore, segment, and cluster the neighborhoods in the city of Toronto which is not readily available and is useful for tourism purpose and also to use public transportation region vise. This can give better idea to the tourists that which areas they can visit in a day that are close to each according to their needs.

**Data**

For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto.

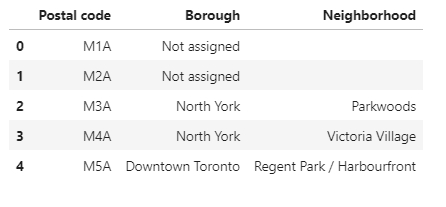
I will scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas dataframe so that it is in a structured format. The location in every neighborhood will be obtained using **Foursquare API**

**Methodology**

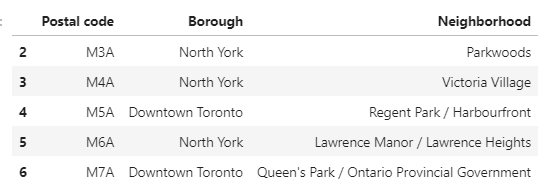
* First import all the required libraries
* Then import requests library and beautiful soup package
* Then read the required data from wiki page and save it in the form of dataframe table
* Then cleaning the dataframe table and only process the cells that have an assigned borough. Ignore cells with a borough that is not assigned
* Then combine rows with the same postal area
* Get geo spatial data
* Create combined dataframe including the geospatial data
* Explore and cluster the neighborhoods in Toronto
* Generating maps to visualize the clusters
* Create a map of Toronto with neighborhoods superimposed on top
* Segment and cluster only the neighborhood in downtown toronto

**Results**

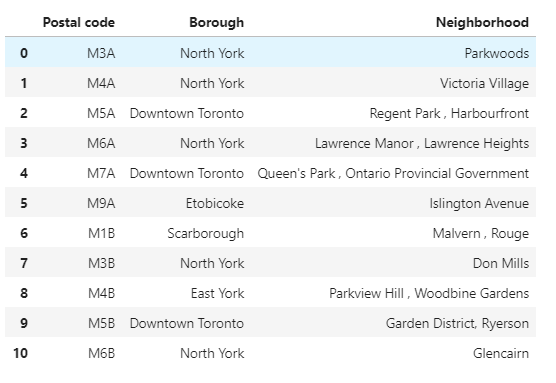
* After reading the required data from wiki page into the dataframe

****

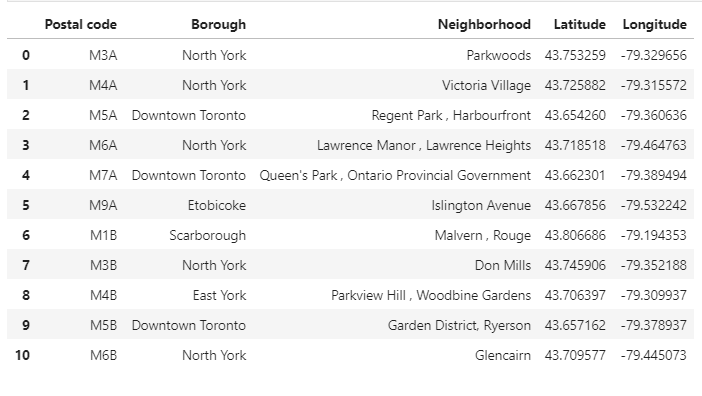
* After cleaning the data



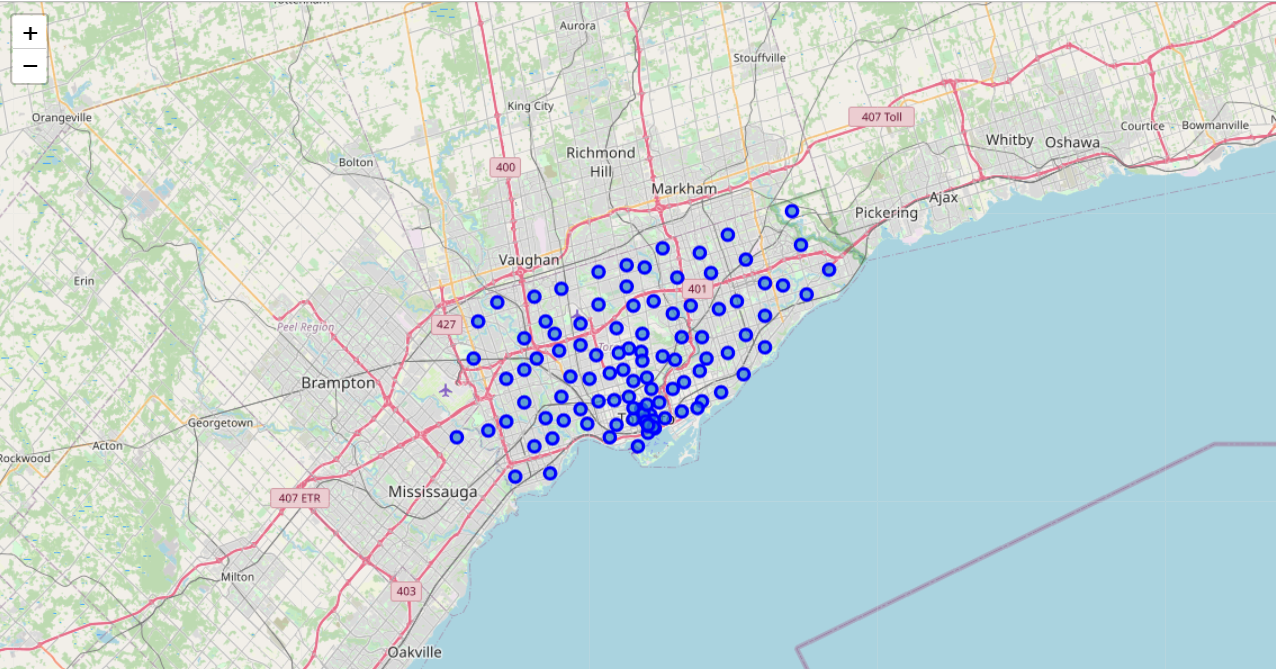
* After combining rows with same postal code



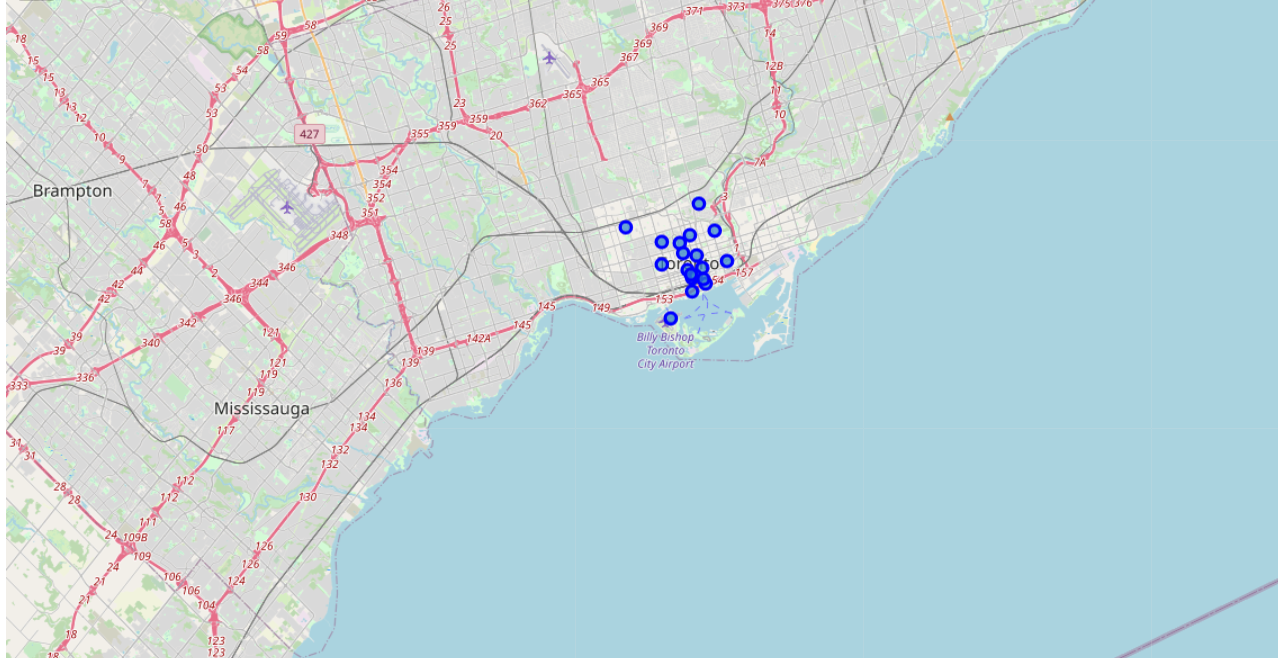
* After combining the dataframe with geospatial data



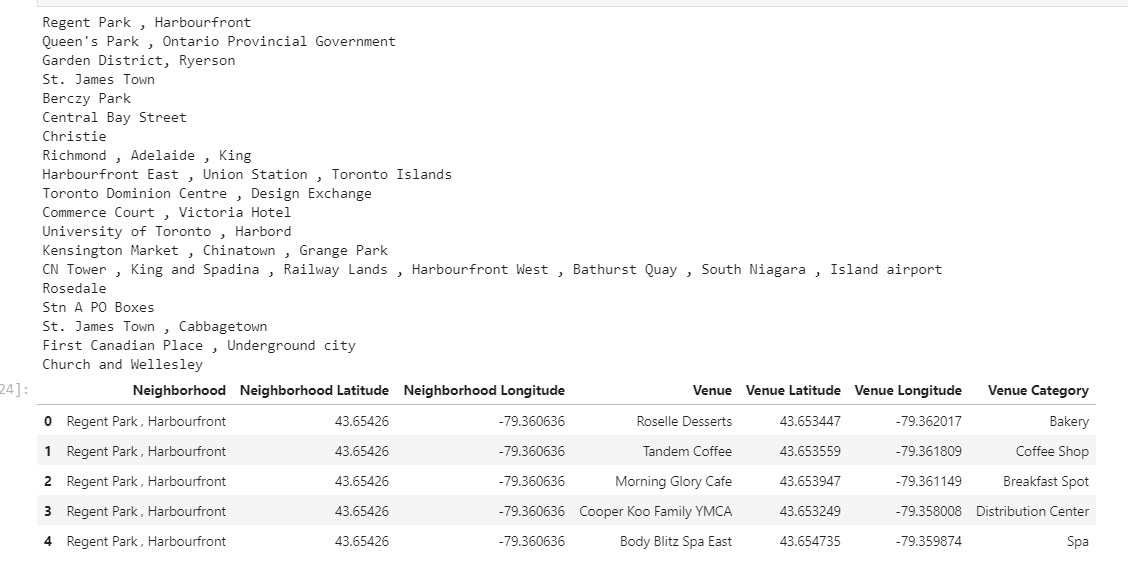
* After creating a map of Toronto on top of neighborhood



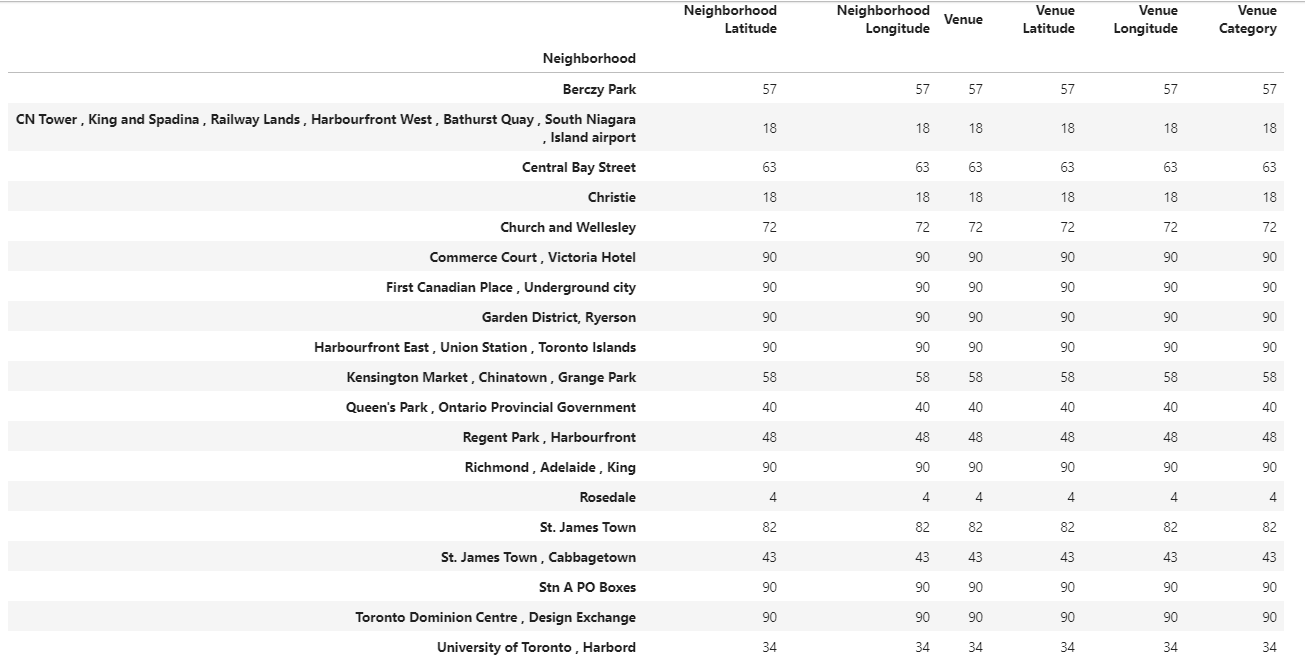
* Visualization of downtown Toronto



* Analyzing Neighborhood



* Checking how many venues were returned for each neighborhood



#### Running and visualizing clusters

