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

**- G.S.Rohitaksh**

**A. Introduction:**

**A.1. Background**

**Toronto** is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 2016. Current to 2016, the Toronto census metropolitan area (CMA), of which the majority is within the Greater Toronto Area (GTA), held a population of 5,928,040, making it Canada's most populous CMA. The city is the anchor of the Golden Horseshoe, an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of Lake Ontario.

Toronto is an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.



**Brief history about Toronto**: People have travelled through and inhabited the Toronto area, located on a broad sloping plateau interspersed with rivers, deep ravines, and urban forest, for more than 10,000 years. After the broadly disputed Toronto Purchase, when the Mississauga surrendered the area to the British Crown, the British established the town of **York** in 1793 and later designated it as the capital of Upper Canada. During the War of 1812, the town was the site of the Battle of York and suffered heavy damage by United States troops. **York** was renamed and incorporated in 1834, as the city of **Toronto**. It was designated as the capital of the province of Ontario in 1867 during Canadian Confederation.

**A.2. Problem Description**:

Now let me explain the context of this Capstone project through a scenario. Say you live on the west side of the city of Toronto in Canada. You love your neighborhood, mainly because of all the great amenities and other types of venues that exist in the neighborhood, such as gourmet fast food joints, pharmacies, parks, graduate schools and so on. Now say 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.

**A.4. Target Audience**

This information provided by this report would be useful for people who are interested in relocating to a different part of the city and are interested in finding new neighborhoods that are highly similar to their existing neighborhood.

**B. Data Description:**

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

**a) 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>

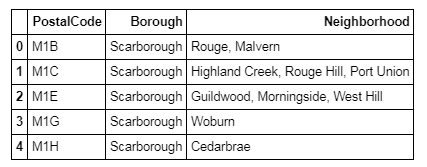
The information obtained i.e. the table of postal codes was transformed into a pandas dataframe for further analysis.

**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>

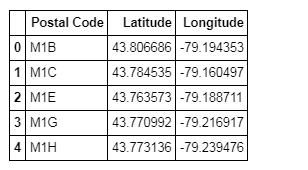
**C. Methodology:**

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

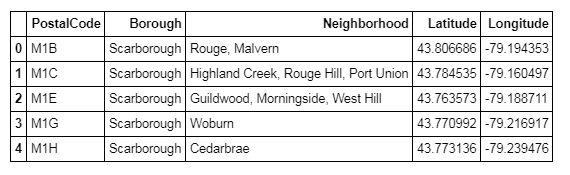
To start with our analysis, we used the **BeautifulSoup** package to transform the data in the table on the Wikipedia page into the below pandas dataframe.



We also fetched the coordinate data for all the neighborhoods in Toronto using the csv file and put it into a dataframe.

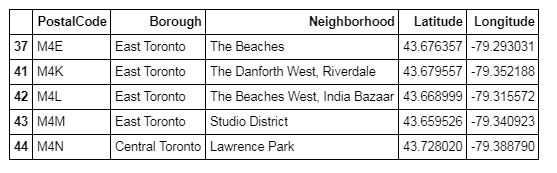


Next, we combine both the dataframes i.e. adding the coordinate data to the original dataframe.

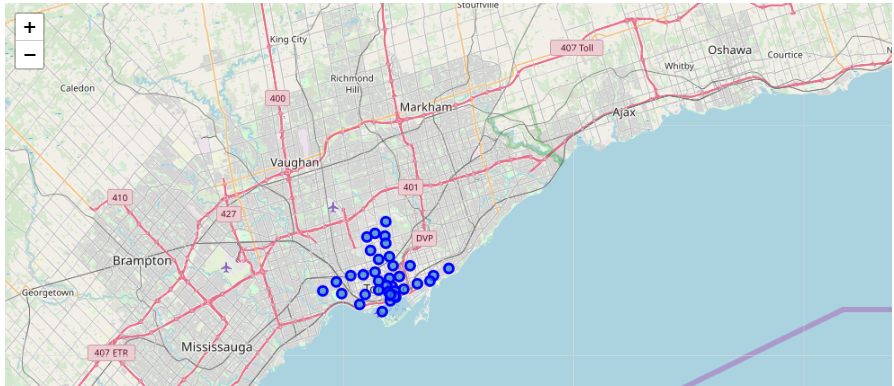


**C.2. Generating a map of Toronto and plotting the Neighborhood data on it**

We first filter the data to find boroughs containing the word “Toronto”,



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:

****

**C.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 500 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.



**C.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 11 to split the neighborhoods into 11 different clusters based on the similarity they have concerning the venues they contain.

**D. Results:**

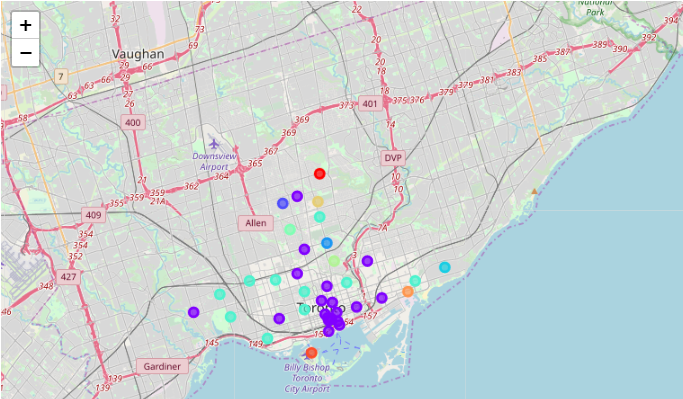
**D.1. Adding the Cluster Labels to the Venue Data**

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



**D.2. Visualizing the resulting Clusters**

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



**E. Discussion:**

The intent with which analysis was carried out was to find out similar neighborhoods for a person relocating within the city.

As we analyze the results section, we can analyze the clusters and see similar neighborhoods in different parts of the city. For example, if we compare the different neighborhoods clustered in cluster 2.



As seen in the table above, if a person wished to move from a suburb in East Toronto to Central Toronto. If a person’s current location were in the Neighborhood of Studio District in East Toronto, which has venues like cafes, bakeries and restaurants nearby, the person, would like to relocate to a neighborhood like North Toronto West in Central Toronto which also has venues like Coffee Shops and Restaurants. This is just one example of how our data analysis can help people relocate from one part of the city to another which similar to their current localities.

**F. Conclusion:**

In a fast moving world, there are many real life problems or scenarios where data can be used to find solutions to those problems. Like seen in the example above, data was used to cluster neighborhoods in Toronto based on the most common venues in those neighborhoods. Similarly, data can also be used to solve other problems, which most people face in metropolitan cities.

**G. References:**

* Wikipedia content: <https://en.wikipedia.org/wiki/Toronto>
* CSV for Coordinate data: [http://cocl.us/Geospatial\_data](https://cocl.us/Geospatial_data)
* Foursquare API