

# **Extensive Study on Venues Around Neighborhoods of Toronto and New York City**

Vismay Shah

May 29<sup>th</sup>, 2020

## **1. Introduction**

### **1.1. Background**

Toronto is Canada's business and financial capital. It is one of the fastest growing financial hub in North America and is in one of the top ten global financial centre. It is competitive in almost every other major business sector from technology and life sciences to green energy; from fashion and design to food and beverage; from film and television production to music and digital media[1]. Toronto's rich industrial diversity has driven immense growth and diversity and attracted lot of immigrants. The number of immigrants have shown growth till date since 2000 and Toronto being the city of receiving enormous amount of immigrants compare to any other city of Canada[2]. Toronto is spread across 243.32 square miles (630.20 km<sup>2</sup>) with the estimated population of 2,731,571 in 2018[3].

On the other hand, there is New York City which is described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports[5]. It is commonly regarded as the finance capital of the world, has been ranked first in the World's Financial Centers by the Global Financial Centers Index (GFCI). The city is a mix of various cultures from across the globe providing a diverse population and workforce [4]. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km<sup>2</sup>), New York is also the most densely populated major city in the United States[5].

### **1.2. Objective**

There are two objectives of this project. First, is to identify the similarities and dissimilarities between the neighborhoods of Toronto and New York City in terms of venues exist around them using K Means Clustering algorithm. Second, is to identify the business opportunities related to food industry in each neighborhood based on the venues exist around it, most common venues of that particular cluster it fall under, and most common venues of the borough it belongs to.

### 1.3. Interest

This comprehensive study will definitely assist those who want to start/expand their business in food industry in either of the cities, who want to relocate from any other city to one of these cities or want to move from one of these cities to the other one, and finally to the international immigrants who will get some prior information about specific neighborhood they are planning to live in.

## 2. Data

The list of Toronto's boroughs and neighborhoods was extracted from Wikipedia page which has it in Tabular format. The link of the page is

[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). Latitude and longitude of all those neighborhoods were provided by IBM in ".csv" format

The dataset of NYC's boroughs and neighborhoods was imported in JSON format using this link [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset). From the dataset imported in the JSON format, useful features like name of the borough, name of the neighborhood, and their latitude and longitude were obtained and put them into pandas dataframe.

Foursquare location data was used to explore the venues exist around each neighborhood of both the cities.

## 3. Methodology

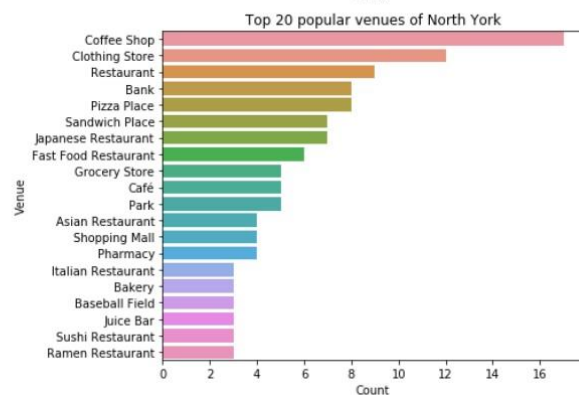
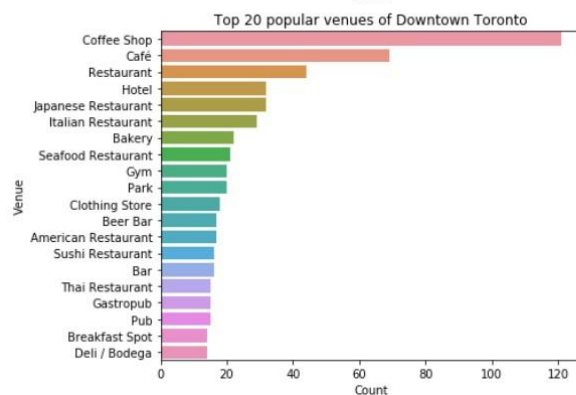
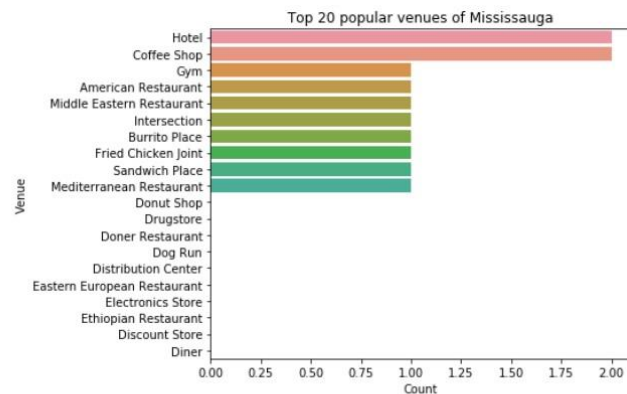
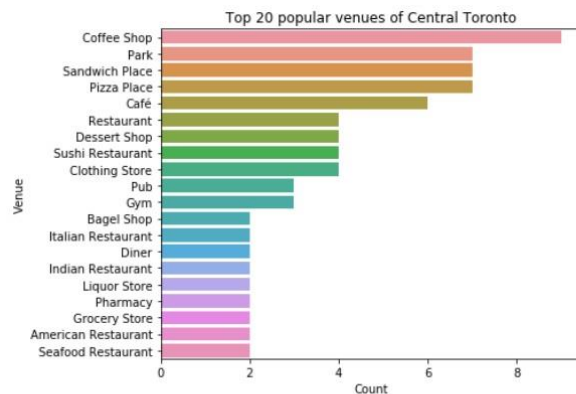
With the information of latitude and longitude of each neighborhood and help of Foursquare location data, venues that exist within 500 meters (0.3 miles) of each neighborhood were explored and extracted into the dataframe. The limit of number of venues to be explored was set to be 100 so no more than 100 venues will show up in the dataframe for any neighborhood. After getting the required data, all the venue categories were converted into dummy variables (0's and 1's). Then groupby() with sum() was implemented on the dataframe and most popular venues of each borough were analyzed to get the big picture of how venues are located in particular region. Then K means clustering was applied on the dataframe to see the difference and similarities of neighborhoods of Toronto and New York City. After that detailed analysis was done on each neighborhood falling under a particular cluster and suggestion were provided to start a new business or expanding the business mostly in the food industry

### 3.1. Exploratory data analysis of boroughs of Toronto

In Toronto's dataset there are 10 boroughs which are Central Toronto, Downtown Toronto, East Toronto, West Toronto, North York, East York, York, Etobicoke, Scarborough. By observing the Fig 3.1.1, we found that

- Few venues almost exist in all the boroughs. These venues are coffee shops, cafe, pizza places, parks, sandwich places, pharmacy, grocery store, bakery, and gym

- Pubs and bars do exist in few of the boroughs but in very less number except in Downtown Toronto
- There are variety of restaurants as well very close to few of the neighborhoods and Italian restaurants are dominating in all of them
- Observing the less common venues or bottom venues of the plots, we see that all the boroughs have different venues around such as liquor store, diner, baseball field, juice bar, pet store, spa, yoga studio, beer store, brewery, gas station, skating ring, shopping plaza etc. These venues will not be important in the final analyses part as we are just interested in most common venues which replicate the characteristics of that neighborhood/borough
- Downtown Toronto is tightly packed with coffee shops, cafe, pubs, bars, bakery, and variety restaurants, hotels, and all other amenities. The interesting thing to observe in Downtown Toronto bar plot is that the pizza place doesn't fall in top 20 venues



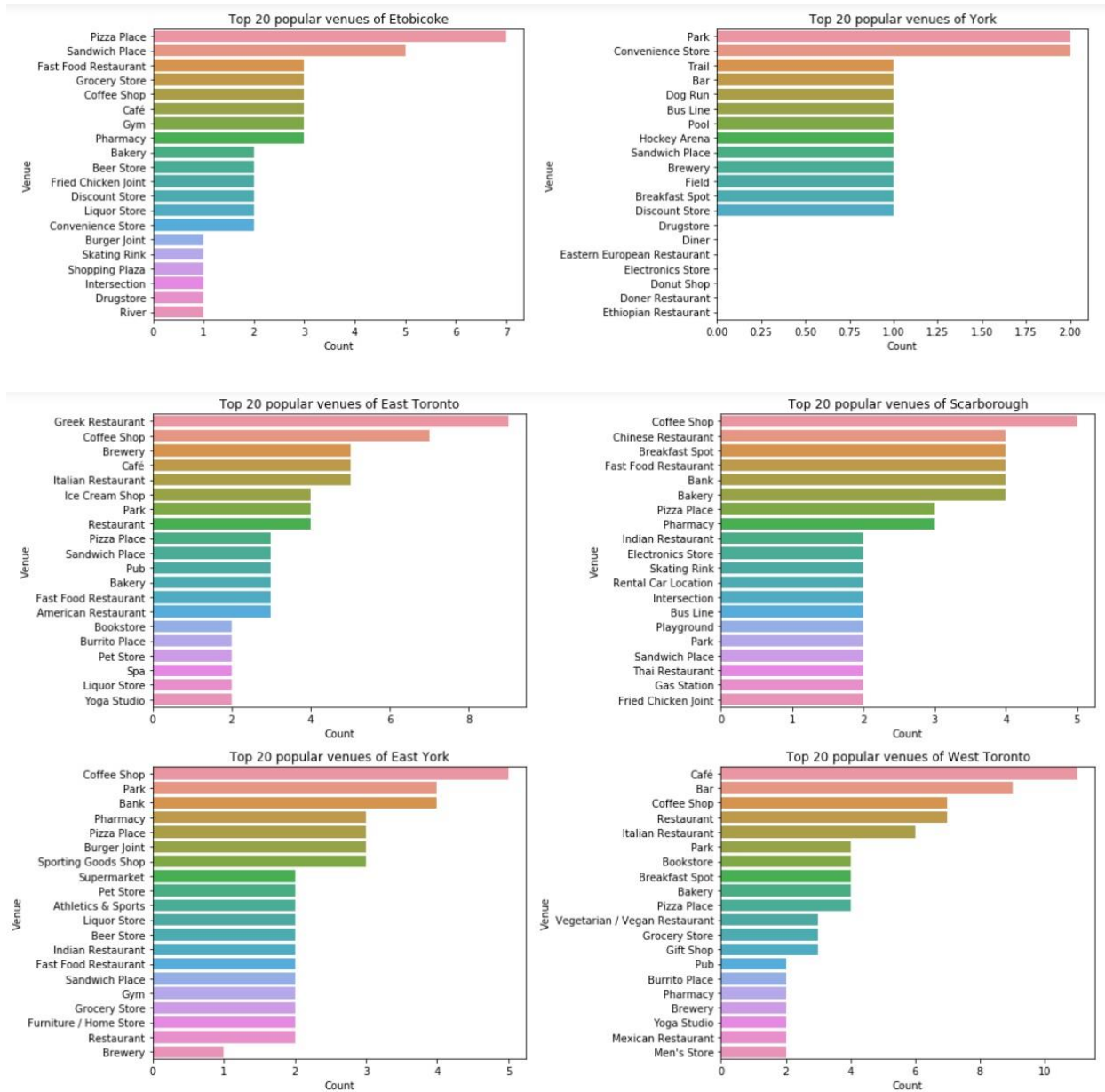


Fig 3.1.1. Top 20 venues of each borough of Toronto city

### 3.2. Exploratory data analysis of neighborhoods of NYC

In NYC's dataset there are 5 boroughs which are Bronx, Brooklyn, Manhattan, Staten Island, and Queens. By observing the Fig 3.2.1, we found that

- All the boroughs are tightly packed with plenty of pizza place, deli/bodega, pharmacy, sandwich place, park, bar, bakery, and variety of restaurants in which Italian, Chinese, and Mexican being most popular ones
- Donut shop, grocery store, bank, bagel shop, and other restaurants are popular in all the boroughs except Manhattan

- Queens has the highest variety of cuisines available including Chinese, Korean, Mexican, Italian, Latin American, and Thai which reflects its vast cultural diversity. It has got the maximum number of beaches as well compare to any other borough
- Overall, most of the boroughs have more or less similar venues except Manhattan. We can clearly see that pharmacy, bank, donut shop, bagel shop or such venues do not fall in top 20 venues of Manhattan because Manhattan downtown is considered NYC's downtown and therefore it has all the eateries and entertainment venues around the neighborhoods

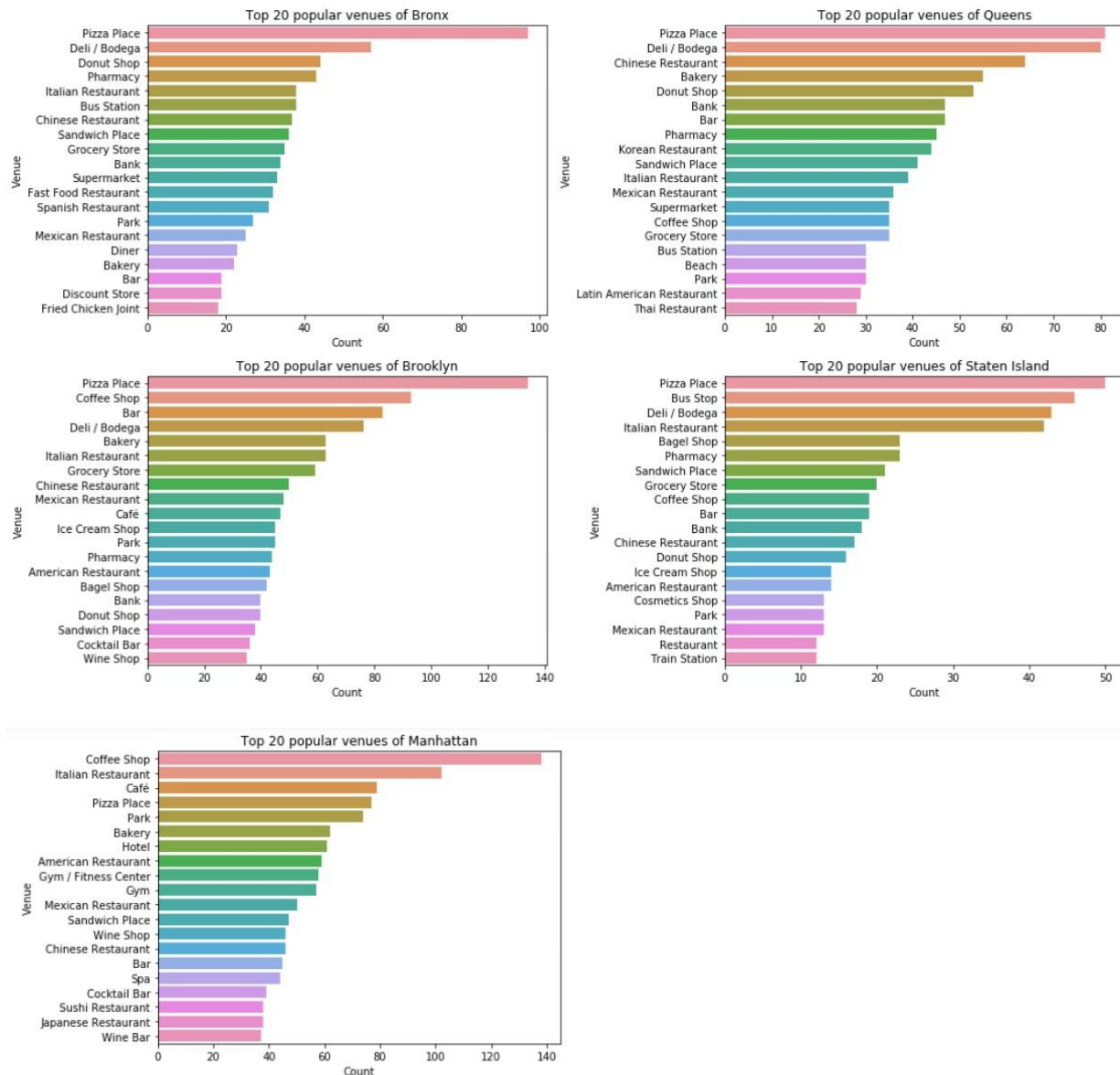


Fig 3.2.1. Top 20 venues of each borough of New York City

### 3.3. Application of K Means clustering

The dataframes of NYC venues and Toronto venues were concatenated to perform the clustering operation. This concatenated dataframe consist of Borough, Neighborhood, Neighborhood latitude, Neighborhood longitude, Venue name, Venue latitude, Venue longitude, and Venue category. To implement clustering we only need venue categories so the data of the venue categories column was converted in 0's and 1's by applying `get_dummies()` function on the column and stored it in a new dataframe. This dataframe with all the venue categories in 0's and 1's was fit in the K means algorithm and elbow method was performed to obtain the best k value between 2 and 20. To obtain the most suitable k value with more confidence, silhouette score was also found for k value between 2 and 20. In both the methods the parameters of the algorithm were kept constant which were (`init= "k-means++"` , `n_init = 12`, `random_state=0`). Two separate dataframes, one for Toronto city with cluster labels and another for NYC with cluster labels, were created to identify and analyze the difference and similarities of the neighborhoods of both these cities. Afterwards each cluster of Toronto city and New York City was analyzed, and suggestions were provided to open up new business based on surrounding venues and characteristics of that particular cluster

## 4. Results and Discussions

### 4.1. Best value of k (Number of clusters)

The resultant plot to find the best value of k is shown below. Inertia is the sum of squared error for each cluster. Therefore, the smaller the inertia the denser the cluster (closer together all the points are). As we can see in the plot that there is no elbow point generating after which the inertia stops decreasing drastically.

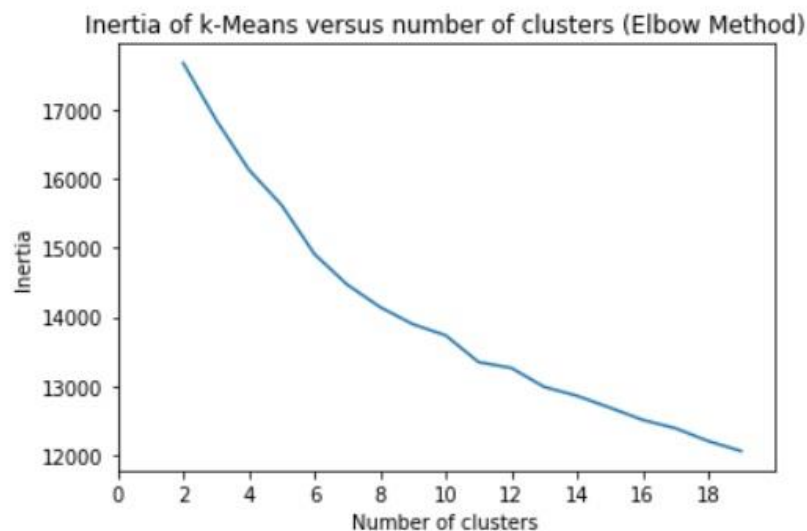


Fig 4.1.1. Plot of Inertia vs. Number of clusters



Below, is the table for different values of k and its Silhouette Score. The Silhouette Score is from -1 to 1 and show how close or far away the clusters are from each other and how dense the clusters are. The closer your silhouette score is to 1 the more distinct your clusters are.

Number of Clusters (k)	Silhouette score	Number of Clusters (k)	Silhouette score	Number of Clusters (k)	Silhouette score
<b>2</b>	0.40	<b>8</b>	0.16	<b>14</b>	0.15
<b>3</b>	0.40	<b>9</b>	0.15	<b>15</b>	0.13
<b>4</b>	0.19	<b>10</b>	0.16	<b>16</b>	0.10
<b>5</b>	0.17	<b>11</b>	0.16	<b>17</b>	0.16
<b>6</b>	0.19	<b>12</b>	0.17	<b>18</b>	0.12
<b>7</b>	0.18	<b>13</b>	0.15	<b>19</b>	0.10

Table 4.1.1. Number of clusters and Silhouette Score

We can see that Silhouette score is pretty much same for all the cluster numbers except  $k = 2$  and  $k = 3$  but using those low number of clusters wouldn't allow to examine the characteristics of the neighborhoods in detail.

So after trying different k value from 2 to 20 and observing the distribution of the cluster labels, I chose number of clusters to be 9 and performed clustering on the dataset with all the venue categories in form of 0's and 1's.

The neighborhoods of Toronto and NYC were plotted on the map to visualize how clusters have formed which we can see in both the figures below.

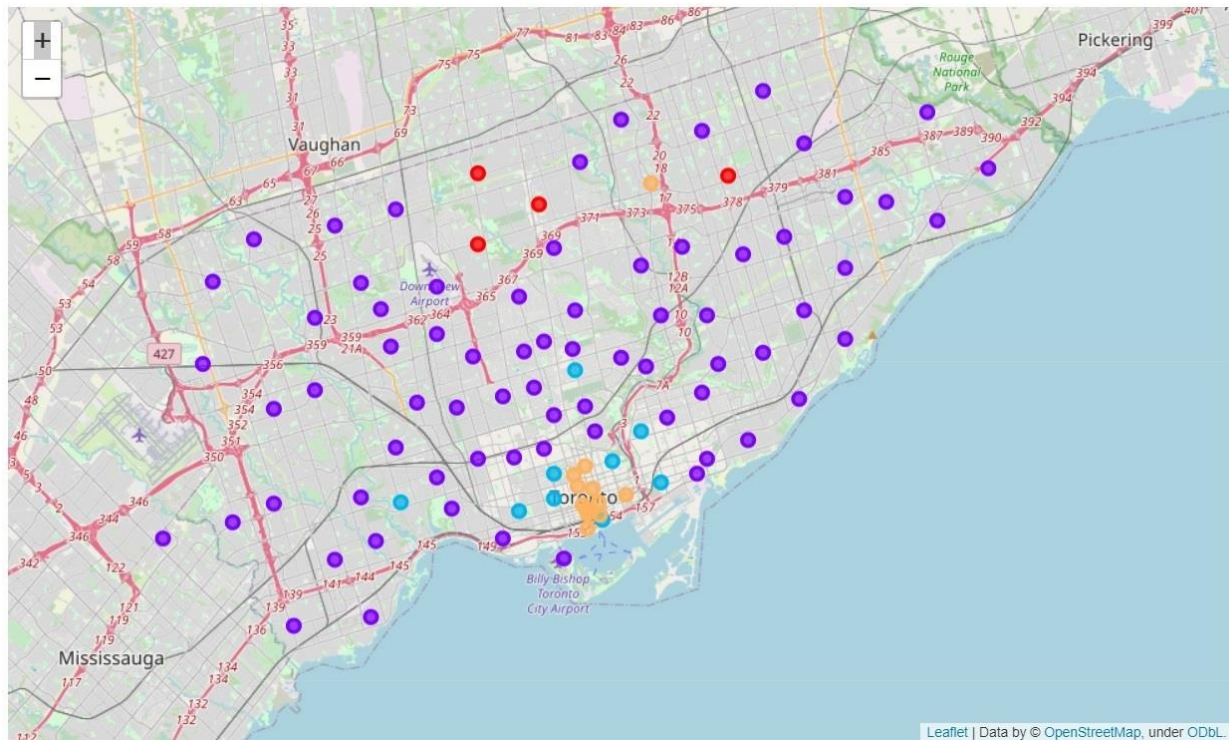


Fig. 4.1.2. Clustering of neighborhoods of Toronto city

From the Fig 4.1.2. it can be clearly observed that most of the neighborhoods of Toronto fall under cluster 1 (violet circles) and the few of the remaining neighborhoods fall under 2-3 different clusters which are cluster 0 (Red circles), cluster 3 (Blue circles), and cluster 7 (orange circles). We can clearly see that Downtown Toronto shows different characteristics than the other neighborhoods of Toronto

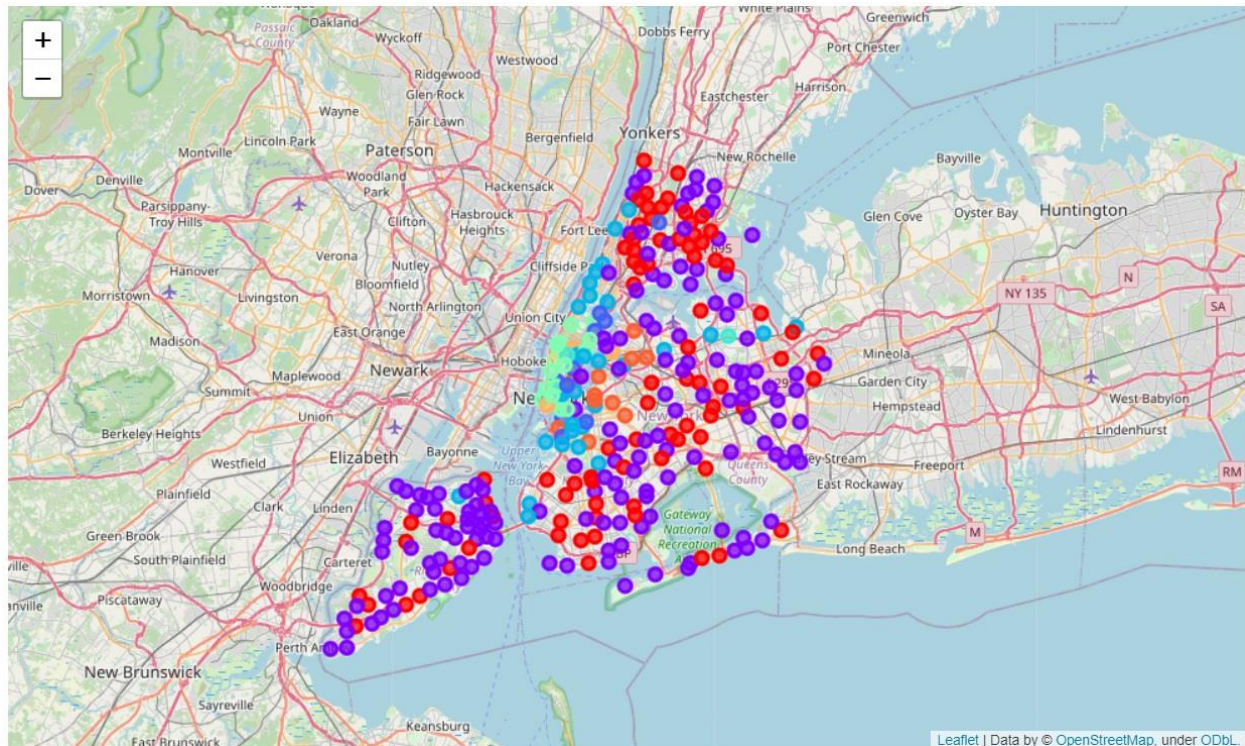


Fig. 4.1.3. Clustering of neighborhoods of New York City

From the Fig. 4.1.3. it can be clearly observed that most of the neighborhoods of NYC fall under cluster 1 (violet circles). There are so many other neighborhoods which fall under cluster 0 (red circles) and the remaining neighborhoods are distributed under different clusters. We can clearly see that the downtown area that is lower Manhattan show different characteristics than the other neighborhoods of NYC

## 4.2. Similarities of neighborhoods of Toronto and New York City

We visualized the neighborhoods of Toronto and New York city in the geographical maps above. We observed that all the different color of circles that exist in the map of Toronto also exist in the map of NYC. So, it is very clear that the neighborhoods which fall in the same cluster share similar characteristics. Below, different bar plots are shown to visualize the data of clusters in more details and identify which venues are most popular and common around these neighborhoods.

From fig 4.2.1. below, we can see that Cluster 0 (Toronto) and Cluster 0 (NYC) has multiple venues exactly same such as pizza place, coffee shop, bank, pharmacy, fast food restaurant,



Chinese restaurant, sandwich place, and ice cream shops. It can be clearly observed that the total number of each venue in cluster 0 (Toronto) is very low which proves that the individual neighborhood falling under this cluster would have very few venues exist around them.

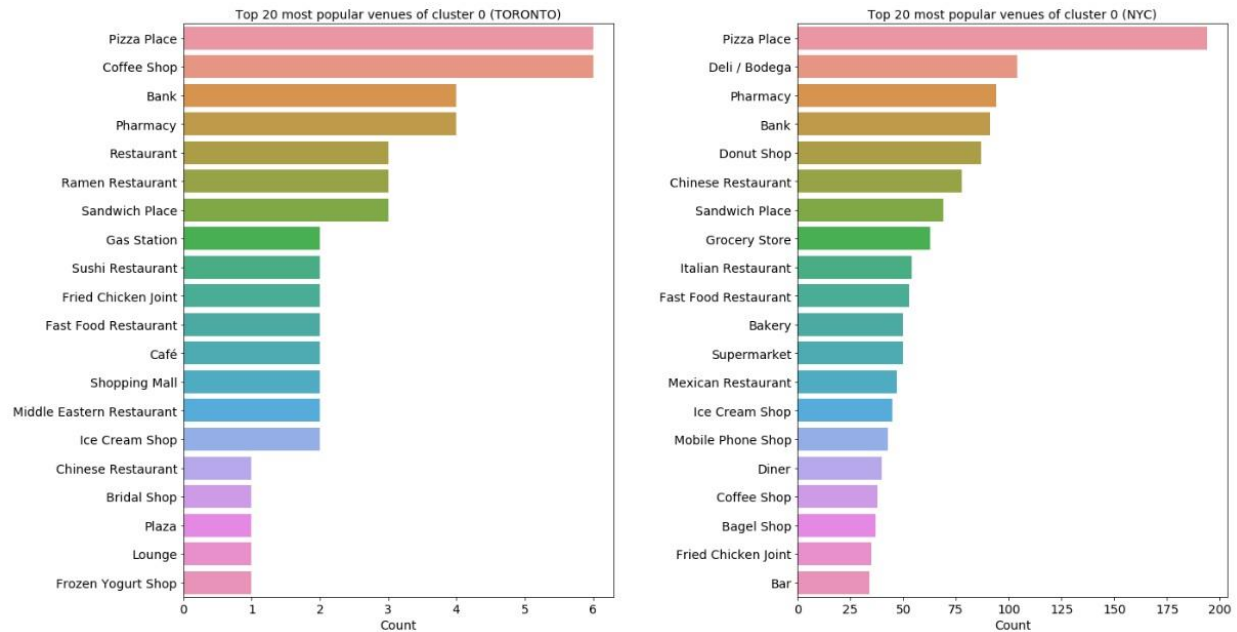


Fig 4.2.1. Neighborhoods of Toronto and NYC falling under cluster 0

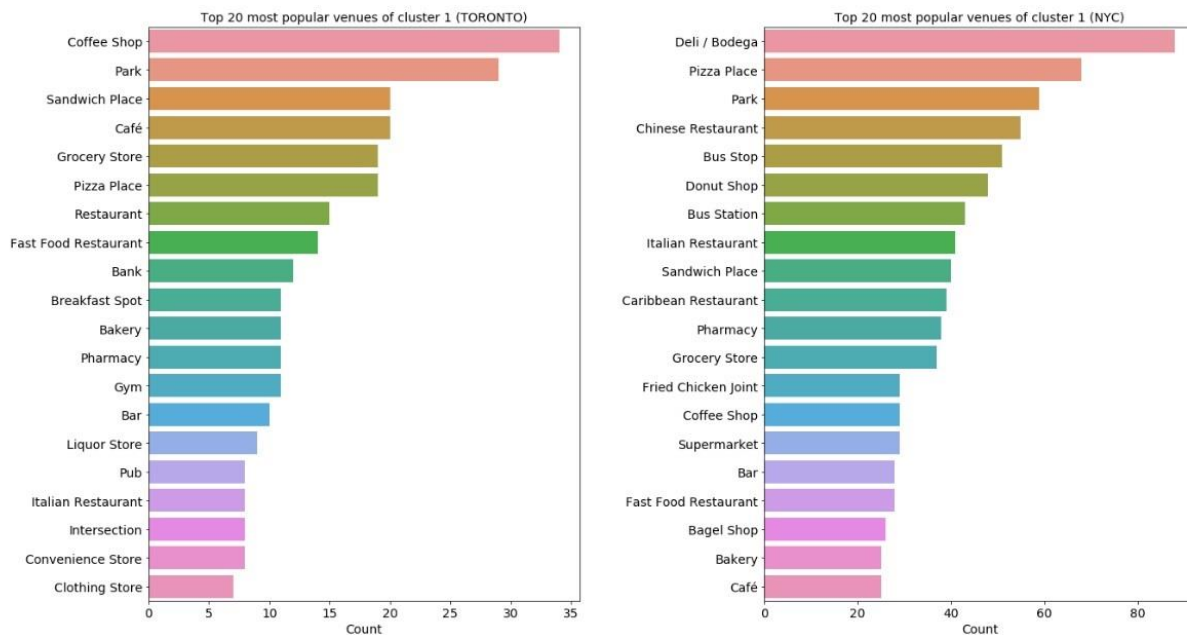


Fig 4.2.2. Neighborhoods of Toronto and NYC falling under cluster 1

Similarly, we can observe from the fig 4.2.2. above that cluster 1 (Toronto) and cluster 1 (NYC) share multiple common venues such as coffee shop, park, sandwich place, cafe, grocery store, pizza place, fast food restaurant, bakery, pharmacy, bar, and Italian restaurants

From the fig 4.2.3. below, we can observe that cluster 3 (Toronto) and cluster 3 (NYC) have cafe, coffee shop, Italian restaurant, pizza place, bakery, bar, park, sandwich place in common

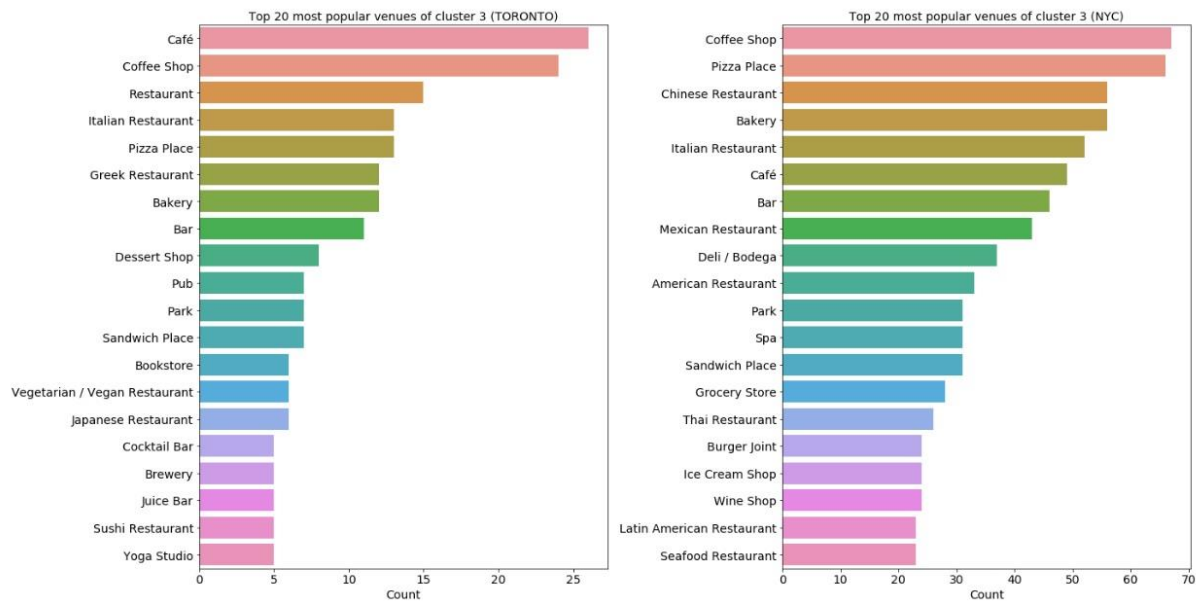


Fig 4.2.3. Neighborhoods of Toronto and NYC falling under cluster 3

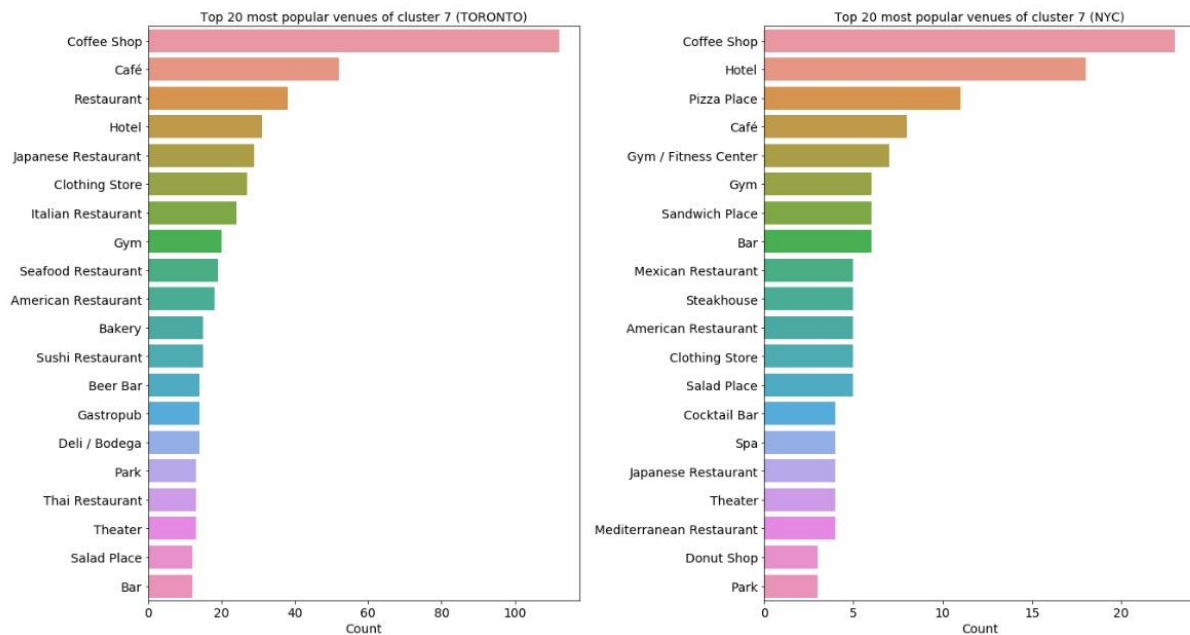


Fig 4.2.4. Neighborhoods of Toronto and NYC falling under cluster 7

From the fig 4.2.4. above, we can observe that cluster 7 (Toronto) and cluster 7 (NYC) share common venues such as coffee shop, cafe, hotel, Japanese restaurant, clothing store, gym, American restaurant, park, salad place, bar.

We can clearly observe that different clusters like cluster 0, cluster 1, cluster 3, and cluster 7 have different topmost venues as well as different bottommost common venues which proves that clustering performed well. The other thing is the venues of Toronto and NYC falling under same clusters do share at least 7-8 venues in common which again proves that clustering performed well, and these neighborhoods do have similar characteristics in terms of venues around them

### 4.3. Differences of neighborhoods of Toronto and New York City

As we saw earlier, the neighborhoods of Toronto just fall under clusters 0, 1, 3, and 7 which itself implies that the neighborhoods of NYC which fall under clusters 2, 4, 5, 6, and 8 have different characteristics than the neighborhoods of Toronto.

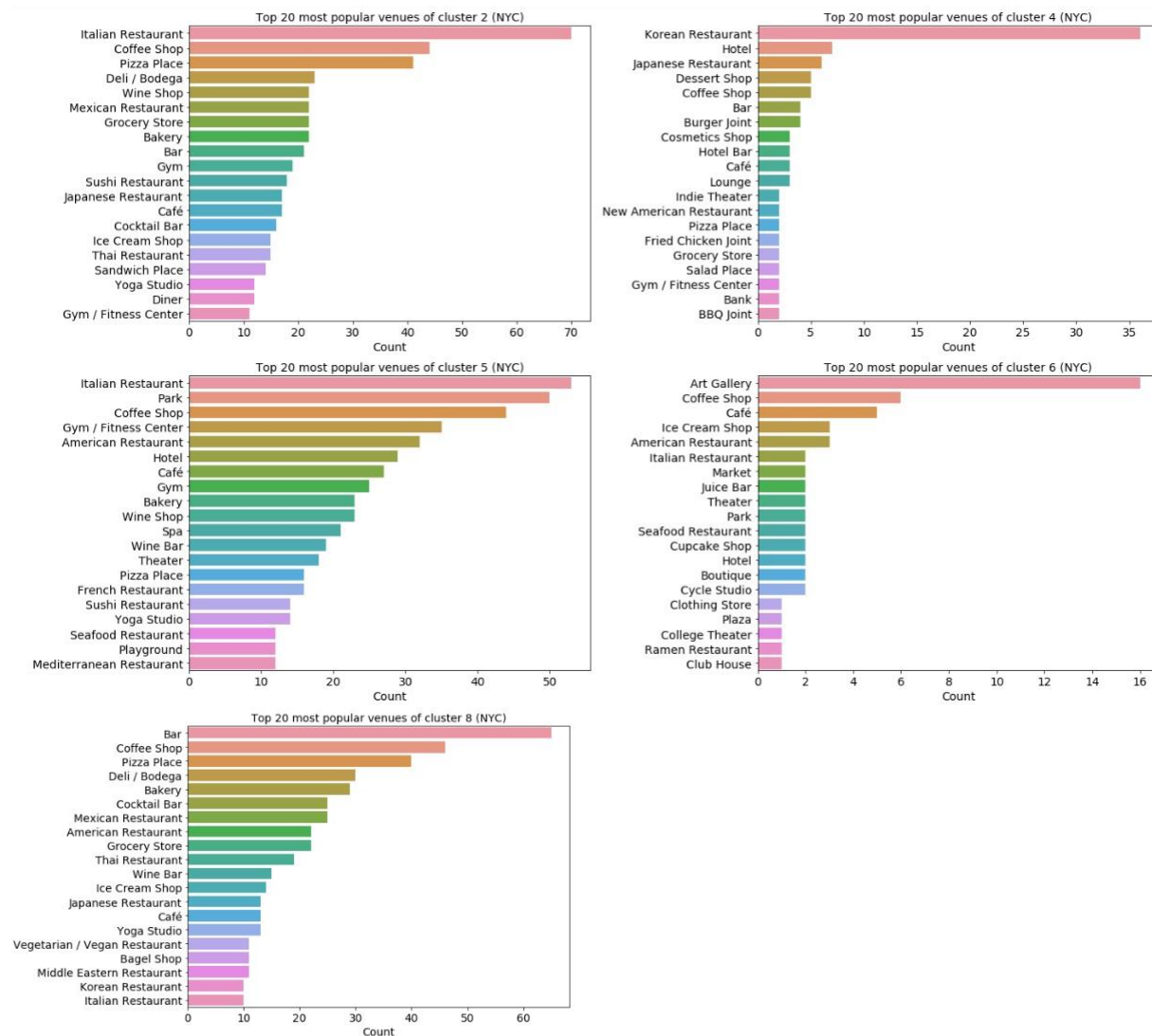


Fig 4.3.1. Neighborhoods of NYC falling under cluster 2, 4, 5, 6, and 8

Fig 4.3.1. has most popular venues of all these clusters. There are two neighborhoods under cluster 4 and just one neighborhood under cluster 6 which have different characteristics than any other neighborhoods of NYC itself, so they had to be different than the neighborhoods of Toronto as well. These two clusters have Art galleries, Korean restaurants, dessert shops, cosmetics shops, juice shops, cupcake shops, plazas, lounges around their neighborhoods.

Cluster 8 definitely has top most common venues similar to some of the neighborhoods of Toronto such as bar, coffee shops, and pizza places but other than that the most common venues include various different types of restaurants such as Mexican, American, Thai, Japanese, Vegan, Korean, Middle Eastern, and Italian. Along with that the cluster also has lot of wine bars, deli/bodega, and bagel shops which all are not seen together in any of the clusters of Toronto.

Cluster 2 has 6 neighborhoods of Manhattan, 2 neighborhoods of Brooklyn, and 1 neighborhood of Bronx. We can see that the Italian restaurant is the most popular venue with very huge number. The other venues seem to be more or less same as the other clusters except the popularity of those venues vary.

All the neighborhoods of cluster 5 are in Lower Manhattan which is downtown of Manhattan. It completely shows different characteristics than the other neighborhoods of Manhattan. That is why we do not see deli/bodega, donut shops, bagel shop, and sandwich place around neighborhoods but more variety of restaurants and pubs, bars, wine bars, cafe, and spa

#### **4.4. Analysis of neighborhoods cluster wise**

In this section, all the neighborhoods will be analyzed and suggestions of starting a new business will be provided based on the venues they are surrounded by and the characteristics of the other neighborhoods of the same cluster. Other parameters such as population of that particular neighborhood, competition with other same businesses, money required to start/expand a new business are not taken into account. On top of that this study is limited to the 500 meters (0.3 miles) of radius from the location of the neighborhood. So, these two things remain as a future scope of this study

As we observed earlier that most common venues across all the boroughs of Toronto are coffee shops, cafe, pizza places, parks, sandwich places, pharmacy, grocery store, bakery, and gym. As pharmacy, bank, gym, bakeries need separate consideration as a business, I have focused more on possibilities of opening up very common and preferable places like coffee shops, café, or pizza places

The most popular venues across different boroughs and different clusters are varying so I have tried to provide suggestions based on the venues surrounding the particular neighborhood and the characteristics of the neighborhood of same cluster. Here, I have tried to focus more on venues such as pizza places, coffee shops, donut shops, bagel shops, sandwich places, and deli/bodega

Providing bar plots for all the neighborhoods was very difficult so I have prepared two separate excel sheets; one for Toronto city which includes 5 most popular venues of each neighborhood of each borough and 10 most popular venues of each borough of Toronto. Another excel sheet for NYC with same concept. They can be found on GitHub by name “FINAL\_TORONTO\_BOROUGH\_NEIGHBORHOOD” and “FINAL\_NYC\_BOROUGH\_NEIGHBORHOOD”

## **Cluster 0:**

Top 5 most common venues of Toronto include pizza place, coffee shop, bank, pharmacy, variety of restaurants, and sandwich places. There are just 3 out of 98 neighborhoods of Toronto fall under this cluster. These three neighborhoods are Bathurst Manor (North York), Willowdale (North York), Clarks Corners (Scarborough). Willowdale has pizza places, coffee shops, and Sandwich places around which means the food industry is already pretty well established there. Bathurst Manor does have coffee shops around but no pizza place or sandwich place. Similarly, Clarks Corners does have pizza places around but not coffee shops or sandwich places. Now as the most common venues of the cluster include pizza places and coffee shops at top than it would be a great idea to open up coffee shop, pizza place, or sandwich place around the neighborhoods which do not have them

82 of 305 neighborhoods of NYC fall under cluster 0. Top 5 most common venues of this cluster are pizza place, deli/bodega, pharmacy, bank, and donut shop.

27 of 51 neighborhoods of Bronx are in this cluster. The coffee shops and cafe are less popular compare to pizza places and donut shops across the neighborhoods of Bronx. Deli/bodega and grocery stores exist everywhere very close by to almost all the neighborhoods of this borough. Pharmacies are also located very close to most of the neighborhoods. We saw earlier that pizza place is the most popular venue in Bronx and donut shop is third most popular venue. Both of them are popular venues of the cluster 0 as well. So here I have tried to find the neighborhood where opening either of them proved beneficial. Concourse, Concourse village, Mount Eden, and Pelham Bay are the neighborhoods where neither pizza places nor donut shops fall in top 5 venues so these neighborhoods would be the best possible locations in Bronx to start the business

20 of 70 neighborhoods of Brooklyn are in this cluster. It was observed that pizza place and donut shop are popular in the neighborhoods of Brooklyn as well. Different neighborhoods have got variety of restaurants close to them where Chinese restaurants being most popular one. One interesting thing to observe is that deli/bodega or grocery stores are not that close to many of these neighborhoods. So, most of the neighborhoods stand a great chance to open up more deli/bodega around them. Other neighborhoods such as Brighton Beach, Ocean Hill, and Rugby do not have any pizza places or donut shops around so those could be great neighborhoods to open up more pizza places or donut shops

Marble Hill (Manhattan) is the only neighborhood in this cluster which has little different qualities than others that it has more coffee shops and sandwich places around than pizza place, donut shop, or deli/bodega

21 of 81 neighborhoods of Queens fall under this cluster. We observed that Deli/bodega being the most popular venue among all the neighborhoods and pizza places and donut shops are behind it. One more interesting observation is that few of the neighborhoods such as Bellerose, College point, Douglaston, Howard beach, Jamaica center, Kew gardens, Kew garden hills, Oakland gardens, Rockaway beach, Woodhaven have got at least two or more restaurants around them. As it is very diversified in terms of its venues it would be pretty difficult to find out the location but still there few neighborhoods such as Douglaston, Glen oaks, Howard beach, and Woodhaven which do not pizza places or donut shops around although they have got many restaurants around them



13 of 62 neighborhoods of Staten Island are part of this cluster. Pizza place being the most popular venue among all the neighborhoods; Italian restaurants, bagel shops and sandwich places are pretty common. Pharmacies are also available in good numbers around the neighborhoods. Donut shop and deli/bodega almost do not have any existence as one of the top 5 venues of any neighborhood except Pleasant Plains and Roseville. So, all these 13 neighborhoods under cluster 1 stand a great chance to start business by opening donut shops or deli/bodega

### **Cluster 1:**

As we already saw on the geographical map of the neighborhoods with different clusters, we observed that most of the neighborhoods of Toronto (72 neighborhoods out of 98) fall under cluster 1. The most common venues of cluster 1 are coffee shop, park, sandwich place, cafe, grocery store, and pizza places. From the data, we observed that there are not so many venues exist closer to the neighborhoods and assuming the less population density behind this. So these neighborhoods which do not have any of these venues closer to them, definitely reflects a great scope of starting a new business or expanding your business.

Almost all the neighborhoods of Central Toronto falling under this cluster 1 which are (Davisville North, Forest Hill North & West, Lawrence park, Moore park, North Toronto west, Roselawn, Summerhill West, The Annex) show great opportunity of opening up cafe, pizza place, and sandwich place as none of the neighborhoods have these venues closer to them

The other neighborhoods of Toronto like CN Tower(Downtown Toronto), Christie (Downtown Toronto), Rosedale (Downtown Toronto), Business reply mail Processing Centre (East Toronto), India Bazaar (East Toronto), The Beaches (East Toronto) stand show lots of opportunities to expand your business in food industry such as opening coffee shops, cafe, pizza places, and sandwich places as well. Starting a new business wouldn't be recommendable as these venues do not seem to be very populated and permanently livable places

The neighborhoods of East York has got few pizza places and coffee shops close by but still there is high scope of opening up cafe, sandwich places, and grocery stores

Sandwich place and pizza place are most popular among the neighborhoods of Etibicoke in this cluster. There are few neighborhoods with coffee shops, park, and grocery store around them. So the neighborhoods like Kingsview village, Mimico NW, New Toronto, Northwest, and South Steeles could have great opportunities of opening up coffee shop, or cafe as most of these neighborhoods do have sandwich places and pizza places already located there which is shows positive sign of growing food industry there

The neighborhoods of North York falling under this cluster seem to have lot of opportunities for food industry to grow. There are 14 neighborhoods of North York in this cluster and very few of them have coffee shops and pizza place closer to them. All the neighborhoods have absolutely uncommon venues like dog run, golf course, fabric shop, gift shop, miscellaneous shop, construction and landscaping and so many like that. Having highest population among all the boroughs of Toronto and second largest land area, it could be possible that the venues are located far away as they have got plenty of land and neighborhoods to reach to. That could be the reason why the stores are not available within 500 meters (0.3 miles) of distance from a particular neighborhood but if that is not true in all the cases than North York stands a great chance of development and offers great opportunities to develop and grow the food industry

Similar theory is applicable to neighborhoods of Scarborough as well. It can be clearly observed that the neighborhoods have variety of restaurants closer to them than the basic coffee shop, cafe, pizza place, grocery store, pharmacy or anything like that. The neighborhoods Birch cliff, Kennedy park, Rouge hill, Steeles west, and Woburn has got either coffee shop, cafe, or a pizza place. So this borough also stands a great chance of opening up any business in the food industry as they already have got the restaurants around. Same thing is applicable to the neighborhoods of West Toronto and York

152 of 305 neighborhoods of NYC fall under cluster 1. Top five most common venues of this cluster are deli/bodega, pizza place, park, Chinese restaurant, and bus stops. Donut shops, Italian restaurants, and sandwich places are behind top five venues. Coffee shops and cafe are among the last 5 venues out of top 20 venues of cluster 1 of NYC.

24 of 51 neighborhoods of Bronx lie in cluster 1. There is no such trend of popular venues can be found in the neighborhoods of Bronx lying in this cluster. There are few of the neighborhoods such as Longwood, Morris Heights, Mount Hope, and Olinville where grocery stores or deli/bodega are more popular and there are few other neighborhoods such as Bronxdale, Castle Hill, Claremont village, Co-op city, and Hunts point where pizza place are popular as they lie in top 5 venues. Coffee shops or donut shops are not found in top 5 except one or two neighborhoods. Another interesting thing to observe is that there are anyways very few venues around the neighborhoods compare to what we observed in cluster 0. So probably all these neighborhoods stand a great chance to start a new business whether it be a coffee shop, pizza place, donut shop, bagel shop, pharmacy, or deli/bodega

28 of 70 neighborhoods of Brooklyn lie in this cluster. Pizza place is more common across different neighborhoods but again it is difficult to see a trend as the top 5 venues of the neighborhoods keep changing and have some uncommon venues as well. Neighborhoods such as Bedford Stuyvesant, Broadway junction, Brownsville, East New York, Flatbush, Gerritsen Beach, Madison, Manhattan beach, Ocean Parkway, Starrett city, Vinegar Hill, and Windsor terrace are the neighborhoods which show ordinary characteristics in terms of venues around it as they have got donut shops, pizza places, bars, pubs, coffee shops, bagel shops, pharmacies around them. There are few unique neighborhoods such as Bergen Beach (venues: harbor, Baseball field and Hockey field), Coney Island (Venues: Baseball stadium, Caribbean restaurant, skating rink), Dyker Heights (venues: Golf course, dance studio, yoga studio), Highland park (venues: Construction and landscaping, big box store, metro station), Marine park (Venues: soccer field, basketball court, baseball field), and Vinegar Hill (venues: Food truck, art gallery, antique shop). So similarly, all these neighborhoods stand a great chance to start a new business whether it be a coffee shop, pizza place, donut shop, bagel shop, or sandwich place.

There are just 3 neighborhoods of Manhattan which fall under cluster 1. They are Central Harlem, Roosevelt Island, and Stuyvesant Town. They have variety of restaurants around them and unique venues such as Indie theatre, Boat or Ferry, Baseball field, and Heliport.

49 out of 81 neighborhoods of Queens fall under cluster 1. Deli/bodega being most popular across so many neighborhoods and pizza place, donut shop, bagel shop, coffee shop, cafe are found here and there in few neighborhoods. There are plenty of bus stops and bus stations around the

neighborhoods. One interesting thing to be observed is that there are very few numbers of venues around each neighborhood if that venue is not being any type of restaurants. There are so many neighborhoods which have got many different types of restaurants within 500 meters of radius. Those neighborhoods are Elmhurst (venues: Thai, Mexican, Vietnam, Chinese restaurants), Pomonoc (venues: Halal and Japanese restaurants), Queensboro Hill (venues: Chinese, Indian, Asian), Ravenswood (venues: Chinese, Brazilian, Japanese), Rochdale (venues: Southern/Sould food restaurant, breakfast spot, fast food restaurant), St. Albans (venues: Carribean and Chinese restaurants). We observed the same thing in cluster 0 which proves that Queens has the most diversified culture compare to other boroughs of NYC. There are four neighborhoods which have got plenty of beaches within 500 meters of radius which are Belle Harbor (3 beaches), Hammels (6 beaches), Neponsit (4 beaches), Roxbury (1 beach). As most of these neighborhoods do not have pizza places, coffee shops, donut shops, or sandwich places close to so many of the neighborhoods, it would be a great idea to open up any of these stores taking the other surroundings and mandatory parameters into consideration

48 out of 62 neighborhoods of Staten Island fall under cluster 1. Deli/bodega being most popular across most of the neighborhoods and then comes Italian restaurants which are seen around many neighborhoods. Pizza place, bagel shop, and American restaurants are the other venues which are found across many neighborhoods as well. Neighborhoods which have two or more top most common venues of cluster 1 are Arden Heights, Arrochar, Chelsea, Concord, Egbertville, Elm park, Fox hills, Graniteville, Grantcity, Huguenot, Manor heights, Mariner's harbor, New Dorp beach, Old Town, Prince's bay, Randall Manor, Richmond town, Shore Acres, South Beach, Tottenville, and Willowbrook. All these neighborhoods show great prospect to open up any business or expand any business related to food industry such as deli/bodega, bagel shop, donut shop, sandwich places, coffee shops or cafes. Other than that all the other neighborhoods have shown one or no top most common venues of cluster 1 so these neighborhoods also stand the chance to open up any of the business related to food industry after careful observation of surrounding venues and other considerable parameters

## **Cluster 2:**

Cluster 2 has got very few neighborhoods, but it involves lot more neighborhoods of Manhattan than cluster 0 and cluster 1. There are 6 neighborhoods of Manhattan, 3 neighborhoods of Brooklyn, and one neighborhood of Bronx. With more neighborhoods of Manhattan, we see the change in top most common venues of the cluster. Italian restaurant is the top most common venue whereas coffee shop and pizza place are other most common venues across cluster 2. One more interesting observation is that all these common venues exist in very large number within 500 meters (0.3 miles) of radius. Belmont of Bronx and Carroll Gardens of Brooklyn is called little Italy and we can clearly analyze that as the Italian restaurants are the most popular venues of these neighborhoods

There are already numerous venues exist within 0.3 miles of radius of all these neighborhoods. We found that there are few donut shops and bagel shops exist around these neighborhoods. Because of already tightly packed neighborhood, the best thing would be not to get into any these neighborhoods to start any new business related to food industry

### **Cluster 3:**

9 of 98 neighborhoods of Toronto fall under this cluster. Observing the venues of the neighborhoods, we found that coffee shop, pizza place, sandwich place, café, bakery and bar are very popular among these neighborhoods.

As we already observed earlier that the pizza places do not fall under top 20 venues of downtown Toronto, it might be a great idea to open up pizza place(s) around Kensington Market, University of Toronto, and Berczy Park. Downtown always has great scope for pubs, bars, clubs and all so opening up pubs or bars in Kensington Market could be a great idea.

Sandwich places were one of the most common venues across all the boroughs of Toronto but here we see that none of the neighborhoods have got sandwich places closer to them except Davisville. So, opening up a sandwich place also stands a great chance of success.

We observed that few neighborhoods of Brooklyn and many neighborhoods of Manhattan (13 out of 40) show little different characteristics and fall under this cluster 3. It is very clear that coffee shops and pizza places are the most common venues of this cluster. The bar plot of top 20 most popular venues of cluster 3 showed the same thing as well and in that plot also we can clearly see that pizza places and coffee shops exist almost in same numbers.

Out of 10 neighborhoods of Brooklyn, few of the neighborhoods have more coffee shops like Boerum Hill, Gowanus, Park Slope, and Williamsburg whereas few neighborhoods have more pizza places like Bay Ridge, Fort Greene, Fort Hamilton, and Prospect Lefferets Gardens and Downtown Brooklyn has got both the places in top 5 most common venues. We can observe few common venues which we have seen earlier like deli/bodega, bagel shop, sandwich places, Chinese and Italian restaurants but they just exist in one or two neighborhoods. From the geographical map it seems that these neighborhoods are around Downtown area and therefore it would be great idea to open up a pizza place or coffee shops in the neighborhoods where they do not exist as described above or open up sandwich places as it exists in few of the neighborhoods but not in all of them. So, neighborhoods like Fort Greene, Gowanus, Park Slope, Prospect Lefferets Gardens, and Williamsburg might be good prospects looking at their surroundings

Coffee shops, Chinese restaurants, and cafe are the most popular venues in the neighborhoods of Manhattan. From the geographical map it can be clearly seen that few of these neighborhoods are around downtown and others are lying near to the shoreline. We can see the two neighborhoods Chinatown and Little Italy which reflect the community located over there. As they already have plenty of local restaurants there and they are tightly packed with them, they shouldn't be the first choice to start any new business. Neighborhoods like Inwood, Manhattan Valley, Lower East Side, Manhattan Valley, Manhattan ville, Morningside heights, Murray Hill, Tudor city, and Upper West side seem to share very similar characteristics with each other according to their surrounding venues so to start any business or expand anything related to food industry than these neighborhoods might be a good prospect

As we have already observed in the neighborhoods of Queens that there are lots of restaurants of different varieties so those places wouldn't be the best places to start any new business in the food industry. Now, there are two neighborhoods which have got plenty of coffee

shops and pizza places around them and they fall in top 5 most common venues as well like Bayside, and Sunnyside.

#### **Cluster 4:**

There are just neighborhoods of NYC which fall under this cluster. One is Midtown South (Manhattan) and Murray Hill (Queens). Here, we found that both of the neighborhoods have got Korean community living there as the Korean restaurants dominate and are in large numbers compare to any other venue of this neighborhood. So, deciding to open up any new business in local community based on the preferences of diversified community would not be a great idea

#### **Cluster 5:**

We found that the neighborhoods falling under this cluster have completely different characteristics than the other neighborhoods of other clusters. The reason behind that is all these neighborhoods belong to core downtown areas of their own Borough.

From the geographical map we can observe that both the neighborhoods of Brooklyn; Dumbo and Fulton Ferry, lie almost at the border of downtown as well as the border of the borough. All the neighborhoods of Manhattan of this cluster can be located in the lower Manhattan area in the geographical map. Lower Manhattan is considered downtown of Manhattan as well as downtown of NYC.

We see parks as the most common venue and then Italian restaurants, coffee shops, gyms, and American restaurants. Pizza place, donut shops, bagel shops do not seem to be very popular around downtown area. Almost all the neighborhoods have three of these popular venues or more except Flatiron, Hudson Yards, Lincoln Square, Tribeca, and Upper East Side of Manhattan which do not have coffee shops in top 5 most common venues. So, looking at the surroundings and other required parameters coffee shops could be one of things to open up or expand

#### **Cluster 6:**

Chelsea, (Manhattan) is the only neighborhood falling under this cluster. It is popular for its art galleries which we could easily see on the bar plot which had 16 art galleries within 500 meters of radius. The other venues around the neighborhood were coffee shop, café, Ice cream shop, and American restaurant.

#### **Cluster 7:**

We found that all the neighborhoods falling under cluster 7 are located in Downtown Toronto except Fairview which is located in North York. We found that there are plenty of coffee shops, cafe, hotel, variety of restaurants where Japanese restaurants being the most popular ones. It is actually pretty difficult to make any conclusion of opening up any new venues or anything as the places are anyways located very close to each other in the downtown. But it would actually be a great idea of opening up those venues which are completely not seen in the top 5 venues of any neighborhoods. First such venue is pizza place as we discussed above that we did not see at all in the Downtown although it is one of the most popular venues of Toronto. Another very common venue which we have seen all across Toronto is Sandwich places which do not fall in top 5 venues of these neighborhoods except one neighborhood that is Central Bay street.



It was found that there are just three neighborhoods of NYC fall under this cluster. Financial District and Midtown of Manhattan, and Long Island City of Queens. They have coffee shops, hotel, and pizza places in common. The bar plot of top 20 venues of cluster 7 also say that these neighborhoods have got sandwich places, donut shops, parks, cafe, and few restaurants around them as well. So, they already have too many venues tightly packed around them so it would be difficult to start any new business over here

#### **Cluster 8:**

There are 7 neighborhoods of Brooklyn, 3 neighborhoods of Queens, and 1 neighborhood of Manhattan fall under this cluster. We found that the most popular venue of this cluster is bar. The other common venues are coffee shops, pizza places, and deli/bodega. We see numerous bars in all the neighborhoods of Brooklyn. Pizza places and coffee shops are other popular venues around these neighborhoods. There are few neighborhoods of Brooklyn and East village of Manhattan which have numerous cocktail bars as well. These neighborhoods have already got plenty of venues tightly packed around them so it is again pretty difficult to start any new business or anything like that but with these many bars around and pizza place and coffee shops being the other popular venues, it would definitely be a great idea to open up pizza place or coffee shop in the neighborhood which doesn't have any of these. Bushwick, East Williamsburg, Prospect Heights of Brooklyn and Astoria, and Woodside of Queens stand a great chance to open up a pizza place. Similarly, Prospect heights of Brooklyn and Astoria and all the three neighborhoods of Queens stand a great chance to open up a new coffee shop

## **5. Future Scope**

- Increase the radius of the neighborhood and see whether the same interpretation is applicable or not
- Find the best value of k (number of clusters) in more efficient way and perform the whole analysis
- Focus on single venue or single borough of any of the cities and filter out other venues and other irrelevant data to analyze a particular borough and neighborhood in much more efficient way

## **6. Conclusion**

There are so many neighborhoods of both the cities which share the same characteristics in terms of venues surrounding the neighborhoods.

NYC has got plenty of venues surrounding almost all the neighborhoods whereas the neighborhoods of Toronto have very small number of venues surrounding the neighborhoods. So, Toronto comes out as a great prospect to start or expand any business related to food industry

## 7. References

- 1) <https://www.toronto.ca/business-economy/invest-in-toronto/strong-economy/>
- 2) <https://www.statista.com/statistics/444906/number-of-immigrants-in-canada/>
- 3) <https://en.wikipedia.org/wiki/Toronto>
- 4) <https://www.investopedia.com/articles/investing/091114/worlds-top-financial-cities.asp>
- 5) [https://en.wikipedia.org/wiki/New\\_York\\_City](https://en.wikipedia.org/wiki/New_York_City)