

## CAPSTONE PROJECT: THE BATTLE OF NEIGHBORHOODS (WEEK 1) \_ RELOCATE WITH CONFIDENCE (By Ajay Mutreja (April 2020))

### **1. BACKGROUND:**

Ever since the US President Donald Trump tightened the visa policies for existing and new immigrants to the USA, the US companies have started to look at Canada as an alternative playground for their businesses, especially the IT companies which depend on skilled workers from all over the world.

Canada Government took the opportunity to offer new visa schemes for skilled workers to immigrate to Canada, Toronto, Ontario as an example has offered many nouvelle benefits for such companies and its workers.

A global relocation company has a huge demand from their customers to provide Canada Destination Information and best match of the neighborhoods for their clients. They have asked me to build a program on their app, in which the clients just need to enter the city from where they are moving into Toronto and the system matches them to the best neighborhood, so that people can satisfactorily settle in quickly and can plug and play into their jobs without worrying.

"Toronto Neighborhood" is a test case, if the system is successful the company may implement it for all Canada and scale up later to the global application.

### **2. BUSINESS PROBLEM:**

Recently, a family approached the relocation company. They wish to move from Mumbai, India to Toronto and they have the following requirements.

1. They want to move to a neighborhood which matches their social needs, be around Indian Community, grocery store, restaurants and around the city center.
2. They want to know the housing prices in that neighborhood, and
3. They want to know about the Schools in the neighborhood for their two children.

### **3. TARGET AUDIENCE:**

Anyone who wishes to move to Toronto from anywhere in the world.

This project aims to create an analysis of features for a people migrating to Toronto to search the best neighborhood as a comparative analysis between neighborhoods.

The features include matching a neighborhood with the lifestyle as close as possible to the lifestyle of the city people are migrating in from, expected housing prices and various school options based on their ratings, it may help people to get the awareness of the area and neighborhood before moving to a new city to start a new fresh life.

### **4. SOLUTION/PROBLEM ADDRESSES:**

Build a program to give a dependable recommendation, based on real-time data analysis.

### **5. THE DATA SCIENCE WORKFLOW & DATA DESCRIPTION:**

This project will rely on public data from Wikipedia and Foursquare.

Canada Neighborhood Data - (Source Identified - Scraped from Wikipedia - Canada, Ontario, Toronto, Postal Code)

Data Link: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

Data wrangling and Cleaning - Data is found but is not in a useable form, so data wrangling and cleaning is performed to create appropriate data frames.

The cleansed data will then be used alongside Foursquare by invoking Foursquare API credentialed location information

Foursquare location data will be leveraged to explore or compare districts around Central New Delhi, India with neighborhoods in Toronto, where consumers go for shopping, dining and entertainment.

Get data about different venues in different neighborhoods of the specific borough.

For each neighborhood, we chose the radius of 100 meters.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes, top get the following information;

The neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venue, Name of the venue e.g. the name of a store or restaurant, Venue Latitude, Venue Longitude, Venue Category.

Data manipulation and analysis to derive subsets of the initial data, segment them, group them, apply K Means clustering algorithm to cluster the neighborhoods similar to Mumbai and use the cluster data frames to output the property prices and school information.

Visualization: Analysis and plotting visualizations, using various mapping libraries.

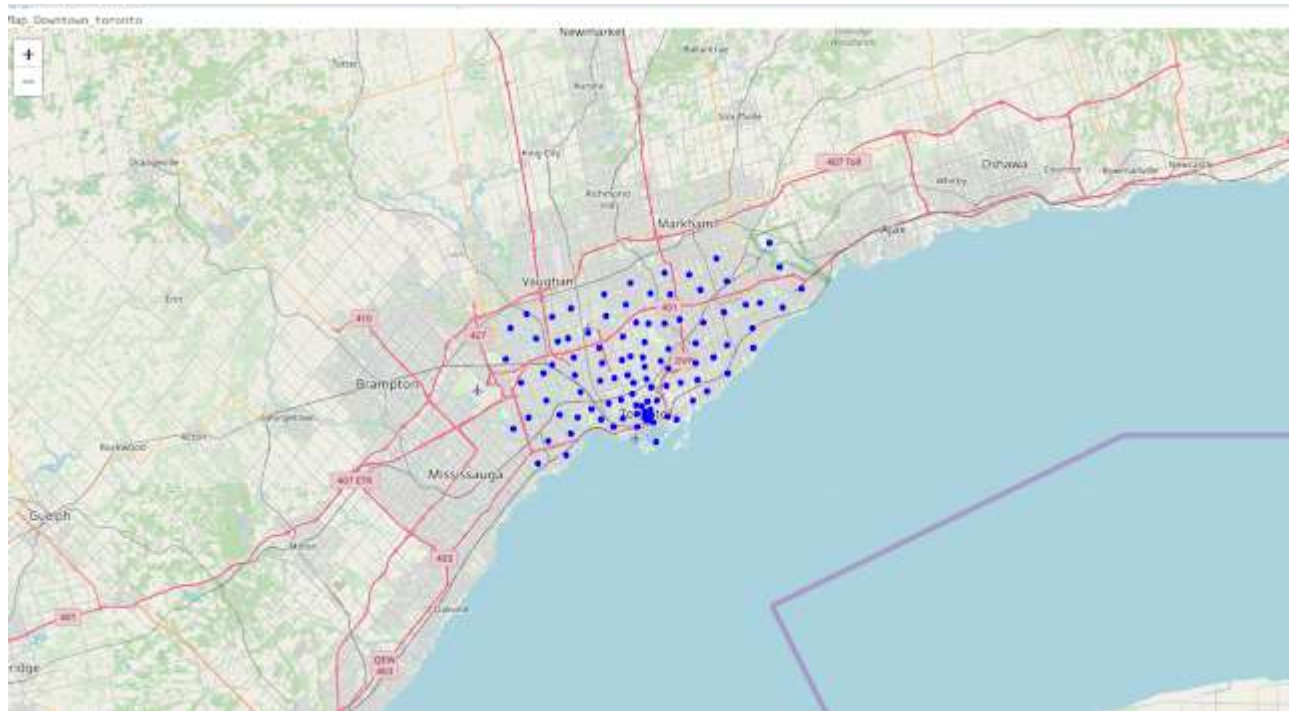
Libraries which are used to develop the Project:

- PANDAS: For creating and manipulating data frames.
- FOLIUM: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
- SCIKIT LEARN: For importing k-means clustering.
- JSON: Library to handle JSON files.
- XML: To separate data from presentation and XML stores data in plain text format.
- GEOCODER: To retrieve Location Data.
- BEAUTIFUL SOUP: To scrap and library to handle HTTP requests.
- MATPLOTLIB: Python Plotting Module.

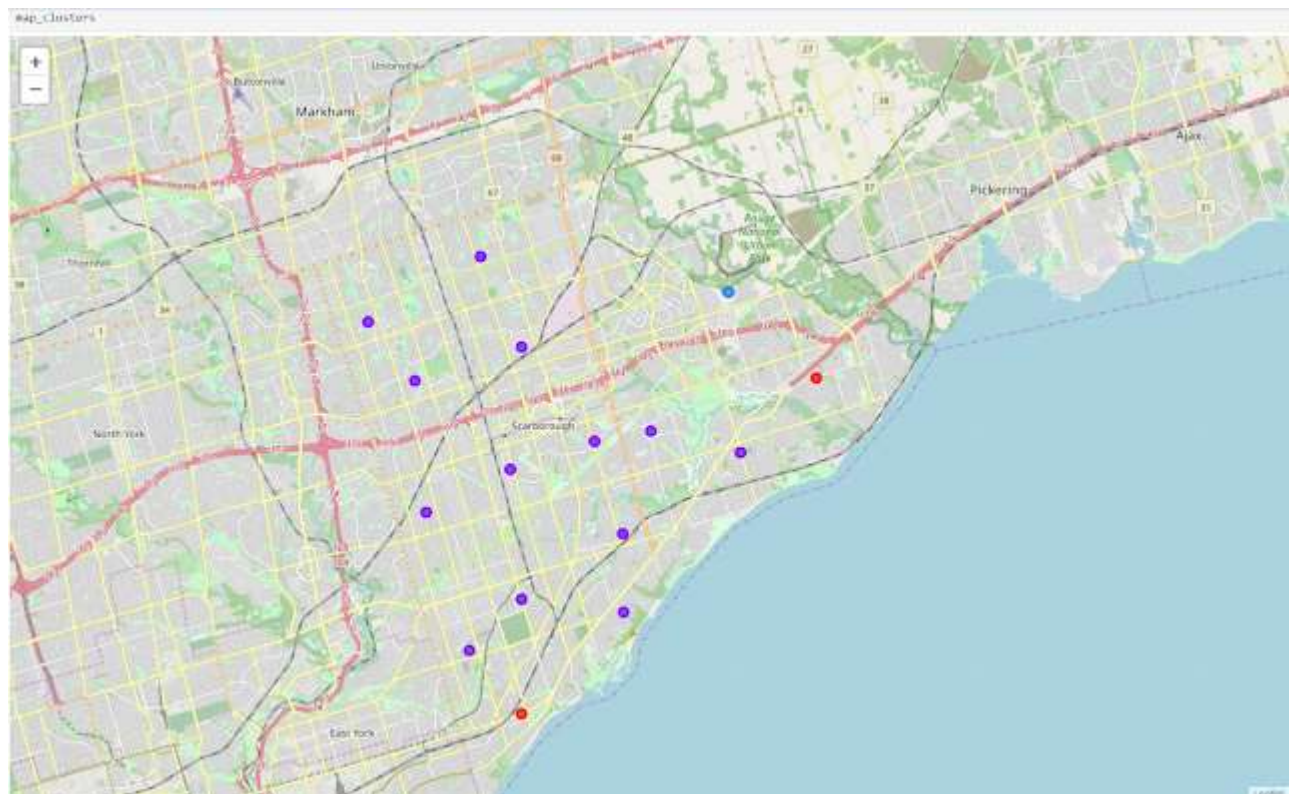
## **6. CONCLUSIONS:**

Using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 103 different latitude and longitude from the dataset, which has very-similar neighborhoods around them.

MAP 1



MAP 2



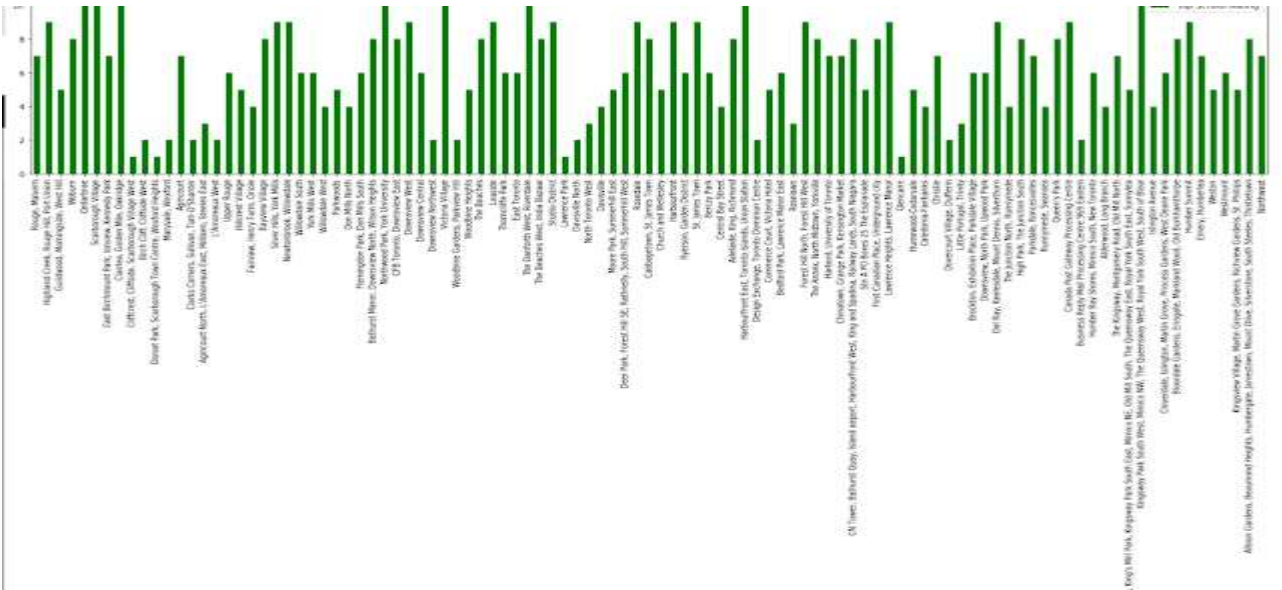
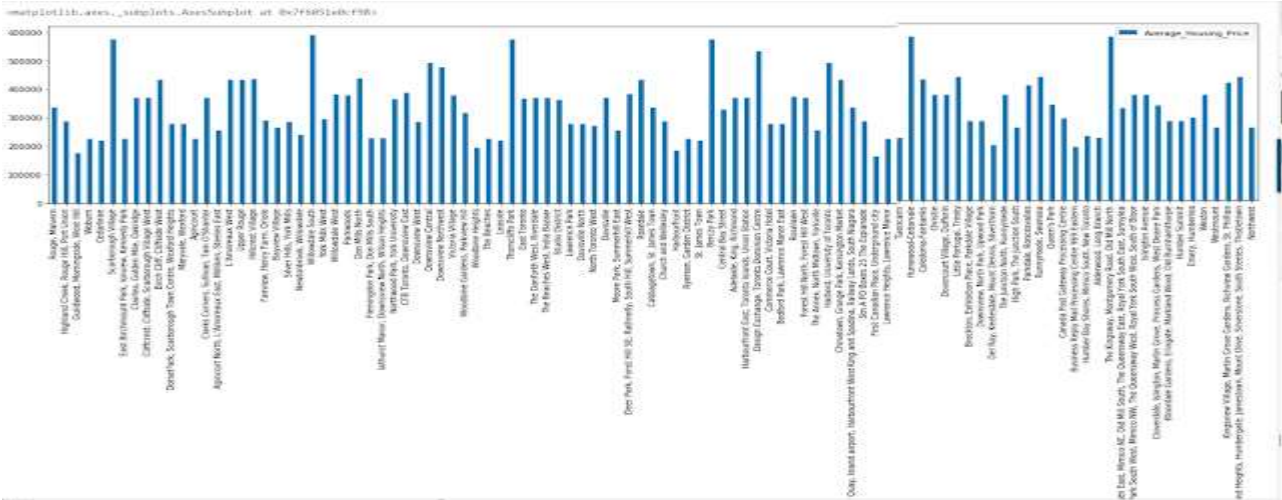
Observe the difference between Map 1 and Map 2, Map 1 Plots all the postal codes in Downtown Toronto however in Map 2 plots only the recommended neighborhoods based on the Indian community density clusters.

Details of such Clusters can be seen in detail in the following table.

	Postcode	Borough	Neighborhood	Latitude	Longitude	cluster_label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	M18	Scarborough	Rouge-Mohawk	43.806626	-79.188913	2	Park	Trail	Women's Store	Farm	Dumping Restaurant	Eastern European Restaurant	Electronics Store	Barterary School	Ethiopian Restaurant	Yahala Restaurant
1	M1C	Scarborough	Highland Creek, Rouge-Hill, Port Union	43.785779	-79.157368	0	Bar	History Museum	Sub & Cham Shop	Park	Farm	Donut Shop	Chumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.765888	-79.185264	1	Pizza Place	Fast Food Restaurant	Park	Supermarket	Bank	Pharmacy	Coffee Shop	Sandwich Place	Burger Joint	Food & Drink Shop
3	M1G	Scarborough	Woburn	43.771545	-79.218138	-1	Coffee Shop	Park	Business Service	Women's Store	Farm	Dumping Restaurant	Eastern European Restaurant	Electronics Store	Barterary School	Ethiopian Restaurant
4	M1H	Scarborough	Cedarbrae	43.765791	-79.238012	1	Gas Station	Athletics & Sports	Bank	The Restaurant	Halls Restaurant	Lounge	Eastern European Restaurant	Electronics Store	Barterary School	Ethiopian Restaurant

Using the charts above results presented to a particular neighborhood based on average house prices and school ratings have been made.

Also, the average housing pricing index gives the average prices of various neighborhoods using the merged data frame as well as the school ratings of various schools in the neighborhood.



Based on this information the immigrant can easily choose where to stay, how much to budget and what to expect. This helps in confidence immigration leading to customer satisfaction.

## **7. DISCUSSION:**

This program can be built for Canada, For Canada Postal Codes Data is available, But data cleaning and converting them into the data frame may not be able to be automated, as the formats of data tables on Wikipedia is non-standard, So a data wrangling program shall be written for each city if Web Scraping Method is used, nevertheless if a source that can provide uniform information be identified, automation may be easier.

The Neighborhood Data may not be available for all the cities of the world, therefore establishing a global program may be a huge challenge.

Average Housing Price index and School ratings for all the cities may not be available, therefore the data will have to be collected through primary surveys

The data may be available but maybe available in a different language, therefore additional ML translation program may have to be invoked for data gathering.

## **8. ACKNOWLEDGEMENT:**

My Work is in continuation of the inspiring work of many capstone project pursuers on the same subject.

For Detailed Project Report please visit my Git Hub Page:

[https://github.com/AjayMutreja/Coursera\\_Capstone\\_2/blob/master/CPW\\_4\\_5-Code.ipynb](https://github.com/AjayMutreja/Coursera_Capstone_2/blob/master/CPW_4_5-Code.ipynb)