The Battle of the Neighborhoods

1. Introduction/Business Problem

Toronto is the most populous city in Canada and the fourth most populous city in North America. Business stakeholders are interested in opening a new coffee shop in Toronto but is concerned about the best neighborhood location for the new venue. Therefore, the goal of this project is to find the most potential location to open a coffee shop in Toronto. We will focus on all the neighborhoods in Toronto.

1.1 Target Audiences

- Entrepreneurs or Business owners who are interested to investing or opening a restaurant.
- Anyone who loves to have their own restaurant as a side business.
- Finding the best location for opening a restaurant.

2. Data

The following data sources will be needed to extract/generate the required information:

- List of all neighborhoods in Canada — https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Coordinates of all neighborhoods and venues — GeoPy Nominatim geocoding
- Retrieved Venues data from Foursquare API and fetch Coffee Roaster, Coffee Shop and Tea Room category. — Foursquare API
- The Venue data will help find which neighborhood is best suitable to open a coffee shop.

3. Methodology

3.1 Data Scraping & Data Cleansing

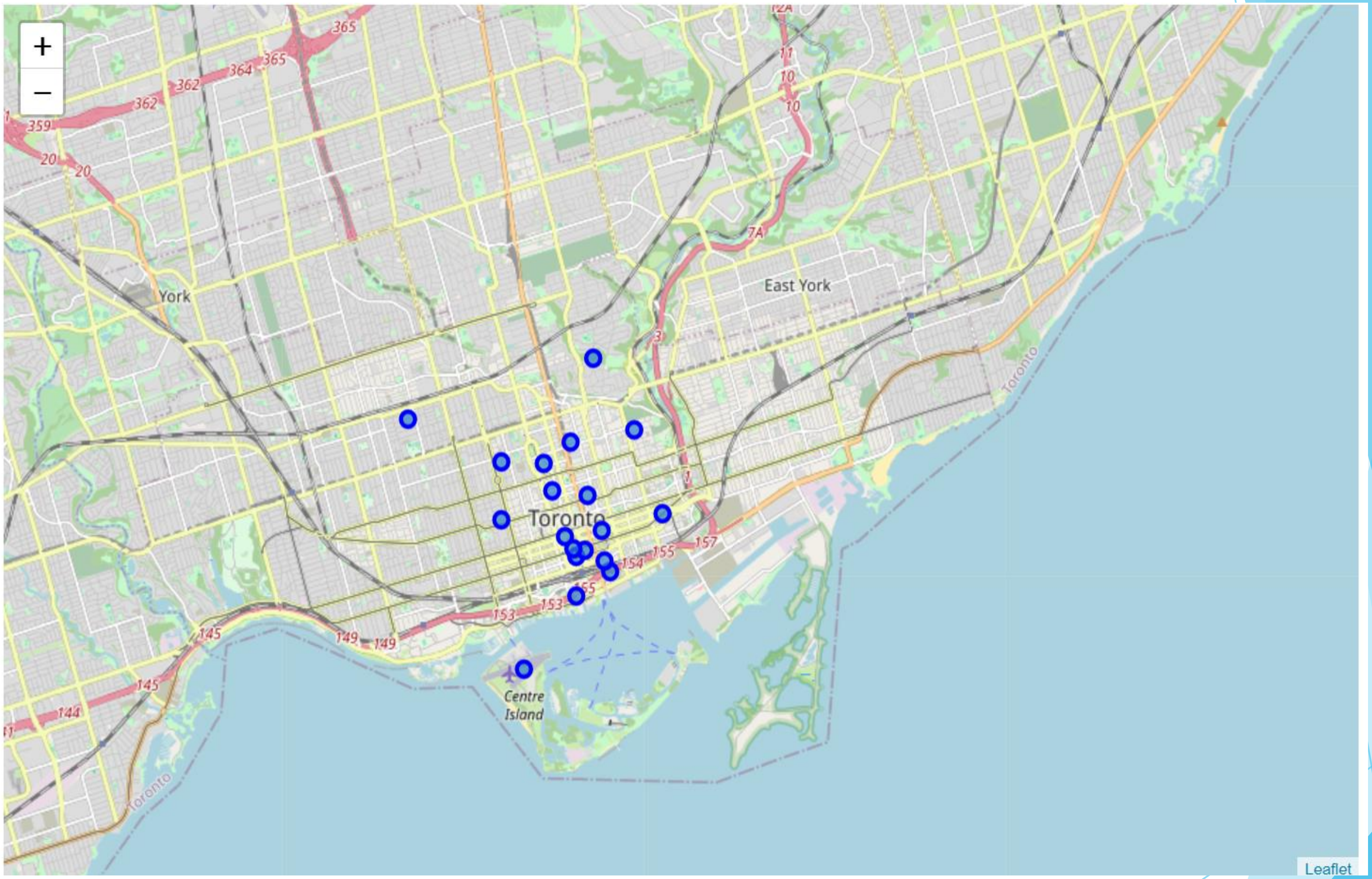
Prior to the start, essential libraries should be loaded and imported. Then, in the first stage of data analysis, data was first collected, cleansed, and transformed into data frames. The Wikipedia link above provides information about the neighborhoods in Canada, including postal codes, boroughs, and the neighborhood names.
Scraping and merging of the data was also required before processing the analysis.

	Postcode	Borough	Neighbourhood
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

3.2 Create Maps and Explore Neighborhoods

After data-cleansing, we then use the provided Geospatial_Coordinates.csv file to get the Geographical coordinates of the neighborhoods with the respective Postal Codes and read the data into a data frame. Next, we will merge the data frame that contains coordinates with the one that contains borough names and choose only the neighborhoods that contain “Downtown Toronto”.

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M4W	Downtown Toronto	Rosedale	43.6796	-79.3775
1	M4X	Downtown Toronto	St. James Town, Cabbagetown	43.668	-79.3677
2	M4Y	Downtown Toronto	Church and Wellesley	43.6659	-79.3832
3	M5A	Downtown Toronto	Regent Park, Harbourfront	43.6543	-79.3606
4	M5B	Downtown Toronto	Garden District, Ryerson	43.6572	-79.3789

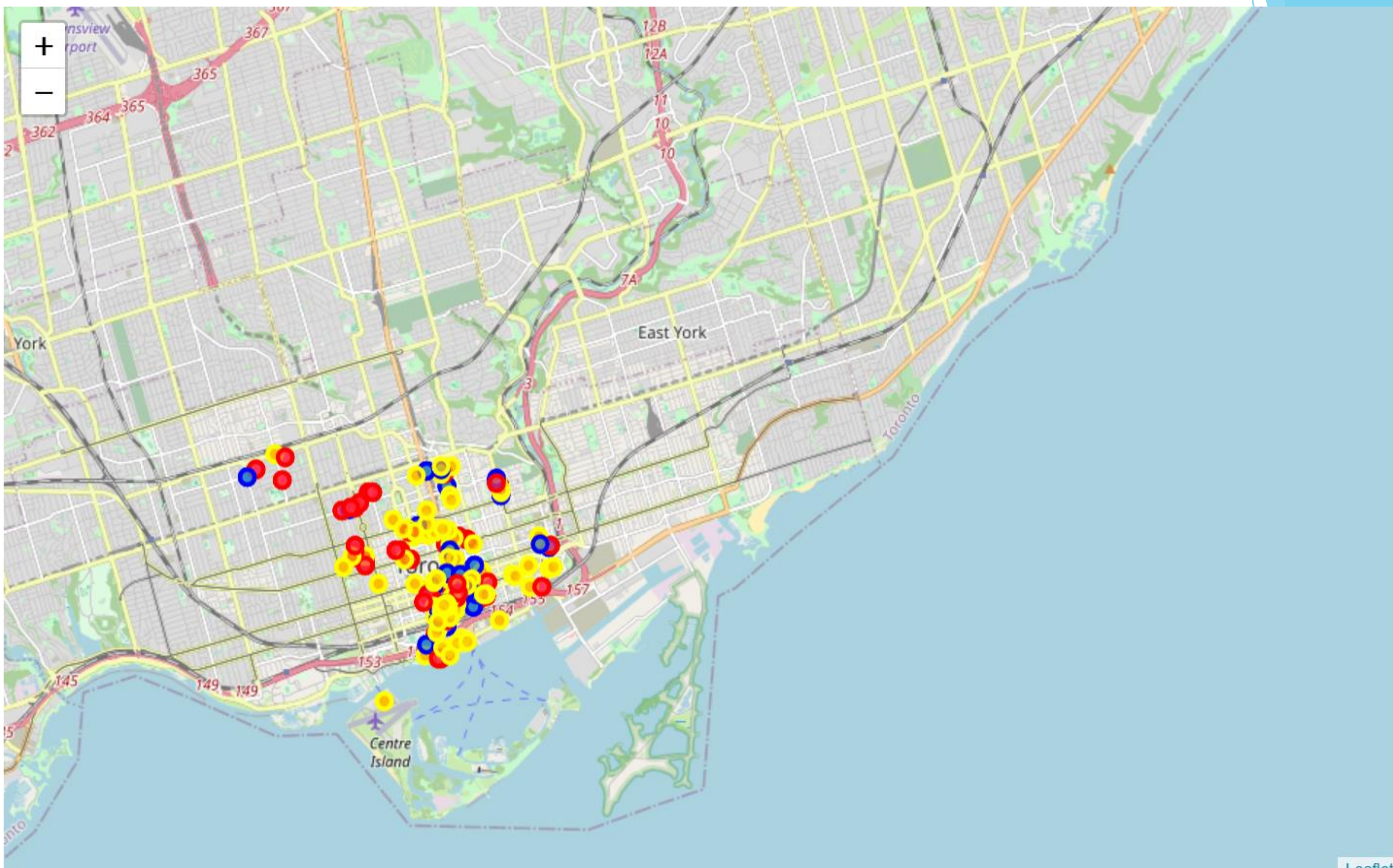


York
East York
Toronto
Centre Island
Leaflet

3.3 Retrieved Venues data using Foursquare API

We now use Foursquare API to fetch Venue lists for the Coffee Roaster, Coffee Shop and Tea Room category. From the Foursquare API, we can find corresponding ID for these categories associated with opening a coffee shop. We will create venue lists based on the categories we chose: Coffee Roaster(5e18993feee47d000759b256), Coffee Shop (4bf58dd8d48988d1e0931735), and Tea Room (4bf58dd8d48988d1dc931735). This data was needed to process the analysis of where to open a new coffee shop in Downtown Toronto.
Combine the 3 venues together and clean up the venue data to only include the top five venue category.

	Venue Category	Count
48	Coffee Shop	390
39	Café	195
167	Restaurant	120
110	Hotel	114
118	Japanese Restaurant	93



+
−
East York
York
Centre
Island
Leaflet

3.4 Analyze each Neighborhood

Then we perform one-hot encoding on the data to process the analysis for each neighborhood. For each neighborhood, frequencies of individual venues were shown at how many of those Venues were in each neighborhood.

	Neighborhood	Café	Coffee Shop	Hotel	Japanese Restaurant	Restaurant
5	St. James Town, Cabbagetown	0	0	0	0	1
7	St. James Town, Cabbagetown	1	0	0	0	0
8	St. James Town, Cabbagetown	1	0	0	0	0
9	St. James Town, Cabbagetown	0	0	0	1	0
23	St. James Town, Cabbagetown	0	0	0	0	1

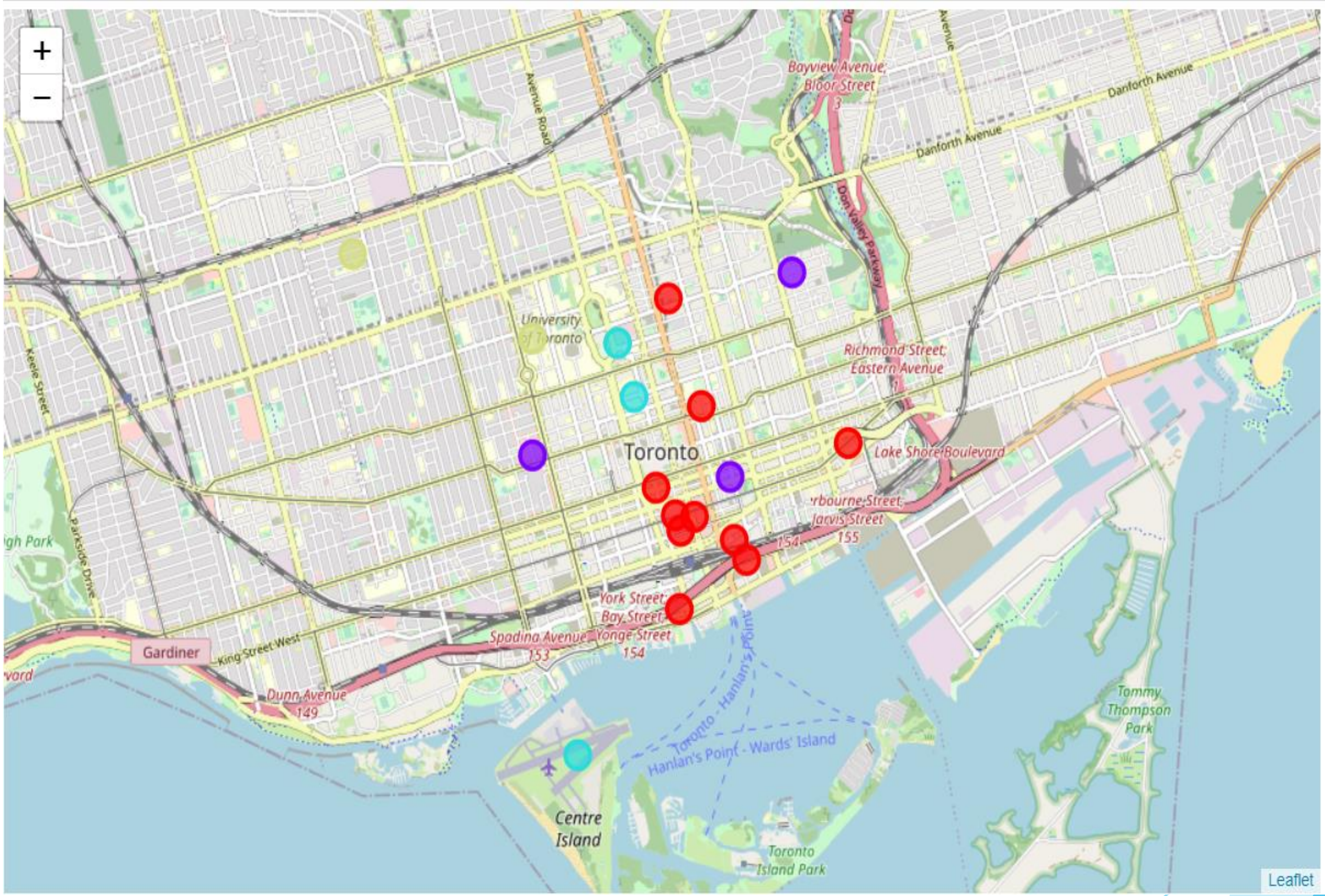
Create a new data frame from the data above with rankings on the top 3 venues. This summarizes the data based on each individual Neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Berczy Park	Coffee Shop	Restaurant	Japanese Restaurant
1	CN Tower, King and Spadina, Railway Lands, Har...	Coffee Shop	Restaurant	Japanese Restaurant
2	Central Bay Street	Coffee Shop	Café	Restaurant
3	Christie	Café	Restaurant	Coffee Shop
4	Church and Wellesley	Coffee Shop	Japanese Restaurant	Restaurant

Kmeans Cluster

Next, we performed a K-means clustering algorithm to partition the neighborhoods into 4 clusters. By applying Machine Learning algorithm, we will be able to cluster the venues based on a list of locations for these three different types of drink venues. We will have a better understanding of the similarities and variations between the neighborhoods and have more insights.

	Postcode	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	M4X	Downtown Toronto	St. James Town, Cabbagetown	43.668	-79.3677	1.0	Restaurant	Coffee Shop	Café
2	M4Y	Downtown Toronto	Church and Wellesley	43.6659	-79.3832	0.0	Coffee Shop	Japanese Restaurant	Restaurant
3	M5A	Downtown Toronto	Regent Park, Harbourfront	43.6543	-79.3606	0.0	Coffee Shop	Restaurant	Café
4	M5B	Downtown Toronto	Garden District, Ryerson	43.6572	-79.3789	0.0	Coffee Shop	Japanese Restaurant	Café
5	M5C	Downtown Toronto	St. James Town	43.6515	-79.3754	1.0	Café	Coffee Shop	Restaurant



Toronto
University of Toronto
Bayview Avenue; Bloor Street
Danforth Avenue
Don Valley Parkway
Richmond Street; Eastern Avenue
Lake Shore Boulevard
Jarvis Street
York Street; Bay Street
Spadina Avenue
Yonge Street
Gardiner
King Street West
Dunn Avenue
Keele Street
Parkside Drive
Centre Island
Toronto Island Park
Tommy Thompson Park
Hanlan's Point - Wards' Island
Leaflet

4. Results

Each cluster can now be analyzed and the venue categories that distinguish each cluster can be determined.

Cluster #1: Coffee Shop is the most common venues

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
2	Church and Wellesley	0	Coffee Shop	Japanese Restaurant	Restaurant
3	Regent Park, Harbourfront	0	Coffee Shop	Restaurant	Café
4	Garden District, Ryerson	0	Coffee Shop	Japanese Restaurant	Café
6	Berczy Park	0	Coffee Shop	Restaurant	Japanese Restaurant
8	Richmond, Adelaide, King	0	Coffee Shop	Café	Restaurant
9	Harbourfront East, Union Station, Toronto Islands	0	Coffee Shop	Hotel	Café
10	Toronto Dominion Centre, Design Exchange	0	Coffee Shop	Hotel	Café
11	Commerce Court, Victoria Hotel	0	Coffee Shop	Restaurant	Hotel
15	Stn A PO Boxes	0	Coffee Shop	Restaurant	Japanese Restaurant
16	First Canadian Place, Underground city	0	Coffee Shop	Café	Hotel

Cluster #2: Café is the most common venues, second is Coffee Shop

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	St. James Town, Cabbagetown	1	Restaurant	Coffee Shop	Café
5	St. James Town	1	Café	Coffee Shop	Restaurant
13	Kensington Market, Chinatown, Grange Park	1	Café	Coffee Shop	Japanese Restaurant

Cluster #3: Coffee Shop is the most common venues

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
7	Central Bay Street	2	Coffee Shop	Café	Restaurant
14	CN Tower, King and Spadina, Railway Lands, Har...	2	Coffee Shop	Restaurant	Japanese Restaurant
18	Queen's Park, Ontario Provincial Government	2	Coffee Shop	Japanese Restaurant	Café

Cluster #4: Café is the most common venues

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
12	University of Toronto, Harbord	3	Café	Japanese Restaurant	Restaurant
17	Christie	3	Café	Restaurant	Coffee Shop

5. Discussions

After reviewing the data of each cluster, the findings are as below: For Cluster 1 and 3, the top 1 most common venue has coffee shop in it. As for Cluster 2 and Cluster 4, coffee shops are only at the 2nd and 3rd most common venue. Therefore, according to the analysis, Cluster 2 and Cluster 4 would be a better choice for anyone who's interested in opening a new coffee shop. Moreover, the neighborhood "University of Toronto, Harbord" in Cluster 4 would be the optimal choice that have a great opportunity. Currently, there's less competition compared to other areas since there's no coffee shops in the top 3 common venue in the neighborhood.

Although there are some of the drawbacks of this analysis— the clustering is completely based on data from the Foursquare API, and it did not take into consideration of other factors such as rent, population, crime rates etc. However, this still provides insights based of data analysis.

6. Conclusion

In conclusion, this is a great opportunity to get hands on experience how a data scientist would approach and perform data analysis on a business problem. We were able to go through the Data Science analysis process from using numerous Python libraries to utilizing Foursquare API. Ideally, a more in-depth analysis would consist of utilizing resources from other databases and external information. Further analysis that takes other factors into account can provide a more thorough view. This project still provides a glimpse how data science projects work with the data available, and a start to tackling business problems.