Final Report- The Battle of Neighborhoods for finding a better place in Scarborough, Toronto

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 decisions on selecting a great neighborhood out of numbers of other neighborhoods in Scarborough, Toronto.

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

This project aims to create an analysis of features for people migrating to Scarborough to search for the 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 an emergency, water resources both fresh and wastewater, and excrement conveyed in sewers and recreational facilities.

It will help people to get the 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.

Data

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

Will use the Scarborough dataset which we scraped 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. 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 every neighborhood. For each neighborhood, we have chosen the radius to be 100 meters.

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

- 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

Map of Scarborough
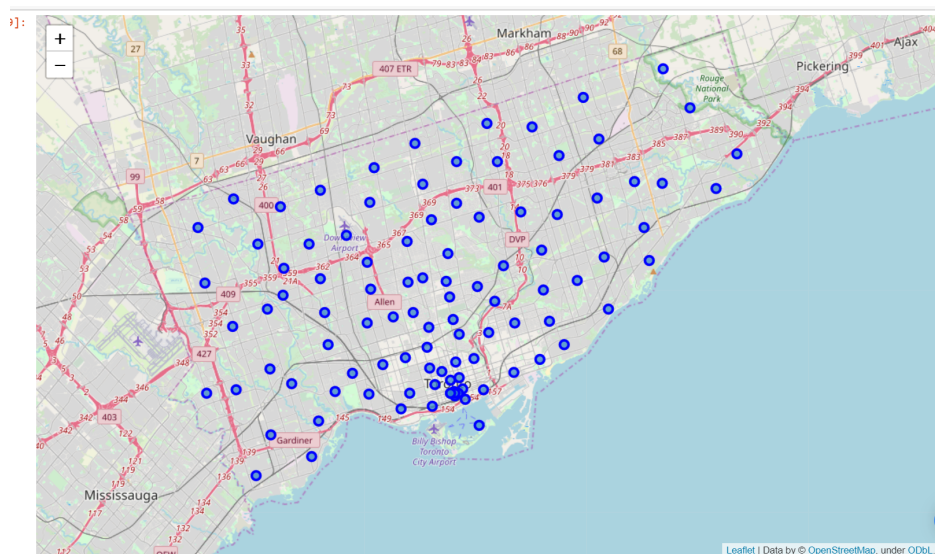
Methodology

Clustering Approach

To compare the similarities of the two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in big cities 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

```
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!
```

t[36]:

orough	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
orough	Rouge, Malvern	43.811525	-79.195517	0	Zoo Exhibit	Financial or Legal Service	Fast Food Restaurant	Construction & Landscaping	Fish & Chips Shop	Filipino Restaurant	Field	Fish Market	Farmers Market	Doner Restaurant
orough	Highland Creek, Rouge Hill, Port Union	43.785665	-79.158725	0	Bar	Falafel Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant	Event Space	Yoga Studio
orough	Guildwood, Morningside, West Hill	43.765815	-79.175193	2	Park	Gym / Fitness Center	Pool	Fried Chicken Joint	Indian Restaurant	Athletics & Sports	Ethiopian Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant
orough	Woburn	43.768369	-79.217590	0	Coffee Shop	Fast Food Restaurant	Business Service	Park	Yoga Studio	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant
orough	Cedarbrae	43.769688	-79.239440	0	Flower Shop	Athletics & Sports	Thai Restaurant	Bank	Bakery	Caribbean Restaurant	Hakka Restaurant	Indian Restaurant	Eastern European Restaurant	Electronics Store

Most Common venues near Neighborhood

```
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'] = 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()
```
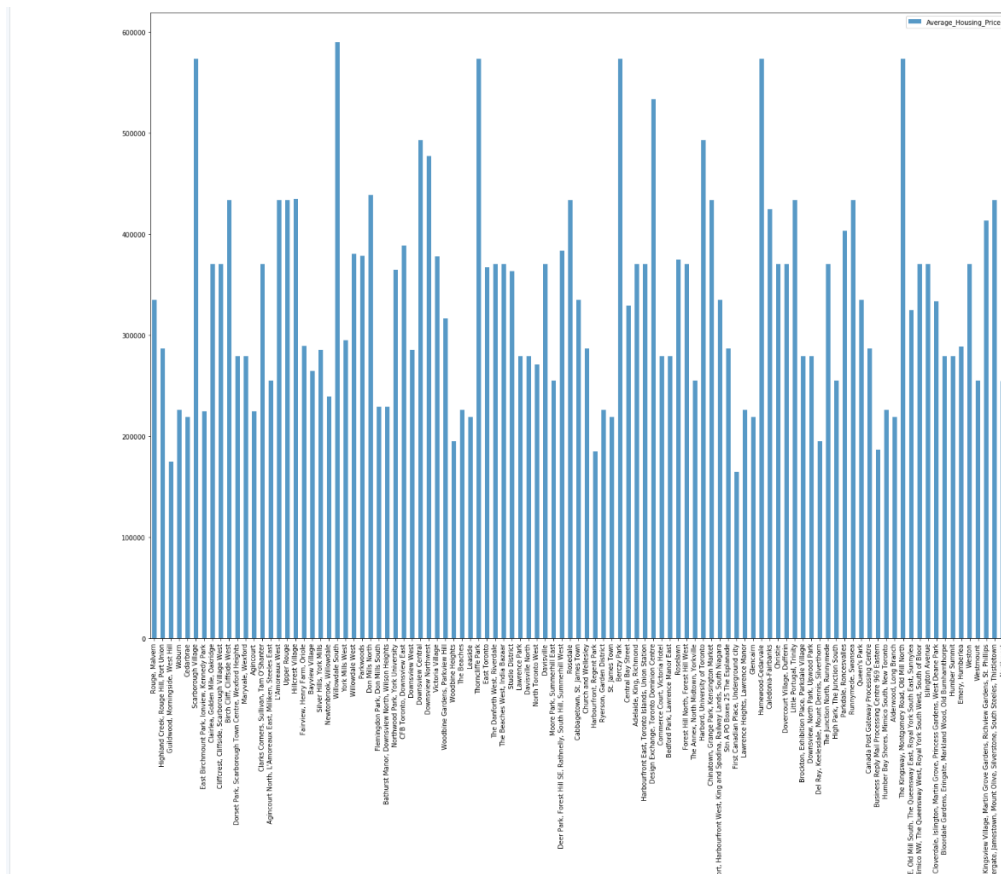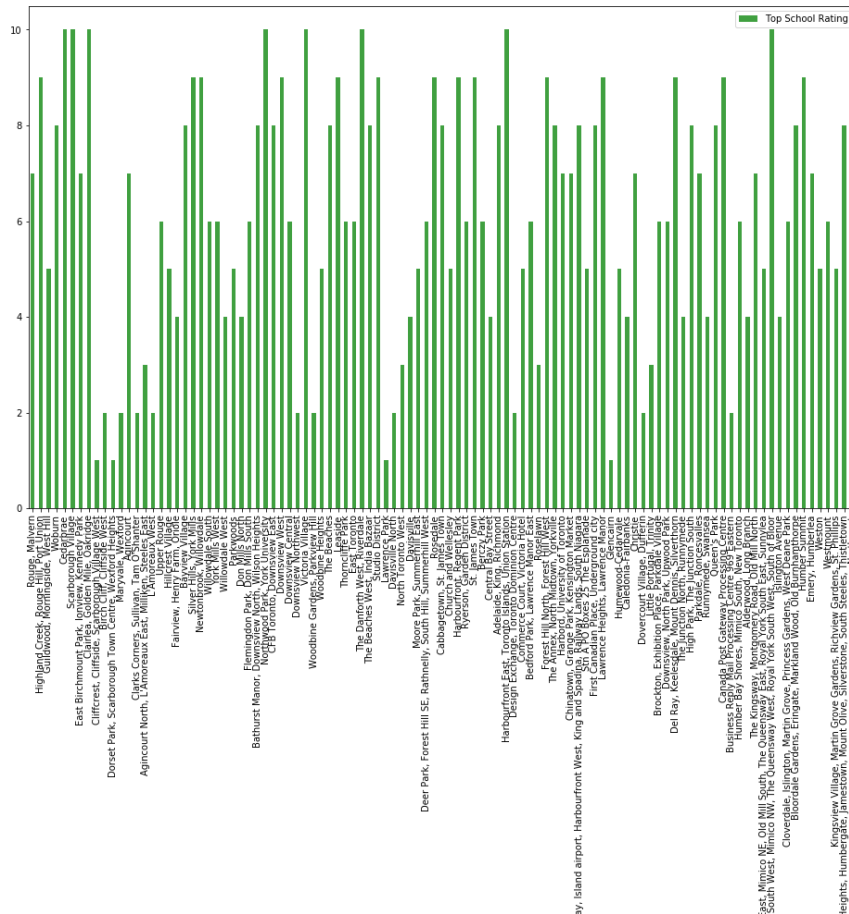
t[34]:

	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	Ethiopian Restaurant

Work Flow

Using credentials of Foursquare API features of nearby 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.

Results

Map of Clusters in Scarborough

Average Housing Price by Clusters in Scarborough

School Ratings by Clusters in Scarborough

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 on the rise.

Foursquare API

This project has 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.

Discussion

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 is 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.

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

Conclusion

In this project, using the k-means cluster algorithm I separated the neighborhood into 10 different clusters and for 103 different latitudes and longitude from the 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.