Final_Report_Battle_of_Neighborhoods

November 25, 2020

1 Final Report for Capstone Project—The Battle of Neighborhoods Finding a Better Place in Scarborough, Toronto

1.1 Name : Kipngeno Kirui

1.2 Table of contents

1.3 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 Scarborough, Toranto.

Scarborough is an administrative division of Toronto, Ontario, Canada. Situated atop the Scarborough Bluffs, it occupies the eastern part of the city. Scarborough is contained within the borders of Victoria Park Avenue on the west, Steeles Avenue to the north, Rouge River and the city of Pickering to the east, and Lake Ontario to the south. It borders Old Toronto, East York and North York in the west and the city of Markham in the north. Scarborough was named after the English town of Scarborough, North Yorkshire.

Lots of people are migrating to various states of Canada and needed lots of research for good housing prices and reputated schools for their children. This project is for these people who are looking for better neighborhoods. For ease of access to Kaffe shops, good schools, supermarkets, medical shops, grocery shops, malls, theatre, hospitals, people with same characters, e.t.c.

This project aims 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.

1.3.1 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.

The vast majority of Scarborough's population is composed of immigrants who have arrived in the last four decades, and their descendants. In 2006, 57% of residents were foreign-born. Visible minorities make up 67.4% of the population. The population is 22.0% South Asian, 19.5% Chinese, 10.3% Black, and 6.5% Filipino. The remaining visible minority groups each represented less than 2% of the population. The immigrant population has created vibrant multicultural locales in various areas of Scarborough.

1.3.2 The Problem

The major purpose of this project is to suggest a better neighborhood in a new city for an individual who would like to live there. We look onto factors such as social presence in society in terms of like minded people, connectivity to the airport, bus stands, city centers, markets and other daily needs things nearby.

- 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

Scarborough, Toronto Page in Wikipedia

1.4 Data

1.4.1 The data

In this project, we will use Scarborough dataset which we scrapped from wikipedia on Week 3. The dataset consists of latitudes and longitudes and zip codes of various neighbourhoods.

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

This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Only the first three characters are listed, corresponding to the Forward Sortation Area.

1.4.2 Foursquare API Data

We will need data about different venues in different neighborhoods for 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 venues and events within an area of interest. Such information includes venue names, locations, menus, ratings 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

1.5 Analysis

The initial dataframe that I acquired consisted of 103 rows and 3 variabes. These variables were **Postalcode**,**Borough** and **Neighborhood**. After various cleaning and adding variables, the shape of the dataframe became 103 by 5 where two columns were added, which are, **Latitude** and **Longitude**, as shown below:

```
In [28]: # Check the first 4 rows of the dataframe
         df2.head(4)

Out[28]:
```

	Postalcode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.81139	-79.19662
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78574	-79.15875
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.76575	-79.17470
3	M1G	Scarborough	Woburn	43.76812	-79.21761

I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as 100 venue and the radius 500 meter for each borough from their given latitude and longitude informations as shown above.
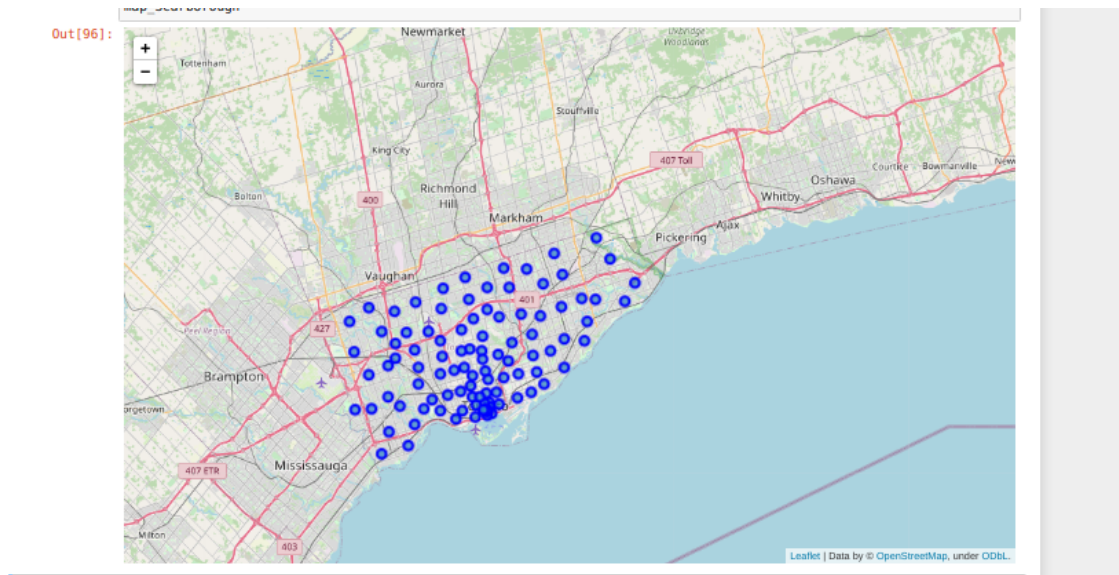
I used python folium library to visualize geographic details of Scarborough and its boroughs and I created a map of Scarborough with boroughs superimposed on top. I used latitude and longitude values to get the visual as below:

`Neighborhoods_Venues_Sorted.head(4)`

	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	Shopping Mall	Pool	Bakery	Sushi Restaurant	Latin American Restaurant	Supermarket	Pharmacy	Bubble Tea Shop	Skating Rink	Shanghai Restaurant
1	Alderwood, Long Branch	Sandwich Place	Pub	Performing Arts Venue	Gym	Coffee Shop	Pharmacy	Gas Station	Print Shop	Pizza Place	Convenience Store
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Mediterranean Restaurant	Arcade	Middle Eastern Restaurant	Mobile Phone Shop	Fried Chicken Joint	Sushi Restaurant	Park	Restaurant	Sandwich Place
3	Bayview Village	Flower Shop	Park	Asian Restaurant	Trail	Gas Station	Yoga Studio	Donut Shop	Eastern European Restaurant	Electronics Store	Elementary School

K-Means Clustering Approach

In summary of this graph with the unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for each borough in below table.

1.6 Clustering

We have some common venue categories in boroughs. In this reason I used unsupervised learning **K-means algorithm** to cluster the boroughs. K-Means algorithm is one of the most common cluster method of unsupervised learning.

Clustering is one of the most common exploratory data analysis technique used to get an intuition about the structure of the data. It can be defined as the task of identifying subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different. In other words, we try to find homogeneous subgroups within the data such that data points in each cluster are as similar as possible according to a similarity measure such as euclidean-based distance or correlation-based distance.

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 Nairobi and Mombasa. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

Kmeans algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.

First, I will run K-Means to cluster the boroughs into 3 clusters because when I analyze the K-Means with elbow method it ensured me the 3 degree for optimum k of the K-Means.

The Elbow Method showing the optimal k
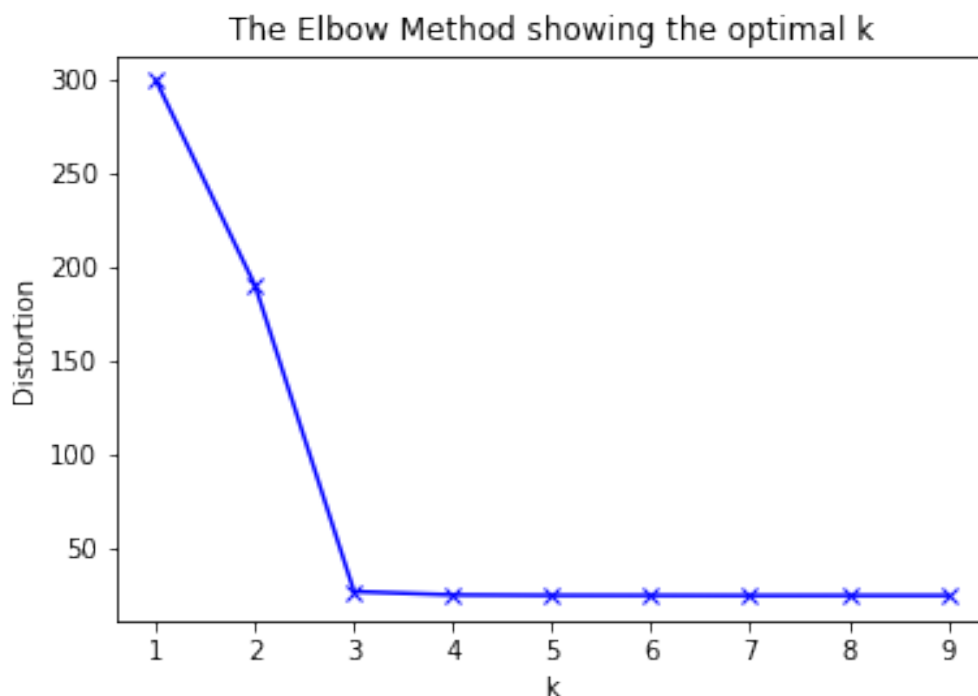
The following shows part of my merged table with cluster labels for each borough:

Out[60]:

	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	M1B	Scarborough	Malvern, Rouge	43.81139	-79.19662	0	Zoo Exhibit	Fast Food Restaurant	Paintball Field	Fish & Chips Shop	Filipino Restaurant	Field	Fish Market	
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78574	-79.15875	2	Bar	Construction & Landscaping	Fish & Chips Shop	Falafel Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.76575	-79.17470	1	Park	Athletics & Sports	Gym / Fitness Center	Gymnastics Gym	Yoga Studio	Dog Run	Donut Shop	R
3	M1G	Scarborough	Woburn	43.76812	-79.21761	1	Park	Chinese Restaurant	Fast Food Restaurant	Coffee Shop	Yoga Studio	Ethiopian Restaurant	Dumpling Restaurant	R

1.6.1 1st Problem : Sorted list of house in terms of housing prices in a ascending or descending order

We can also examine the **Average Housing Price by Clusters in Scarborough**. Thus, we can create a bar chart which may help us to find proper labels for each cluster.

1.6.2 2nd Problem: Sorted list of schools in terms of location, fees, rating and reviews
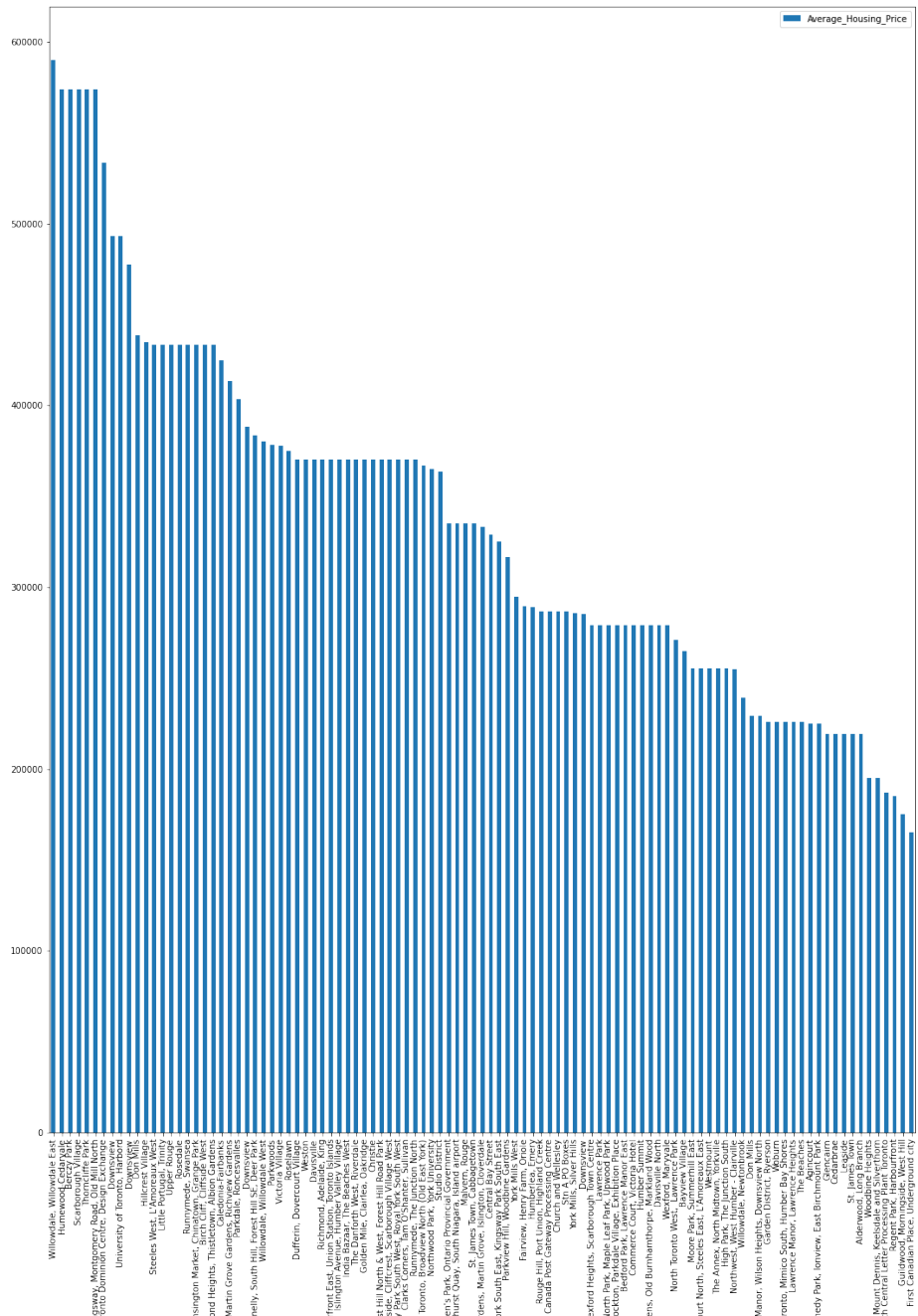
Here we are going to get the barchat of **School Ratings by Clusters in Scarborough** as shown below:
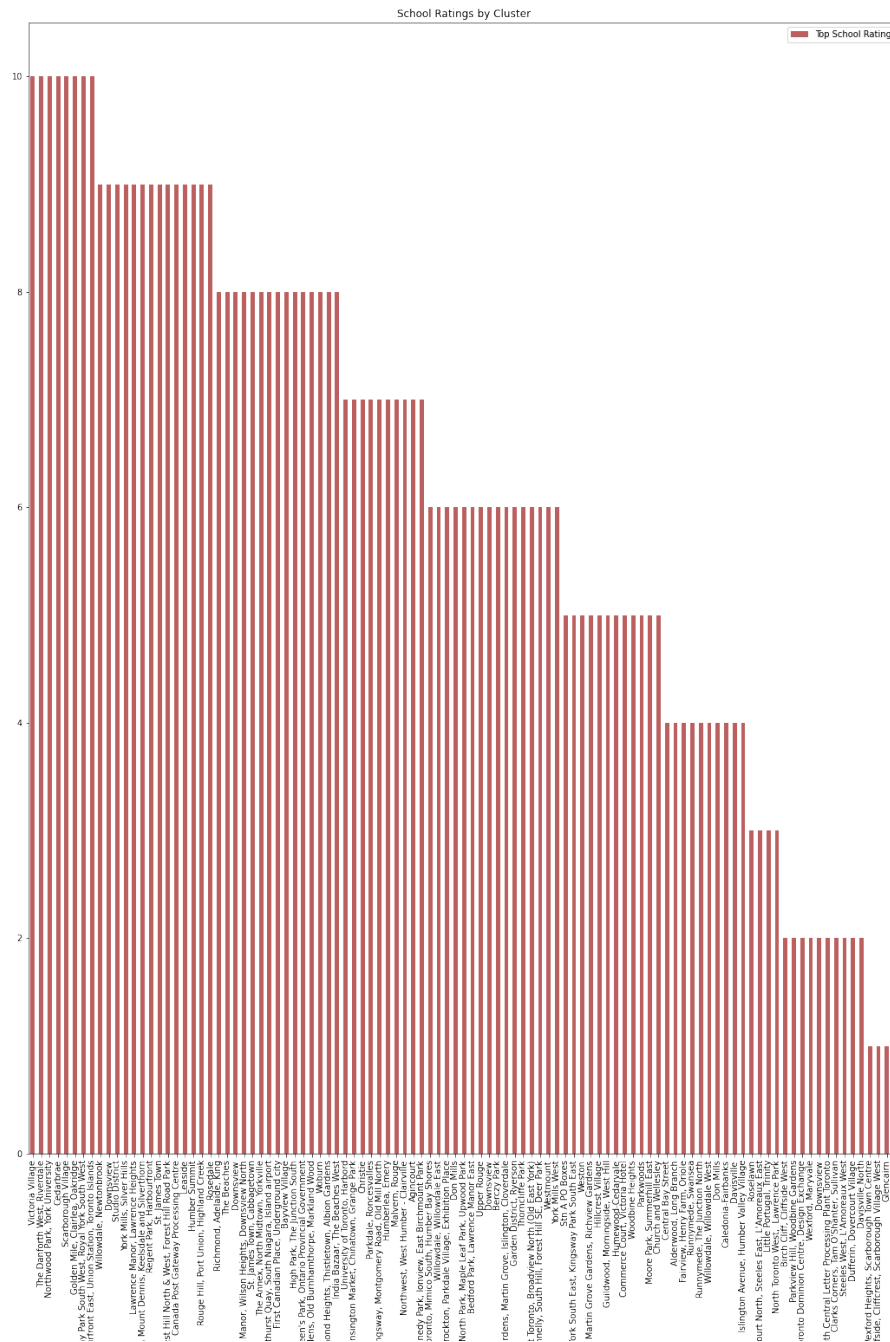
1.7 Discussion

As indicated before, Scarborough is a very large city and attracts many tourists. In addition, Llts of people are migrating to various states of Canada and needed lots of research for good housing prices and reputated schools for their children.

There is such a complexity, very different approaches can be tried in clustering and classification studies. Moreover, it is obvious that not every classification method can yield the same high quality results for this city.

The major purpose of this project was to suggest a better neighborhood in a new city for the person who are shiffting there using factors such as social presence in society in terms of like minded people, connectivity to the airport, bus stands, city centers, markets and other daily needs things nearby.

We used the data from Scarbolough city in Toronto and tried to solve the following two problems:

```
1. Sort the list of house in terms of average housing prices in a ascending or descending order,
```

```
2. Sort the list of schools of Scarborough in terms of location, fees, rating and reviews.
```

With this information, a person moving in to Scarborough will already know the average house prices in various neighborhoods and also the high performing schools in the different neighborhoods. This information is very critical especially for a person moving in with children who needs to go to school and also, since it is a large family, he/she will already be aware of the prices of the houses he/she is supposed to pick and plan a priori.

I used the Kmeans algorithm as part of this clustering study. When I tested the Elbow method, I set the optimum k value to 3.

1.8 Conclusion

In this project, by using the K-means cluster algorithm, I was able to separate the neighborhood into 10(Ten) different clusters and for 103 different latitudes and longitudes from dataset, which have very-similar neighborhoods around them.

As shown in the charts above, neighborhoods based on average house prices and school rating have been made. This information is very important to an individual who has children and would like to migrate to Scarborugh.

People are turning to big cities to start a business or work. For this reason, people can achieve better outcomes through their access to the platforms where such information is provided a priori.

Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.

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 Work   This work can be extended to finding the best hospital care, supermarkets with best rates and also stores with best rates for buying clothing. This will ensure that an individual moving to Scarborough will have the best informstion for the *basic needs* a human being needs to have, i.e., food, shelther, clothing and also water, in extension.

1.9   References

- [1] Scarborough, Toronto, Wikipedia Page
- [2] Foursquare Developers Website
- [3] List of postal codes in Canada Data in Wikipedia
- [4] Google Website
- [5] Great Schools Website

[ ]: