

---

C ∠ (P) (S) ↗ ↘ □ ∅ ○ P ⊙ □ ⊕ ○ ▭ ↗ ↘ -

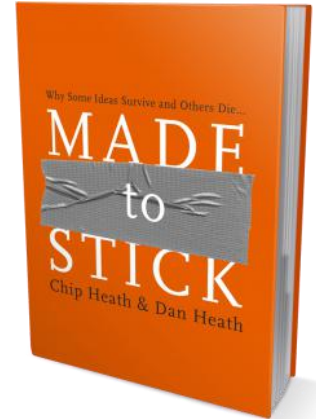
## The Battle of Neighborhoods Finding a Better Place in Scarborough, Toronto

A Presentation by Ananya

---

# GOOOAL!

Lots of people are migrating to various states of Canada and need 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. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, like minded people, etc.





# 1. Intro

→ **DATA COLLECTION**

Highlight what's new, unusual, or surprising.

→ **METHODOLOGY**

A little behind the scenes

→ **RESULTS**

Final outcome

---

**We will use Scarborough data from Toronto Dataset which we scrapped from wikipedia . Dataset consisting of latitude and longitude, Postal Code of the neighborhood.**

**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)

---

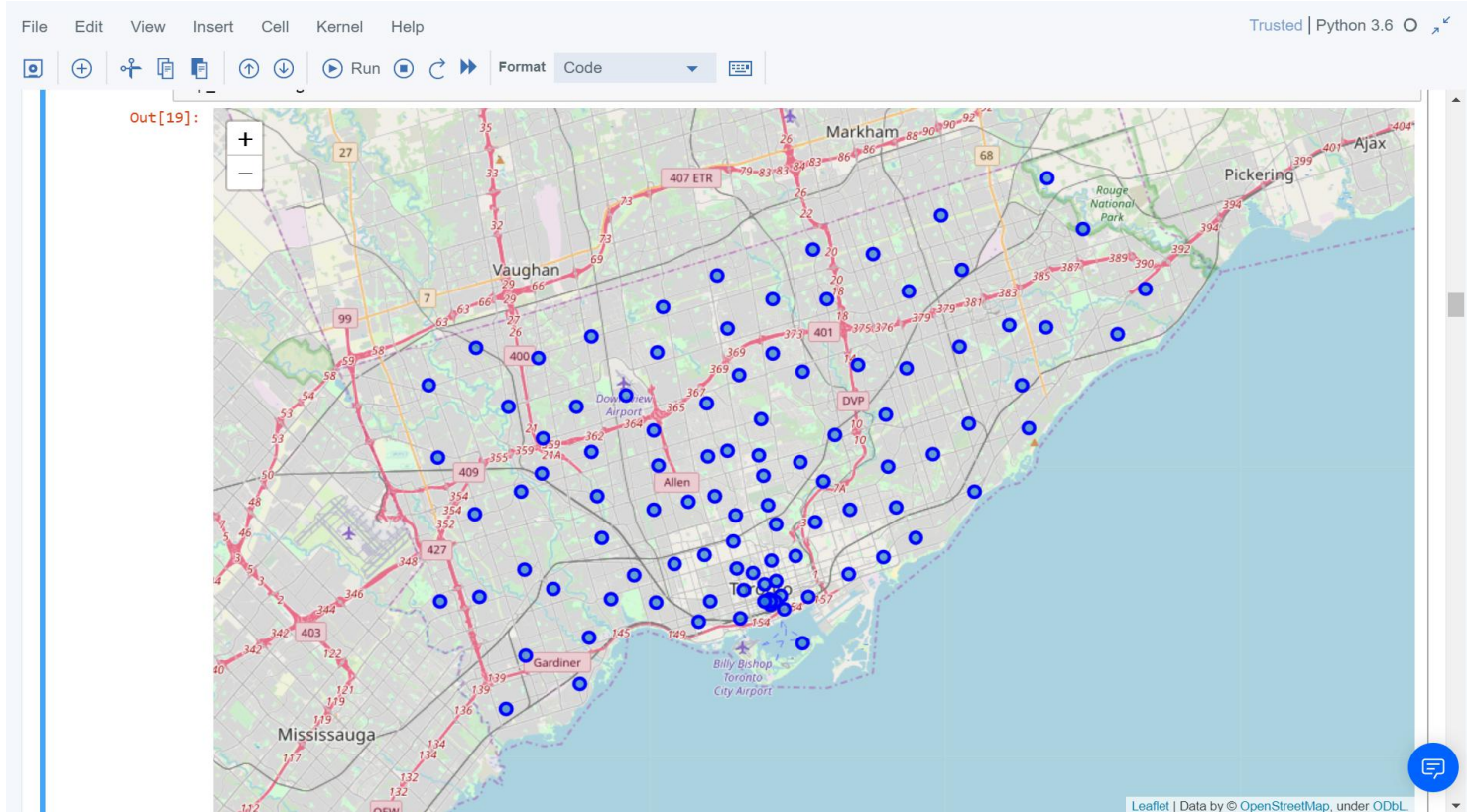
### **Foursquare API Data:**

We will need data about different venues in different neighborhoods of the specific borough - Scarborough. In order to gain that information we will use "Foursquare" location 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.

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

# Map Marking All Scarborough Neighborhoods



## DATASET OF SCARBOROUGH

	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
4	M1H	Scarborough	Cedarbrae	43.76944	-79.23892
5	M1J	Scarborough	Scarborough Village	43.74446	-79.23117
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	43.72582	-79.26461
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.71289	-79.28506
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	43.72360	-79.23496
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.69510	-79.26466



# DATASET OBTAINED FROM FOURSQUARE API

	name	categories	lat	lng
0	Disney Store	Toy / Game Store	43.775537	-79.256833
1	American Eagle Store	Clothing Store	43.776012	-79.258334
2	St. Andrews Fish & Chips	Fish & Chips Shop	43.771865	-79.252645
3	SEPHORA	Cosmetics Shop	43.775512	-79.258288
4	Shoppers Drug Mart	Pharmacy	43.773305	-79.251662

### **3. Methodology Section**

#### **Clustering Approach:**

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

## MOST COMMON VENUE IN NEIGHBORHOOD

### Most Common venues near neighborhood

```
In [31]: 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))

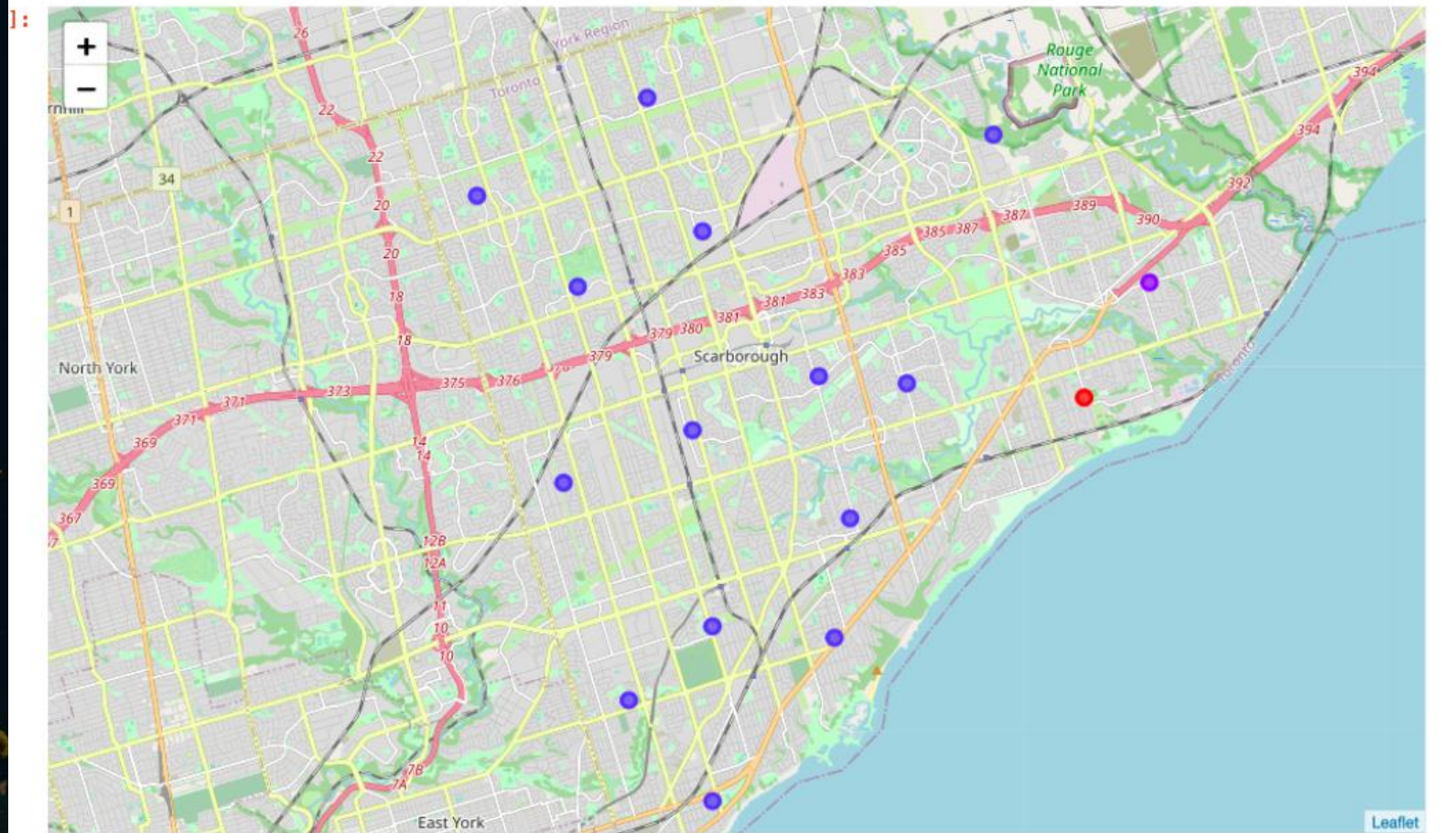
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = Scarborough_grouped['Neighborhood']

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_
neighborhoods_venues_sorted.head()
```

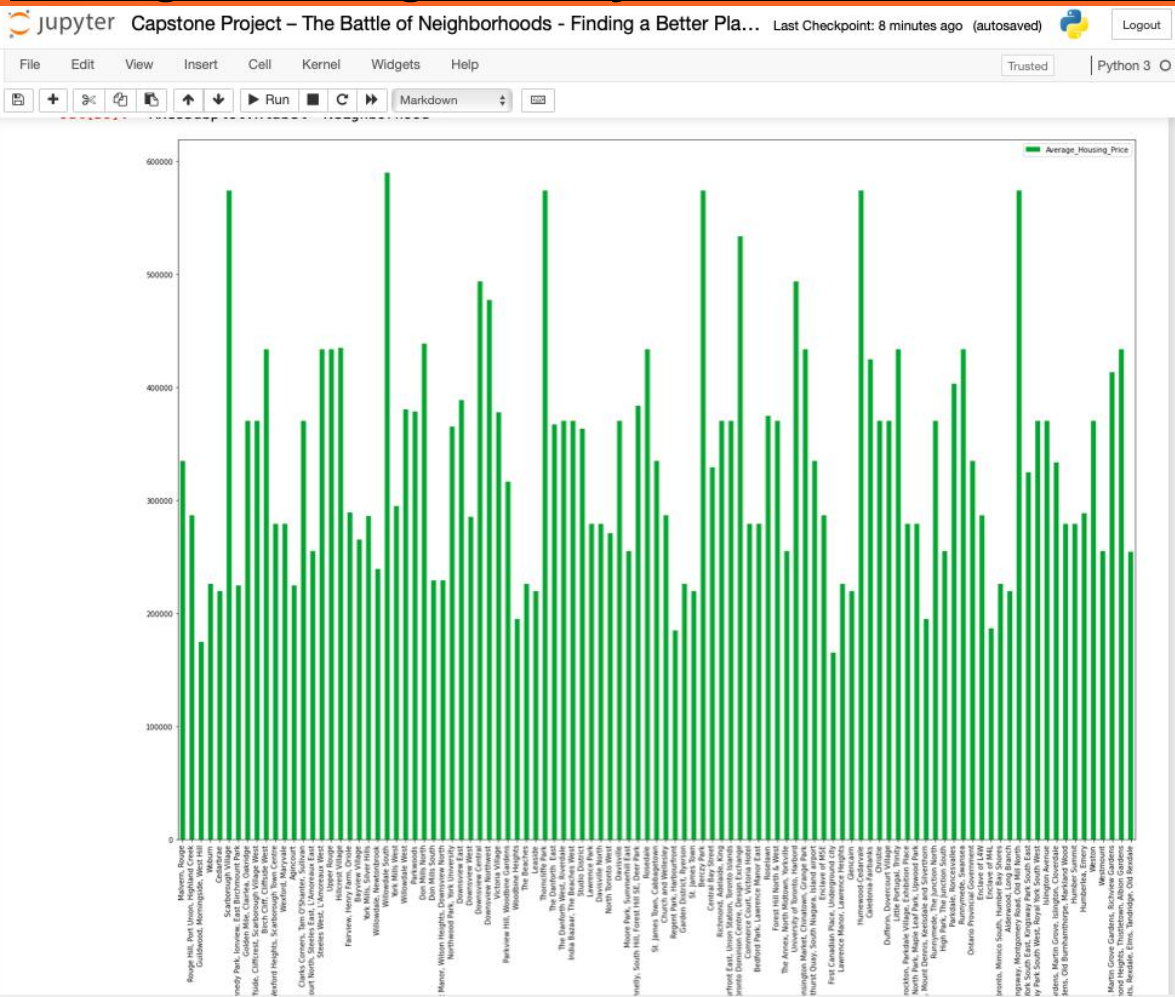
Out[31]:

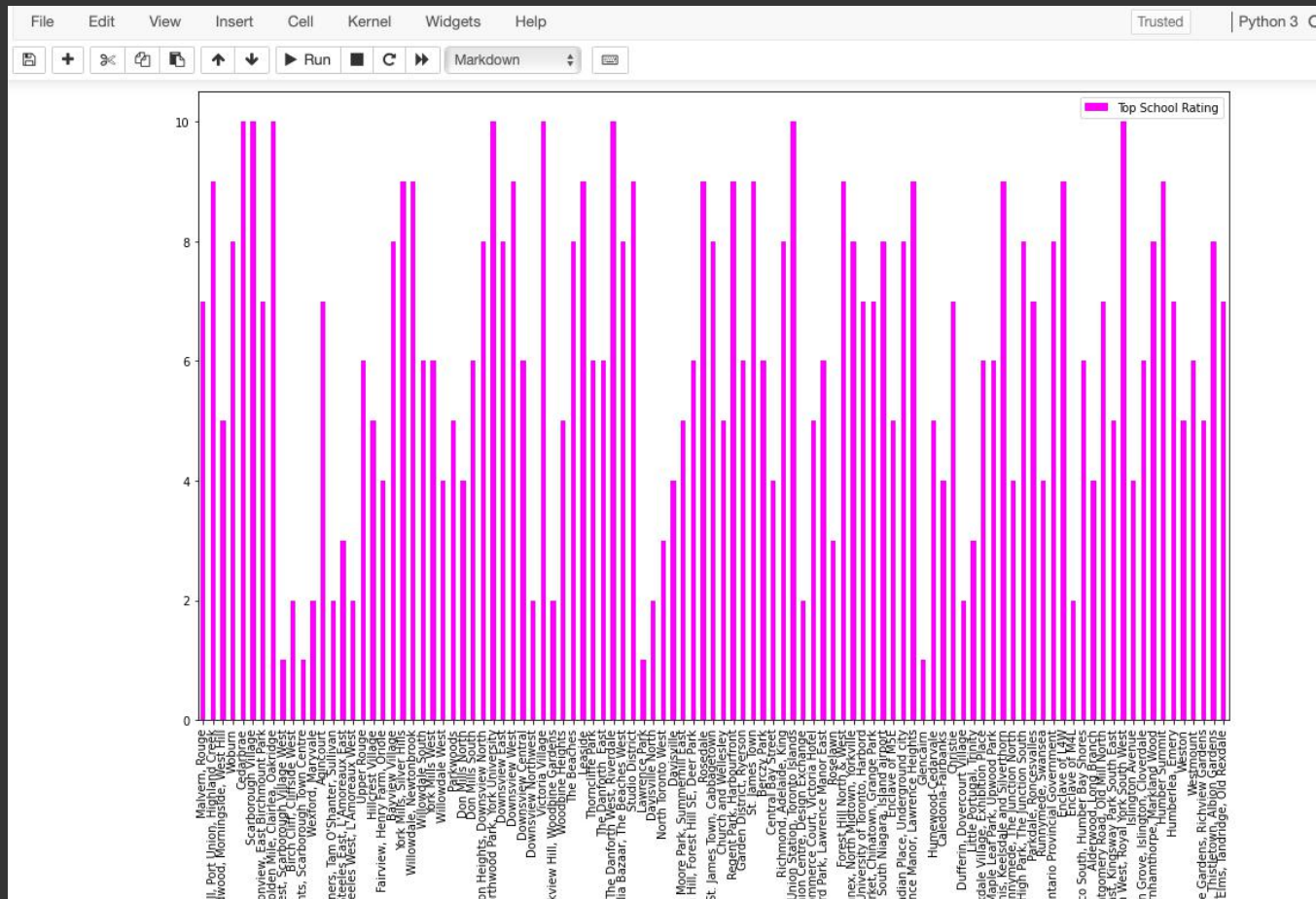
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Shopping Mall	Pharmacy	Clothing Store	Bank	Coffee Shop	Park	Supermarket	Sushi Restaurant	Skating Rink	Newsagent
1	Alderwood, Long Branch	Sandwich Place	Pizza Place	Print Shop	Pub	Gas Station	Coffee Shop	Gym	Zoo Exhibit	Music Store	Nightclub
2	Bathurst Manor, Wilson Heights, Downsview North	Park	Coffee Shop	Middle Eastern Restaurant	Intersection	Mediterranean Restaurant	Fried Chicken Joint	Sushi Restaurant	Mobile Phone Shop	Restaurant	Sandwich Place
3	Bayview Village	Asian Restaurant	Construction & Landscaping	Dog Run	Trail	Park	Gas Station	Zoo Exhibit	Office	Nightclub	Newsagent
4	Bedford Park, Lawrence Manor East	Italian Restaurant	Coffee Shop	Sandwich Place	Sushi Restaurant	Juice Bar	Thai Restaurant	Greek Restaurant	Grocery Store	Restaurant	Fast Food Restaurant

## 4. Results Section: Map of Clusters in Scarborough



## Average Housing Price by Clusters in Scarborough







## 6. Conclusion Section

In this project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 103 different latitude and logitude 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.

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 Works:**

This 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.



# Libraries Which were Used to Develop the Project:

Pandas: For creating and manipulating dataframes.

Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Data.

Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.