Capstone Project - The Battle of Neighborhoods (Week 2)

Sanujit Senapati

1 INTRODUCTION: BUSINESS PROBLEM

A real estate development and investment company is trying to identify and shortlist retail opportunities in the **Greater Toronto** area based on trends and popularity. The company realizes the importance and relevance of social media in understanding the pulse of the market and seeks to use data as a key driver in decision making.

How can the company use social trends to select popular venues, understand and identify characteristics of the venues, and select new locations with similar characteristics which would have high growth potential?

In this study, as a Data Scientist, I provide a point of view of how data can be acquired, cleansed, curated and analyzed through machine learning technique to better drive the decision-making process.

2 DATA

To drive the understanding and analysis in this data science project, I have used the following data sets:

1. **Toronto neighborhoods data** from Wikipedia, which was also used in the weeks 3 assignment. This data set includes the Postal Codes, Boroughs and Neighborhood in the Toronto area starting with the letter M.
2. The above data set was augmented with **geo codes** for each postal code from the data set provided by **Cognitive Class** at http://cocl.us/Geospatial_data. Upon merging the data sets, the resulting data set included geo coordinates, i.e. latitude and longitude, for each postal code.
3. **Foursquare Places API** for Venues – Foursquare provides various Regular and Premium API endpoints. Regular endpoints include basic venue firmographic data, category, and ID. Premium endpoints include rich content such as ratings, URLs, photos, tips, menus, etc. For the analysis, I have used the "**explore**" Regular API endpoint to get venue recommendations via https://developer.foursquare.com/docs/venues/explore.

The data sets used were already curated and did not require any additionally preparation such as reformatting. The only preparation steps were merging and reshaping of the data frames during the analysis.

3 METHODOLOGY

3.1 NEIGHBORHOOD CANDIDATE SELECTION

I started with data acquisition from the various data sources, followed by data shaping, transformation, and visualization. In the subsections below, I describe the steps were executed during the candidate selection process.

3.1.1 Fetch and merge neighborhood and geocode data

1. Fetch the neighborhood data from Wikipedia and read it into a DataFrame. Next filter and transform records per specifications. [FIGURE 1]
2. Fetch geocode file and read it into a DataFrame. [FIGURE 2]
3. Merge the neighborhood and geocode DataFrames. [FIGURE 3]

	PostalCode	Borough	Neighborhoc
0	M1B	Scarborough	Rouge, Malve
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Unic
2	M1E	Scarborough	Guildwood, Morningside, West H
3	M1G	Scarborough	Wobu
4	M1H	Scarborough	Cedarbr

Figure 1: DataFrame representation of Postal Code, Borough and Neighborhood mapping

	PostalCode	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Figure 2: DataFrame representation of Postal Code and Geo Code (latitude and longitude) mapping

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	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

Figure 3: DataFrame representation of the combined Postal Code, Borough, Neighborhood and Geo Code mapping

3.1.2 Visualize neighborhoods on map

4. Get coordinates of Toronto using the OpenStreetMap Nominatim API.
5. Visualize the neighborhoods as markers overlaid in a map of Toronto created using Folium. [FIGURE 4]

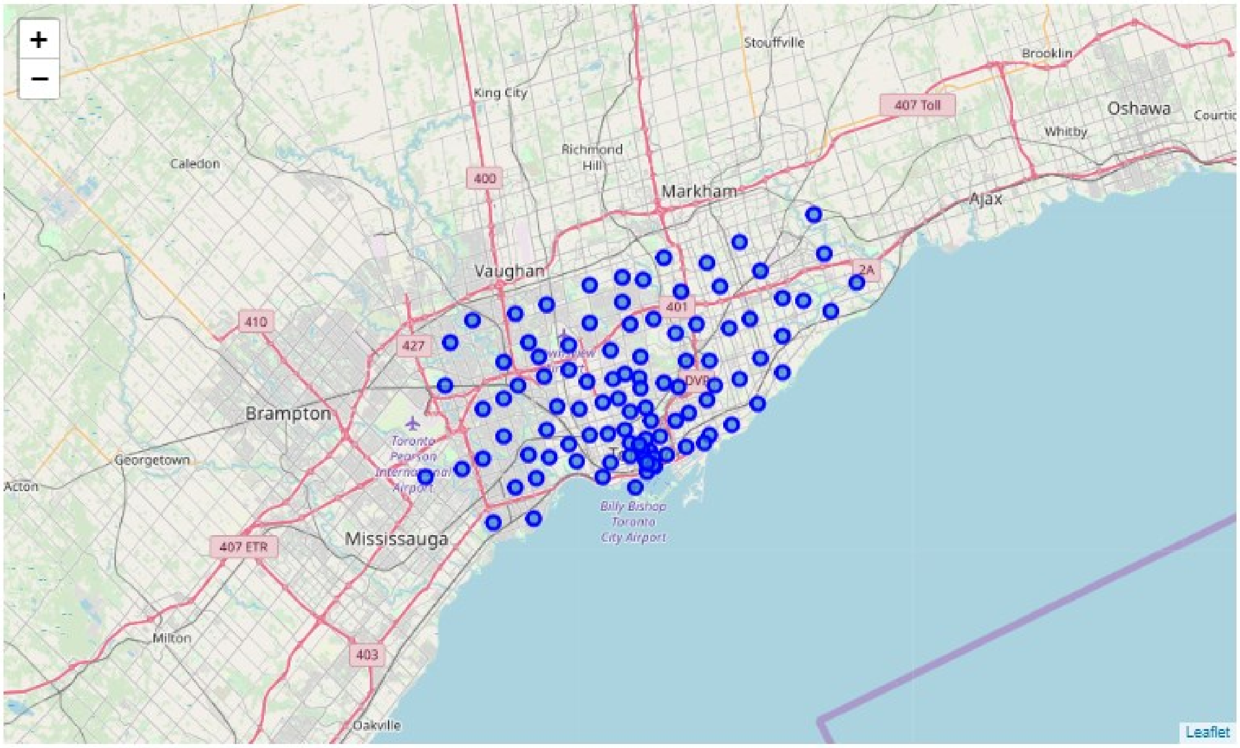

Figure 4: Neighborhoods plotted on map based on Geo Code using Folium

3.1.3 Fetch venue data from Foursquare

6. Fetch venue data from Foursquare for each neighborhood using the "explore" API endpoint. Aggregate the data into a master venues DataFrame. [FIGURE 5]

Note: The explore API sets a limit of 50 results (venues) per request.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue ID	Venue	Venue Latitude	Venue Longitude	Venue Category
883	Adelaide, King, Richmond	43.650571	-79.384568	5227bb01498e17bf485e6202	Downtown Toronto	43.653232	-79.385296	Neighborhood
896	Adelaide, King, Richmond	43.650571	-79.384568	50684f27e4b08e43cf97522d	Momofuku Nikai	43.649464	-79.386005	Cocktail Bar
897	Adelaide, King, Richmond	43.650571	-79.384568	4fccaa8fe4b05a98df3d9417	Sam James Coffee Bar (SJCB)	43.647881	-79.384332	Café
898	Adelaide, King, Richmond	43.650571	-79.384568	559a8f5a498e31f945041245	Maman	43.648309	-79.382253	Café
899	Adelaide, King, Richmond	43.650571	-79.384568	59603f86112c6c70931c9401	King Taps	43.648476	-79.382058	Gastropub

Figure 5: DataFrame representation of Venues

3.2 EXPLORATORY DATA ANALYSIS

After the data from various sources were aggregated, an exploratory data analysis was conducted using the following process steps.

7. Generate statistics from the venue data such as:
    a. Venue counts by neighborhood [FIGURE 6]
    b. Top 20 neighborhoods by venue count [FIGURE 7]

8. Visualize the venue data by plotting the number of neighborhoods for each venue count and range of venue counts. [FIGURE 8, FIGURE 9]
9. Visualize the top 10 categories across the entire set of venues (all neighborhoods) and as well as the top 20 neighborhoods. [FIGURE 10, FIGURE 11]
10. One hot encode the data.
11. Create a DataFrame containing the top venues by neighborhood based on the one hot encoded data and visualize the data. [FIGURE 12]

	Venue Count by Neighborhood
count	99.000000
mean	17.090909
std	17.253861
min	1.000000
25%	4.000000
50%	9.000000
75%	24.000000
max	50.000000

Figure 6: Venue counts by neighborhood

Neighborhood
Adelaide, King, Richmond
Central Bay Street
Stn A PO Boxes 25 The Esplanade
St. James Town
Ryerson, Garden District
Little Portugal, Trinity
First Canadian Place, Underground city
Fairview, Henry Farm, Oriole
Design Exchange, Toronto Dominion Centre
Commerce Court, Victoria Hotel
Church and Wellesley
Chinatown, Grange Park, Kensington Market
Harbourfront East, Toronto Islands, Union Station
Berczy Park
Harbourfront, Regent Park
Cabbagetown, St. James Town
The Danforth West, Riverdale
Queen's Park
Studio District
Runnymede, Swansea

Figure 7: Top 20 neighborhoods by venue count

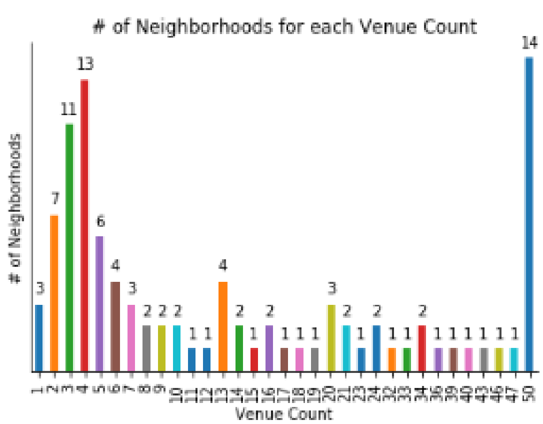

Figure 8: Number of neighborhoods for each "venue count"

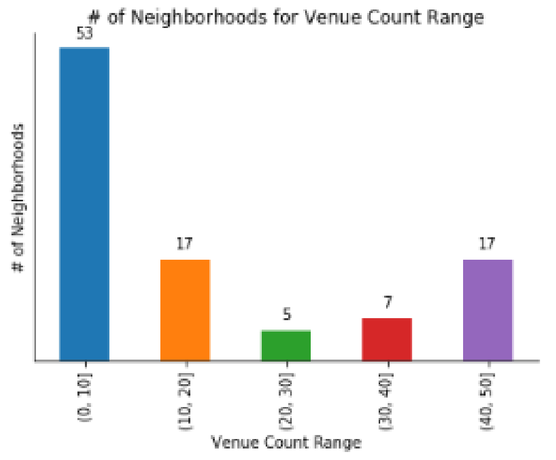

Figure 9: Number of neighborhoods for "venue count range"

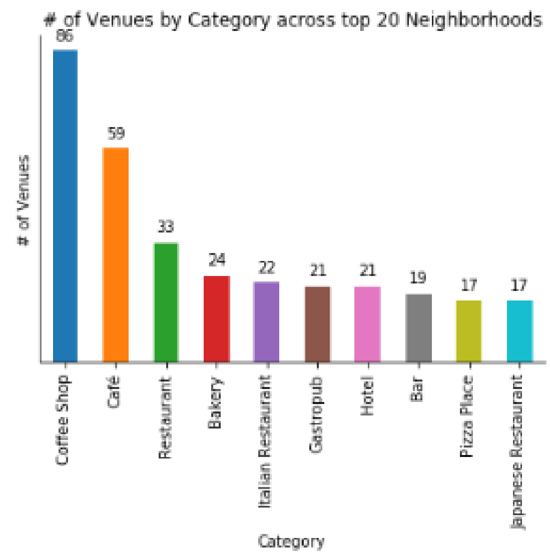

Figure 11: Number of venues by category across top 20 neighborhoods

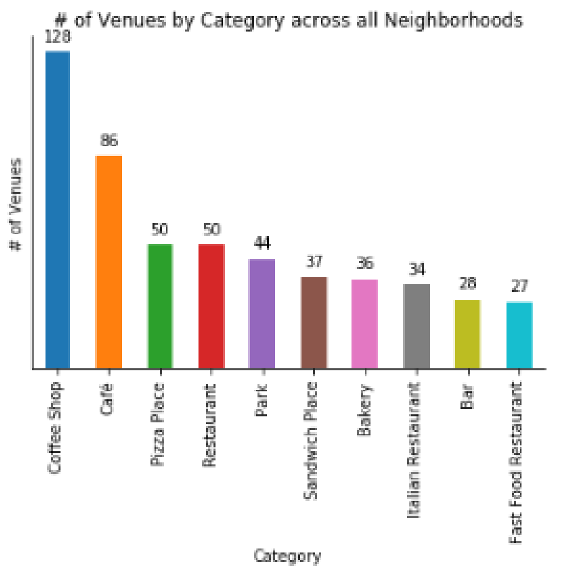

Figure 10: Number of venues by category across all neighborhoods

	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	Adelaide, King, Richmond	Coffee Shop	Steakhouse	Café	Restaurant	Breakfast Spot	Pizza Place	Gastropub	Asian Restaurant	Hotel	American Restaurant
1	Agincourt	Lounge	Clothing Store	Skating Rink	Breakfast Spot	Deli / Bodega	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run
2	Agincourt North, L'Amoreaux East, Milliken, St...	Playground	Park	Curling Ice	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run	Dive Bar	Discount Store
3	Albion Gardens, Beaumond Heights, Humbergate, ...	Grocery Store	Pizza Place	Pharmacy	Beer Store	Japanese Restaurant	Fried Chicken Joint	Sandwich Place	Fast Food Restaurant	Discount Store	Video Store
4	Alderwood, Long Branch	Pizza Place	Pharmacy	Coffee Shop	Pub	Sandwich Place	Skating Rink	Gym	Gastropub	Garden Center	Discount Store

Figure 12: Top 10 venues by neighborhood

3.3 ANALYSIS (MACHINE LEARNING)

Finally, as part of the analysis, I use a clustering based unsupervised learning approach to group the neighborhood and venue data and understand the characteristics of each group. This process used the following steps:

12. Based on the data available for neighborhoods and venues, we can define venue categories as features for machine learning. Given there are approximately 166 categories across a data set of 99 neighborhoods, use of k-means clustering to cluster the neighborhoods sounds like a reasonable approach. The 166 categories naturally map to features used in the k-means model. An initial value of 'k' was set to 7 [square root of 49.5 (99 divided by 2)]. This generates the cluster labels for each of the neighborhoods.
13. Next, we combine the cluster labels into neighborhood and top 10 venue data set. [FIGURE 13]
14. Visualize the resulting clusters overlaid on a map of Toronto created using Folium. [FIGURE 14]
15. Cluster analysis is provided in the Results section.

	PostalCode	Borough	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 Mo Commo Venu
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353	6.0	Fast Food Restaurant	Print Shop	Women's Store	Curling Ice	Eastern European Restaurant	Dumpling Restaurant	Drugstor
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	0.0	Bar	Golf Course	History Museum	Women's Store	Department Store	Dessert Shop	Dim Su Restaura
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	6.0	Electronics Store	Rental Car Location	Spa	Medical Center	Breakfast Spot	Mexican Restaurant	Pizz Plac
3	M1G	Scarborough	Woburn	43.770992	-79.216917	5.0	Coffee Shop	Korean Restaurant	Women's Store	Dance Studio	Electronics Store	Eastern European Restaurant	Dumplin Restaura
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	6.0	Athletics & Sports	Lounge	Hakka Restaurant	Fried Chicken Joint	Thai Restaurant	Bakery	Bar

Figure 13: DataFrame representation of clustered neighborhoods and top 10 venues in each neighborhood

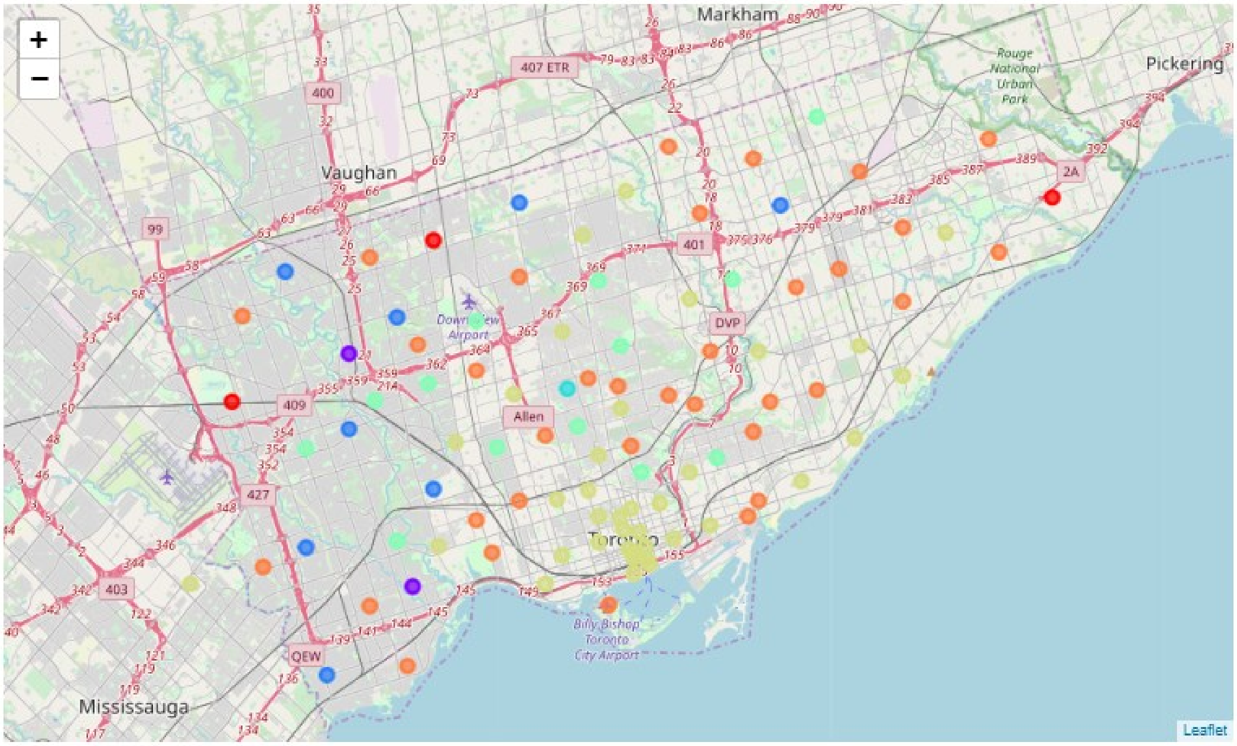

Figure 14: Clusters overlaid on a map of Toronto

4 RESULTS

In this section we examine the clusters and view cluster specific details. Key metrics from the analysis are presented below.

4.1 DISTRIBUTION OF NEIGHBORHOODS AND VENUES BY CLUSTER

Figure 15 and Figure 16 present the distribution of 99 neighborhoods and 1692 venues across 7 clusters.

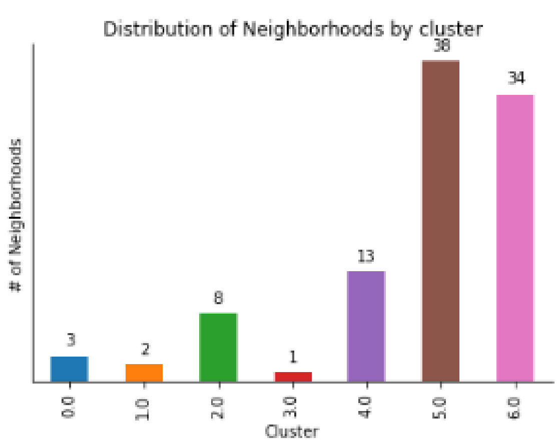

Figure 15: Distribution of neighborhoods by cluster

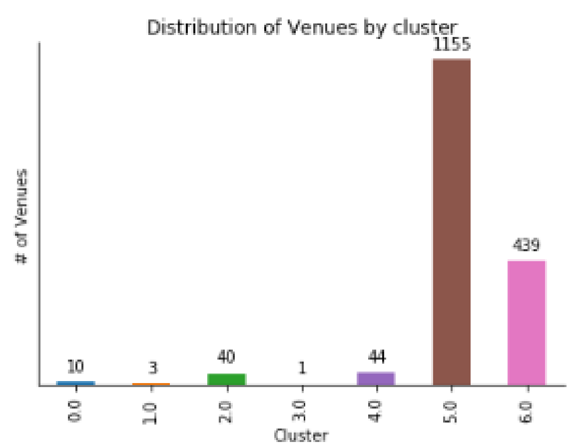

Figure 16: Distribution of venues by cluster

Each cluster includes neighborhoods that have commonality based on the feature set which equates to the categories of the venues in the neighborhood. A key point to notice is the uneven distribution of the neighborhoods and venues across the clusters indicating the similarity or cohesiveness in the clusters 5 and 6.

4.2 MOST COMMON VENUES BY CLUSTER

Next, we look at the most common venue in each cluster in Figure 17 below. Coffee Shops and Fast Food Restaurants are the most common venues in clusters 5 and 6 respectively.

	Cluster Labels	Venue Category	Count
0	0.0	Bar	3
8	1.0	Baseball Field	2
26	2.0	Pizza Place	9
32	3.0	Garden	1
46	4.0	Park	15
99	5.0	Coffee Shop	105
309	6.0	Fast Food Restaurant	20

Figure 17: Most common venue category in each cluster

4.3 CLUSTER SPECIFIC ANALYSIS

4.3.1 Cluster 0

Cluster 0 has 3 neighborhoods which are geographically apart (Scarborough on the east, North York in the northcentral, and Etobicoke) with a suburban flavor, bars, coffee shops, open areas like golf course and dog run, and women's stores.

	Borough	Neighborhood	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
1	Scarborough	Highland Creek, Rouge Hill, Port Union	0.0	Bar	Golf Course	History Museum	Women's Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run
29	North York	Northwood Park, York University	0.0	Coffee Shop	Massage Studio	Caribbean Restaurant	Bar	Women's Store	Dive Bar	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store
102	Etobicoke	Northwest	0.0	Drugstore	Bar	Rental Car Location	Women's Store	Dance Studio	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Dog Run	Dive Bar

Figure 18: Cluster 0 neighborhoods and most common venues

4.3.2 Cluster 1

Cluster 1 has 2 neighborhoods which are in the same boroughs as cluster 0, and with similar characteristics. While outdoor spaces like Parks and Golf Course are a common thread, the key difference is the Baseball Field which is a common thread in this cluster.

	Borough	Neighborhood	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
91	Etobicoke	Humber Bay, King's Mill Park, Kingsway Park So...	1.0	Baseball Field	Construction & Landscaping	Women's Store	Deli / Bodega	Empanada Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run
97	North York	Emery, Humberlea	1.0	Baseball Field	Women's Store	Event Space	Empanada Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run	Dive Bar

Figure 19: Cluster 1 neighborhoods and most common venues

4.3.3 Cluster 2

Cluster 2 has 8 neighborhoods that are geographically apart like cluster 0. Pizza Place and Grocery Stores are the most common venues.

	Borough	Neighborhood	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 I Com V
13	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter	2.0	Pizza Place	Pharmacy	Fast Food Restaurant	Italian Restaurant	Thai Restaurant	Fried Chicken Joint	Chinese Restaurant	Noodle House	Discount Store	Curlin
24	North York	Willowdale West	2.0	Grocery Store	Coffee Shop	Butcher	Pharmacy	Pizza Place	College Stadium	Comfort Food Restaurant	Electronics Store	Eastern European Restaurant	Dum Resta
31	North York	Downsview West	2.0	Grocery Store	Bank	Hotel	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dive Bar	D S
81	York	The Junction North, Runnymede	2.0	Grocery Store	Convenience Store	Bus Line	Pizza Place	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run	D S
89	Etobicoke	Alderwood, Long Branch	2.0	Pizza Place	Pharmacy	Coffee Shop	Pub	Sandwich Place	Skating Rink	Gym	Gastropub	Garden Center	Disc S
94	Etobicoke	Cloverdale, Islington, Martin Grove, Princess ...	2.0	Bank	Women's Store	Event Space	Empanada Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run	Dive
96	North York	Humber Summit	2.0	Empanada Restaurant	Pizza Place	Curling Ice	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run	Dive Bar	Disc S
99	Etobicoke	Westmount	2.0	Pizza Place	Playground	Middle Eastern Restaurant	Sandwich Place	Chinese Restaurant	Coffee Shop	Dog Run	Dive Bar	Discount Store	D

Figure 20: Cluster 2

4.3.4 Cluster 3

Cluster 3 has 1 neighborhood which indicates characteristics that are unique to the venue categories. It also indicates that the neighborhood is small and does not have enough venues to fit other clusters.

	Borough	Neighborhood	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
63	Central Toronto	Roselawn	3.0	Garden	Women's Store	Dance Studio	Empanada Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run	Dive Bar

Figure 21: Cluster 3

4.3.5 Cluster 4

Cluster 4 has 13 neighborhoods which cover most of the boroughs. A common thread is the number of parks, playgrounds, and trails around the Don River Valley. Access to public transportation such as Bus Line is a key feature. Restaurants serving multi-cultural cuisine is common in this cluster.

	Borough	Neighborhood	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
14	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...	4.0	Playground	Park	Curling Ice	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run	Dive Bar
23	North York	York Mills West	4.0	Park	Bank	Women's Store	Dance Studio	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run
25	North York	Parkwoods	4.0	Bus Stop	Food & Drink Shop	Fast Food Restaurant	Park	Diner	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant
30	North York	CFB Toronto, Downsview East	4.0	Airport	Park	Bus Stop	Women's Store	Discount Store	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant
40	East York	East Toronto	4.0	Convenience Store	Coffee Shop	Park	Women's Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store
44	Central Toronto	Lawrence Park	4.0	Bus Line	Park	Swim School	Discount Store	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant	Diner
50	Downtown Toronto	Rosedale	4.0	Park	Playground	Trail	Diner	Dance Studio	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant
64	Central Toronto	Forest Hill North, Forest Hill West	4.0	Bus Line	Trail	Park	Sushi Restaurant	Jewelry Store	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant
74	York	Caledonia-Fairbanks	4.0	Park	Women's Store	Market	Pharmacy	Fast Food Restaurant	Diner	Deli / Bodega	Department Store	Dessert Shop
79	North York	Maple Leaf Park, North Park, Upwood Park	4.0	Park	Basketball Court	Construction & Landscaping	Bakery	Women's Store	Dog Run	Dim Sum Restaurant	Diner	Discount Store
90	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	4.0	River	Pool	Park	Women's Store	Dim Sum Restaurant	Dance Studio	Deli / Bodega	Department Store	Dessert Shop
98	York	Weston	4.0	Park	Convenience Store	Women's Store	Dance Studio	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run
100	Etobicoke	Kingsview Village, Martin Grove Gardens, Richv...	4.0	Pizza Place	Park	Bus Line	Dance Studio	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run

Figure 22: Cluster 4

4.3.6 Cluster 5

Cluster 5 is the largest cluster with 38 neighborhoods and includes 1155 venues. It covers most of the boroughs. Although the most common venue is Coffee Shops and Cafes, this cluster has a wide coverage of restaurants serving multi-cultural cuisine.

	Borough	Neighborhood	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
3	Scarborough	Woburn	5.0	Coffee Shop	Korean Restaurant	Women's Store	Dance Studio	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run	Dive Bar
6	Scarborough	East Birchmount Park, Ionview, Kennedy Park	5.0	Discount Store	Hobby Shop	Coffee Shop	Department Store	Dive Bar	Deli / Bodega	Dessert Shop	Dim Sum Restaurant	Diner	Women's Store
8	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	5.0	Movie Theater	American Restaurant	Motel	Women's Store	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run	Dive Bar
9	Scarborough	Birch Cliff, Cliffside West	5.0	College Stadium	Café	Skating Rink	General Entertainment	Women's Store	Discount Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner

Figure 23: Cluster 5

4.3.7 Cluster 6

Cluster 6 is the second largest cluster with 34 neighborhoods and includes 439 venues. Like cluster 5, it covers most of the boroughs. The most common venues is Fast Food Restaurant. Apart from restaurants, this cluster includes a wide range of retails outlets.

	Borough	Neighborhood	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
0	Scarborough	Rouge, Malvern	6.0	Fast Food Restaurant	Print Shop	Women's Store	Curling Ice	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run	Dive Bar	Discount Store
2	Scarborough	Guildwood, Morningside, West Hill	6.0	Electronics Store	Rental Car Location	Spa	Medical Center	Breakfast Spot	Mexican Restaurant	Pizza Place	Concert Hall	Comfort Food Restaurant	College Gym
4	Scarborough	Cedarbrae	6.0	Athletics & Sports	Lounge	Hakka Restaurant	Fried Chicken Joint	Thai Restaurant	Bakery	Bank	Caribbean Restaurant	Diner	Dim Sum Restaurant
5	Scarborough	Scarborough Village	6.0	Playground	Convenience Store	Dance Studio	Empanada Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run	Dive Bar

Figure 24: Cluster 6

5 DISCUSSION

I observed the following during the analysis of the results:

1. Predominance of 2 clusters across the neighborhoods and venue categories which indicates similarity or commonality of features. The remaining 4 clusters had more distinguishing or unique features.
2. The k-means clustering approach relied of frequency of a category across the 255 unique categories. The feature set may be large compared to the number of samples, i.e. number of neighborhoods (99).
3. I tried multiple values of k in the k-means clustering. For lower values of k, the larger clusters coalesced into a single cluster. For higher values of k, the number of smaller clusters increased but the larger clusters did not break up noticeably any further.
4. The Foursquare data is primarily social and is crowdsourced. I noticed the API calls returned slightly different data sets when executed at various times of the day or day of the week.

Based on the results, I have the following recommendations:

1. I had planned initially to use the Premium Endpoint to fetch ratings but was unable to because of the daily limits of API calls. This extended data could have provided a social dimension, but the data would change frequently.
2. Running the analysis and comparing results over a period as opposed to a snapshot would stabilize the findings.
3. Consider other unsupervised learning methods for comparative analysis.
4. Augment demographic data for neighborhoods to get additional insights.

6 CONCLUSION

In conclusion, this study was a positive step for the stakeholders to understand how data from various sources can be used via powerful tools and visualization techniques to derive insights.

From a personal perspective, it provided me with exposure to the data science methodology from a business problem, analysis, data acquisition, preparation, feature selection, model creation, train/fit and test/analyze results. The libraries for data acquisition, preparation, and visualization demonstrated the value of data science.