

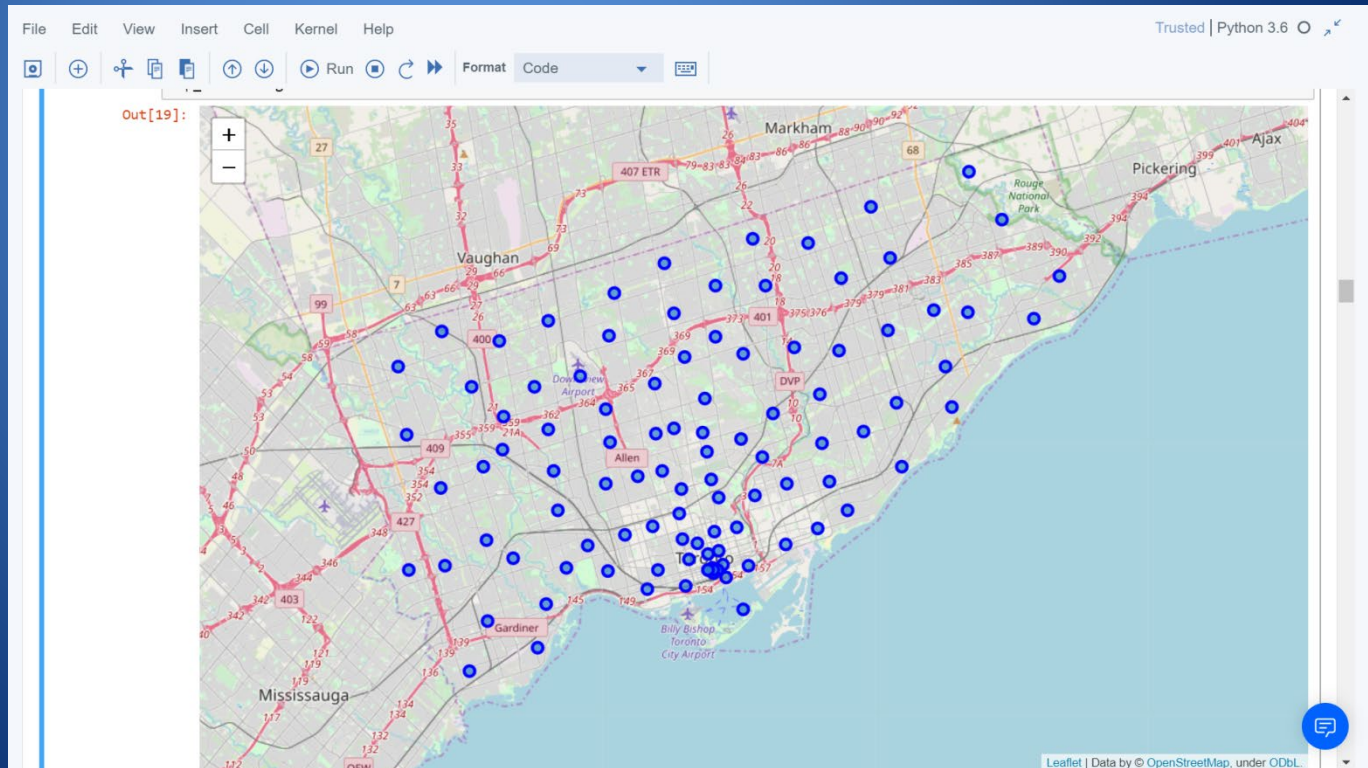
# Finale : Battle of the Neighbourhoods

## Introduction

The purpose of this Project is to, through the use of Foursquare Location Data, help people in exploring better facilities around their neighbourhood. Specifically, we take a look at the various neighbourhoods in Scarborough, Toronto.

## The Location

Scarborough, Toronto



# Foursquare API

This project would use Four-square API as its prime data gathering source as it has a database of millions of places.

For area/borough specific data in Toronto, 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.

We are connecting to the Foursquare API to gather information about venues inside each and every neighbourhood. The information obtained per venue include but are not limited to:

1. Neighbourhood
2. Neighbourhood Latitude
3. Neighbourhood 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

## **We will be using these Python Libraries for this Project:**

1. Pandas: For creating and manipulating dataframes.
2. Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
3. Scikit Learn: For importing k-means clustering.
4. JSON: Library to handle JSON files.
5. XML: To separate data from presentation and XML stores data in plain text format.
6. Geocoder: To retrieve Location Data.
7. Beautiful Soup and Requests: To scrap and library to handle http requests.
8. Matplotlib: Python Plotting Module.

# Data Section

We are primarily making use of Toronto dataset we scrapped in Week 3. Data

Link: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

## Methodology Section

Clustering Approach: To compare the similarities of two cities, we decided to explore neighbourhoods, segment them, and group them into clusters to find similar neighbourhoods 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.

We will be using K-Means Clustering Approach to find the most common venues near Neighbourhood

Out[36]:

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
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
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
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
Scarburn	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
Cedarbrae	43.769688	-79.230000	0	Flower Shop	Athletics & Sports	Thai Restaurant	Bank	Bakery	Caribbean Restaurant	Hakka Restaurant	Indian Restaurant	Eastern European Restaurant	Electronics Store

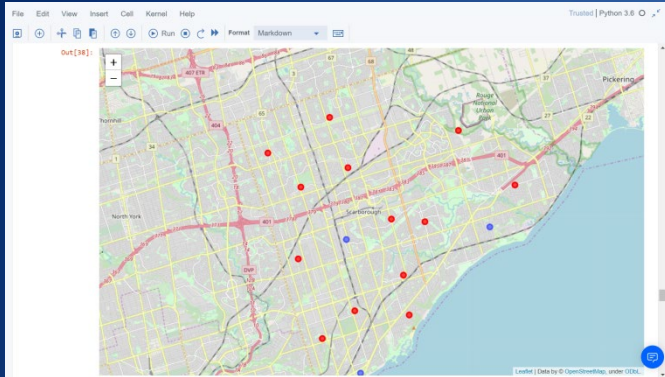
Map of Clusters

7]: kclusters = 10

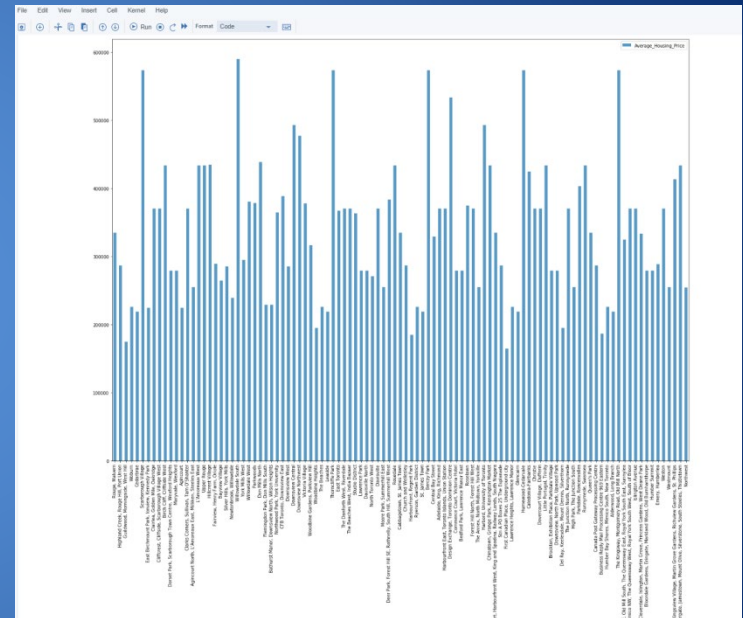
Work Flow: Using credentials of Foursquare API features of near-by places of the neighbourhoods would be mined. Due to http request limitations the number of places per neighbourhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.



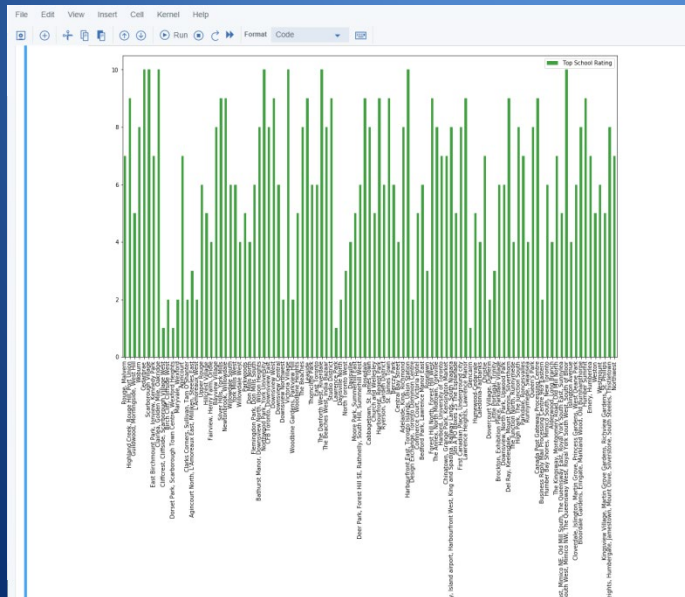
## Map of Clusters in Scarborough



## Average Housing Prices



## School Ratings



## Discussion Section

The major purpose of this project is to shortlist a good neighbourhood in a new city for the person who are shifting there.

1. We have since sorted list of house in terms of housing prices in a ascending or descending order
2. And sorted list of schools in terms of location, fees, rating and reviews

## Conclusion

Using k-means cluster algorithm, we separated the neighbourhood into different clusters and have short-listed the various latitude and longitude from the dataset. Based on the ranking of house prices and proximity to schools, users will be able to make preliminary analysis on the suitability of locating their next home.