

Capstone Project Final Report

The Competition between Neighborhoods

1. Introduction

In order to help people explore the equipment and facilities better in their neighborhood, this project will achieve this goal by assisting people making smart and effective decisions to select excellent neighborhood from a number of other neighborhoods in Scarborough, Toronto.

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.

The reason to choose a better neighborhood is mainly for the kids. Most people believe that a neighborhood with excellent reputation and reputed education would really benefit their kids. Moreover, appropriate house prices, convenience to shops and ease of accessing hospitals are also significant variants of decision making.

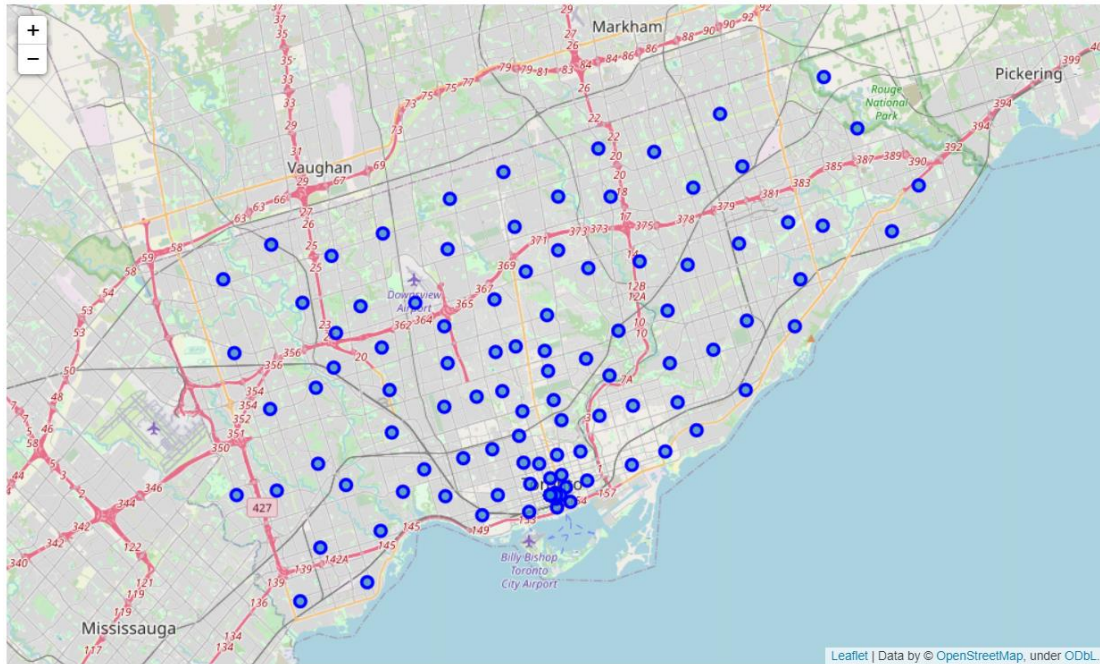
Based on comparative analysis between different neighborhoods, the goal of the project is to create an analysis among various features influencing people to migrate to Scarborough. Features include house price, school rating, crime rates, traffic connectivity, weather conditions, emergency management, water supply and waste treatment.

2. Data Description

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

Data about different venues in different neighborhoods of that specific borough is required for the project. In order to gain that data we will use Foursquare locational information, including venue names, locations, menus and even photos.

First of all, we should make a list of neighborhoods we need. After that we connect to the Foursquare API to gather information about venues inside every neighborhood, setting the radius to be 100 meters. The map of Scarborough, Toronto is shown below:



The detailed information are as follows:

1. Neighborhood names
2. Neighborhood coordinates(latitude, longitude)
3. Venue names
4. Venue coordinates(latitude, longitude)
5. Venue category

3. Problems and solutions

Core and major purpose of this project is to give a suggestion of a better neighborhood for people migrating here. So there are two problems to solve:

1. House prices listed in a sorted way(ascending or descending).
2. Location, tuition fees, rating and reviews of schools listed in a sorted way(ascending or descending).

This project would use Foursquare 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.

We decided to explore neighborhoods by comparing the similarities of two cities. By segmenting and grouping them into clusters, we could find similar neighborhoods in a big city like New York and Toronto. We will apply k-means clustering algorithm to cluster data, which is a form of unsupervised machine learning. As for work flow and data scope, 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 500km. K-Means approach is shown as below:

```
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
Scarborough_merged = df_2.iloc[:,16,:]

# merge toronto_grouped with toronto_data to add Latitude/Longitude for each neighborhood
Scarborough_merged = Scarborough_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')
Scarborough_merged.head()# check the last columns!
```

	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
0	M1A\n	Not assigned\n	Not assigned\n	43.64869	-79.38544	0	Coffee Shop	Hotel	Café	Restaurant	Japanese Restaurant	Beer Bar	Theater
1	M1B\n	Scarborough\n	Malvern, Rouge	43.81139	-79.19662	0	Zoo Exhibit	Construction & Landscaping	Fast Food Restaurant	Paintball Field	Escape Room	Doner Restaurant	Donut Shop
2	M1C\n	Scarborough\n	Rouge Hill, Port Union, Highland Creek	43.78574	-79.15875	1	Bar	Yoga Studio	Ethiopian Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store
3	M1E\n	Scarborough\n	Guildwood, Morningside, West Hill	43.76575	-79.17470	2	Park	Gym / Fitness Center	Athletics & Sports	Gymnastics Gym	Yoga Studio	Dog Run	Doner Restaurant
4	M1G\n	Scarborough\n	Woburn	43.76812	-79.21761	2	Fast Food Restaurant	Chinese Restaurant	Park	Coffee Shop	Yoga Studio	Elementary School	Doner Restaurant

Here are the results of the most common venues near the neighborhood using clustering method, shown as below.

```
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'] = Scarborough_grouped['Neighborhood']

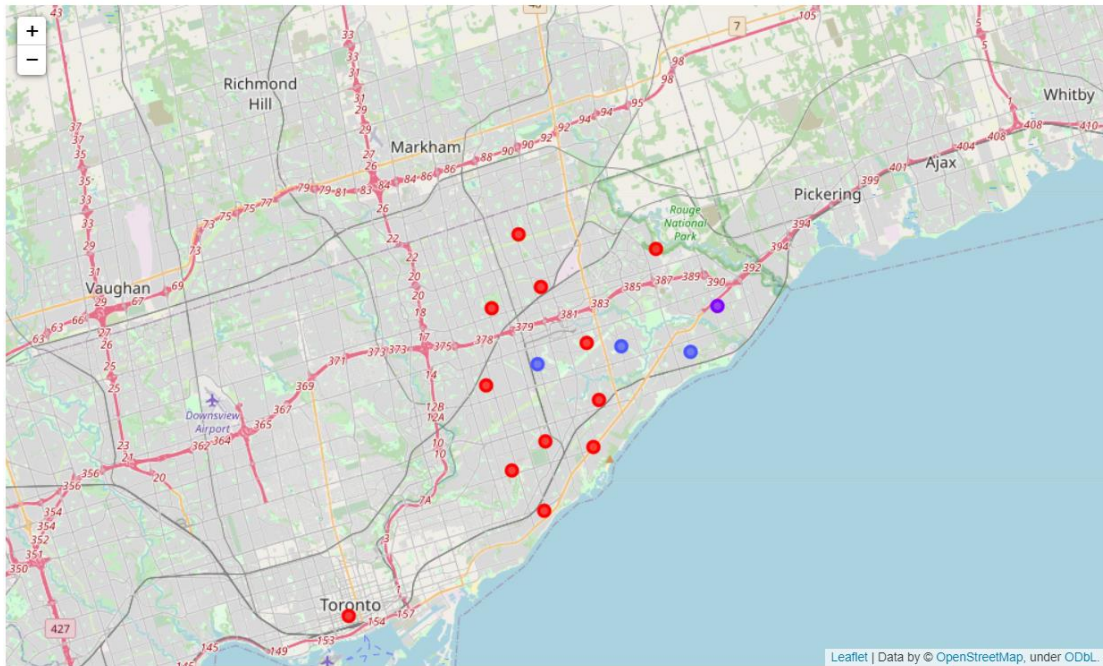
for ind in np.arange(Scarborough_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(Scarborough_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

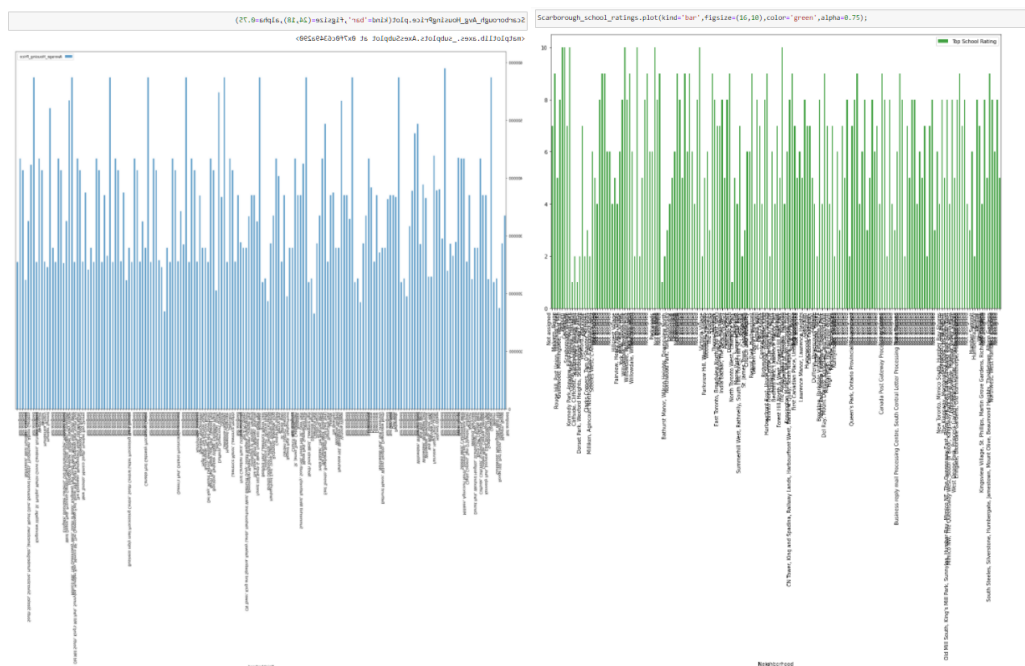
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Chinese Restaurant	Shopping Mall	Hong Kong Restaurant	Sandwich Place	Sushi Restaurant	Supermarket	Latin American Restaurant	Pharmacy	Bubble Tea Shop	Newsagent
1	Alderwood, Long Branch	Gas Station	Sandwich Place	Pizza Place	Pub	Pharmacy	Coffee Shop	Gym	Eastern European Restaurant	Distribution Center	Dive Bar
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Park	Mediterranean Restaurant	Fried Chicken Joint	Convenience Store	Restaurant	Deli / Bodega	Sandwich Place	Middle Eastern Restaurant	Men's Store
3	Bayview Village	Dog Run	Flower Shop	Gas Station	Trail	Park	Asian Restaurant	Electronics Store	Doner Restaurant	Donut Shop	Dumpling Restaurant
4	Bedford Park, Lawrence Manor East	Coffee Shop	Sandwich Place	Italian Restaurant	Pet Store	Juice Bar	Restaurant	Sports Club	Thai Restaurant	Intersection	Pub

4. Results

Here is the map of clusters in the neighborhood of Scarborough, Toronto, shown as below:



And these two charts are average housing price by clusters in Scarborough and school ratings by clusters in Scarborough.



5. Conclusion

Using k-means cluster algorithm, I separated the neighborhood into 10 different clusters and for 180 different latitude and longitude from dataset, which have very-similar neighborhoods around them. As the charts above, the results show 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.

6. Future Works:

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.