Coursera Capstone Project - Optimal Locations for Coffee Shops in Toronto

A Python workbook created by Catherine Gao

In this capstone project, I will leverage the Foursquare location data to solve a real-world business problem.

1. Introduction/Business Problem

In this capstone project, I will identify the optimal locations to open a coffee shop in the city of Toronto, Ontario. Toronto is one of the major cities in Canada. Being one of the country's populous metropolitan areas, Toronto is an international centre for business and finance, and home to numerous tourists from all over the world.

Opening a coffee shop in this city represents business opportunity but also strong competition at the same time. This project will explore clusters of neighborhoods in Toronto to recommend locations for a good coffee shop business.

The target audience of this project will be those who desire to open a coffee shop in the city of Toronto. Those who are interested in opening a business in Toronto or other metropolitan areas can apply the same analysis to their choice of business.

2. Data

This project will leverage the following two data sets:

1. Toronto Postall Codes Wikipedia Page - I will scrape the following Wikipedia page, [https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M), in order to obtain the data on postal codes and boroughs in the city of Toronto.
2. Foursquare location data - an API that provides multi-dimensional data on geolocations, venues, and attributes. This data will help to identify types of business in a specific area and their associated ranks. A link to connect to Foursquare's API platform can be found here: [https://developer.foursquare.com/ (https://developer.foursquare.com/)](https://developer.foursquare.com/).
3. Geolocation csv file - a csv file that has the geographical coordinates of each postal code: [http://cocl.us/Geospatial_data (http://cocl.us/Geospatial_data)](http://cocl.us/Geospatial_data). This data will be used to visualize the clusters of neighborhoods on a folium map.

3. Exploratory Data Analysis

which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

3.1 Data Processing

First, scrape data and clean up neighborhood data for Toronto. Put it into an acational pandas dataframe.

More than one neighborhood can exist in one postal code area. For example, in the table on the Wikipedia page, you will notice that M5A is listed twice and has two neighborhoods: Harbourfront and Regent Park. These two rows will be combined into one row with the neighborhoods separated with a comma.

Only process the cells that have an assigned borough. Ignore cells with a borough that is Not assigned. If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough.

	PostalCode	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, Humber Valley Village
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

3.2 Loading Geolocation Data

Use the Geocoder csv file for the geographical coordinates of each postal code.

	Postal Code	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

Merge the two data sets on matching Postal Code.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
7	M3B	North York	Don Mills	43.745906	-79.352188
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937

3.3 Display Boroughs in Toronto on a Map

First, I will use geopy library to get the latitude and longitude values of Toronto, Ontario. Then I will add markers on the map to represent the locations of boroughs.

```
The geograpical coordinate of Toronto are 43.6534817, -79.3839347.
```
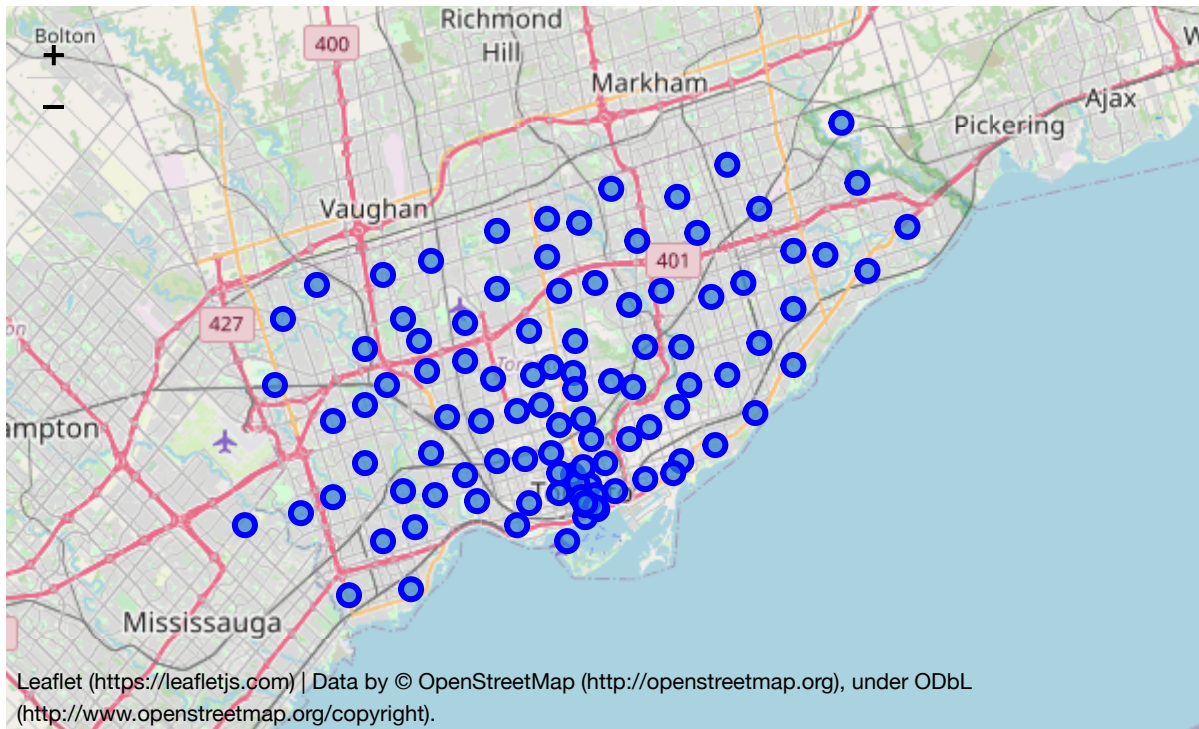
3.4 Use Foursquare Data to Understand the Businesses in the Neighborhoods

Let's get the top 100 venues that are in these neighborhoods within a radius of 500 meters.

```
Your credentails:
CLIENT_ID: Removed for privacy.
CLIENT_SECRET: Removed for privacy.
```

Parkwoods
Victoria Village
Regent Park, Harbourfront
Lawrence Manor, Lawrence Heights
Queen's Park, Ontario Provincial Government
Islington Avenue, Humber Valley Village
Malvern, Rouge
Don Mills
Parkview Hill, Woodbine Gardens
Garden District, Ryerson
Glencairn
West Deane Park, Princess Gardens, Martin Grove, Islington, Cloverdale
Rouge Hill, Port Union, Highland Creek
Don Mills
Woodbine Heights
St. James Town
Humewood-Cedarvale
Eringate, Bloordale Gardens, Old Burnhamthorpe, Markland Wood
Guildwood, Morningside, West Hill
The Beaches
Berczy Park
Caledonia-Fairbanks
Woburn
Leaside
Central Bay Street
Christie
Cedarbrae
Hillcrest Village
Bathurst Manor, Wilson Heights, Downsview North
Thorncliffe Park
Richmond, Adelaide, King
Dufferin, Dovercourt Village
Scarborough Village
Fairview, Henry Farm, Oriole
Northwood Park, York University
East Toronto, Broadview North (Old East York)
Harbourfront East, Union Station, Toronto Islands
Little Portugal, Trinity
Kennedy Park, Ionview, East Birchmount Park
Bayview Village
Downsview
The Danforth West, Riverdale
Toronto Dominion Centre, Design Exchange
Brockton, Parkdale Village, Exhibition Place
Golden Mile, Clairlea, Oakridge
York Mills, Silver Hills
Downsview
India Bazaar, The Beaches West
Commerce Court, Victoria Hotel
North Park, Maple Leaf Park, Upwood Park
Humber Summit
Cliffside, Cliffcrest, Scarborough Village West
Willowdale, Newtonbrook
Downsview
Studio District
Bedford Park, Lawrence Manor East

Del Ray, Mount Dennis, Keelsdale and Silverthorn
Humberlea, Emery
Birch Cliff, Cliffside West
Willowdale, Willowdale East
Downsview
Lawrence Park
Roselawn
Runnymede, The Junction North
Weston
Dorset Park, Wexford Heights, Scarborough Town Centre
York Mills West
Davisville North
Forest Hill North & West, Forest Hill Road Park
High Park, The Junction South
Westmount
Wexford, Maryvale
Willowdale, Willowdale West
North Toronto West,  Lawrence Park
The Annex, North Midtown, Yorkville
Parkdale, Roncesvalles
Canada Post Gateway Processing Centre
Kingsview Village, St. Phillips, Martin Grove Gardens, Richview Gardens
Agincourt
Davisville
University of Toronto, Harbord
Runnymede, Swansea
Clarks Corners, Tam O'Shanter, Sullivan
Moore Park, Summerhill East
Kensington Market, Chinatown, Grange Park
Milliken, Agincourt North, Steeles East, L'Amoreaux East
Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst
Quay, South Niagara, Island airport
New Toronto, Mimico South, Humber Bay Shores
South Steeles, Silverstone, Humbergate, Jamestown, Mount Olive, Beaumon
d Heights, Thistletown, Albion Gardens
Steeles West, L'Amoreaux West
Rosedale
Stn A PO Boxes
Alderwood, Long Branch
Northwest, West Humber - Clairville
Upper Rouge
St. James Town, Cabbagetown
First Canadian Place, Underground city
The Kingsway, Montgomery Road, Old Mill North
Church and Wellesley
Business reply mail Processing Centre, South Central Letter Processing
Plant Toronto
Old Mill South, King's Mill Park, Sunnylea, Humber Bay, Mimico NE, The
Queensway East, Royal York South East, Kingsway Park South East
Mimico NW, The Queensway West, South of Bloor, Kingsway Park South Wes
t, Royal York South West


There are 273 uniques categories.

By display the top 10 venue categories in Toronto, we notice there are many coffee shops and cafes in the area. This represents substantial competition to any new coffee shop businesses.



Top 10 Venue Category in Toronto



Top 20 Toronto Neighborhoods with Most Venue Categories
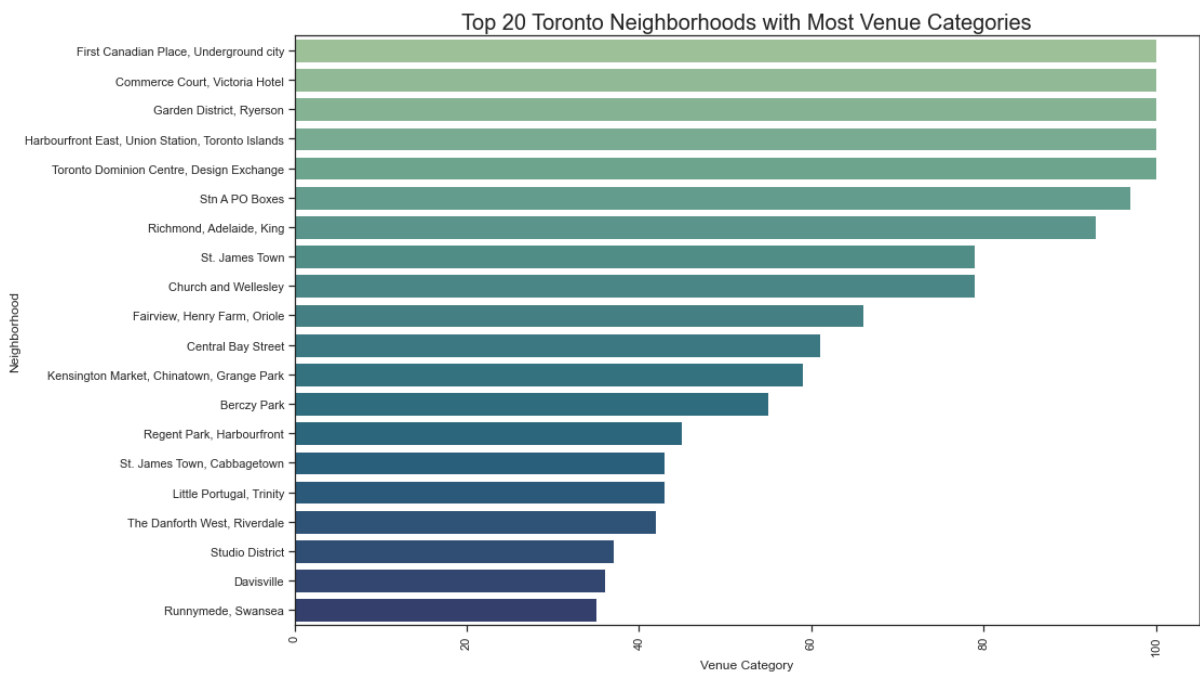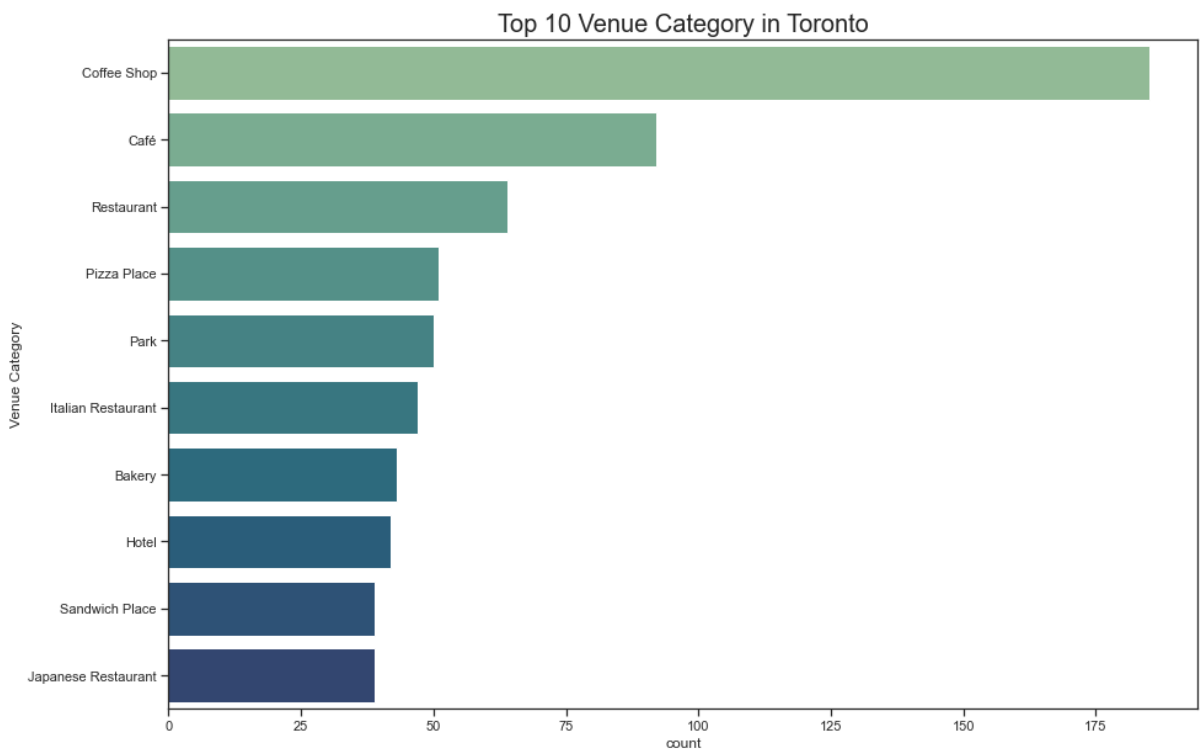
Identify the top 10 venue categories for each neighborhood.

| | 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 | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Agincourt | Skating Rink | Lounge | Latin American Restaurant | Breakfast Spot | Donut Shop | Diner | Discount Store | D |
| 1 | Alderwood, Long Branch | Pizza Place | Gym | Coffee Shop | Skating Rink | Pub | Discount Store | Deli / Bodega | D |
| 2 | Bathurst Manor, Wilson Heights, Downsview North | Coffee Shop | Bank | Pharmacy | Fried Chicken Joint | Shopping Mall | Bridal Shop | Diner | |
| 3 | Bayview Village | Japanese Restaurant | Café | Bank | Chinese Restaurant | Dessert Shop | Dim Sum Restaurant | Diner | |
| 4 | Bedford Park, Lawrence Manor East | Italian Restaurant | Coffee Shop | Sandwich Place | Women's Store | Butcher | Restaurant | Café | |

Now let's compile a list of neighborhoods that have "Coffee Shop" or "Cafe" as the 1st most common venue.

| | 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 | 7 C |
|---|---|---|---|---|---|---|---|---|
| 2 | Bathurst Manor, Wilson Heights, Downsview North | Coffee Shop | Bank | Pharmacy | Fried Chicken Joint | Shopping Mall | Bridal Shop | |
| 5 | Berczy Park | Coffee Shop | Cocktail Bar | Bakery | Beer Bar | Farmers Market | Seafood Restaurant | |
| 7 | Brockton, Parkdale Village, Exhibition Place | Café | Coffee Shop | Breakfast Spot | Bakery | Stadium | Burrito Place | Fu |
| 11 | Canada Post Gateway Processing Centre | Coffee Shop | Hotel | Mediterranean Restaurant | American Restaurant | Intersection | Food Truck | Sა |
| 13 | Central Bay Street | Coffee Shop | Italian Restaurant | Sandwich Place | Café | Salad Place | Bubble Tea Shop | |
| 15 | Church and Wellesley | Coffee Shop | Sushi Restaurant | Japanese Restaurant | Gay Bar | Restaurant | Yoga Studio | |
| 18 | Commerce Court, Victoria Hotel | Coffee Shop | Restaurant | Café | Hotel | Gym | American Restaurant | Re |
| 29 | First Canadian Place, Underground city | Coffee Shop | Café | Hotel | Restaurant | Gym | Japanese Restaurant | |
| 31 | Garden District, Ryerson | Coffee Shop | Clothing Store | Middle Eastern Restaurant | Italian Restaurant | Bubble Tea Shop | Japanese Restaurant | |
| 35 | Harbourfront East, Union Station, Toronto Islands | Coffee Shop | Aquarium | Café | Hotel | Italian Restaurant | Fried Chicken Joint | |
| 43 | Kensington Market, Chinatown, Grange Park | Café | Coffee Shop | Vegetarian / Vegan Restaurant | Vietnamese Restaurant | Mexican Restaurant | Bar | |
| 55 | North Toronto West, Lawrence Park | Coffee Shop | Clothing Store | Yoga Studio | Italian Restaurant | Pet Store | Chinese Restaurant | |
| 62 | Queen's Park, Ontario Provincial Government | Coffee Shop | Sushi Restaurant | College Cafeteria | Bar | Beer Bar | Smoothie Shop | Sა |
| 63 | Regent Park, Harbourfront | Coffee Shop | Bakery | Pub | Park | Breakfast Spot | Restaurant | |
| 64 | Richmond, Adelaide, King | Coffee Shop | Café | Restaurant | Gym | Thai Restaurant | Deli / Bodega | |

| | 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 | 7 C |
|---|---|---|---|---|---|---|---|---|
| 68 | Runnymede, Swansea | Coffee Shop | Café | Sushi Restaurant | Pizza Place | Italian Restaurant | Pub | |
| 72 | St. James Town | Coffee Shop | Café | Cocktail Bar | Gastropub | American Restaurant | Clothing Store | M Re |
| 73 | St. James Town, Cabbagetown | Coffee Shop | Bakery | Restaurant | Café | Pizza Place | Pub | Re |
| 75 | Stn A PO Boxes | Coffee Shop | Seafood Restaurant | Japanese Restaurant | Restaurant | Café | Beer Bar | |
| 76 | Studio District | Coffee Shop | Café | Brewery | Bakery | Gastropub | American Restaurant | |
| 77 | Summerhill West, Rathnelly, South Hill, Forest... | Coffee Shop | Pizza Place | Pub | Light Rail Station | Supermarket | Bank | Re |
| 78 | The Annex, North Midtown, Yorkville | Café | Sandwich Place | Coffee Shop | Pizza Place | Indian Restaurant | Pharmacy | |
| 83 | Toronto Dominion Centre, Design Exchange | Coffee Shop | Hotel | Café | Seafood Restaurant | Italian Restaurant | Restaurant | A Re |
| 84 | University of Toronto, Harbord | Café | Bookstore | Bakery | Bar | Japanese Restaurant | Yoga Studio | |
| 93 | Woburn | Coffee Shop | Convenience Store | Korean BBQ Restaurant | Doner Restaurant | Dim Sum Restaurant | Diner | D |

## 3.5 Clustering the Neighborhoods

Run k-means to cluster the neighborhood into 6 clusters using the Foursquare venue data.

| | PostalCode | Borough | Neighborhood | Latitude | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3... |
|---|---|---|---|---|---|---|---|---|---|
| 0 | M3A | North York | Parkwoods | 43.753259 | -79.329656 | 2.0 | Park | Food & Drink Shop | |
| 1 | M4A | North York | Victoria Village | 43.725882 | -79.315572 | 0.0 | Pizza Place | Coffee Shop | |
| 2 | M5A | Downtown Toronto | Regent Park, Harbourfront | 43.654260 | -79.360636 | 0.0 | Coffee Shop | Bakery | |
| 3 | M6A | North York | Lawrence Manor, Lawrence Heights | 43.718518 | -79.464763 | 0.0 | Clothing Store | Women's Store | |
| 4 | M7A | Downtown Toronto | Queen's Park, Ontario Provincial Government | 43.662301 | -79.389494 | 0.0 | Coffee Shop | Sushi Restaurant | |

| | PostalCode | Borough | Neighborhood | Latitude | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue |
|---|---|---|---|---|---|---|---|---|
| 5 | M9A | Etobicoke | Islington Avenue, Humber Valley Village | 43.667856 | -79.532242 | NaN | NaN | NaN |
| 45 | M2L | North York | York Mills, Silver Hills | 43.757490 | -79.374714 | NaN | NaN | NaN |
| 95 | M1X | Scarborough | Upper Rouge | 43.836125 | -79.205636 | NaN | NaN | NaN |

These rows have NA because no venue data was returned for these neighborhood from Foursquare. We will label these as "6" in clusters.

| | PostalCode | Borough | Neighborhood | Latitude | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3 C |
|---|---|---|---|---|---|---|---|---|---|
| 0 | M3A | North York | Parkwoods | 43.753259 | -79.329656 | 2 | Park | Food & Drink Shop | V |
| 1 | M4A | North York | Victoria Village | 43.725882 | -79.315572 | 0 | Pizza Place | Coffee Shop | |
| 2 | M5A | Downtown Toronto | Regent Park, Harbourfront | 43.654260 | -79.360636 | 0 | Coffee Shop | Bakery | |
| 3 | M6A | North York | Lawrence Manor, Lawrence Heights | 43.718518 | -79.464763 | 0 | Clothing Store | Women's Store | |
| 4 | M7A | Downtown Toronto | Queen's Park, Ontario Provincial Government | 43.662301 | -79.389494 | 0 | Coffee Shop | Sushi Restaurant | ( |

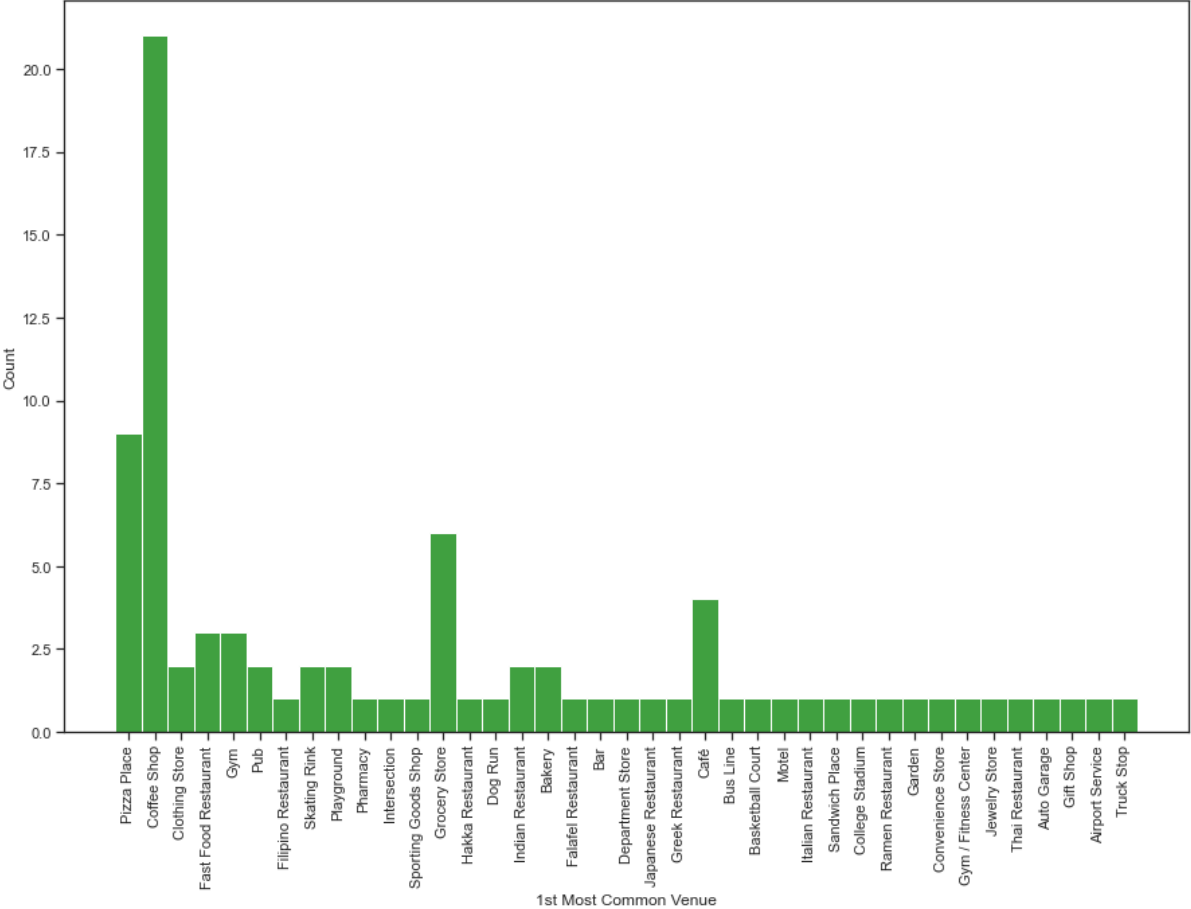| | Cluster Labels | count |
|---|---|---|
| 0 | 0 | 85 |
| 1 | 1 | 1 |
| 2 | 2 | 10 |
| 3 | 3 | 2 |
| 4 | 4 | 1 |
| 5 | 5 | 1 |
| 6 | 6 | 3 |

Visualize the clusters!
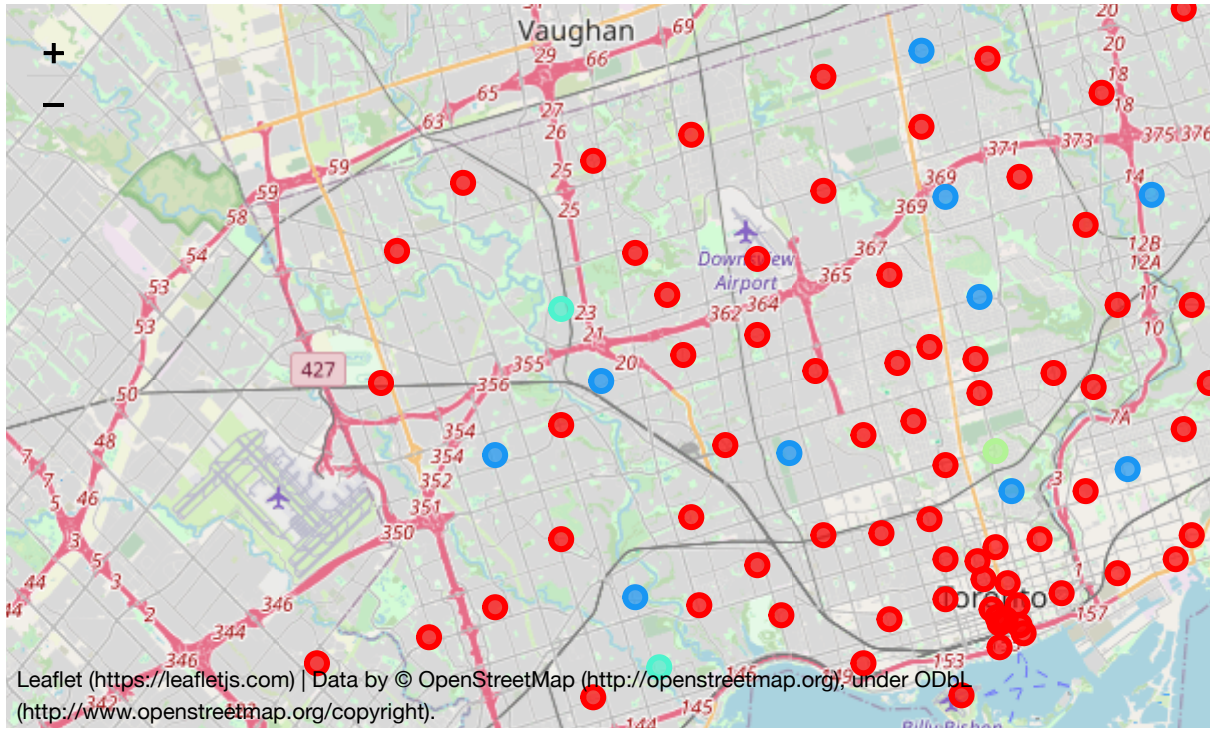
## 3.6 Explore the Clusters
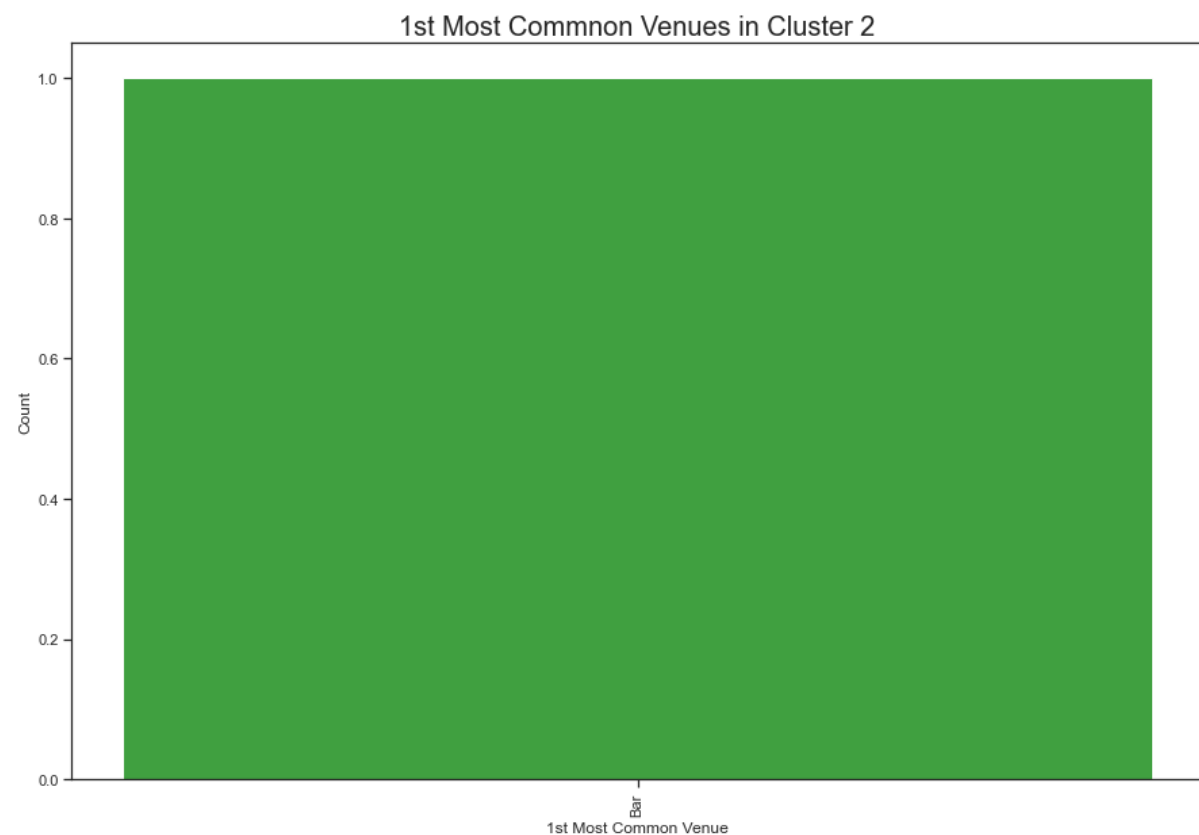
Cluster 1

| | 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 | |
|---|---|---|---|---|---|---|---|---|---|
| **1** | North York | 0 | Pizza Place | Coffee Shop | Hockey Arena | Portuguese Restaurant | Intersection | Dumpling Restaurant | F |
| **2** | Downtown Toronto | 0 | Coffee Shop | Bakery | Pub | Park | Breakfast Spot | Restaurant | |
| **3** | North York | 0 | Clothing Store | Women's Store | Sporting Goods Shop | Boutique | Miscellaneous Shop | Furniture / Home Store | |
| **4** | Downtown Toronto | 0 | Coffee Shop | Sushi Restaurant | College Cafeteria | Bar | Beer Bar | Smoothie Shop | |
| **6** | Scarborough | 0 | Fast Food Restaurant | Print Shop | Women's Store | Dog Run | Dessert Shop | Dim Sum Restaurant | |

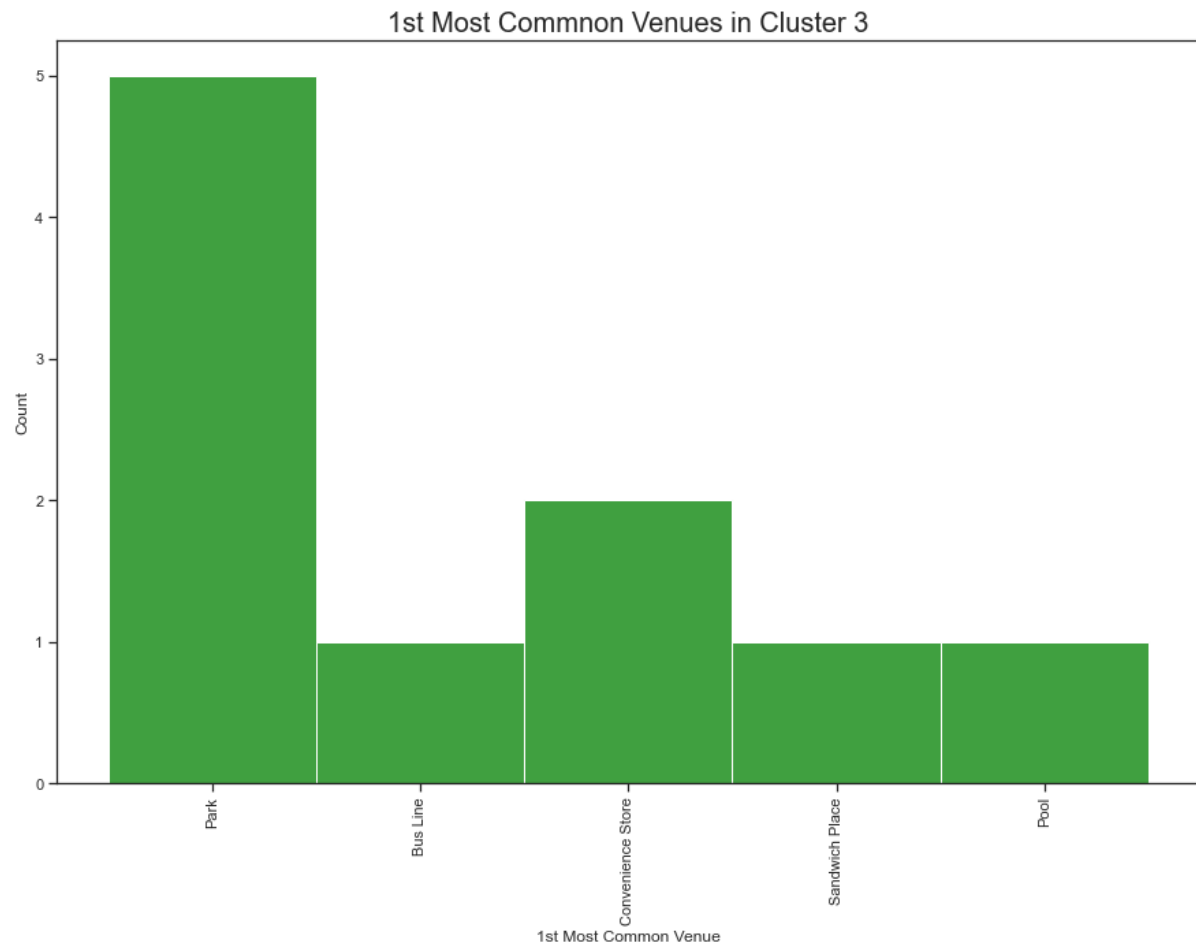1st Most Commnon Venues in Cluster 1

Cluster 2

| | 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 |
|---|---|---|---|---|---|---|---|---|---|
| 12 | Scarborough | 1 | Bar | Women's Store | Dim Sum Restaurant | Diner | Discount Store | Distribution Center | Dog R |

## 1st Most Commnon Venues in Cluster 2
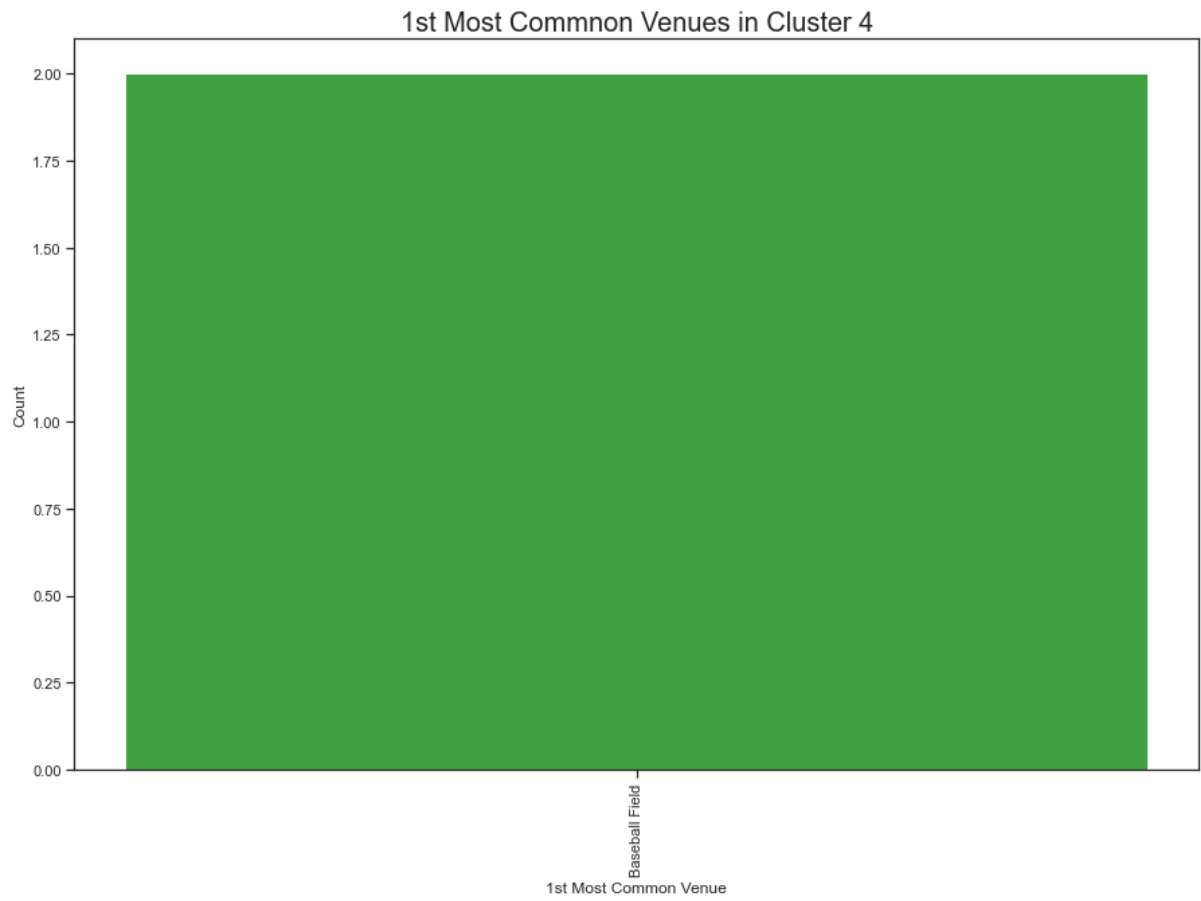


Cluster 3

| | 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 |
|---|---|---|---|---|---|---|---|---|---|
| **0** | North York | 2 | Park | Food & Drink Shop | Women's Store | Dog Run | Dessert Shop | Dim Sum Restaurant | |
| **21** | York | 2 | Park | Women's Store | Pool | Dog Run | Department Store | Dessert Shop | Dim Resta |
| **35** | East York | 2 | Park | Convenience Store | Metro Station | Donut Shop | Dim Sum Restaurant | Diner | Dis |
| **52** | North York | 2 | Park | Women's Store | Doner Restaurant | Dessert Shop | Dim Sum Restaurant | Diner | Dis |
| **61** | Central Toronto | 2 | Bus Line | Park | Swim School | Dog Run | Dessert Shop | Dim Sum Restaurant | |

**1st Most Commnon Venues in Cluster 3**
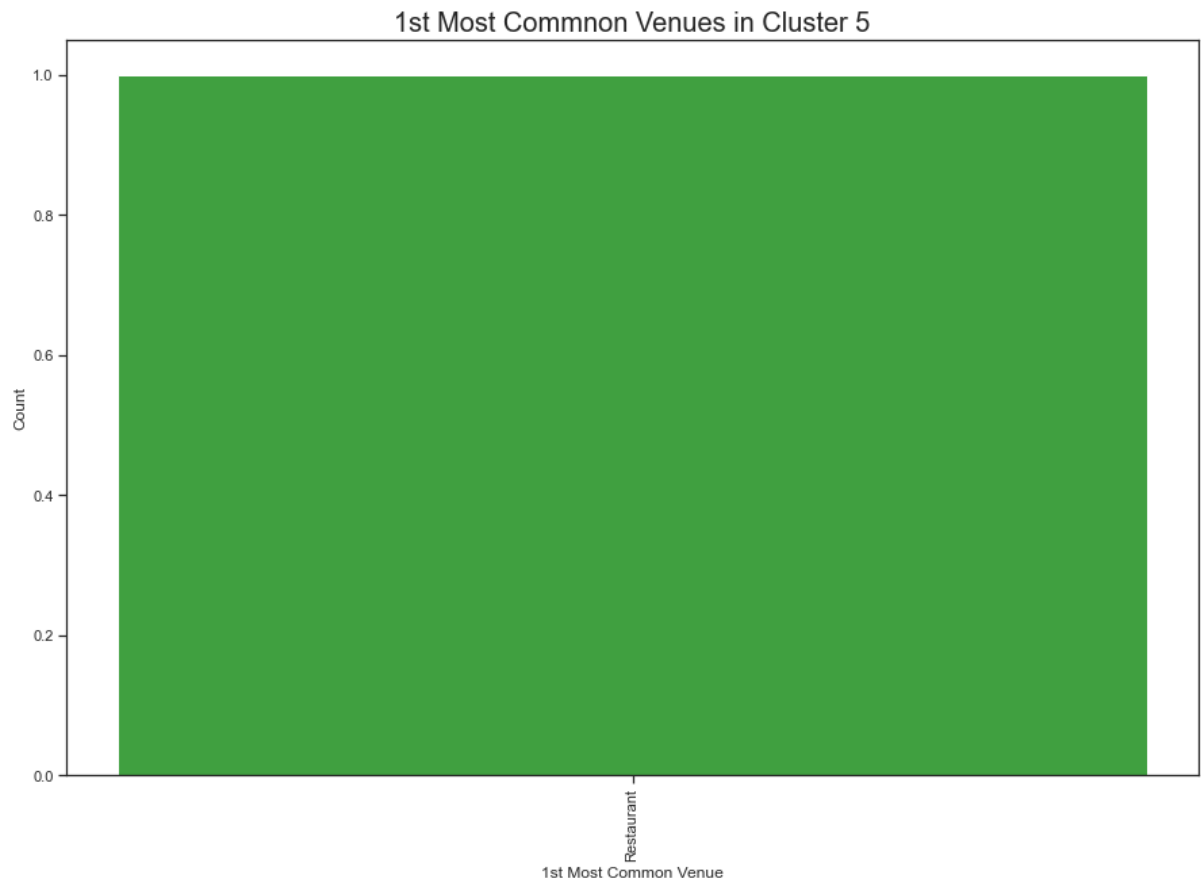
Cluster 4

| | 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 |
|---|---|---|---|---|---|---|---|---|---|
| **57** | North York | 3 | Baseball Field | Paper / Office Supplies Store | Women's Store | Dim Sum Restaurant | Diner | Discount Store | Distribution Center |
| **101** | Etobicoke | 3 | Baseball Field | Donut Shop | Dim Sum Restaurant | Diner | Discount Store | Distribution Center | Dog |

## 1st Most Commnon Venues in Cluster 4
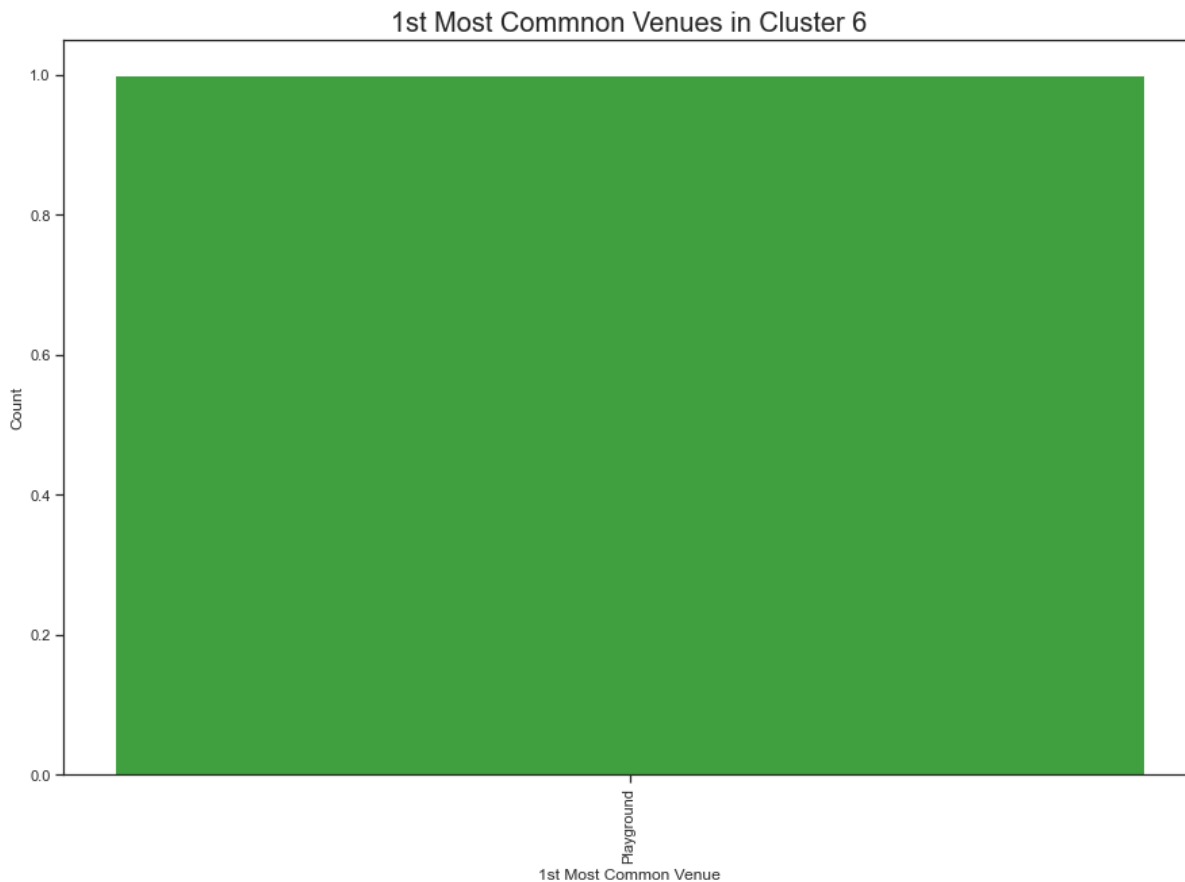


Cluster 5

| | 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 |
|---|---|---|---|---|---|---|---|---|---|
| 83 | Central Toronto | 4 | Restaurant | Summer Camp | Women's Store | Dog Run | Dessert Shop | Dim Sum Restaurant | Diner |

## 1st Most Commnon Venues in Cluster 5



Cluster 6

| | 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 Com V |
|---|---|---|---|---|---|---|---|---|---|
| **32** | Scarborough | 5 | Playground | Spa | College Gym | College Rec Center | Ethiopian Restaurant | Electronics Store | Ea Euro Resta |

1st Most Commnon Venues in Cluster 6

## 4. Results and Discussion

Cluster 1 contains the most diverse profile of venue categories, and coffee shop is the most common venue category in this cluster. From the sample detail of Cluster 1, these areas are most likely the financial/shopping districts in Toronto with many shops and event spaces. Cluster 3 also has a diverse profile of business types, but the most common venue category is park. The sample data of Cluster 3 shows that the area is mainly consisted of parks, transportaion stations, and distribution centers. Cluster 3 is mostly lcoated in the remote areas of Toronto. The other clusters only have 1 member.

## 5. Discussion

This project identifies the optimal location for a new coffee shopp business based on the popular venue category in eacch neighborhood. However, other factors should also be accounted for when making business decisions. For example, the type of neighborhood will create different customer traffic in different times of the day. A commercial neighborhood will have more traffic during the weekdays while a residential neighborhood will likely to havee more traffic on the weekends. The demoograrphics are also important, given people from one culture might prefer drinking coffee than people from other cultures.

# 6. Conclusion

Based on the cluster analysis, we can conclude that neighborhoods in cluster 1 will have most customer trafic but also strong business competition. Cluster 3 is represents a good opportunith with less competiton, but might also have slow business growth for a coffee shop business. Other factors such as demographic components and neighborhood type should also be taken into consideration.