

Final Report Capstone Project –Finding a Better Place in Etobicoke, Toronto

1. Introduction:

The purpose of this Project is to help people in exploring better facilities around their neighborhood. It will help people making smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in Etobicoke, Toronto.

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, like minded people, etc.

This Project aim to create an analysis of features for a people migrating to Scarborough to search a best neighborhood as a comparative analysis between neighborhoods. The features include median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, good management for emergency, water resources both freash and waste water and excrement conveyed in sewers and recreational facilities.

It will help people to get awareness of the area and neighborhood before moving to a new city, state, country or place for their work or to start a new fresh life.

2. Data Section

Data

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

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

Foursquare API Data:

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.

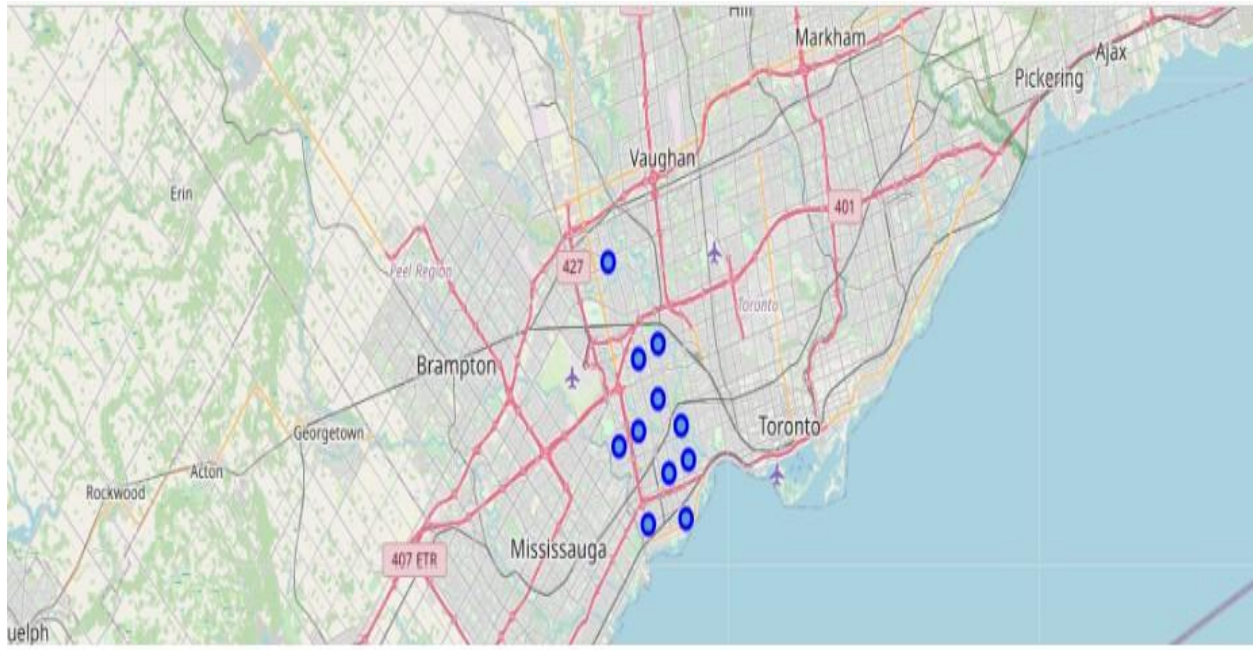
After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each

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 venue as follows:

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 Etobicoke



3. Methodology Section

Clustering Approach:

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

```
In [56]: neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

Etoibcoke_merged = df_2.iloc[:16,:]

# merge toronto_grouped with toronto_data to add Latitude/Longitude for each neighborhood
Etoibcoke_merged = Etoibcoke_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

Etoibcoke_merged.head()# check the last columns!
```

Out[56]:

	Postalcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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
5	M9A	Etoibcoke	Islington Avenue	43.667856	-79.532242	0	Pharmacy	Café	Skating Rink	Park	Playground	Shopping Mall	Grocery Store	Bank	Baseball Field	Coffee Shop
11	M9B	Etoibcoke	West Deane Park, Princess Gardens, Martin Gr...	43.650943	-79.554724	2	Pizza Place	Theater	Convenience Store	Yoga Studio	Golf Course	Gas Station	Fried Chicken Joint	Food & Drink Shop	Flea Market	Fast Food Restaurant
17	M9C	Etoibcoke	Eringate, Bloorale Gardens, Old Burnhamthor...	43.643515	-79.577201	0	Pizza Place	Shopping Plaza	Pharmacy	Coffee Shop	Café	Convenience Store	Liquor Store	Park	Electronics Store	Beer Store
70	M9P	Etoibcoke	Westmount	43.696319	-79.532242	0	Pizza Place	Golf Driving Range	Intersection	Chinese Restaurant	Golf Course	Middle Eastern Restaurant	Discount Store	Sandwich Place	Coffee Shop	Ice Cream Shop
77	M9R	Etoibcoke	Kingsview Village, St. Phillips, Martin Grov...	43.688905	-79.554724	0	American Restaurant	Sandwich Place	Intersection	Coffee Shop	Mobile Phone Shop	Chinese Restaurant	Pharmacy	Bus Line	Pizza Place	Gas Station

Most Common venues near Neighborhood

```
In [54]: import numpy as np
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{} {} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = Etoibcoke_grouped['Neighborhood']

for ind in np.arange(Etoibcoke_grouped.shape[0]):
    neighborhood_venues_sorted.iloc[ind, 1:] = return_most_common_venues(Etoibcoke_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

Out[54]:

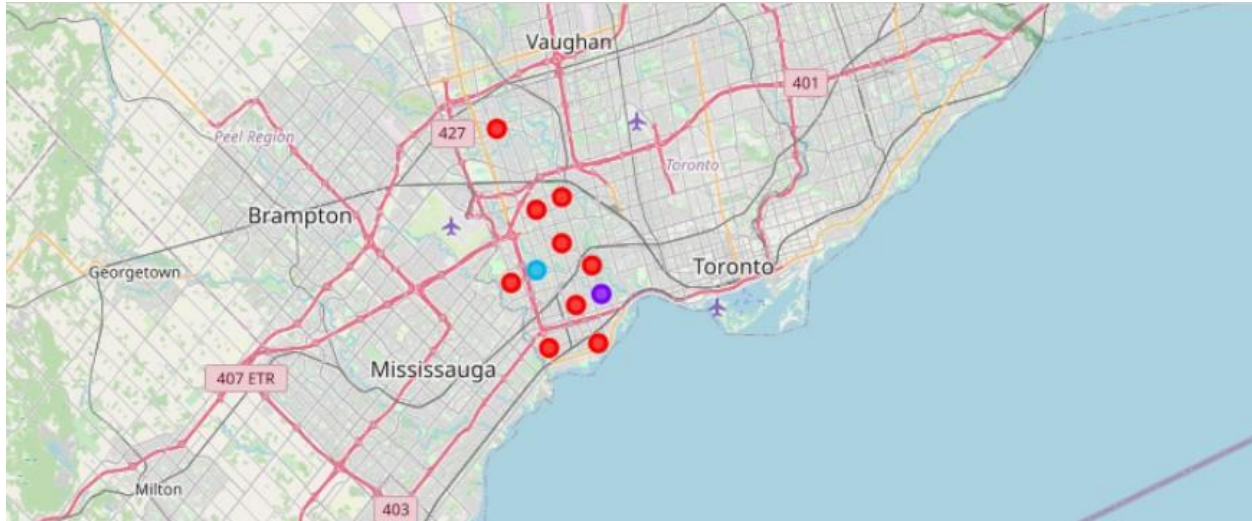
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Aldwood, Long Branch	Pizza Place	Convenience Store	Sandwich Place	Gym	Gas Station	Coffee Shop	Pub	Skating Rink	Dessert Shop	Discount Store
1	Eringate, Bloorale Gardens, Old Burnhamthor...	Pizza Place	Shopping Plaza	Pharmacy	Coffee Shop	Café	Convenience Store	Liquor Store	Park	Electronics Store	Beer Store
2	Islington Avenue	Pharmacy	Café	Skating Rink	Park	Playground	Shopping Mall	Grocery Store	Bank	Baseball Field	Coffee Shop
3	Kingsview Village, St. Phillips, Martin Grov...	American Restaurant	Sandwich Place	Intersection	Coffee Shop	Mobile Phone Shop	Chinese Restaurant	Pharmacy	Bus Line	Pizza Place	Gas Station
4	Mimico NW, The Queensway West, South of Bloo...	Gym / Fitness Center	Italian Restaurant	Liquor Store	Bakery	Bank	Burger Joint	Café	Convenience Store	Discount Store	Fast Food Restaurant

Work Flow:

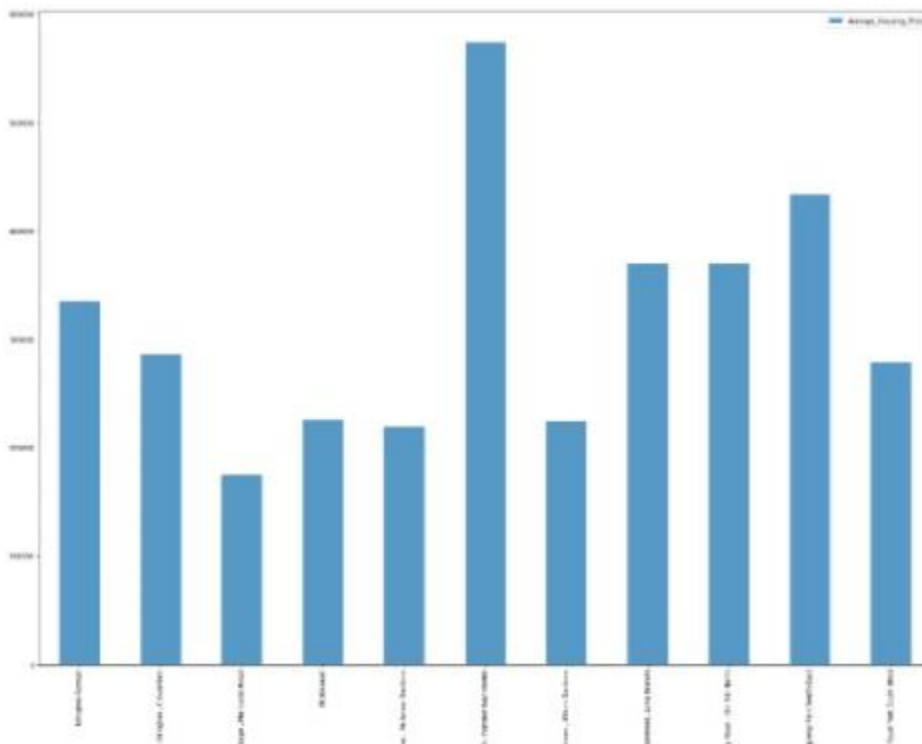
Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

4. Results Section

Map of Clusters in Etobicoke



Average Housing Price by Clusters in Etobicoke



Foursquare API:

This project has been made possible by the API on its online data gathering, and also its use as a data source of

5. Discussion Section

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 shifting 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.

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

6. Conclusion Section

In this project, using k-means cluster algorithm I separated the neighborhood into 5(Five) different clusters and for 11 different latitude 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.

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 Works:

This project can be continued for making it more precise in terms to find best house in Etobicoke. 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.

Libraries Which are Used to Develop the Project:

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.

Blog Link: <https://2018itom121.medium.com/capstone-project-finding-a-better-place-in-etobicoke-toronto-5fa500dad076>