

The Battle of Neighborhoods

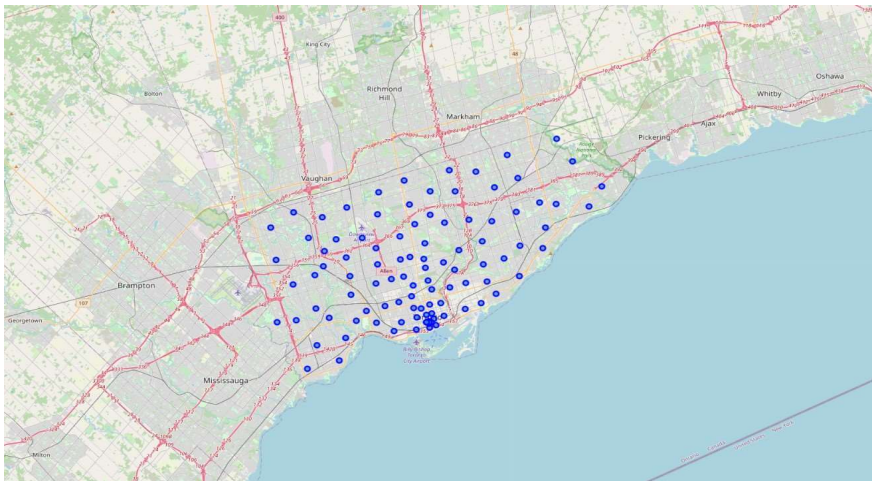
1. Introduction

- ▶ The purpose of this Capstone 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, 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.

2. Data Section

- ▶ 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.
- ▶ 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.
- ▶ 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: Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venue, Name of the venue, Venue Latitude, Venue Longitude, Venue Category.

Map of Scarborough



Using K-Means Clustering Approach | Most Common Venue

```

In [36]: neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
         Scarborough_merged = df_2.iloc[:,1:]
         # 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!

Out[36]:

```

rough 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
rough Midland	43.811525	-79.195517	0	Zoo	Financial or Legal Service	Fast Food Restaurant	Construction & Landscaping	Fish & Chips Shop	Filipino Restaurant	Field	Fish Market	Farmers Market	Donut Restaurant
rough Highland Creek, Rouge Hill, Port Union	43.785665	-79.158725	0	Bar	Faithful Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant	Event Space	Yoga Studio
rough Guildwood, Montclare, 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
rough Wilsdon	43.788369	-79.217590	0	Coffee Shop	Fast Food Restaurant	Business Service	Park	Yoga Studio	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant
rough Cedarbrae	43.789588	-79.239440	0	Flower Shop	Athletics & Sports	Thai Restaurant	Bank	Bakery	Caribbean Restaurant	Hakka Restaurant	Indian Restaurant	Eastern European Restaurant	Electronics Store

Map of Clusters

```

In [37]: kclusters = 18

```

Most Common Venues near Neighborhood

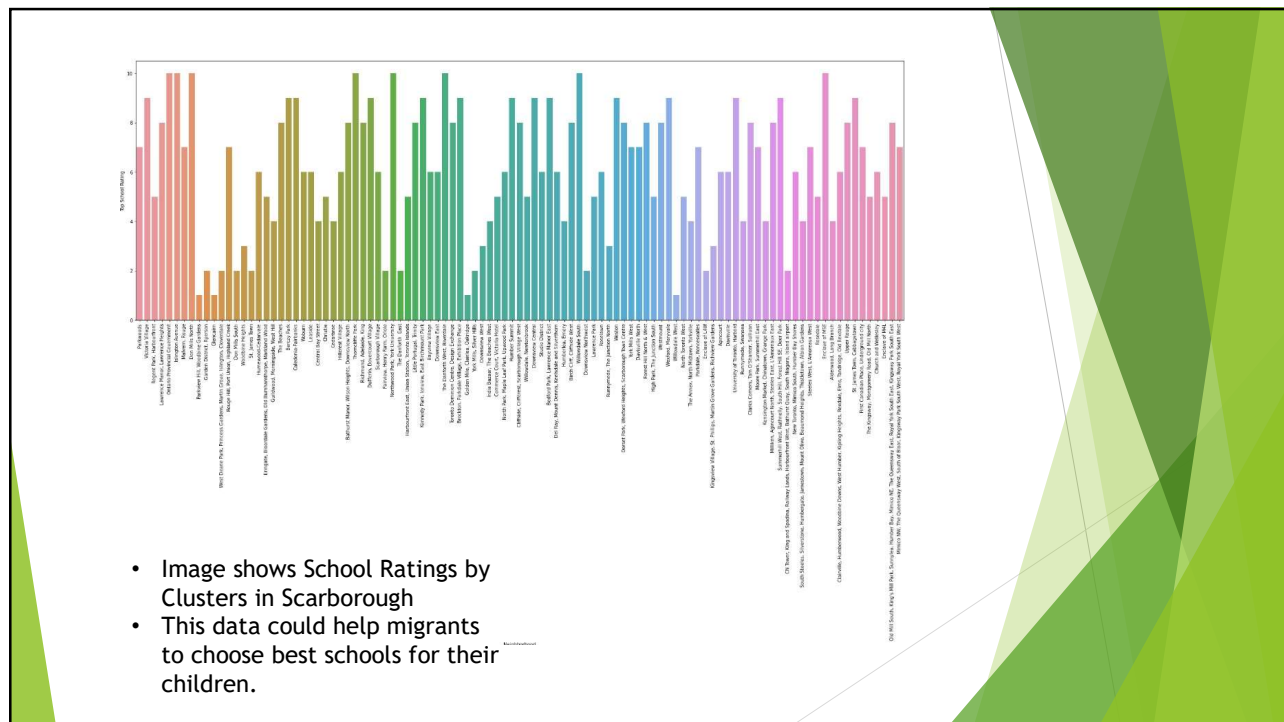
```

In [34]: import numpy as np
         num_top_venues = 18
         indicators = ['1st', 'nd', 'rd']
         columns = ['Neighborhood']
         for ind in np.arange(num_top_venues):
             try:
                 columns.append('({})th 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()

Out[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	Adeleide, King, Richmond	Coffee Shop	Cafe	Hotel	Gastropub	Burger Joint	Asian Restaurant	Bar	Restaurant	American Restaurant	Breakhouse
1	Agincourt	Chinese Restaurant	Shopping Mall	Pizza Place	Supermarket	Sushi Restaurant	Breakfast Spot	Pet Shop	Mediterranean Restaurant	Coffee Shop	Pool
2	L'Amoreux East, Milliken, St.	Pharmacy	Sandwich Place	Sushi Restaurant	Donut Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant
3	Alton Gardens, Beaumont Heights, Scarborough	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	Faithful Restaurant



Conclusion

- ▶ In this Capstone project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 103 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.
- ▶ This Capstone project can be continued for making it more precise in terms to find best house in Scarborough.
- ▶ Other Clustering algorithms can be used and compared with K-mean algorithm.