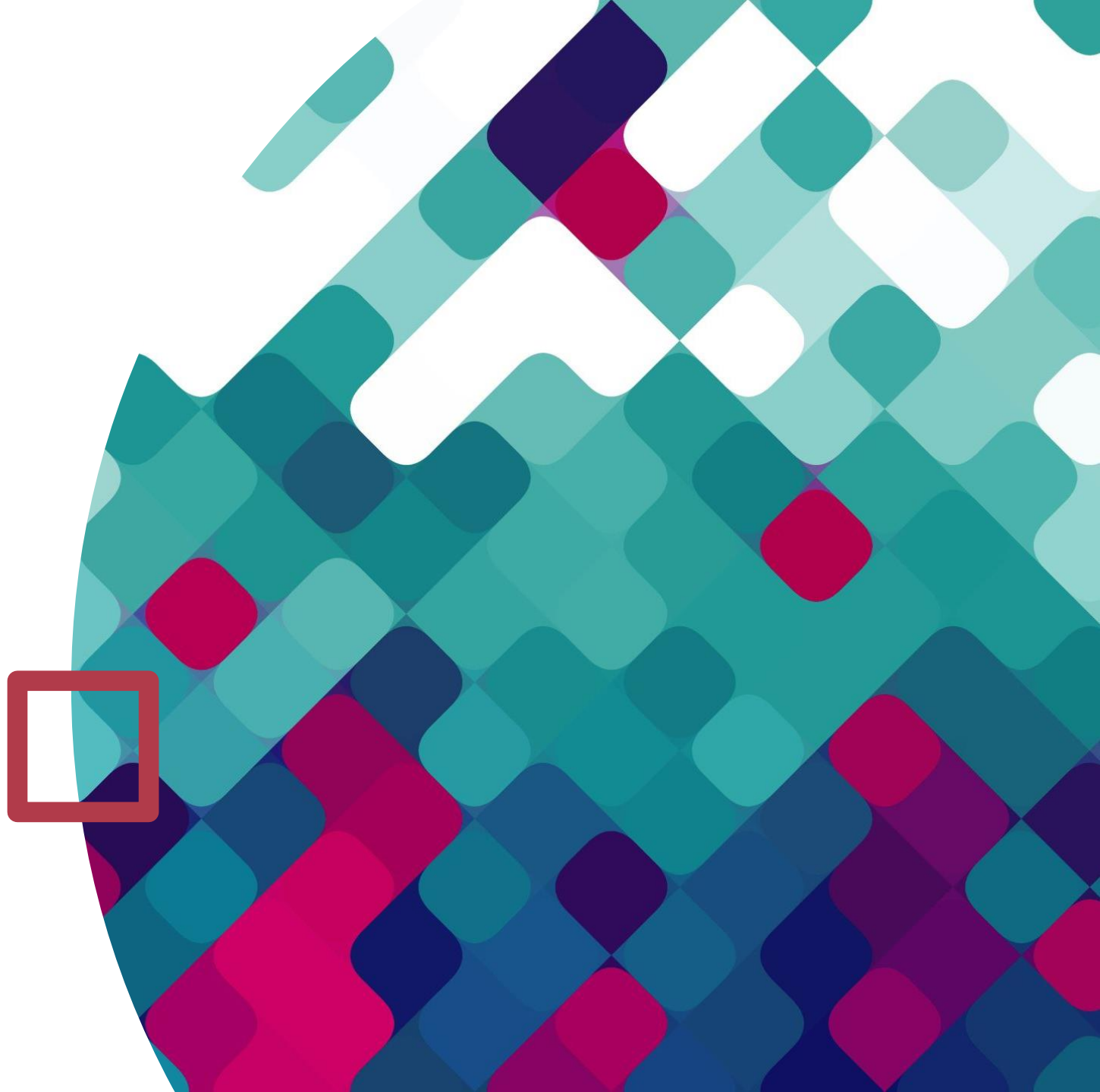# Capstone Project - The Battle of Neighborhoods

- IBM Data Science Professional Certificate
- Heris Agimi
- May 2020

Background

- Toronto is a city with scores of neighborhoods, some of which have identities that have been imposed upon them by realtors, whereas others are of much longer standing and have a more distinctive character.

- The main premise of our research will center on investigating and determining the ideal locale in Toronto, Canada, for the establishment of a new dining experience.

Business Problem

- Opening a successful restaurant is all about location; physical brick-and-mortar venues matter even in today's world of virtual reality. Location can make or break a restaurant.

- The object of this project is to identify the ideal location to start a new restaurant in Toronto, Canada.

Target Audience

- Our target audience are prospective investors and entrepreneurs who consider to establish a restaurant in Toronto. One such critical aspect of an exercise of this nature, is selecting a location, that is, often what type of community, city, state or even country to locate their business in.

Data Required

- List of neighborhoods in Toronto, Canada
- Latitude and longitude coordinates of the neighborhoods
- Venue data

Sources of Data

- Neighborhoods in Toronto is provided in Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M).

- I will use Python Requests and Beautiful-soup packages to scrape and extract the data table from the Wikipedia page.

- Then I used Toronto_Geo list (https://cocl.us/Geospatial_data) to get latitude and longitude into the Toronto neighborhood dataset based on postal code.

- Finally, I will connect Python with Foursquere API to get venue data for Toronto neighborhoods.

Data Cleaning and Wrangling

- I removed "Not Assigned" cells and only kept those boroughs which have "Toronto" written in the names in order to clean the data.

- As a result, there are 103 postal codes, 10 boroughs, and 98 neighborhoods left in the dataset.

	Postal code	Borough	Neighborhood
count	103	103	103
unique	103	10	98
top	M4G	North York	Downsview
freq	1	24	4

Methodology

- An unsupervised machine learning algorithm – K-means clustering will be conducted to help determine where is the best place to run what kind of business. For this project the focus will be on looking for the cluster of neighborhood which has more restaurants than the others.

Data Analysis

- • Cluster the Neighborhoods into groups based on the frequency and attributes using K-means Clustering Algorithm.

Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
East Toronto	The Beaches	43.676357	-79.293031	0	Park	Neighborhood	Trail
East Toronto	Riverdale, The Danforth West	43.679557	-79.352188	8	Greek Restaurant	Italian Restaurant	Ice Cream Shop
East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572	7	Pizza Place	Sandwich Place	Fast Food Restaurant
East Toronto	Studio District	43.659526	-79.340923	7	Café	Coffee Shop	Bakery
Central Toronto	Lawrence Park	43.728020	-79.388790	7	Park	Swim School	Bus Line

Decision Making and Reporting Results

- Let us use Folium to populate the clusters on the map. The orange cluster represents Cluster 8 which had more restaurants than the other clusters.

- Cluster 8 is located around East Toronto area which I would highly recommend to consider opening new restaurants here.
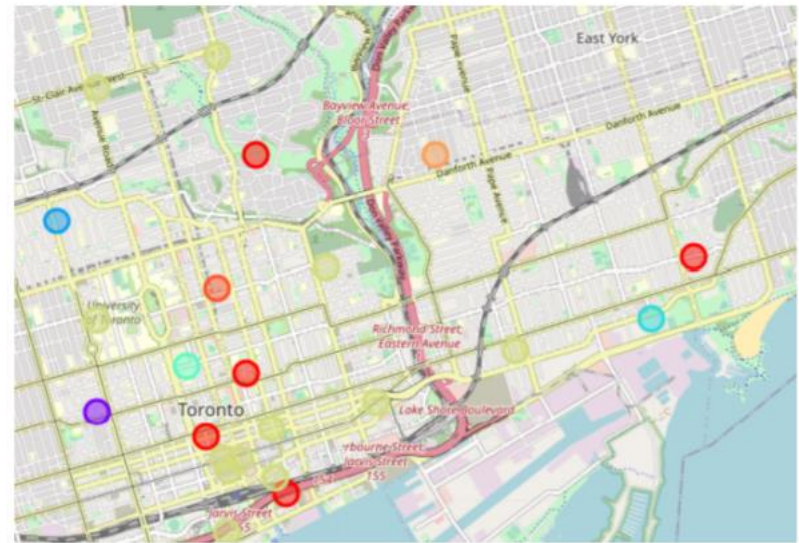
Discussion

- Density of a certain kind of business is an important factor to consider when looking for a location. Higher density means better popularity but also means more competition in this region. To make a restaurant outstanding among the competition, investors and entrepreneurs also need to continue analyzing more details such as types of surrounding restaurants and local culture in order to determine target customers and marketing strategy.

Conclusion

- Machine learning provides useful algorithms to explore a greater volume and variety of data quickly. Due to the size and complexity of these data sets, machine learning can help unlock value from all this data in a way that humans cannot. Decision makers can leverage technologies to drive efficiency. However, decision-making steps still need humans involved to make sure the results support the initial purposes of the project.

THANK YOU