Opening a Coffee shop in Toronto, CA

IBM Data Science Capstone Project

Vaibhav Goswami - 27 April 2020

Introduction & Business Problem

Toronto is Canada's largest city and a world leader in such areas as business, finance, technology, entertainment and culture. Its large population of immigrants from all over the globe has also made Toronto one of the most multicultural cities in the world.

Forbes Magazine tracked the "World's Most Economically Powerful Cities" and included Toronto in its top ten list. Toronto is well-known for its healthy arts scene, numerous sporting events, a vibrant nightlife, and restaurants, bistros and eateries featuring cuisine from just about any nationality one could think of.

It is a blooming place to start a food business such as coffee shop.


The objective of this capstone project is to find the most suitable location for a budding investor to open a new coffee shop in the city of Toronto, Canada. By using the data science methods and tools along with machine learning algorithms such as clustering.

This project aims to find the solution to the problem - "In Toronto, if a person wants to open a coffee shop, where should they consider opening it?"

Data

For the Toronto neighbourhood data, a Wikipedia page exists that has all the information needed to explore and cluster the neighbourhoods in Toronto. We will also be required to scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas dataframe so that it is in a structured format.

We will also be using Foursquare, a local search-and-discovery app which provides personalized recommendations of places to go near a specific location. To get this information I signed up for a Foursquare developer account to be able to use their API. I will then make calls to the API using a list of Toronto neighbourhoods.

The data set contains the following columns:

- Postal Code : code identifying each area
- Borough : local administration division
- Neighbourhood : The locality in the city

	Postal code	Borough	Neighborhood
0	M1B	Scarborough	Malvern / Rouge
1	M1C	Scarborough	Rouge Hill / Port Union / Highland Creek
2	M1E	Scarborough	Guildwood / Morningside / West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
...	...	...	...
98	M9N	York	Weston
99	M9P	Etobicoke	Westmount
100	M9R	Etobicoke	Kingsview Village / St. Phillips / Martin Grov...
101	M9V	Etobicoke	South Steeles / Silverstone / Humbergate / Jam...
102	M9W	Etobicoke	Northwest

Methodology

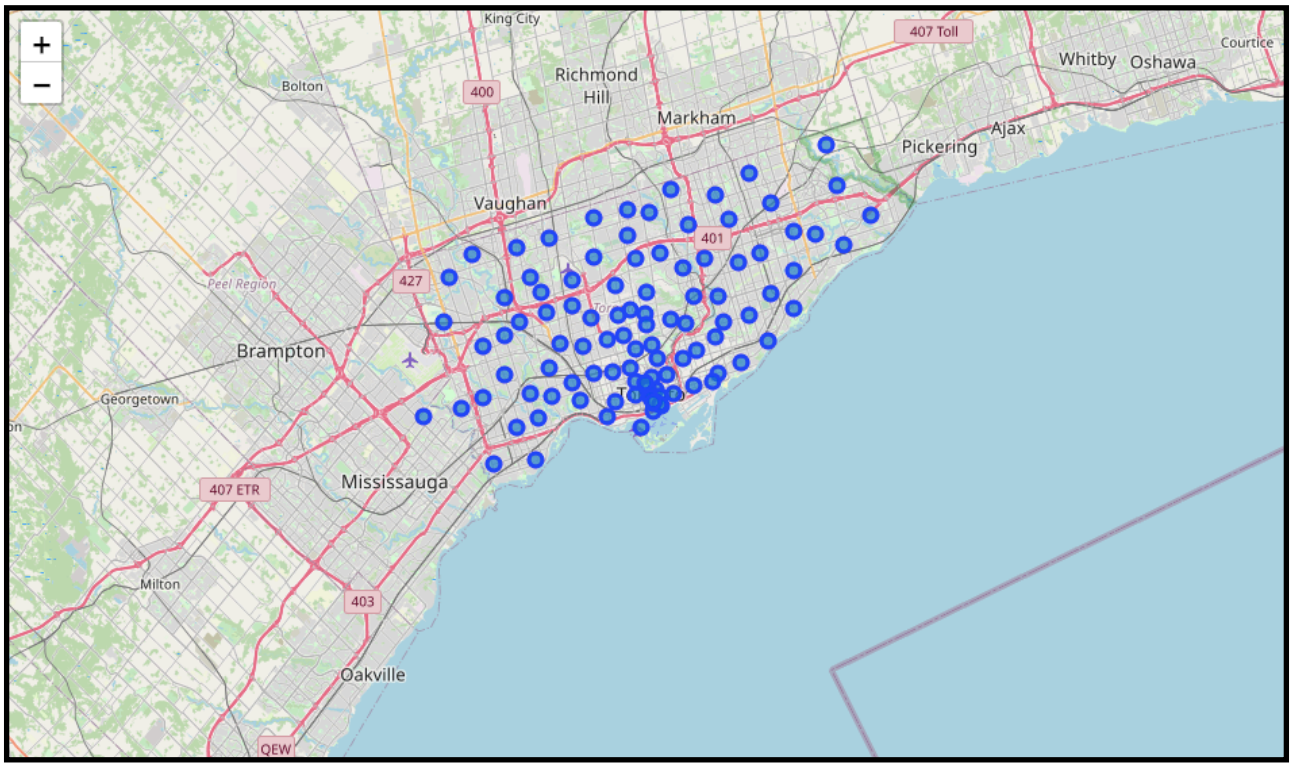
After the data is collected, cleaned and organized, we try to plot the longitude and latitude data of the locations on a geographic map with Folium map. We set the starting longitude and latitude, the magnitude of the zoom and even the tile type. We then plot all the locations on the Folium map with pop ups which provide all information about the various sites.

We use the information regarding latitudes and longitudes for each of the neighbourhood in Toronto and combine with the previously available data.
This will be really helpful in creation of maps using Folium and better visualisation.

	Postal code	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill / Port Union / Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood / Morningside / West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
...	...	...	...	...	...
98	M9N	York	Weston	43.706876	-79.518188
99	M9P	Etobicoke	Westmount	43.696319	-79.532242
100	M9R	Etobicoke	Kingsview Village / St. Phillips / Martin Grov...	43.688905	-79.554724
101	M9V	Etobicoke	South Steeles / Silverstone / Humbergate / Jam...	43.739416	-79.588437
102	M9W	Etobicoke	Northwest	43.706748	-79.594054

Folium is a powerful Python library that helps you create several types of Leaflet maps. The fact that the Folium results are interactive makes this library very useful for dashboard building. It provides real life and accurate maps very easily, with zoom in and zoom out option.

After that, we call the FourSquare API with our unique credentials to find locations and venues surrounding our sites and put that information in a JSON file. We then group the data in the JSON file into various columns and create a dataframe for them. Information about the Explore calls can be found at- https://developer.foursquare.com/docs/api/venues/explore.



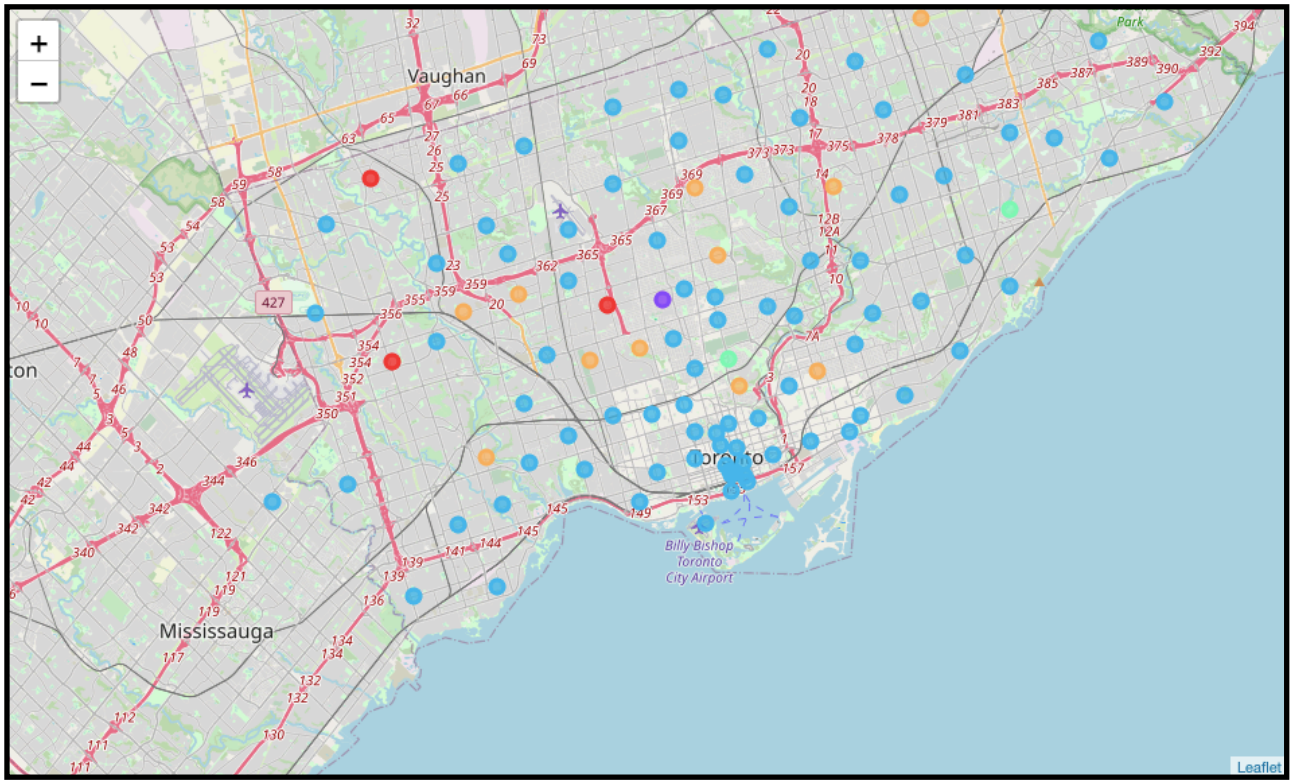
As a machine learning algorithm, we use k-means clustering algorithm for the purpose of clustering data available. We then apply the K-Means clustering algorithm to the data set.

```
kclusters = 5

toronto_grouped_clustering = toronto_grouped.drop('Neighborhood', 1)

kmeans = KMeans(n_clusters=kclusters, random_state=6).fit(toronto_grouped_clustering)

kmeans.labels_
```

This will help in creating group of clusters which will help in analysing the location of different institutions and businesses across the city.

Result

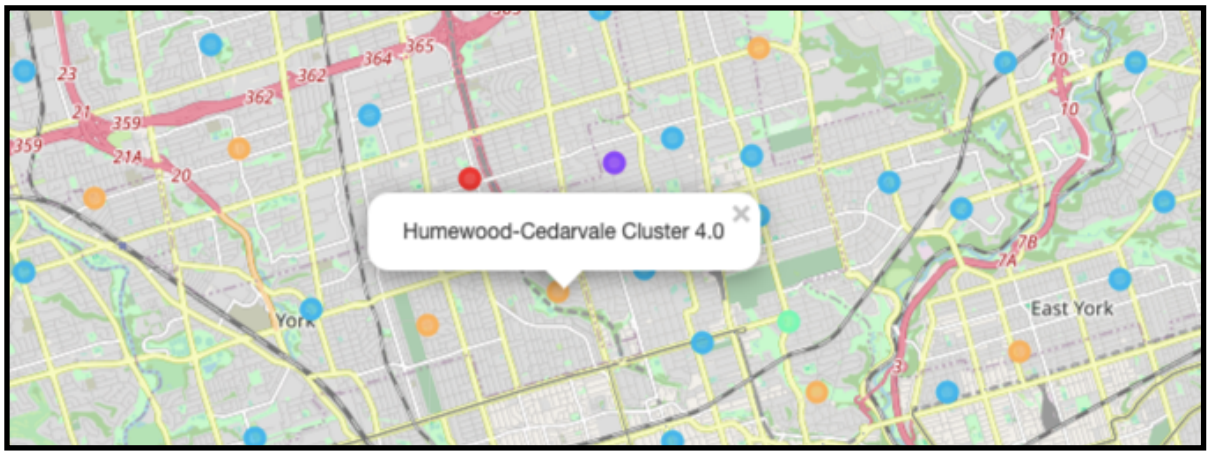
After running the K-means clustering we can access each cluster created to see which neighbourhoods were assigned to each of the clusters.



The cluster 2 is the biggest cluster with 87 of the 102 clusters in the boroughs in Toronto. Upon closely examining these neighbourhoods we can see that the most common venues in these neighbourhoods are Restaurants, Pubs, Cafe, Supermarkets, and stores.

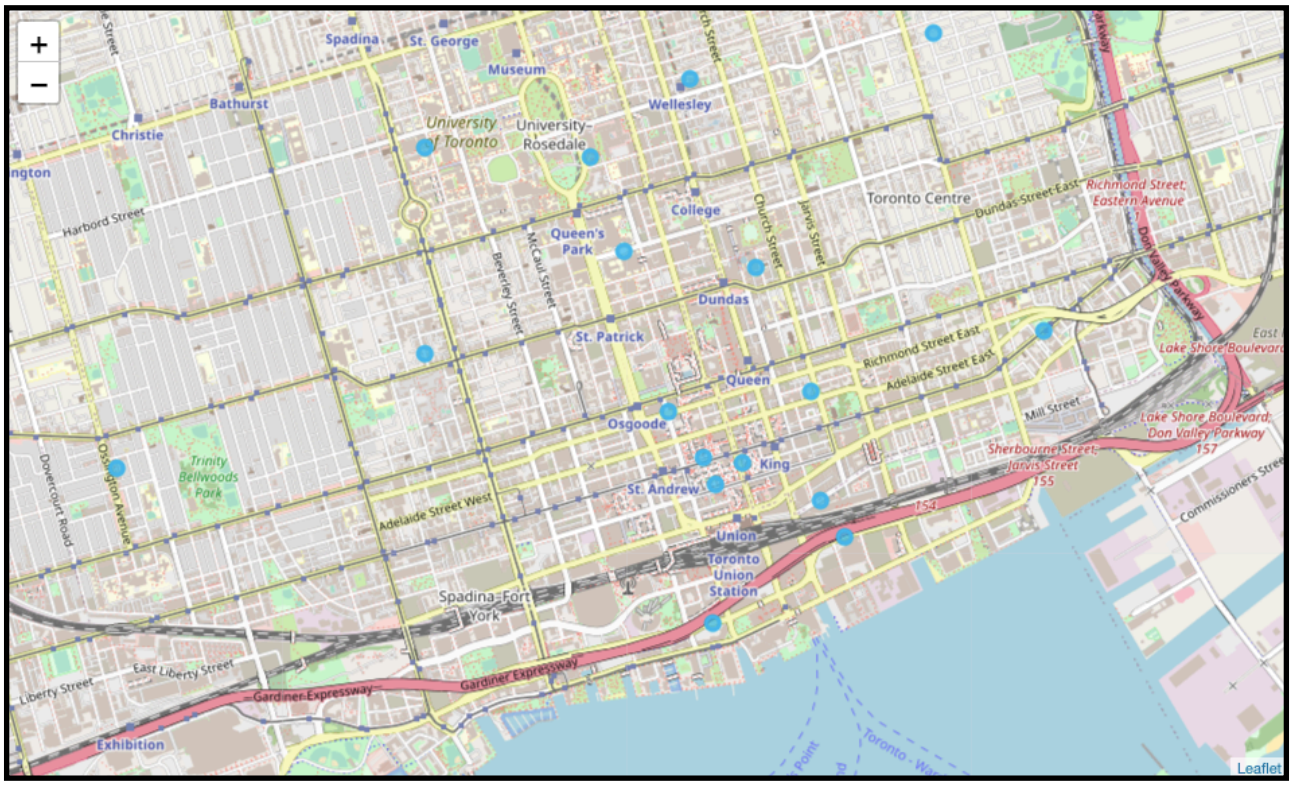
Coffee Shop	17
Café	12
Grocery Store	7
Pizza Place	4
Japanese Restaurant	3
Bar	2
Indian Restaurant	2
Baseball Field	2
Park	2
Clothing Store	2
Fast Food Restaurant	2
Bakery	1
Sandwich Place	1

Each cluster is colour coded for the ease of presentation, we can see that majority of the neighbourhood falls in the blue cluster which is the first cluster. Three neighbourhoods have their own cluster (Orange, Purple and Red), these are other clusters in the city of Toronto.



Discussion

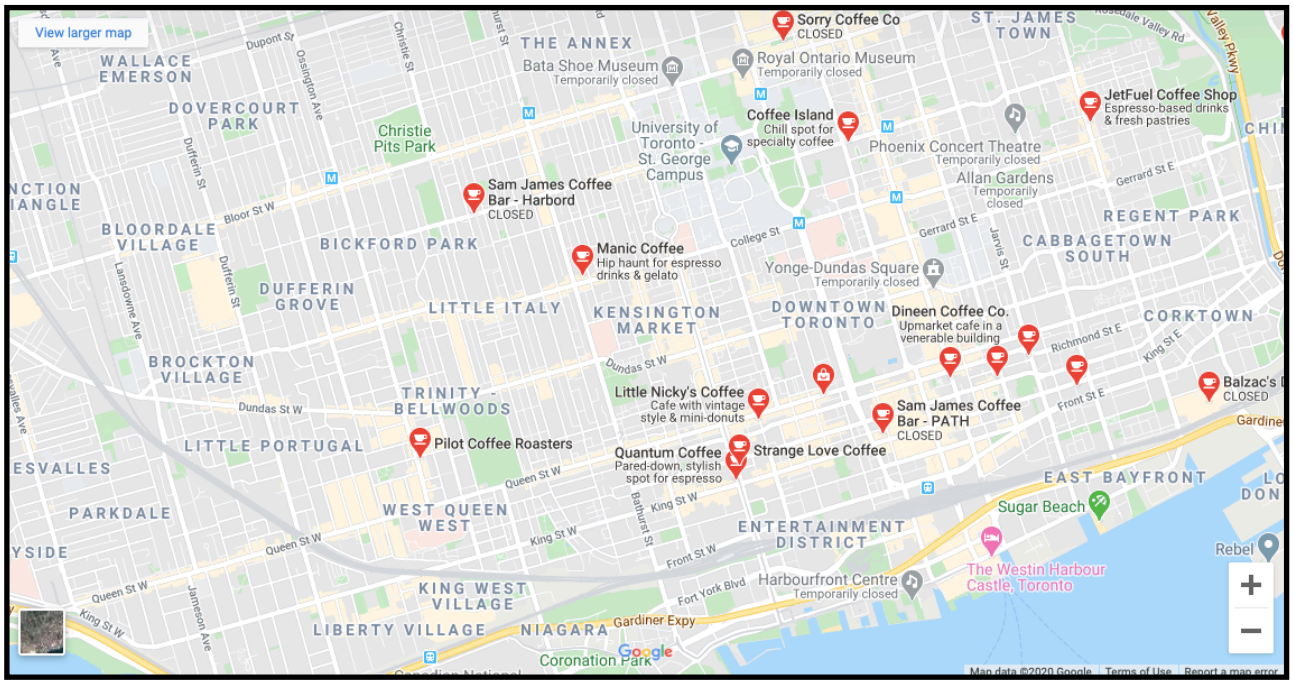
The aim of this project is to help people who want to open a new coffee shop in the city of Toronto, expats can chose the neighbourhoods to which they want to settle based on the most common profitable places in the city.



For example, it can be seen from the above map that the area near The University of Toronto is a possible place to open a new coffee shop. There can be many reasons behind this, such as student crowd, population, large number of hostels and rentals, etc.

Also, the area near the subway and railways stations is well suited for a coffee shop business because of the large number of people travelling in and out of the city for work.

Also below is the actual map of the city of Toronto showing the locations of different coffee shops in the city and the statements made above hold true by analysing it.



Conclusion

We began this project with the hope that we could find some linkages between Toronto's demographic data and Foursquare data for venues to open new coffee shop in Toronto. After comparing the results with the current scenario in the city, we can say that the results were promising. This project helps a person get a better understanding of the neighbourhoods with respect to the most common venues in that neighbourhood.

However, the results show general trends, but cannot be used to predict exact amounts. There is still too much variability that is not explained by the predictors used here, both for predicting demographic information and Foursquare information. However, we do believe these results can be used for a quick check for new opportunities of starting a coffee shop to see if their other research matches with the results found in this project.