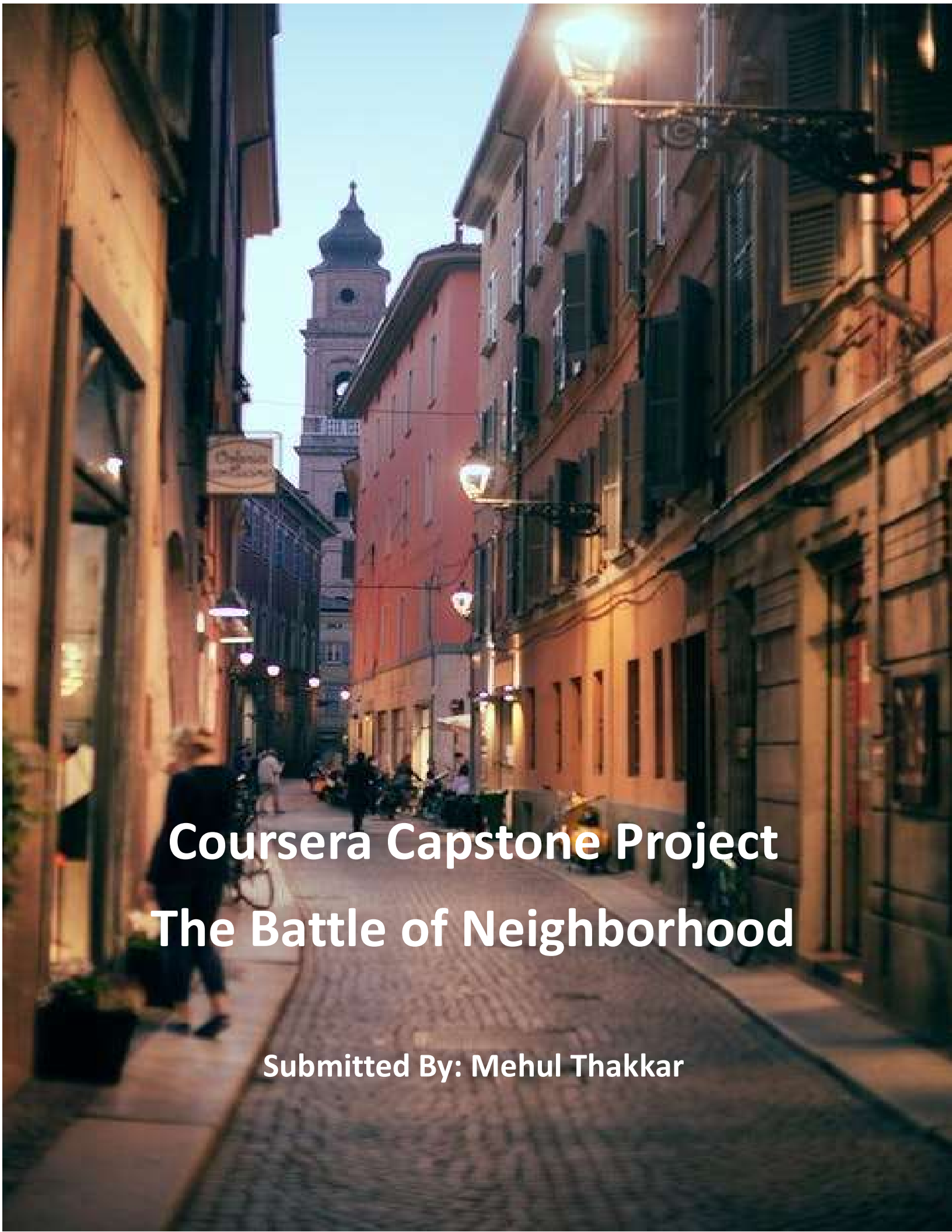


Coursera Capstone Project
The Battle of Neighborhood

Submitted By: Mehul Thakkar

Objective of this Project Report is to provide an overview of the Coursera Capstone Project. As part of the project work, my objective is to provide an overview of the entire project including

1. Project overview
2. Business Problem
3. Proposed Solution & Methodology
4. Data used
5. Results
6. Conclusion
7. References

1. Project Overview

As mentioned above this project has been undertaken as part of the Coursera IBM Data Science Specialization. Key objective of Data Science is to identify a problem and then use Data to find innovative yet impactful solution to the problem. This requires understanding the User needs, exploring various data available and defining a methodology that solves the problem in an effective manner. For this project, we are looking to solve the problem of selecting a neighborhood to live when you move to a new country or city. I hope the sections outlined below succinctly summarize the problem, proposed solution, methodology and outcome. Happy reading !

Business Problem

2. Business Problem

When you move to a new country, one of the biggest questions is to find a place to live. Everyone has their unique taste of what they like, and their lifestyle could be different from others. Due to this, you need to do your own research to identify a neighborhood that you would love to live in. The purpose of this Project is to help users in exploring better facilities around their residence. It will help them make a well-informed decision supported by data. For purpose of demonstration, we are using the Scarborough county in Toronto as an example but this model can be used for any county, suburb or district in any city around the world as long as there are data points available to run this model.

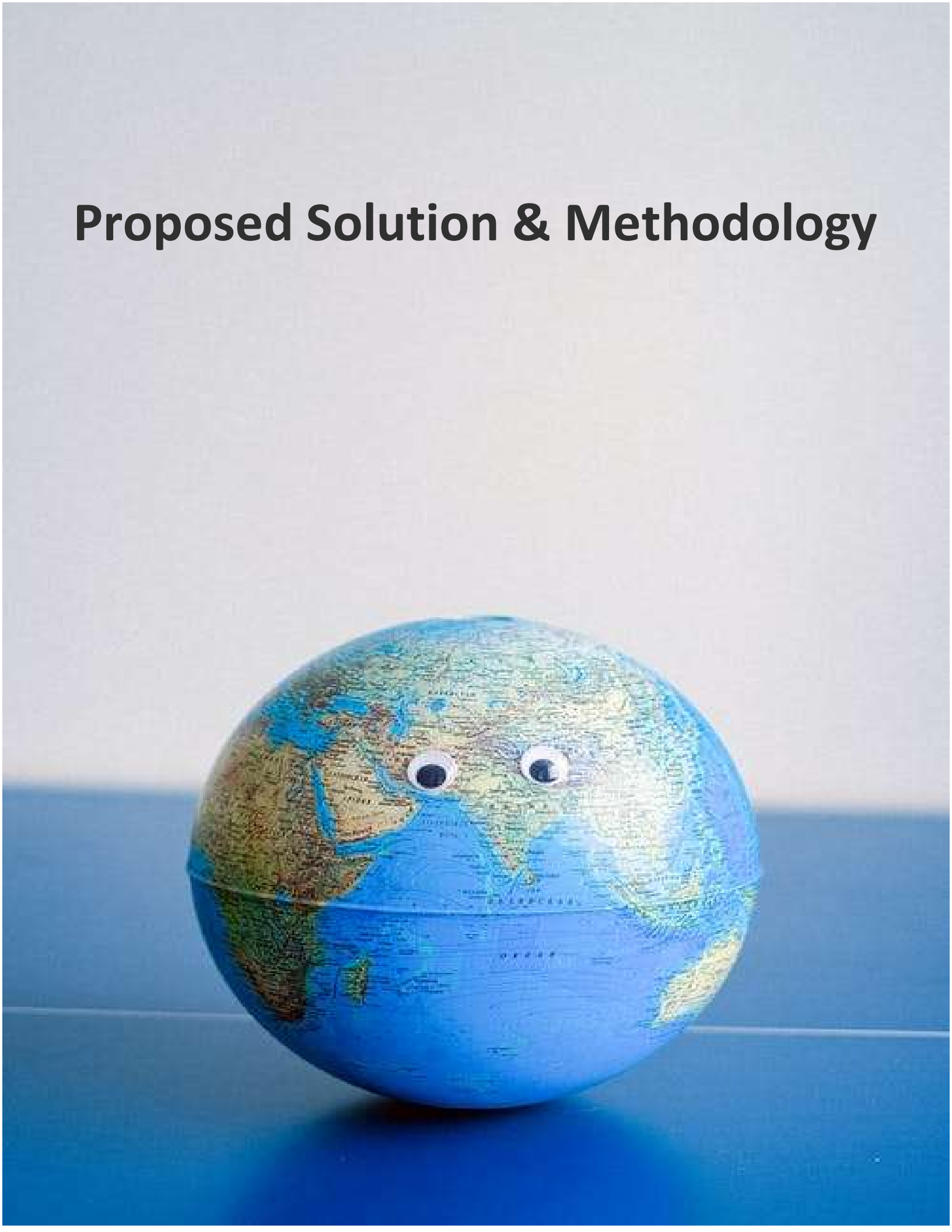
Lots of people are migrating to various states of Canada and needed lots of research for good housing prices and reputed schools for their children. This project is for those people who are looking for better neighborhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, likeminded people, etc.

This Project aim to create an analysis of salient features for the user to search a best neighborhood as a comparative analysis between neighborhoods. The features include

- Median housing price
- Good school according to ratings
- Crime rates
- Road connectivity
- Weather conditions
- Good management for emergency
- Recreational facilities
- Water resources

This Project is aimed to solve this problem and help user to assess a particular neighborhood before making the decision to rent or buy an apartment.

Proposed Solution & Methodology



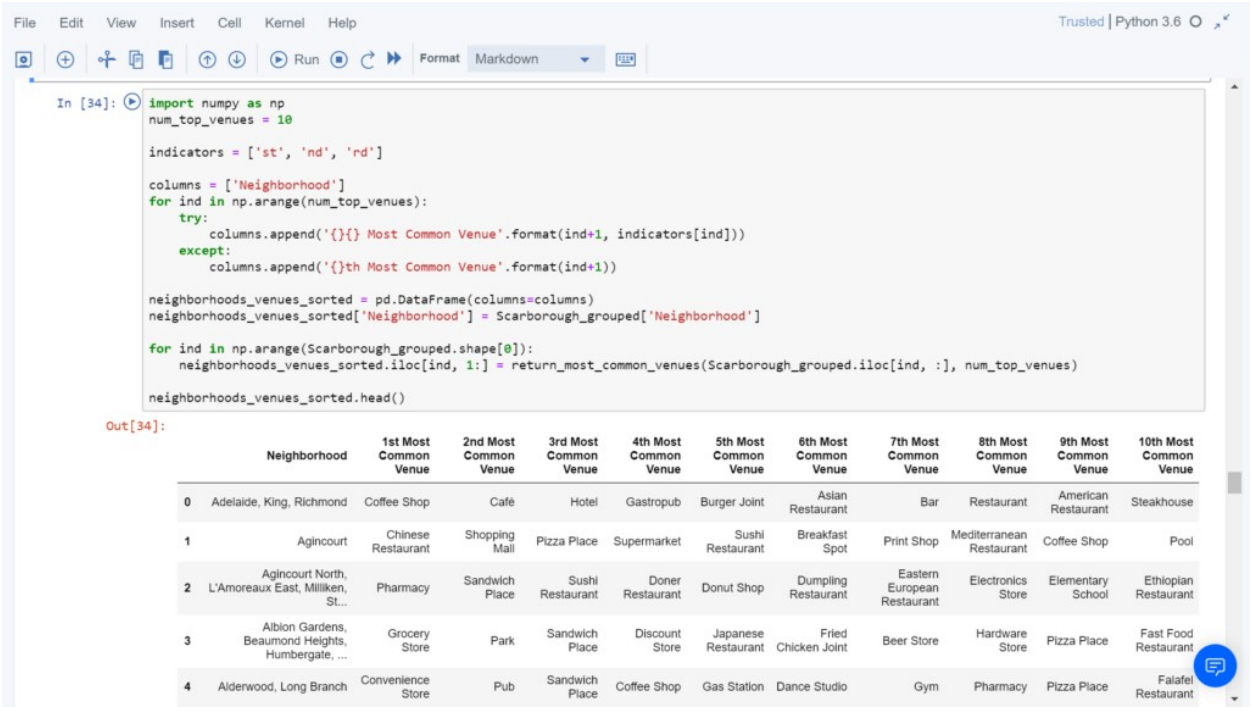
3. Proposed Solution & Methodology

We will use Foursquare API to source various information related to various neighborhoods. We will then use K-Means Clustering approach to segment the neighborhood.

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

Using K-Means Clustering Approach | Most Common Venue

Most Common Venues near Neighborhood | Using Clustering



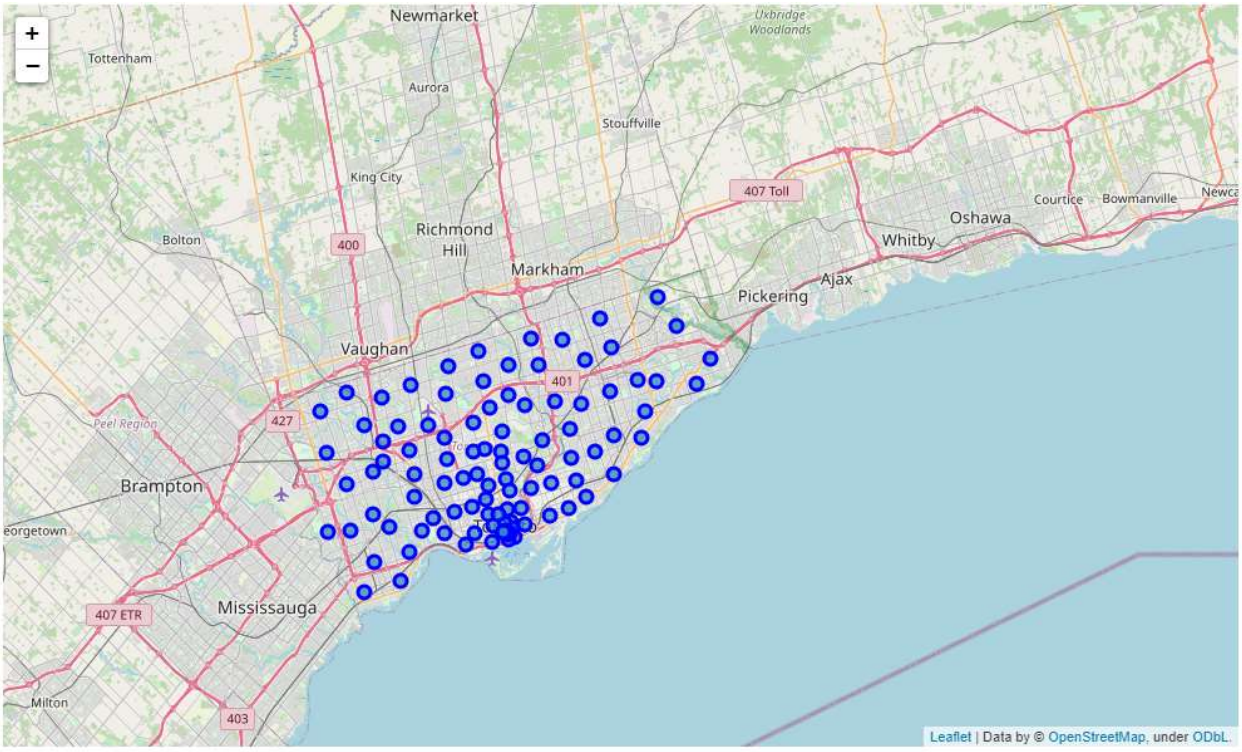
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Café	Hotel	Gastropub	Burger Joint	Asian Restaurant	Bar	Restaurant	American Restaurant	Steakhouse
1	Agincourt	Chinese Restaurant	Shopping Mall	Pizza Place	Supermarket	Sushi Restaurant	Breakfast Spot	Print Shop	Mediterranean Restaurant	Coffee Shop	Pool
2	Agincourt North, L'Amoreaux East, Milliken, St...	Pharmacy	Sandwich Place	Sushi Restaurant	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethioplan Restaurant
3	Albion Gardens, Beaumond Heights, Humbergate, ...	Grocery Store	Park	Sandwich Place	Discount Store	Japanese Restaurant	Fried Chicken Joint	Beer Store	Hardware Store	Pizza Place	Fast Food Restaurant
4	Alderwood, Long Branch	Convenience Store	Pub	Sandwich Place	Coffee Shop	Gas Station	Dance Studio	Gym	Pharmacy	Pizza Place	Falafel Restaurant

Work Flow:

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

1 Neighborhood
2 Neighborhood Latitude
3 Neighborhood Longitude
4 Venue
5 Name of the venue e.g. the name of a store or restaurant
6 Venue Latitude
7 Venue Longitude
8 Venue Category

Map of Scarborough



Data used for the Project

4. Data used for the Project

We have used various data sources as part of our project. We enlist those data sources below.

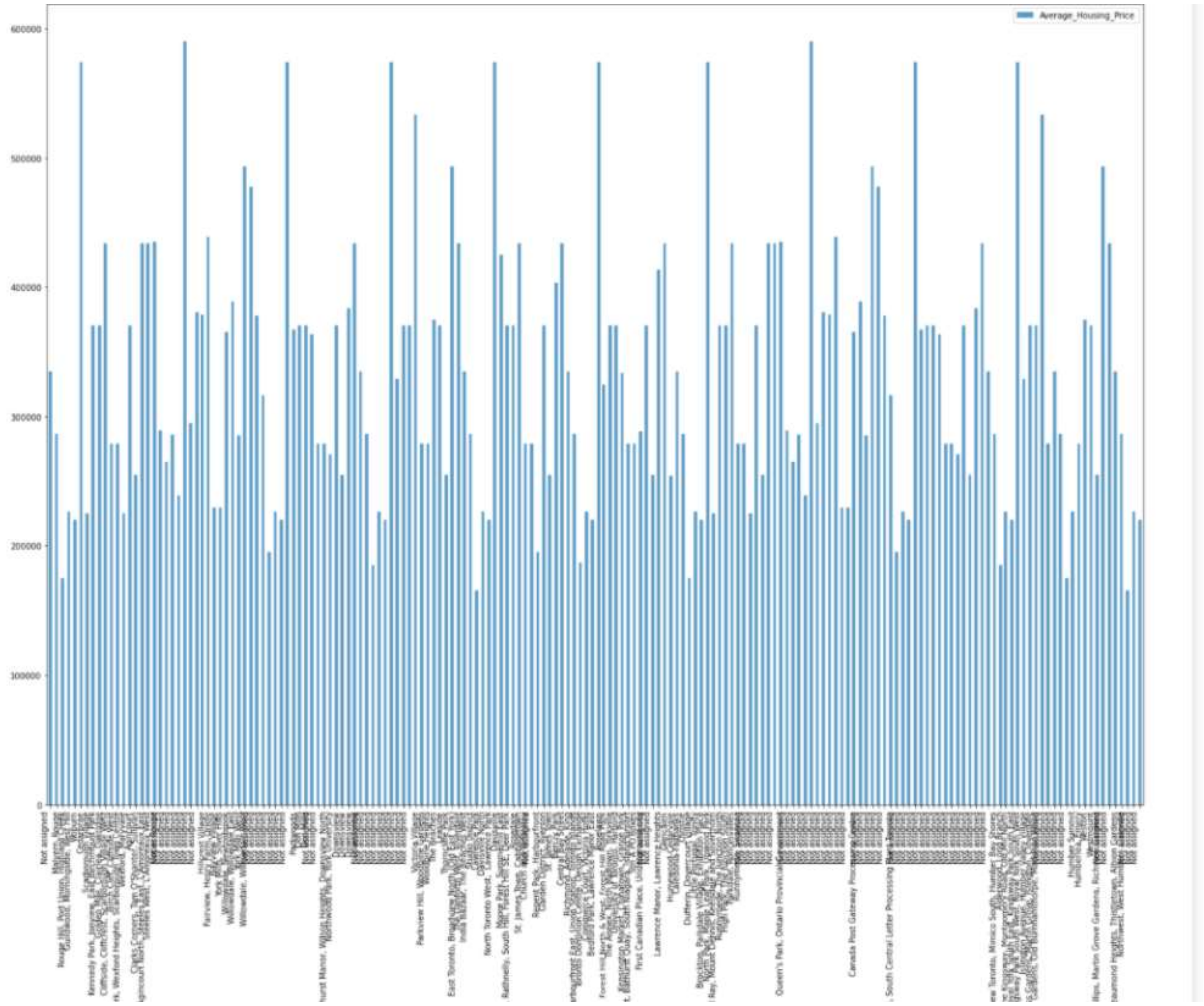
Wikipedia

We have used Wikipedia for the various information related to the counties within the target locations. You can refer to the data sourcne from the link shared below.

Data Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Will use Scarborough dataset which we scrapped from wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes.

Average Housing Price by Clusters in Scarborough



Foursquare API Data:

This Capstone project have used Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

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

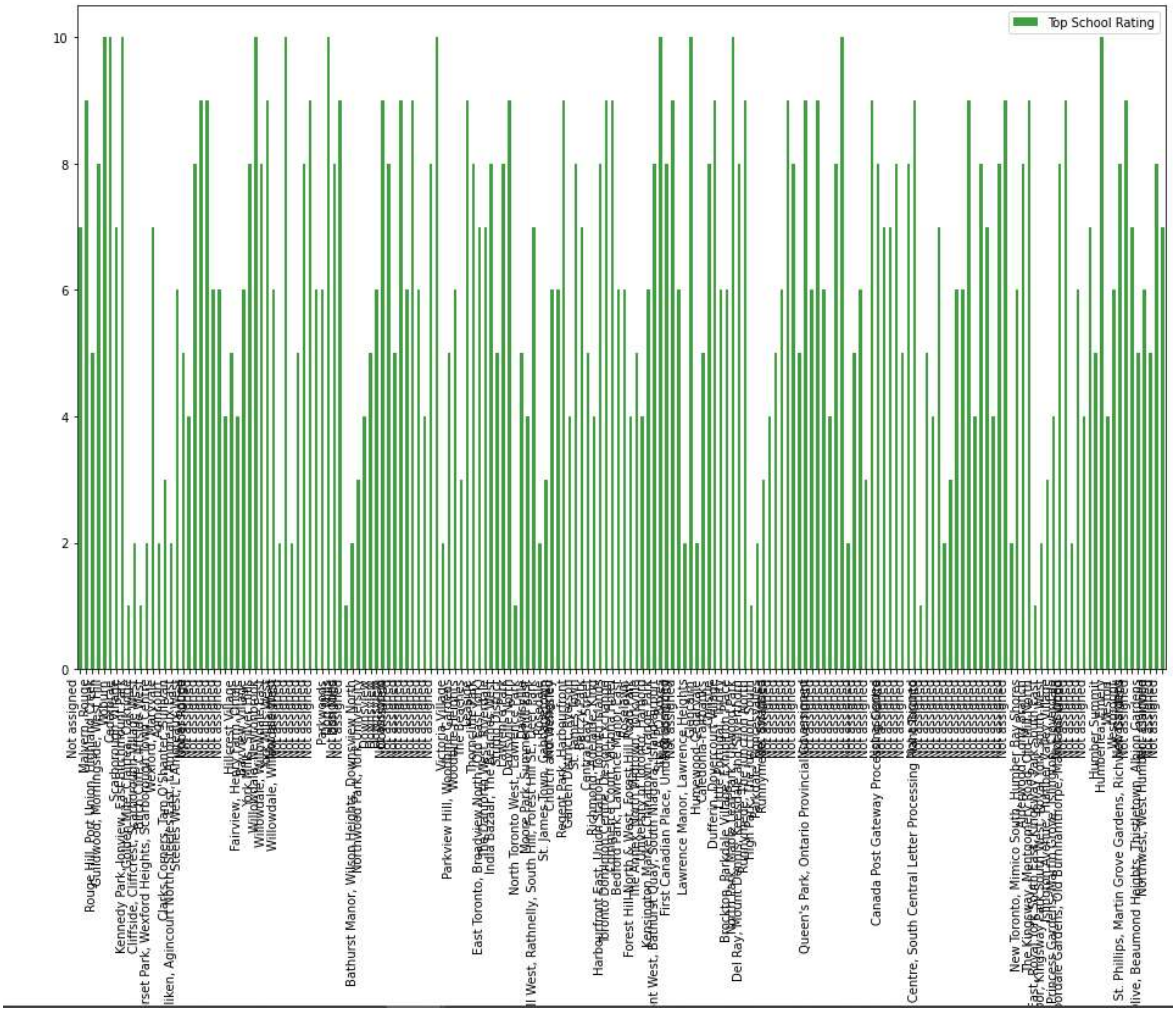
We will need data about different venues in different neighborhoods of that specific borough.

In order to gain that information we will use “Foursquare” locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

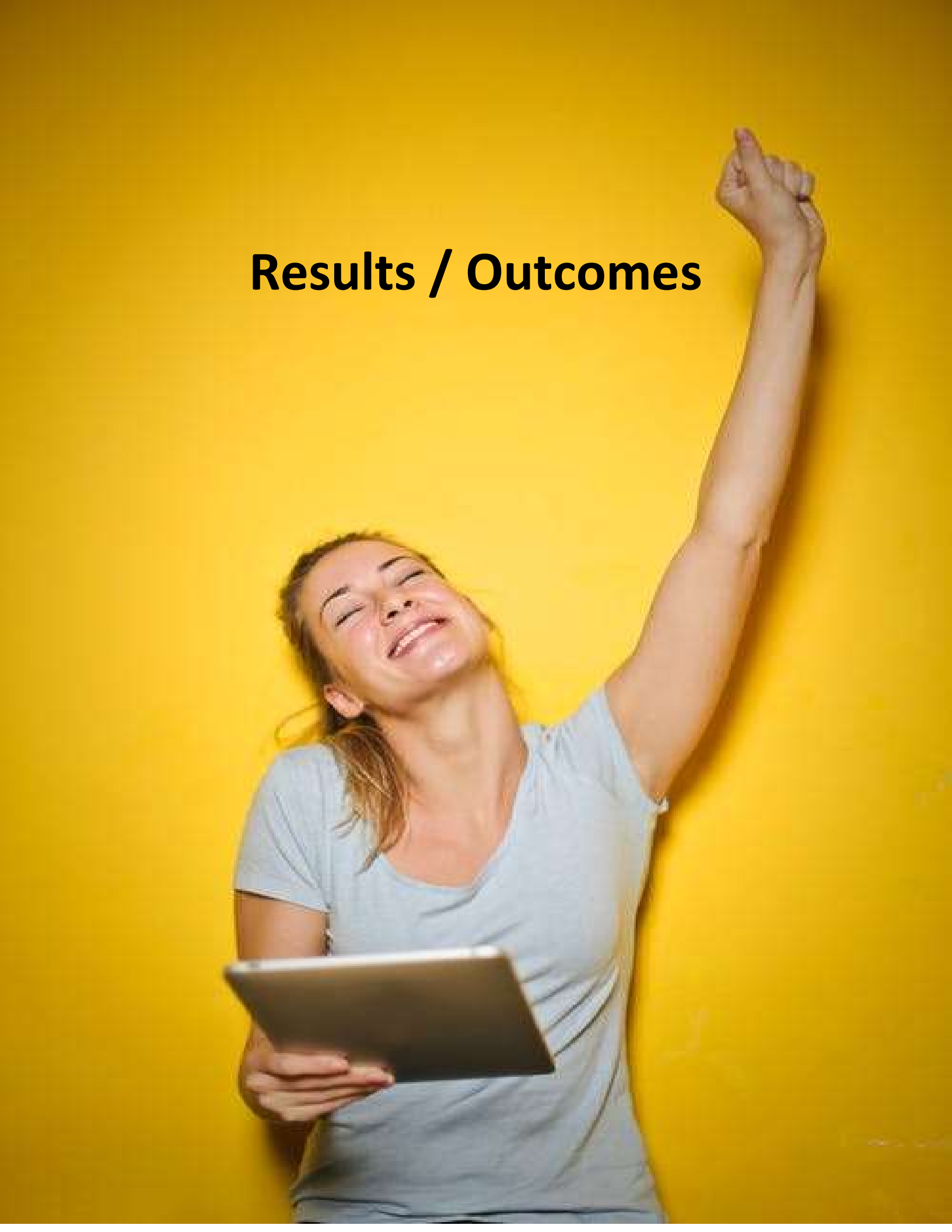
After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venues.

School Ratings by Clusters in Scarborough

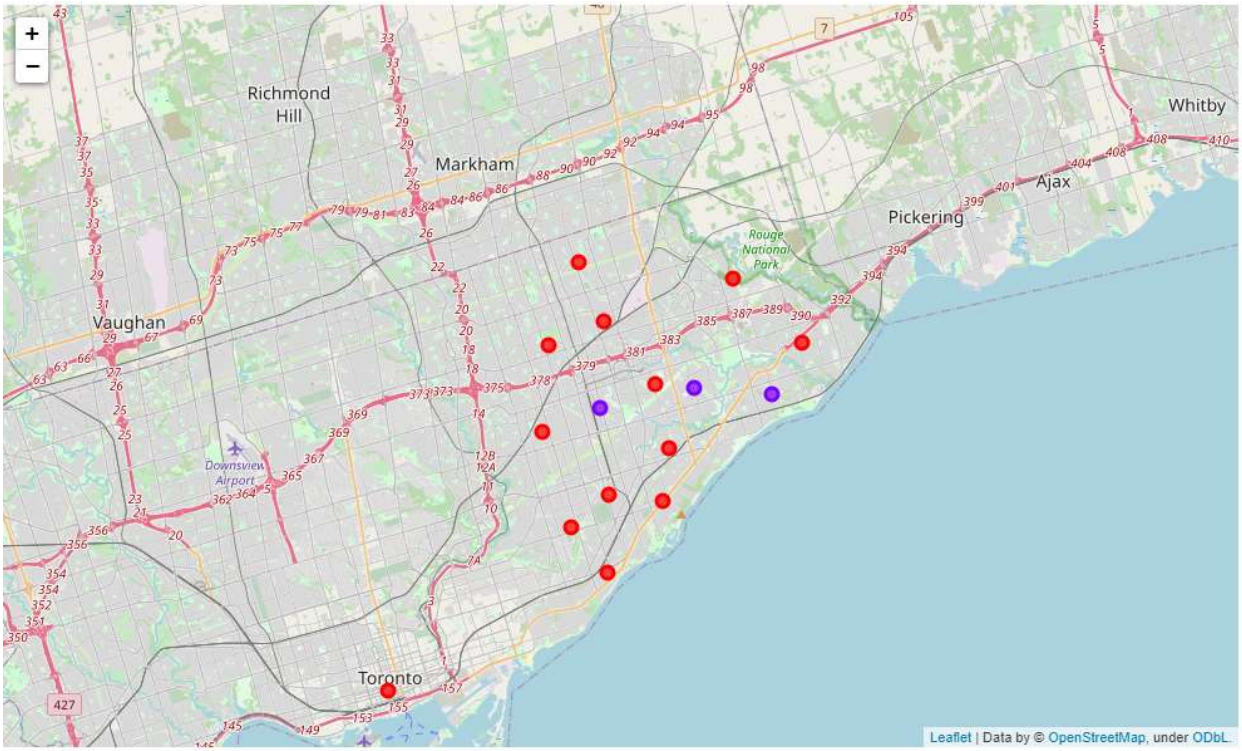


Results / Outcomes



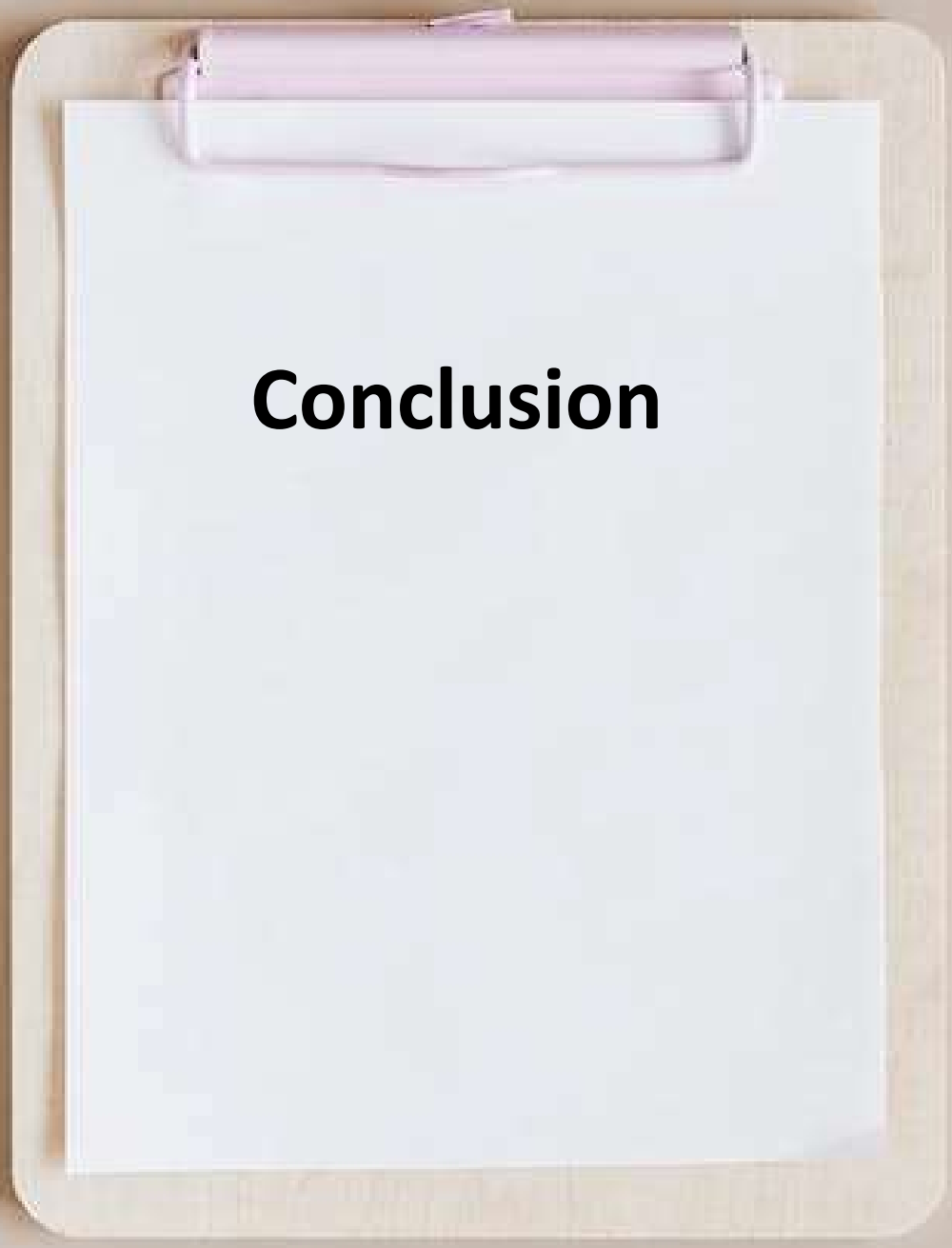
5. Results

Map of Clusters in Scarborough



The Location:

Scarborough is a popular destination for new immigrants in Canada to reside. As a result, it is one of the most diverse and multicultural areas in the Greater Toronto Area, being home to various religious groups and places of worship. Although immigration has become a hot topic over the past few years with more governments seeking more restrictions on immigrants and refugees, the general trend of immigration into Canada has been one of on the rise.


Conclusion

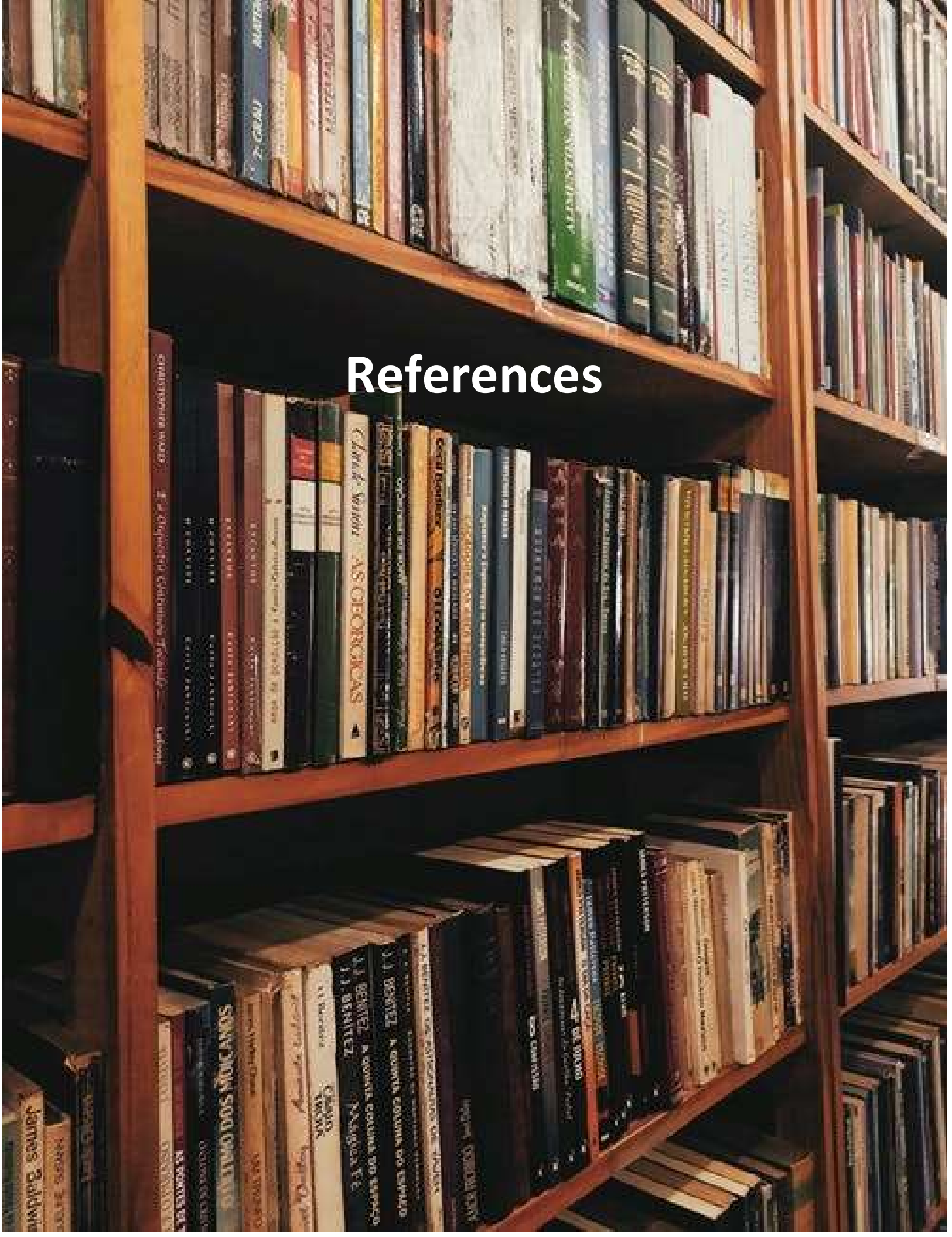
6. Conclusion

In this Capstone project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 103 different lattitude and logitude 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.

I feel rewarded with the efforts and believe this course with all the topics covered is well worthy of appreciation.
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.

Future Development

This Capstone project can be continued for making it more precise in terms to find best house in Scarborough. Best means on the basis of all required things(daily needs or things we need to live a better life) around and also in terms of cost effective.



References

7. References

List of Python Libraries and Functions deployed

- Pandas: For creating and manipulating dataframes.
- 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 and Requests: To scrap and library to handle http requests.
- Matplotlib: Python Plotting Module.

Github Link for reference

GitHub Link of Complete Project can be found at – https://github.com/MehulTh12/Coursera_Capstone/blob/main/Capstone_TBON_Wk2.ipynb

Blogpost

I have published a blogpost summarizing my approach and finding. You can fins the Blog post at the link below

https://mehulblog754758905.wordpress.com/2020/12/30/coursera-capstone-battle-of-neighborhood-project/