Capstone Project - The Battle of the Neighborhoods – Toronto

Applied Data Science Capstone by IBM/Coursera

Rimma Dzhusupova

Table of Contents

1. Introduction: Business Problem

The battle of the neighborhood is an important data science project targeting the clusters of similar types to provide a user to exploit this as a tool for an ample number of business choices. This project is based on the application of different data science techniques learned in the entire course.

The purpose of this project was to identify the best venues in Toronto Downtown borough to aid stakeholders in narrowing down the search for the optimal type for a new venue.

2. Data

Based on the problem statement:

- Following data sources will be needed to extract/generate the required information: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Geocoder
- Number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API
- Coordinates of Toronto center will be obtained using Geocoder of well known Toronto location
- The data frame will consist of three columns: PostalCode, Borough, and Neighborhood
- Needs to process only those cells that have an assigned Borough. Ignore cells with a Borough that is Not assigned.
- More than one neighborhood can exist in one postal code area. For example, in the table on the Wikipedia page, one can see 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.
- If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough. So for the 9th cell in the table on the Wikipedia page, the value of the borough and the Neighborhood columns will be Queen's Park.
- Needs to clean the data using appropriate methods.

3. Methodology

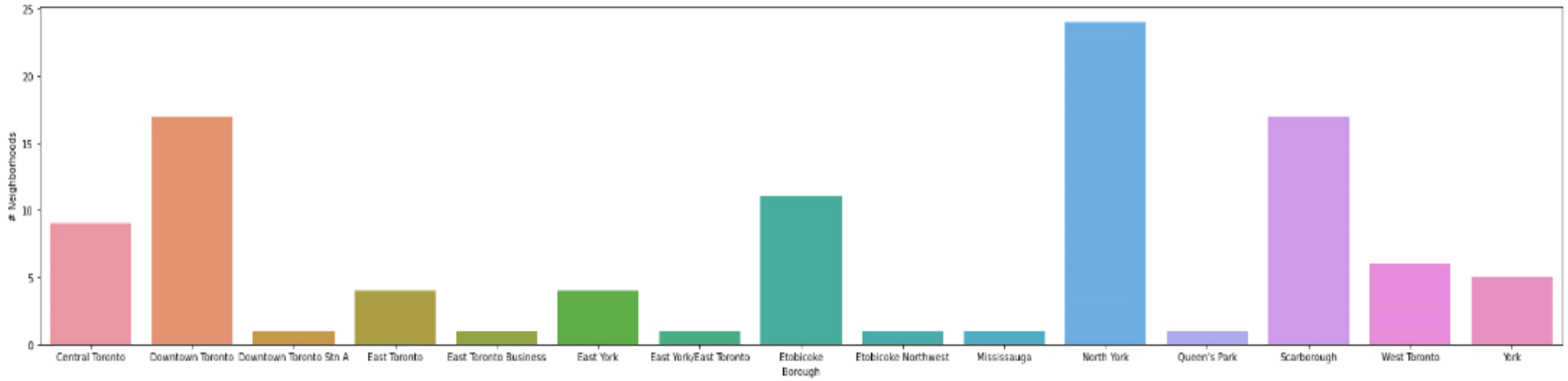
Data Collection Method: Data collection has been completed using a methodology based on the BeautifulSoup library to extract data frames from the existing webpage. This is the initial stage to collect the data. This data has been utilized for generating word cloud a special feature of Python.
Geographic Location Data Collection Method: The initial data frame has been used to get longitude and latitude of neighborhoods on the Geopy library. Because of frequent data traffic from our kernel (103 POST and 103 GET), and traffic limitation this call does not make a significant impact. Therefore the location data for postal codes have been collected from the Canada Government website.
Map Exploration Method: The Folium library has been utilized to get the map of Toronto.

K-means clustering- K means cluster is applied to the dataset for the exploration of different venues. Here the algorithm formed 5 clusters with the trends. We have also presented the top 10 venue categories for each cluster.

4. Analysis

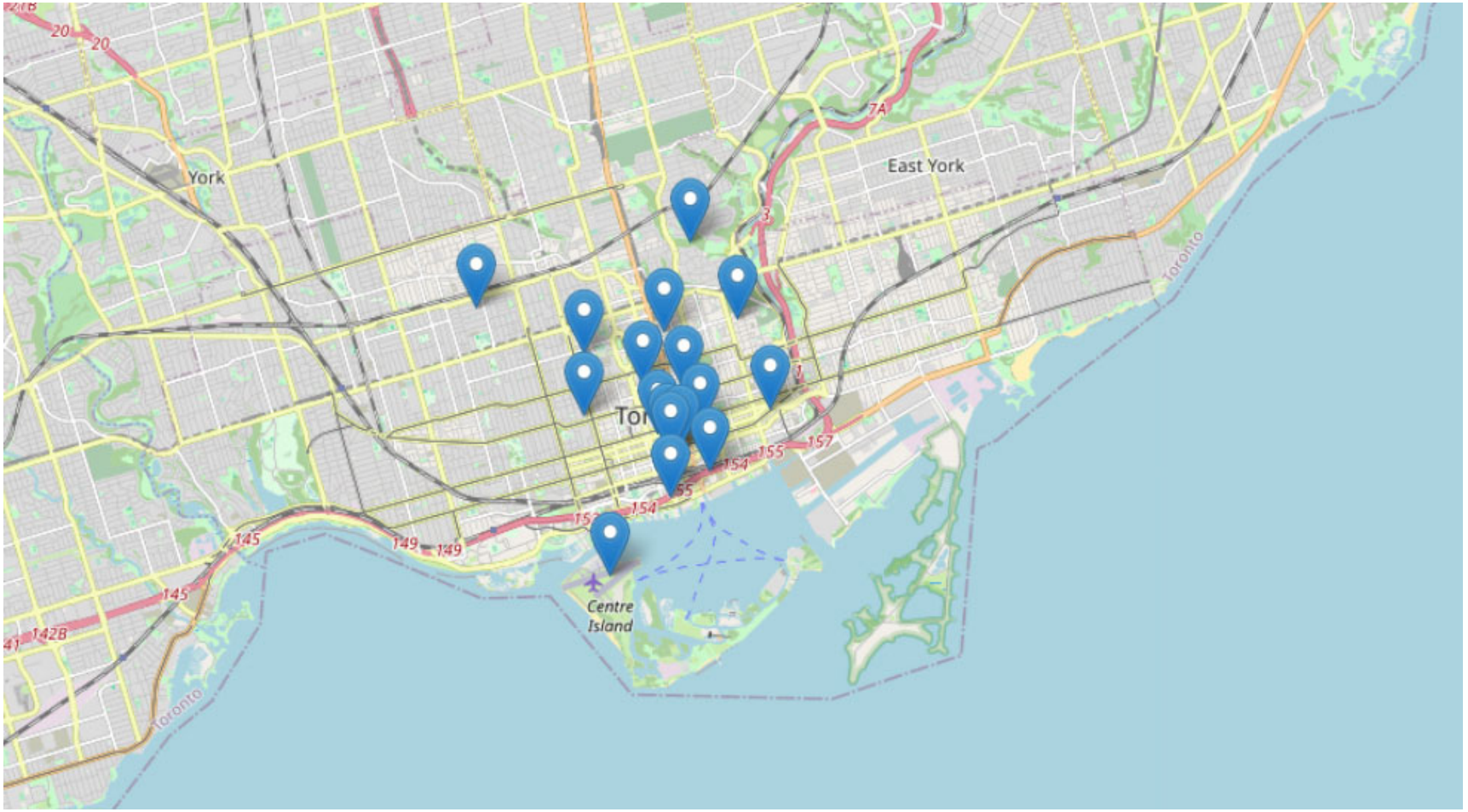
Let's perform some basic explanatory data analysis and derive some additional info from our raw data. First, let's count the **number of neighborhoods in each borough.**

The following graph explores the number of neighborhoods present at a given borough and neighborhood.



We choose Toronto Downtown borough for neighborhoods analysis
The geograpical coordinates of Downtown Toronto are 43.6541737, -79.38081164513409.

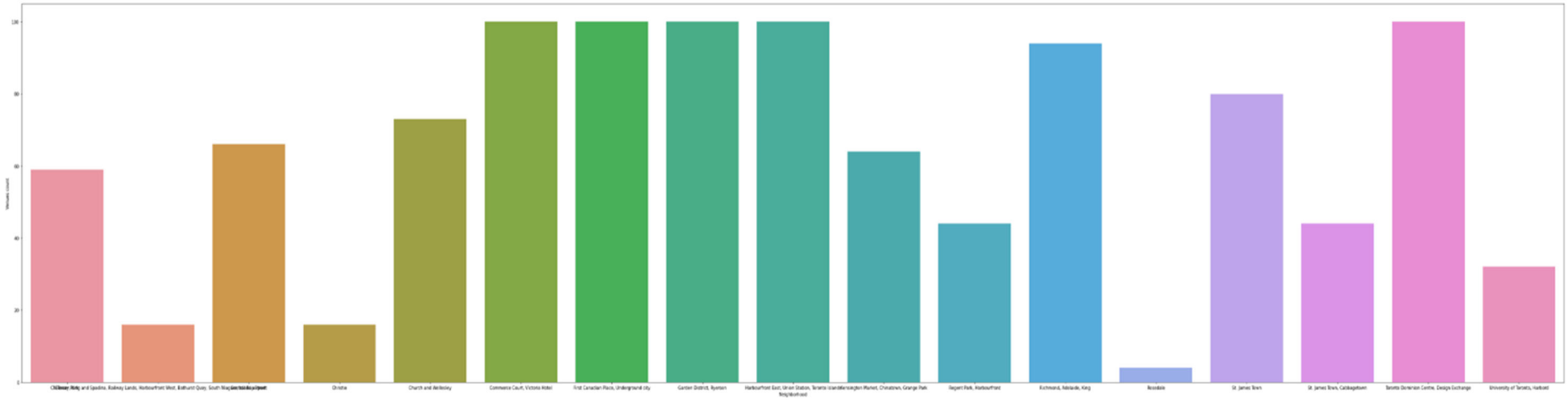
Let's visualize Downtown Toronto the neighborhoods in it.



4.1 Explore Neighborhoods in Downtown Toronto

We are going to start utilizing the Foursquare API to explore the neighborhoods and segment them.

There are 204 uniques venues in Toronto Downtown. The number of unique venues in each neighborhood is visualized below:



4.2 Analyze Each Neighborhood

Now we can identify the top 10 venues for each neighborhood.

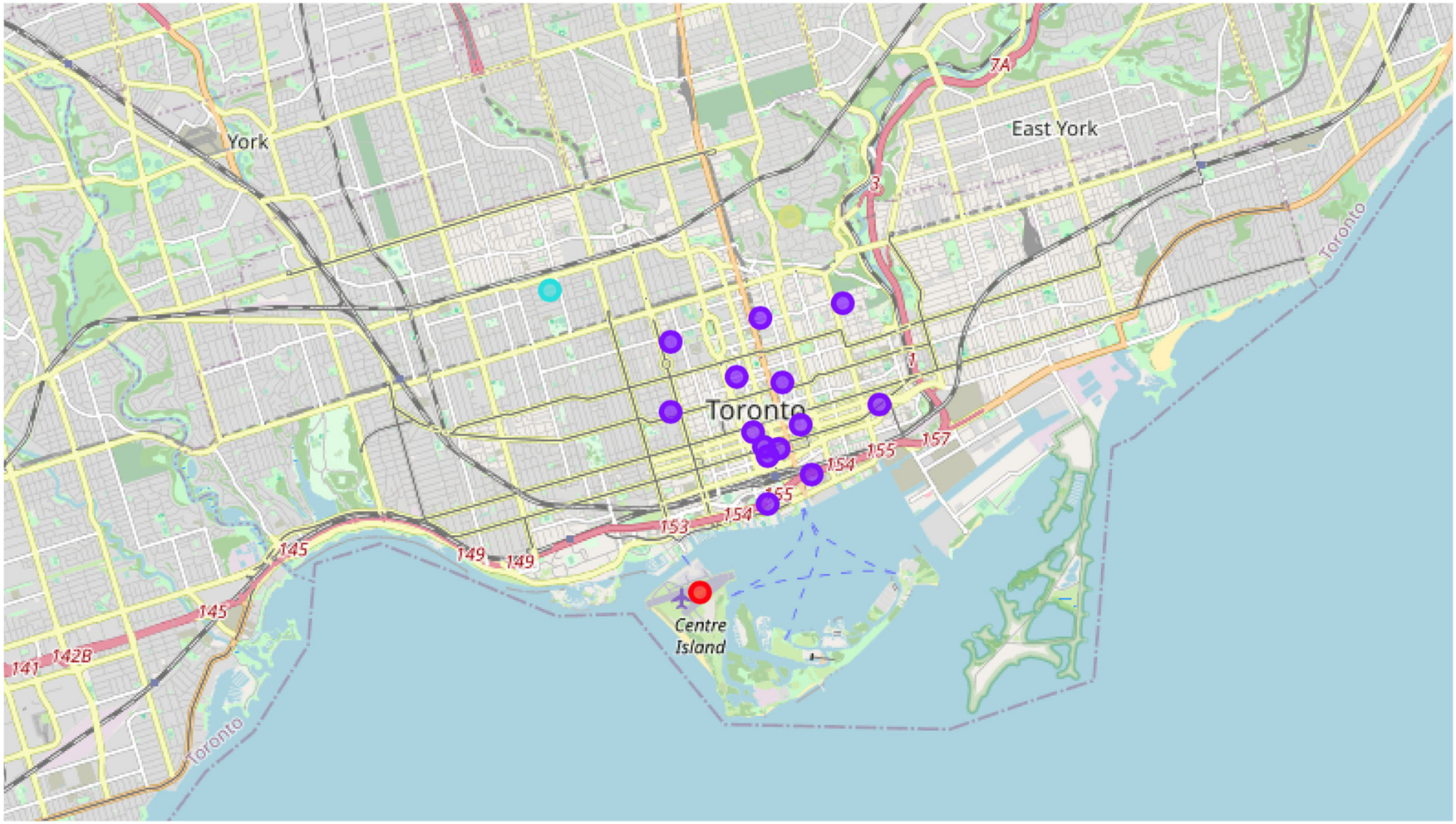
Neighbor-hood	*1st Most Common Venue*	*2nd Most Common Venue*	*3rd Most Common Venue*	*4th Most Common Venue*	*5th Most Common Venue*	*6th Most Common Venue*	*7th Most Common Venue*	*8th Most Common Venue*	*9th Most Common Venue*	*10th Most Common Venue*
Berczy Park	Coffee Shop	Cocktail Bar	Bakery	Beer Bar	Restaurant	Cheese Shop	Seafood Restaurant	Pharmacy	Farmers Market	Diner

CN Tower, King and Spadina, Railway Lands, Har...	*Airport Service*	*Airport Lounge*	*Airport Terminal*	*Boat or Ferry*	*Airport*	*Boutique*	*Plane*	*Sculpture Garden*	*Rental Car Location*	*Coffee Shop*
Central Bay Street	Coffee Shop	Sandwich Place	Italian Restaurant	Café	Bubble Tea Shop	Salad Place	Restaurant	Thai Restaurant	Burger Joint	Department Store
Christie	Grocery Store	Café	Park	Coffee Shop	Nightclub	Baby Store	Athletics & Sports	Restaurant	Italian Restaurant	Candy Store
Church and Wellesley	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Restaurant	Hotel	Mediterranean Restaurant	Men's Store	Grocery Store	Gay Bar	Pub
Commerce Court, Victoria Hotel	Coffee Shop	Restaurant	Hotel	Café	Italian Restaurant	Gym	Japanese Restaurant	Seafood Restaurant	Deli / Bodega	American Restaurant
First Canadian Place, Underground city	Coffee Shop	Café	Hotel	Japanese Restaurant	Gym	Restaurant	Seafood Restaurant	American Restaurant	Salad Place	Deli / Bodega
Garden District, Ryerson	Clothing Store	Coffee Shop	Middle Eastern Restaurant	Cosmetics Shop	Café	Bubble Tea Shop	Ramen Restaurant	Burger Joint	Diner	Bookstore
Harbourfront East, Union Station, Toronto Islands	Coffee Shop	Aquarium	Café	Hotel	Fried Chicken Joint	Restaurant	Scenic Lookout	Brewery	Sporting Goods Shop	Italian Restaurant

Kensington Market, Chinatown, Grange Park	Café	Coffee Shop	Vietnamese Restaurant	Mexican Restaurant	Vegetarian / Vegan Restaurant	Bakery	Gaming Cafe	Park	Farmers Market	Bar
Regent Park, Harbourfront	Coffee Shop	Pub	Park	Bakery	Breakfast Spot	Theater	Café	Restaurant	Event Space	Chocolate Shop
Richmond, Adelaide, King	Coffee Shop	Café	Restaurant	Clothing Store	Gym	Thai Restaurant	Deli / Bodega	Salad Place	Pizza Place	Steakhouse
Rosedale	Park	Playground	Trail	Music Venue	Mediterranean Restaurant	Men's Store	Mexican Restaurant	Middle Eastern Restaurant	Miscellaneous Shop	Modern European Restaurant
St. James Town	Coffee Shop	Café	Cosmetics Shop	Cocktail Bar	Department Store	Farmers Market	Moroccan Restaurant	Beer Bar	Seafood Restaurant	Hotel
St. James Town, Cabbagetown	Coffee Shop	Pizza Place	Italian Restaurant	Bakery	Pub	Restaurant	Café	Caribbean Restaurant	Beer Store	Market
Toronto Dominion Centre, Design Exchange	Coffee Shop	Hotel	Café	Japanese Restaurant	Italian Restaurant	Seafood Restaurant	Restaurant	Salad Place	Concert Hall	Sushi Restaurant
University of Toronto, Harbord	Café	Bookstore	Italian Restaurant	Japanese Restaurant	Bar	Bakery	Restaurant	French Restaurant	Dessert Shop	Sandwich Place

4.3 Cluster neighborhoods

Run k-means to cluster the neighborhood into 4 clusters.



Now, one can examine each cluster and determine the discriminating venue categories that distinguish each cluster.

Cluster 0

Borough	Neighborhood	Cluster #	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
Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	0	Airport Service	Airport Lounge	Airport Terminal	Boat or Ferry	Airport	Boutique	Plane	Sculpture Garden	Rental Car Location	Coffee Shop

Cluster 1

Borough	Neighborhood	Cluster #	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
Downtown Toronto	Regent Park, Harbourfront	1	Coffee Shop	Pub	Park	Bakery	Breakfast Spot	Theater	Café	Restaurant	Event Space	Chocolate Shop
Downtown Toronto	Garden District, Ryerson	1	Clothing Store	Coffee Shop	Middle Eastern Restaurant	Cosmetics Shop	Café	Bubble Tea Shop	Ramen Restaurant	Burger Joint	Diner	Bookstore
Downtown Toronto	St. James Town	1	Coffee Shop	Café	Cosmetics Shop	Cocktail Bar	Department Store	Farmers Market	Moroccan Restaurant	Beer Bar	Seafood Restaurant	Hotel
Downtown Toronto	Berczy Park	1	Coffee Shop	Cocktail Bar	Bakery	Beer Bar	Restaurant	Cheese Shop	Seafood Restaurant	Pharmacy	Farmers Market	Diner

Downtown Toronto	Central Bay Street	1	Coffee Shop	Sandwich Place	Italian Restaurant	Café	Bubble Tea Shop	Salad Place	Restaurant	Thai Restaurant	Burger Joint	Department Store
Downtown Toronto	Richmond, Adelaide, King	1	Coffee Shop	Café	Restaurant	Clothing Store	Gym	Thai Restaurant	Deli / Bodega	Salad Place	Pizza Place	Steakhouse
Downtown Toronto	Harbourfront East, Union Station, Toronto Islands	1	Coffee Shop	Aquarium	Café	Hotel	Fried Chicken Joint	Restaurant	Scenic Lookout	Brewery	Sporting Goods Shop	Italian Restaurant
Downtown Toronto	Toronto Dominion Centre, Design Exchange	1	Coffee Shop	Hotel	Café	Japanese Restaurant	Italian Restaurant	Seafood Restaurant	Restaurant	Salad Place	Concert Hall	Sushi Restaurant
Downtown Toronto	Commerce Court, Victoria Hotel	1	Coffee Shop	Restaurant	Hotel	Café	Italian Restaurant	Gym	Japanese Restaurant	Seafood Restaurant	Deli / Bodega	American Restaurant
Downtown Toronto	University of Toronto, Harbord	1	Café	Bookstore	Italian Restaurant	Japanese Restaurant	Bar	Bakery	Restaurant	French Restaurant	Dessert Shop	Sandwich Place
Downtown Toronto	Kensington Market, Chinatown, Grange Park	1	Café	Coffee Shop	Vietnamese Restaurant	Mexican Restaurant	Vegetarian / Vegan Restaurant	Bakery	Gaming Cafe	Park	Farmers Market	Bar
Downtown Toronto	St. James Town, Cabbagetown	1	Coffee Shop	Pizza Place	Italian Restaurant	Bakery	Pub	Restaurant	Café	Caribbean Restaurant	Beer Store	Market

Downtown Toronto	First Canadian Place, Underground city	1	Coffee Shop	Café	Hotel	Japanese Restaurant	Gym	Restaurant	Seafood Restaurant	American Restaurant	Salad Place	Deli / Bodega
Downtown Toronto	Church and Wellesley	1	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Restaurant	Hotel	Mediterranean Restaurant	Men's Store	Grocery Store	Gay Bar	Pub

Cluster 2

Borough	*Neighborhood*	*Cluster #*	*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*
Downtown Toronto	Christie	2	Grocery Store	Café	Park	Coffee Shop	Nightclub	Baby Store	Athletics & Sports	Restaurant	Italian Restaurant	Candy Store

Cluster 3

Borough	*Neighborhood*	*Cluster #*	*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*
Downtown Toronto	Rosedale	3	Park	Playground	Trail	Music Venue	Mediterranean Restaurant	Men's Store	Mexican Restaurant	Middle Eastern Restaurant	Miscellaneous Shop	Modern European Restaurant

5. Results and Discussion

The battle of a neighborhood project explored different categories of venues. In this analysis, initially, a sum of 103 unique postal codes has been observed for 10 different boroughs. Using geocoder of Geopy library for Toronto, North York, Parkwoods we get the raw results that are comrehensive and self explanatory in nature.

Based on the above result we shave done an analysis of each neighborhood in Toronto Downtown. The result of one-hot encoding identified 204 types of places. Followed by one-hot encoding we explore the top 10 common venues in each of the neighborhoods.

After one hot encoding and venue exploration, we applied the k-means algorithm for k=4. Clusters 0, 2 and 3 are having only one neighborhood, cluster 1 is having the maximum number of neighborhoods.

6. Conclusion

The purpose of this project was to identify the best venues in Toronto Downtown borough to aid stakeholders in narrowing down the search for the optimal type for a new venue. Clustering of those locations was then performed to create major zones of interest (containing the greatest number of potential locations) of those zone centers were created to be used as starting points for final exploration by stakeholders. Based on analysis and identification the most common venues are coffeeshops.