Identifying Unique Local Venues in Toronto

1. Introduction

1.1 Background

Toronto is the largest and the most populous city in Canada, with a population of nearly 3 million people. It is one of the most diverse and multicultural metropolitan areas in the world. It is also one of Canada's leading tourism destinations and has a rich cultural life. In 2017, the Toronto-area received 44 million tourists, of whom over 10 million were domestic visitors and 3 million were from the United States, spending a total of $8.84 billion.

As is generally the case with such cities, Toronto has a rich food culture that includes a vast array of international cuisines influenced by the city's diverse demographics. It is home to tens of thousands of restaurants and several hundreds of thousands of independent as well as chain coffee shops, bakeries, food stores, etc.

As such, when tourists make plans to visit Toronto or new residents from different parts of Canada or immigrants from other countries plan their moves to Toronto, there is quite a lot of research material available with regards to popular food and tourist destinations. However, some of the other venues that are an integral part of people's day-to-day lives receive less focus and attention. Some examples of these are pet stores, local outdoor recreation facilities, fitness centers, to name a few.

So, this project will attempt to identify top local venues in four preselected venue categories, i.e., fitness, pets, outdoor recreation, and transportation. It will exclude restaurants and other food-related venues returned as a part of the venue analysis, as this data is not relevant to the problem statement being addressed by this project. Additionally, it will limit the number of boroughs and/or neighborhoods that will be considered in scope, to run a practical and meaningful analysis.

The project will utilize data science tools to analyze Toronto neighborhood data and venue data from Foursquare API, and focus on four venue types listed above, and will aim to determine the presence and distribution of venues of these four types across the neighborhoods of Toronto.

1.2 Problem

As introduced above, the aim of this project is to find venues and local establishments in Toronto, Canada, that specifically fall into one of the four select categories, i.e., fitness, pets, outdoor recreation, and transportation. These venue types were selected keeping in mind the target stakeholders such as people who have newly moved to Toronto and are looking to find local spots that they can frequent as a part of their new routine, as these locations are not commonly covered in tourist brochures or materials. The project aims to solve this problem by leveraging data on Toronto neighborhood venues, their types, frequencies, etc., thus, providing beneficial information to people looking for a unique experience in a large and diverse city, such as Toronto.

1.3 Interest

While a lot of information is readily available on top restaurants and common tourist attractions, people who have newly moved to the city of Toronto from different parts of Canada or from international locations, for temporary assignments or permanent relocation, can greatly benefit from finding local spots like gyms/fitness centers, bookshops, pet stores etc. which are not commonly featured in visitor resources. It can also benefit tourists and visitors who may be interested in checking out local establishments.

2. Data acquisition and cleaning

2.1 Data sources

This assignment requires exploring, segmenting, and clustering the neighborhoods in the city of Toronto. While this neighborhood data is not readily available on the internet, a Wikipedia page exists for the Toronto neighborhood data, that has all the information needed to explore and cluster the neighborhoods in Toronto. This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Specifically, this table of Toronto postal codes also contains the names of Boroughs and Neighborhoods in Toronto.

Next, geographical coordinates for each of the Toronto postal codes will be extracted from a csv file: Toronto Geospatial Data. Finally, venue categories and details required for the last part of the analysis will obtained via Foursquare API.

2.2 Data cleaning

Once the dataset is scraped from the above Wikipedia page, it will need to be wrangled, cleaned, and read into a *pandas* dataframe so that it is in a structured format for further analysis.

After reading the data from the Wikipedia page, we notice that the headers, *Postal code, Borough,* and *Neighborhood* have trailing new line characters ("\n"), which need to be removed. Also, each of the data fields have a similar issue, where they have a trailing new line character. Therefore, these fields are also cleaned to ensure a consistent format.

Additionally, there are records where the Boroughs are "Not Assigned". For the purpose of this analysis, these records need to be dropped. Similarly, there are records where the Neighborhood column is blank. To address this, we identify such records and assign the value of the *Borough* field to the corresponding *Neighborhood* field, for consistency.

Once the data is in a structured format, we can start the analysis of the dataset to explore and cluster the neighborhoods in the city of Toronto.

2.3 Feature Selection

After the first step of data cleaning described above, there were 181 samples and three features in the dataset. After completing the second step of data cleaning, that is, dropping rows where the *Borough* field has a "Not Assigned" value, we end up with 104 samples (and three features).

Upon examining each of the three features, it is clear that all three features present in the dataset, *Postal code, Borough,* and *Neighborhood,* are essential to our analysis. Therefore, we proceed to the next step with these features.

3. Methodology

3.1 Explore Dataset

After reading the data from the Wikipedia page and cleaning it, it is transformed into a *pandas* dataframe. Before performing this step, we import necessary libraries such as *pandas* and *numpy.*

This dataframe captures the three essential features discussed above: *Postal code, Borough,* and *Neighborhood.*

	Postal code	Borough	Neighborhood
0	M1A	Not assigned	
1	M2A	Not assigned	
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park / Harbourfront
5	M6A	North York	Lawrence Manor / Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government
7	M8A	Not assigned	
8	M9A	Etobicoke	Islington Avenue
9	M1B	Scarborough	Malvern / Rouge

Figure 1: **A snapshot of a subset of records from the *pandas* dataframe**

To prepare the dataset for further analyses, as previously described in the section above, (1) we drop records where the *Borough* field has a "Not Assigned" value, and, (2) assign the value from the *Borough* column to the *Neighborhood* column, where the *Neighborhood* field is blank.

	Postal code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park / Harbourfront
3	M6A	North York	Lawrence Manor / Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government
5	M9A	Etobicoke	Islington Avenue
6	M1B	Scarborough	Malvern / Rouge
7	M3B	North York	Don Mills
8	M4B	East York	Parkview Hill / Woodbine Gardens
9	M5B	Downtown Toronto	Garden District, Ryerson

Figure 2: **A snapshot of a subset of records from the *pandas* dataframe after 'Not Assigned' Boroughs are dropped**

Finally, the Neighborhood records are grouped based on Postal codes and Boroughs, resulting in a clean and consolidated dataset which is conducive to performing the next step, i.e., exploring the neighborhoods in Toronto.

	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
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	Kennedy Park , Ionview , East Birchmount Park
7	M1L	Scarborough	Golden Mile , Clairlea , Oakridge
8	M1M	Scarborough	Cliffside , Cliffcrest , Scarborough Village West
9	M1N	Scarborough	Birch Cliff , Cliffside West

***Figure 3:* A snapshot of a subset of records from the consolidated dataframe, that is grouped by postal codes and boroughs**

Then, we need to use **geopy** package to get the geographical coordinates of the Toronto neighborhoods. However, this package can sometimes function in an unreliable fashion and may require multiple calls to get the geographical coordinates of the neighborhoods.

Therefore, this project used a **.csv** file that has the geographical coordinates of each postal code found in Toronto: http://cocl.us/Geospatial_data

This data was read into a *pandas* dataframe, cleaned to ensure that it is ready for merging with the initial dataframe. After the data cleaning step, both dataframes were merged, such that the resulting dataframe contains geographical coordinates (latitude, longitude) for each postal code in the original dataframe. This merged dataframe now has five columns or features: *Postal code, Borough, Neighborhood, Latitude,* and *Longitude*.

	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
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	Kennedy Park , Ionview , East Birchmount Park	43.727929	-79.262029
7	M1L	Scarborough	Golden Mile , Clairlea , Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffside , Cliffcrest , Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff , Cliffside West	43.692657	-79.264848
10	M1P	Scarborough	Dorset Park , Wexford Heights , Scarborough To...	43.757410	-79.273304
11	M1R	Scarborough	Wexford , Maryvale	43.750072	-79.295849
12	M1S	Scarborough	Agincourt	43.794200	-79.262029
13	M1T	Scarborough	Clarks Corners , Tam O'Shanter , Sullivan	43.781638	-79.304302
14	M1V	Scarborough	Milliken , Agincourt North , Steeles East , L'...	43.815252	-79.284577

***Figure 4:* A snapshot of a subset of records from the merged dataframe that includes geographical coordinates for each postal code**

In the next step, we import and use the ***folium*** visualization library in conjunction with the geographical coordinates from the above step, to create a map of Toronto neighborhoods. Note: to reduce complexity, this project takes into account only those boroughs that contain the term "Toronto".

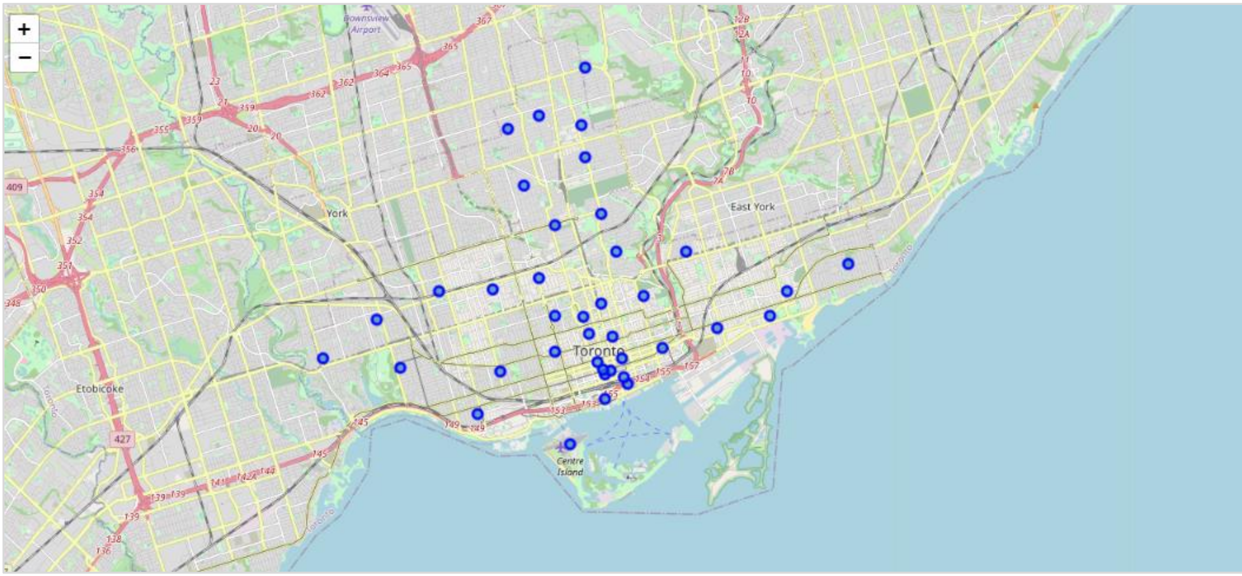

Figure 5: **Map of Toronto neighborhoods created using *folium* visualization library**

3.2 Explore Neighborhoods in Toronto

In this stage, the **Foursquare API** will be utilized to explore the neighborhoods in Toronto and segment them. For this, it is required to define our Foursquare credentials and version as a starting point. We then get a neighborhood's name and geographical coordinates to start the neighborhood exploration process. We then set a limit on the number of venues returned by the Foursquare API to 100 as well as define a radius of 500 meters. Once the API returns the results in the form of a JSON file, it is cleaned and structured into a *pandas* dataframe.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Glen Stewart Park	43.675278	-79.294647	Park
4	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood

Figure 6: **A snapshot of the dataframe with venue details returned by the Foursquare API**

We can create a user defined function, **getNearbyVenues**, to repeat the above process for all neighborhoods in Toronto. 226 unique venue categories were identified across 39 Toronto boroughs as a result of this process.

3.3 Analyze Each Neighborhood

Using **one-hot encoding** technique, we further analyze each neighborhood in the boroughs of Toronto city. Next, rows are grouped by neighborhood, by taking the mean of the frequency of occurrence of each venue category.

	Neighborhood	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Auto Workshop	BBQ Joint	Baby Store	Bagel Shop	Bakery	Bank	Bar
0	Berczy Park	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	0.017241	0.000000	0.000000	0.000000	0.000000	0.017241	0.000000	0.017241	0.034483	0.000000	0.000000
1	Brockton , Parkdale Village , Exhibition Place	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.041667	0.000000	0.041667
2	Business reply mail Processing CentrE	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.058824	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	CN Tower , King and Spadina , Railway Lands , ...	0.0625	0.0625	0.0625	0.125	0.1875	0.125	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.062500

Figure 7: **A snapshot showing Toronto neighborhoods with the different venue categories and their frequency of occurrence**

In the next step, the top 5 (or 10) venue categories can be printed out for each neighborhood, which is a crucial part of this project. These should then be converted into a *pandas* dataframe for the next step of the analysis.

```
----Berczy Park----
          venue  freq
0   Coffee Shop  0.07
1  Cocktail Bar  0.05
2        Bakery  0.03
3      Beer Bar  0.03
4    Restaurant  0.03


----Brockton , Parkdale Village , Exhibition Place----
                  venue  freq
0                  Café  0.12
1           Coffee Shop  0.08
2        Breakfast Spot  0.08
3             Nightclub  0.08
4  Performing Arts Venue  0.04
```

Figure 8: **An example of neighborhoods and their top venue categories**

3.4 Cluster Neighborhoods

Once the dataframe with Toronto neighborhoods and their top venue categories is created, we use the **k-means Clustering algorithm**, for clustering Toronto neighborhoods. The k-value was set at 5, and the algorithm was run to cluster the Toronto neighborhoods into 5 clusters.

We then create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood, as shown below.

	Postal code	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Park	Pub	Trail	Health Food Store	Yoga Studio	Diner	Deli / Bodega	Department Store	Dessert Shop	Distribution Center
1	M4K	East Toronto	The Danforth West , Riverdale	43.679557	-79.352188	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bookstore	Furniture / Home Store	Frozen Yogurt Shop	Cosmetics Shop	Brewery	Bubble Tea Shop
2	M4L	East Toronto	India Bazaar , The Beaches West	43.668999	-79.315572	0	Fast Food Restaurant	Pizza Place	Sandwich Place	Italian Restaurant	Restaurant	Steakhouse	Ice Cream Shop	Board Shop	Brewery	Burrito Place
3	M4M	East Toronto	Studio District	43.659526	-79.340923	0	Café	Coffee Shop	Gastropub	Bakery	Brewery	American Restaurant	Convenience Store	Sandwich Place	Cheese Shop	Pet Store
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	0	Park	Bus Line	Swim School	Deli / Bodega	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run	Distribution Center

Figure 9: **A snapshot of the merged dataframe that includes cluster information in addition to the neighborhood venue categories**

Finally, these neighborhood clusters are visualized with the help of **folium** visualization library as described in a previous section.

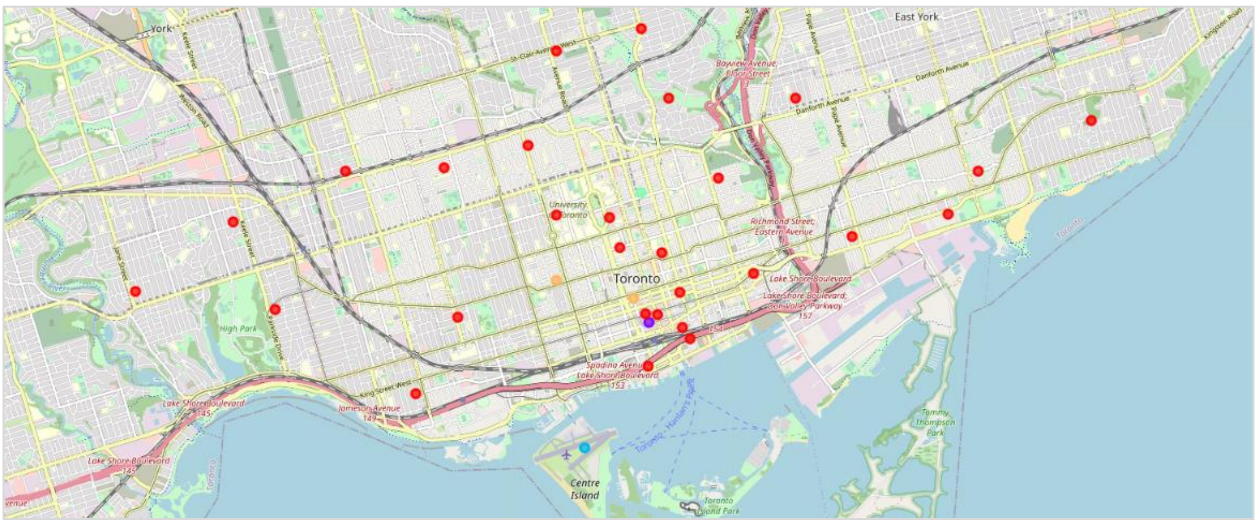

Figure 10: **A map of Toronto neighborhood clusters generated using the *folium* library**

3.5 Examine Clusters

In this stage, each of the five neighborhood clusters are examined to determine the venue categories that distinguish each cluster from the others. Based on the defining categories, different names have been assigned to each of these neighborhood clusters in Toronto.

For example, in the first neighborhood cluster, the key venue categories are restaurants and coffee shops while the third cluster has a unique venue category of airport that distinguishes it from all other clusters.

	Borough	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	East Toronto	0	Park	Pub	Trail	Health Food Store	Yoga Studio	Diner	Deli / Bodega	Department Store	Dessert Shop	Distribution Center
1	East Toronto	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bookstore	Furniture / Home Store	Frozen Yogurt Shop	Cosmetics Shop	Brewery	Bubble Tea Shop
2	East Toronto	0	Fast Food Restaurant	Pizza Place	Sandwich Place	Italian Restaurant	Restaurant	Steakhouse	Ice Cream Shop	Board Shop	Brewery	Burrito Place
3	East Toronto	0	Café	Coffee Shop	Gastropub	Bakery	Brewery	American Restaurant	Convenience Store	Sandwich Place	Cheese Shop	Pet Store
4	Central Toronto	0	Park	Bus Line	Swim School	Deli / Bodega	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run	Distribution Center
5	Central Toronto	0	Dance Studio	Hotel	Park	Breakfast Spot	Gym	Sandwich Place	Department Store	Food & Drink Shop	Donut Shop	Doner Restaurant
6	Central Toronto	0	Clothing Store	Coffee Shop	Yoga Studio	Bagel Shop	Mexican Restaurant	Rental Car Location	Spa	Restaurant	Chinese Restaurant	Toy / Game Store
7	Central Toronto	0	Sandwich Place	Dessert Shop	Pizza Place	Café	Gym	Coffee Shop	Italian Restaurant	Sushi Restaurant	Deli / Bodega	Bar
8	Central Toronto	0	Playground	Yoga Studio	Dance Studio	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run	Distribution Center	Discount Store
9	Central Toronto	0	Coffee Shop	Pub	Supermarket	Bank	Sports Bar	Sushi Restaurant	Pizza Place	Liquor Store	Fried Chicken Joint	Light Rail Station
10	Downtown Toronto	0	Park	Trail	Playground	Yoga Studio	Cupcake Shop	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run	Distribution Center

Figure 11: **A snapshot of the dataframe showing the first Toronto neighborhood cluster where the top venue categories are restaurants and coffee shops**

	Borough	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
27	Downtown Toronto	2	Airport Service	Airport Lounge	Airport Terminal	Airport	Harbor / Marina	Coffee Shop	Sculpture Garden	Boutique	Bar	Boat or Ferry

Figure 12: **A snapshot of the dataframe showing the third Toronto neighborhood cluster with the distinguishing venue category of airport and related services**

4. Results

The objective of this project was to help stakeholders identify some of the unique local spots in the city of Toronto, that fall in one of the four select categories. This has been achieved by first scraping the Wikipedia page that contains the necessary neighborhood details for Toronto, Foursquare API to get different venue categories and venue details, and applying k-means clustering algorithm to segment and cluster the various neighborhoods in the city of Toronto.

By following the methodology described in the section above and performing an analysis of the Toronto neighborhood dataset, the following results were achieved.

As noted previously, only those boroughs containing 'Toronto' in their name were considered in scope for the purpose of this analysis. This resulted in 39 boroughs in Toronto being grouped into 5 clusters, for which 226 unique venue categories were returned by the Foursquare API. This section will discuss each of these 5 neighborhood clusters in detail and leverage the venue information returned by the analysis to address the questions this project set out to answer.

4.1 Cluster 1

When k-means clustering algorithm was applied, with k=5, the first cluster returned was the following. The Toronto boroughs covered in this cluster include, *East Toronto*, *Central Toronto*, *Downtown Toronto*, and *West Toronto.* As is evident from the dataframe included below, while setting aside the most popular venue categories of 'restaurants and coffee shops', this neighborhood offers popular alternate venues of the four preselected venue types as follows:

Venue Type (Result Category)	Foursquare Venue Category	Borough(s)
Fitness	Gym	Central Toronto, Downtown Toronto, West Toronto
Fitness	Dance Studio	Central Toronto, Downtown Toronto
Fitness	Yoga Studio	East Toronto, Central Toronto, Downtown Toronto, West Toronto
Fitness	Swim School	Central Toronto
Pets	Pet Store	East Toronto, Downtown Toronto
Pets	Dog Run	Central Toronto, Downtown Toronto, West Toronto
Outdoor Recreation	Playground	Central Toronto, Downtown Toronto
Outdoor Recreation	Park	East Toronto, Central Toronto, Downtown Toronto
Outdoor Recreation	Trail	East Toronto, Central Toronto, Downtown Toronto
Transport	Light Rail Station	Central Toronto, East Toronto
Transport	Bus Line	Central Toronto

Table 1: **Summary of results from Cluster 1 analysis**

	Borough	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	East Toronto	0	Park	Pub	Trail	Health Food Store	Yoga Studio	Diner	Deli / Bodega	Department Store	Dessert Shop	Distribution Center
1	East Toronto	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bookstore	Furniture / Home Store	Frozen Yogurt Shop	Cosmetics Shop	Brewery	Bubble Tea Shop
2	East Toronto	0	Fast Food Restaurant	Pizza Place	Sandwich Place	Italian Restaurant	Restaurant	Steakhouse	Ice Cream Shop	Board Shop	Brewery	Burrito Place
3	East Toronto	0	Café	Coffee Shop	Gastropub	Bakery	Brewery	American Restaurant	Convenience Store	Sandwich Place	Cheese Shop	Pet Store
4	Central Toronto	0	Park	Bus Line	Swim School	Deli / Bodega	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run	Distribution Center
5	Central Toronto	0	Dance Studio	Hotel	Park	Breakfast Spot	Gym	Sandwich Place	Department Store	Food & Drink Shop	Donut Shop	Doner Restaurant
6	Central Toronto	0	Clothing Store	Coffee Shop	Yoga Studio	Bagel Shop	Mexican Restaurant	Rental Car Location	Spa	Restaurant	Chinese Restaurant	Toy / Game Store
7	Central Toronto	0	Sandwich Place	Dessert Shop	Pizza Place	Café	Gym	Coffee Shop	Italian Restaurant	Sushi Restaurant	Deli / Bodega	Bar
8	Central Toronto	0	Playground	Yoga Studio	Dance Studio	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run	Distribution Center	Discount Store
9	Central Toronto	0	Coffee Shop	Pub	Supermarket	Bank	Sports Bar	Sushi Restaurant	Pizza Place	Liquor Store	Fried Chicken Joint	Light Rail Station
10	Downtown Toronto	0	Park	Trail	Playground	Yoga Studio	Cupcake Shop	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run	Distribution Center

Figure 13: **A snapshot showing the first Toronto neighborhood cluster**

4.2 Cluster 2

The second cluster returned by the algorithm includes only one Toronto borough, i.e., Downtown Toronto. And the analysis of this cluster shows that all top venue categories provided by the Foursquare API for this neighborhood align with the restaurants or food places category. Thus, this cluster does not yield any results that are necessary to address the problem this project aims to solve. i.e., it does not return any venue that fall under one of the four chosen venue types.

	Borough	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
20	Downtown Toronto	1	Coffee Shop	Hotel	Café	Restaurant	Japanese Restaurant	Seafood Restaurant	Salad Place	Deli / Bodega	American Restaurant	Asian Restaurant

Figure 14: **A snapshot showing the second Toronto neighborhood cluster**

4.3 Cluster 3

The third cluster returned by the algorithm also includes only one Toronto borough, i.e., Downtown Toronto. The analysis of this cluster shows that this neighborhood includes the Airport, which falls in one of the four venue categories in scope for this project's problem statement. The presence of Airport and relates service options in this neighborhood cluster makes it unique and differentiates it from all the other neighborhoods.

	Borough	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
27	Downtown Toronto	2	Airport Service	Airport Lounge	Airport Terminal	Airport	Harbor / Marina	Coffee Shop	Sculpture Garden	Boutique	Bar	Boat or Ferry

Figure 15: **A snapshot showing the third Toronto neighborhood cluster**

Venue Type (Result Category)	Foursquare Venue Category	Borough(s)
Transport	Airport	Downtown Toronto
Transport	Boat or Ferry	Downtown Toronto

Table 2: **Summary of results from Cluster 3 analysis**

4.4 Cluster 4

Similar to the previous two clusters, the fourth cluster below also includes only one Toronto borough, i.e., Downtown Toronto. The analysis of this cluster shows that majority of the top venue categories returned by the Foursquare API for this neighborhood fall in venue categories that are not in scope for this project, i.e., restaurants/coffee shops/bars. However, there is one venue category, *Yoga Studio*, in this cluster that can help address the problem statement at hand.

	Borough	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
12	Downtown Toronto	3	Sushi Restaurant	Japanese Restaurant	Coffee Shop	Gay Bar	Restaurant	Pub	Gastropub	Hotel	Yoga Studio	Men's Store

Figure 16: **A snapshot showing the fourth Toronto neighborhood cluster**

Venue Type (Result Category)	Foursquare Venue Category	Borough(s)
Fitness	Yoga Studio	Downtown Toronto

Table 3: **Summary of results from Cluster 4 analysis**

4.5 Cluster 5

Since the k in the k-means algorithm used for this project was set at 5, the following is the final cluster that was analyzed for this project. It covers two neighborhoods within Downtown Toronto borough. Analyzing this final cluster provides one meaningful result which is tabulated below.

	Borough	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
18	Downtown Toronto	4	Coffee Shop	Café	Restaurant	American Restaurant	Clothing Store	Hotel	Thai Restaurant	Gym	Deli / Bodega	Cosmetics Shop
26	Downtown Toronto	4	Café	Coffee Shop	Mexican Restaurant	Vietnamese Restaurant	Bakery	Dessert Shop	Gaming Cafe	Grocery Store	Bar	Vegetarian / Vegan Restaurant

Figure 17: **A snapshot showing the fifth Toronto neighborhood cluster**

Venue Type (Result Category)	Foursquare Venue Category	Borough(s)
Fitness	Gym	Downtown Toronto

Table 4: **Summary of results from Cluster 5 analysis**

4.6 Summary

The results obtained by analyzing the five neighborhood clusters returned by the k-means clustering algorithm are summarized here below. This project looks at four select categories: *Fitness*, *Pets*, *Outdoor Recreation*, and *Transport*. It does not include venue categories such as restaurants, coffee shops, and other food related venues.

Results from Tables 1 through 4 above are consolidated in Table 5 below. Redundant records were dropped to avoid reporting duplicate results.

Venue Type (Result Category)	**Foursquare Venue Category**	**Borough(s)**
Fitness	Gym	Central Toronto, Downtown Toronto, West Toronto
Fitness	Dance Studio	Central Toronto, Downtown Toronto
Fitness	Yoga Studio	East Toronto, Central Toronto, Downtown Toronto, West Toronto
Fitness	Swim School	Central Toronto
Pets	Pet Store	East Toronto, Downtown Toronto
Pets	Dog Run	Central Toronto, Downtown Toronto, West Toronto
Outdoor Recreation	Playground	Central Toronto, Downtown Toronto
Outdoor Recreation	Park	East Toronto, Central Toronto, Downtown Toronto
Outdoor Recreation	Trail	East Toronto, Central Toronto, Downtown Toronto
Transport	Light Rail Station	Central Toronto, East Toronto
Transport	Bus Line	Central Toronto
Transport	Airport	Downtown Toronto
Transport	Boat or Ferry	Downtown Toronto

Table 5: **Summary of results consolidated from analysis of 5 Toronto Neighborhood clusters**

5. Discussion

This project set out to identify unique local establishments that fall in four select venue categories: *Fitness*, *Pets*, *Outdoor Recreation*, and *Transport.* Segmenting and exploring different boroughs and neighborhoods in the city of Toronto by bringing together location data and venue data (Foursquare), and analyzing it with the help of a clustering algorithm such as k-means, helped find answers to the problem this project set out to solve.

To discuss the results summarized above, while most top venues in a majority of the Toronto neighborhoods are restaurants, coffee shops, bars, and other food/drink places, we were able to identify other local establishments and locations that are offered by the different neighborhoods of the city, including gyms / fitness centers, yoga studios, dance studios, parks, dog runs, and pet stores that are critical to providing a fulfilling experience to Toronto residents (new and current) as well as visitors and tourists.

As information on restaurants and other food-related venues is more readily available than information on other types of venues, the above analysis uncovers the fact that there are several neighborhoods in a large and diverse metropolitan city like Toronto where other venue categories, such as the four selected for this project, consistently fall in the top venue categories. This information can be easily accessed by the temporary visitors and new residents of the city to make these locations a part of their new life or routine in the city of Toronto.

6. Conclusion

In conclusion, this project report explored and analyzed the different neighborhoods in the city of Toronto, Canada, using the postal codes, boroughs, neighborhoods, and geographical coordinates data with the objective of identifying popular local establishments of four preselected venue types. It segmented and clustered each of the Toronto neighborhoods with the help of k-means Clustering algorithm. And then studied the different venue categories, specifically focusing on popular yet unique local venues such as fitness centers, pet stores, etc., while excluding some of the conventionally popular venue types such as restaurants and coffee shops. It helped uncover insights on local spots that are not generally highlighted in other readily-available tourist resources and materials, and can be leveraged by new residents of this city as well as visitors to have a unique experience during their time in the city of Toronto.