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

Toronto, Ontario, incorporated as a city in 1834, population 2,731,571 (2016 census), 2,615,060 (2011 census). Toronto is Ontario’s capital city, Canada’s largest municipality and the fourth largest city in North America. It is made up of the former cities of Toronto, North York, Scarborough, York and Etobicoke, and the former borough of East York. The city is home to a large immigrant population, and is a national and international hub for finance, communications and cultural life.

**A Brief History**

The city officially became Toronto on March 6, 1834, but its roots are much older. In the early 1600s a Frenchman named Etienne Brûlé was sent into the not-yet-Canadian wilderness by the famous explorer Samuel de Champlain to see what he could discover. He found the river and portage routes from the St. Lawrence to Lake Huron, possibly Lakes Superior and Michigan, and eventually Lake Ontario. The native Huron peoples had known this area between the Humber and Don rivers for centuries—and had long called it "Toronto," believed to mean "meeting place."

A bustling village called Teiaiagon grew up here, which became the site of a French trading post. After the British won the Seven Years' War, the trading post was renamed York in 1793. More than 40 years later the city again took the name Toronto. Following an unsuccessful American invasion in 1812, several devastating fires, and a rebellion in 1837, there was a slow but steady increase in the population of white Anglo-Saxon Protestants leading into the 20th century. Since World War II, Toronto has attracted residents from all over the world. Unlike the American "melting pot," Toronto is more of a "tossed salad" of diverse ethnic groups.

**A.2. Problem Description**:

In the context of this Capstone project, let us imagine that you live on the west side of the city of Toronto in Canada. And you receive a job offer from a great company on the other side of the city with great career prospects. However, given the far distance from your current place you unfortunately must move if you decide to accept the offer.

Wouldn't it be great if you are able to determine neighborhoods on the other side of the city that are the same as your current neighborhood, and if not perhaps similar neighborhoods that are at least closer to your new job?

**A.3. Objective**

The aim of this report is to study and analyze the neighborhoods of Toronto city and group them into similar clusters and, to analyze those clusters to gather meaningful information. That information can be used to find out neighborhoods that are same as your current neighborhood or at least similar. This report would be useful for people who are interested in relocating to North York

**B. Data Description:**

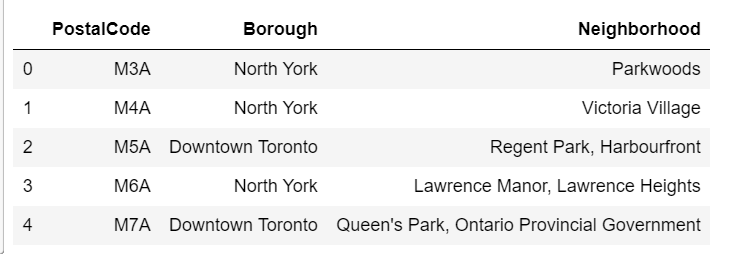
To consider the objective stated above, we can list the below data sources used for the analysis.

1. **Toronto Neighborhood Data:** The following Wikipedia page was scraped to pull out the necessary information: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
2. b) Coordinate data for each Neighborhood in Toronto: The following csv file gave us the geographical coordinates of each postal code: http://cocl.us/Geospatial\_data

**C. Methodology:**

1. Scrape the Wikipedia page and gathering data into a Pandas dataframe

To start with our analysis, we used the BeautifulSoup package to transform the data in the table on the Wikipedia page into the below pandas dataframe.

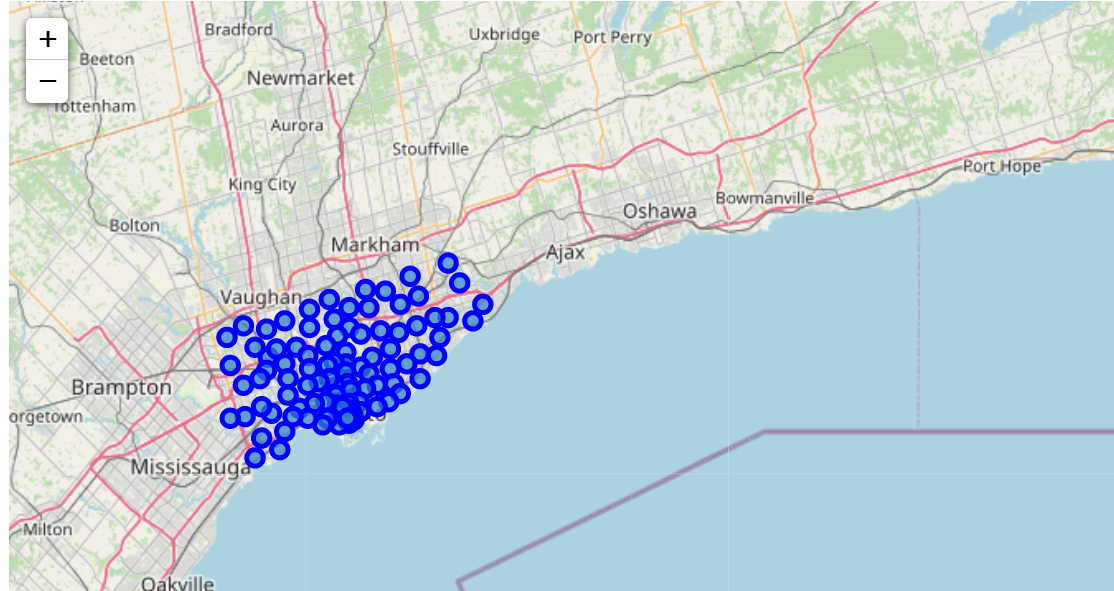


Next, combined both the dataframes by adding the coordinate data to the original dataframe.



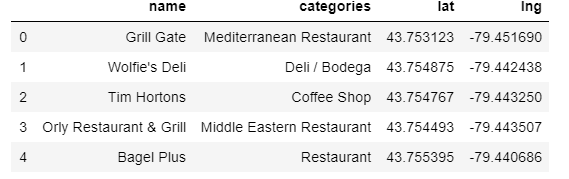
**2. Generating a map of Toronto and plotting the Neighborhood data**

We then use the python folium library to visualize geographic details of Toronto and its boroughs. I created a map of Toronto with boroughs superimposed on top using the latitude and longitude values to get the visual as below:

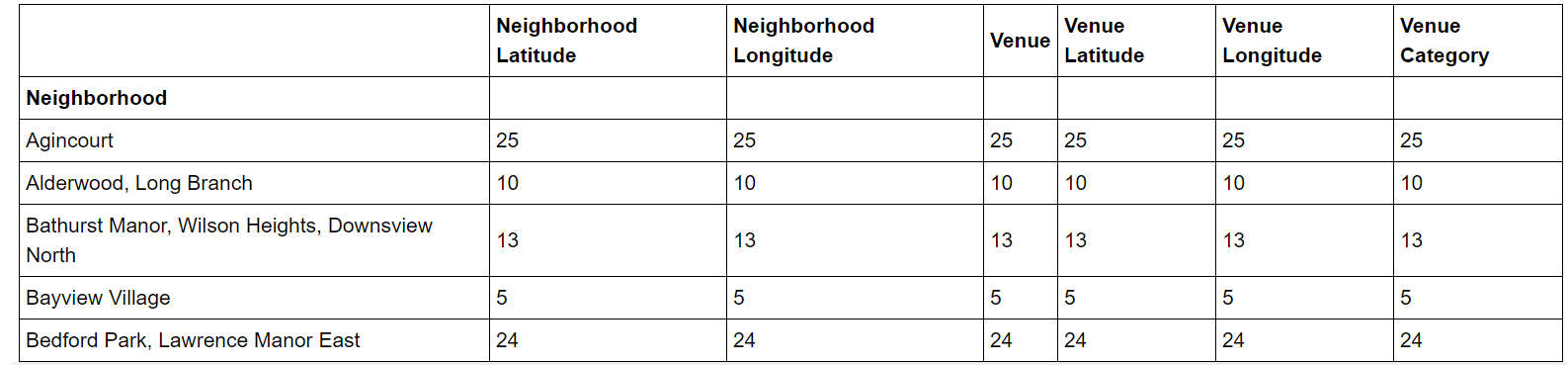


**3. Utilizing Foursquare API to explore the neighborhoods**

Next, we are going to start utilizing the Foursquare API to explore the neighborhoods and segment them. We set the LIMIT parameter to 100, which would limit the number of venues returned by the Foursquare API and the radius of 700 meter. Here is a head of the list of Nearby Venues for the first neighborhood i.e. The Beaches.

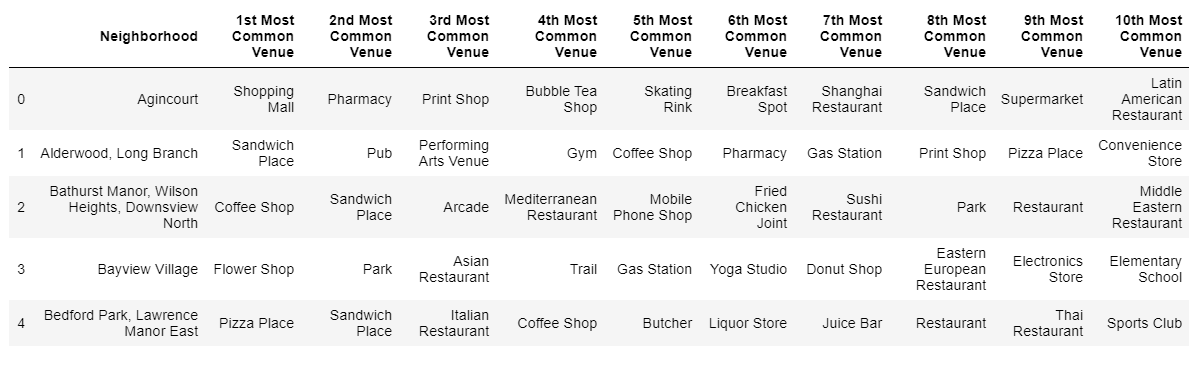


We create a new function that will repeat the process above for all the neighborhoods in Toronto. This function will give us a list of all venues present in Toronto city. Here is a head () value of this dataframe.



**4. Analyze each neighborhood**

We use One Hot Encoding, use the neighborhood to group data, and find out the top ten venues present in each neighborhood.



We have some common venue categories in the neighborhoods. We use the unsupervised learning K-means algorithm to cluster the neighborhoods. K-Means algorithm is one of the most common method for clustering in unsupervised learning.

We use a k\_cluster value of 10 to split the neighborhoods into 10 different clusters based on the similarity they have concerning the venues they contain.

**D. Results:**

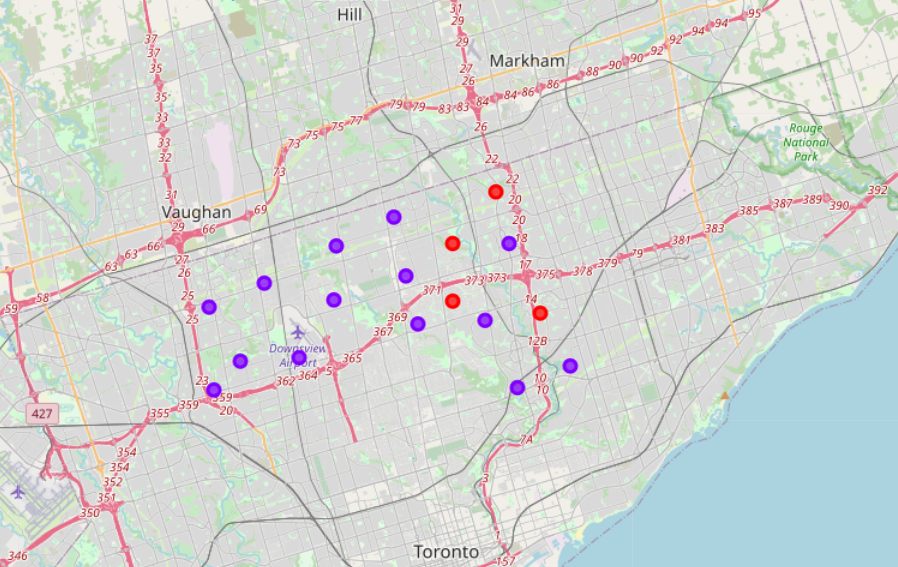
**D.1. Adding the Cluster Labels to the Venue Data**

The below table depicts the clustered data along with the top 10 most common venues in that cluster.

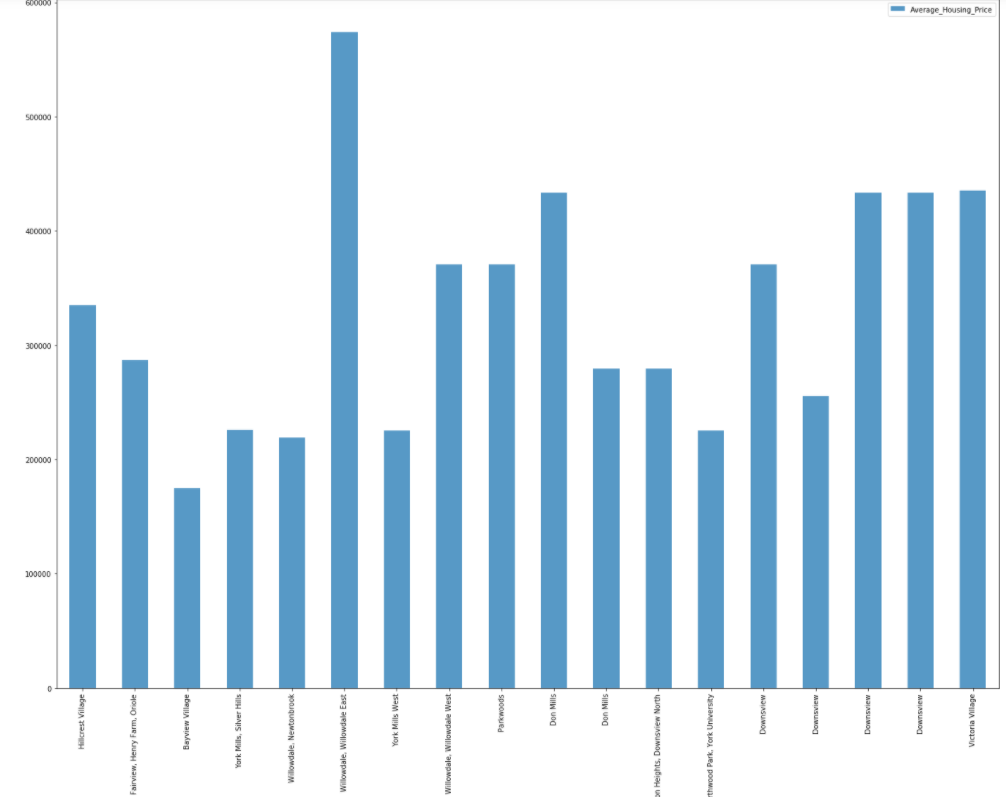


**2. Visualizing the resulting Clusters**

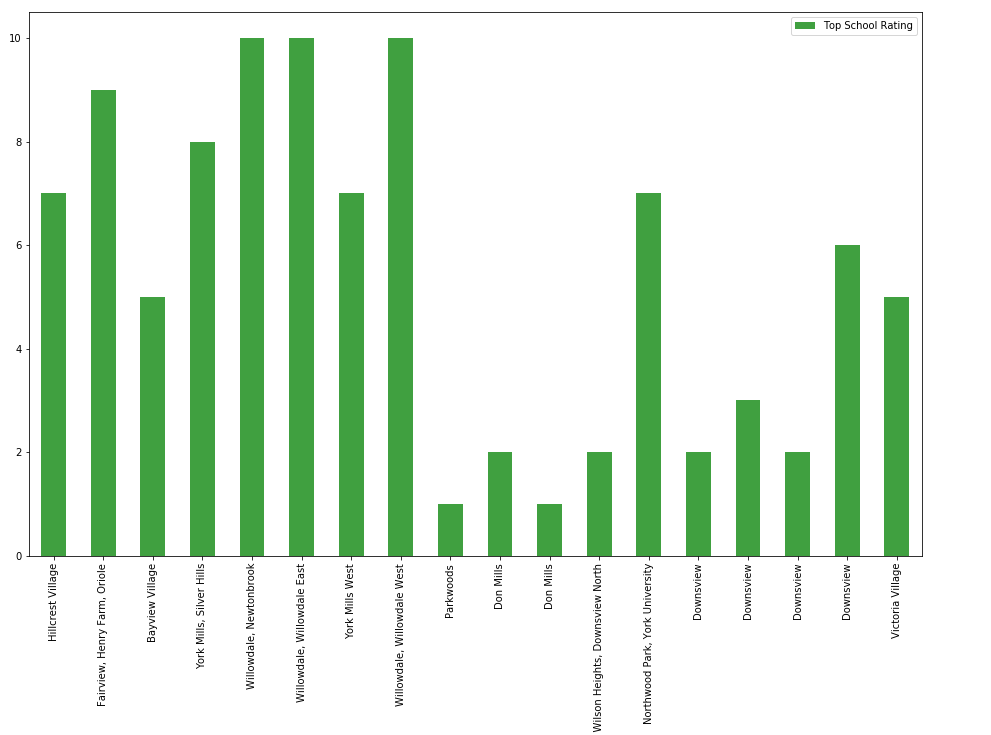
We use the matplotlib and folium packages to visualize the clusters on a map of Toronto.



**Average Housing Price by Clusters in North York**



**School Ratings by Clusters in North York**



**E. Discussion:**

Problem Which Tried to Solve:

The major purpose of this project is to suggest a better neighborhood in a new city for the person who are shiffting there. Social presence in society in terms of like minded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

* Sorted list of house in terms of housing prices in an ascending or descending order
* Sorted list of schools in terms of location, fees, rating and reviews

**Conclusion Section**

In this Capstone project, using k-means cluster algorithm I separated the neighborhood into 2 different clusters and for 103 different lattitude and longitude from dataset, which have very-similar neighborhoods around them. Using the charts above results presented to a particular neighborhood based on average house prices and school rating have been made.

This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools.

The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.