1. Introduction

1.1 Background

The city of Toronto is the largest in Canada, as well as one of the most culturally diverse cities in the world. With its official motto, “Diversity is our strength,” Toronto stands apart from multicultural cities like New York, Los Angeles, or Singapore in many ways like having a proportion of foreign-born residents of approximately 47%. This statistic is higher than any metro region in the United States, London, and Paris. When it comes to diversity, Toronto has the largest Latin American community in Canada, and within that community, a lot of people come from Colombia. Like most immigrants from other countries, Colombians choose to move to Canada in search of security and a better quality of life, such as getting a job opportunity, move in with family, start a business, or open a restaurant.

1.2 Problem

Toronto is a large city with over 100 neighborhoods to choose from as a potential location for a new Colombian restaurant. Fortunately, there is data available to help us narrow down possible neighborhood locations from 140 to 5. With some graphs and statistical analysis, we can gather insights in specific areas in terms of city demographics, neighborhood venues such as restaurants or cafes, and crime rates. Choosing a good location is a fundamental component to succeed in a business. In addition to a good menu and professional staff, a good location can give a substantial path to success. This project will take these variables into account to generate neighborhood clusters by familiar places that will help select the best locations for a Colombian restaurant. These clusters will not only consist of neighborhood demographics but also neighborhood diversity.

1.3. Interest

Any immigrants from Colombia looking to share their culture with not only fellow compatriots but with everyone living in or visiting Toronto would be interested in looking for a neighborhood to open their restaurant.

2. Data Acquisition and Cleaning

This analysis will use four different datasets. The first dataset, the Toronto neighborhood data, comes from the City of Toronto's Open Data Portal. The original dataset consists of 16 columns and 140 observations. I dropped eight columns, including neighborhood geometry and parent area ID. The columns used for the rest of the analysis are area (neighborhood), area code, latitude, longitude, shape area in square meters, shape length in meters, etc.

The second dataset is the average income for the 140 neighborhoods. I noticed that a measure of income could be essential to the selection of a restaurant location. The income data is part of the 2016 Toronto Neighborhood profiles, which is based on the Toronto Census and provides demographic, social, and economic characteristics of people and households in the city. To get the income, I filtered this large excel file to the income topic and determined that the best measure would be the average household income for each neighborhood. Once I imported this dataset, I joined it to the neighborhood data.

The third dataset for the analysis comes from the Toronto Police Service Portal. The Toronto neighborhood crime rates dataset consists of the 140 neighborhoods as rows and 61 columns. Most of the columns are crime rates and incident counts for crimes such as assault, auto theft, breaking/entering, and homicide from 2014 to 2019. Other variables in the dataset are neighborhood population, shape area, and shape length. To incorporate this broad dataset in the analysis, I calculated the overall crime rates by year from 2017 to 2019. The overall crime rate for a given year consists of dividing the total crime incidents for that year and the neighborhood population and multiplying the result by 100,000. After that, I dropped all of the variables except the neighborhood name, population, neighborhood ID, and the calculated crime rates. Then, I joined this resulting dataset to the neighborhood data by neighborhood ID. This combined dataset now contains neighborhood coordinates combined with some crime variables as well as neighborhood population, income, neighborhood area, etc.

The fourth dataset, the core of this analysis, comes from the Foursquare Places API. It offers real-time access to a global database of venue data in several cities around the world. For this analysis, I sourced the API with credentials such as client ID, client secret, and version number to gather venue data for all Toronto neighborhoods with the help of neighborhood name, latitude, and longitude variables from the combined Toronto dataset. The gathered data from the Foursquare API has 7 columns and 5684 observations, which represent each venue. Its variable names are neighborhood, neighborhood coordinates, venue name, venue coordinates, and venue category. This dataset had only one duplicate observation that has been dropped. Also, there are 348 venue categories including restaurants, parks, cafes, arcades, museums, etc.

3. Exploratory Data Analysis

3.1. Correlations

Before mapping the neighborhoods, I made a correlation plot of all the variables from the combined Toronto dataset. According to **figure 1,** four out of seven variables seem to have high correlations. Specifically, the three crime rates, as well as shape length and area, are highly correlated with each other.

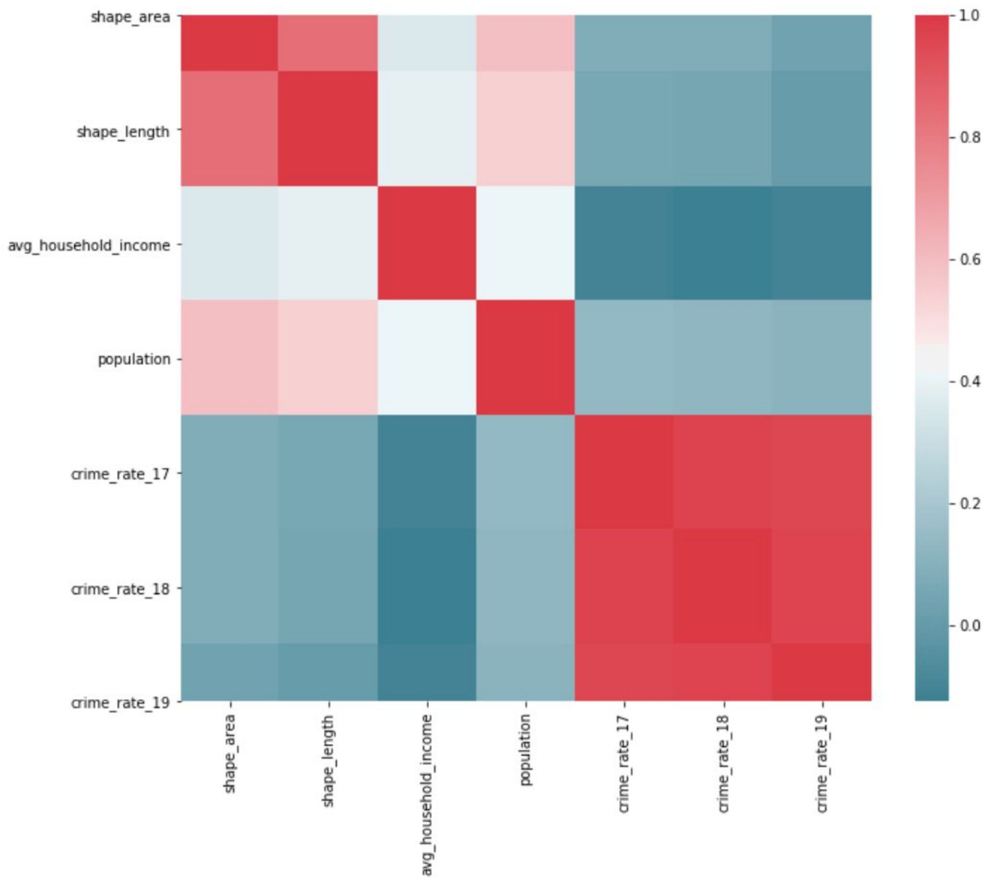


Figure 1

In order to accurately model the data, they cannot have high multicollinearity. Therefore, I used a metric named variance inflation factor (IVF) to determine which variables need to be excluded. If a variable scores greater than 10, then that variable should not be used in the analysis due to its high correlation with one or more variables.

	VIF	features
0	5.900379	shape_area
1	11.755830	shape_length
2	4.182230	avg_household_income
3	7.473740	population
4	83.820920	crime_rate_17
5	101.617077	crime_rate_18
6	76.704289	crime_rate_19

Table 1

According to table 1, the high variance inflation factor for neighborhood shape length reflects its high correlation with neighborhood area. Since the VIF for shape length is twice as high as and less important than area, shape length should be dropped from the data. Also, the high VIFs for the annual crime rates indicate that two of them are excluded from the analysis. Since crime rate is a variable of interest, only the 2019 variable should be part of the analysis.

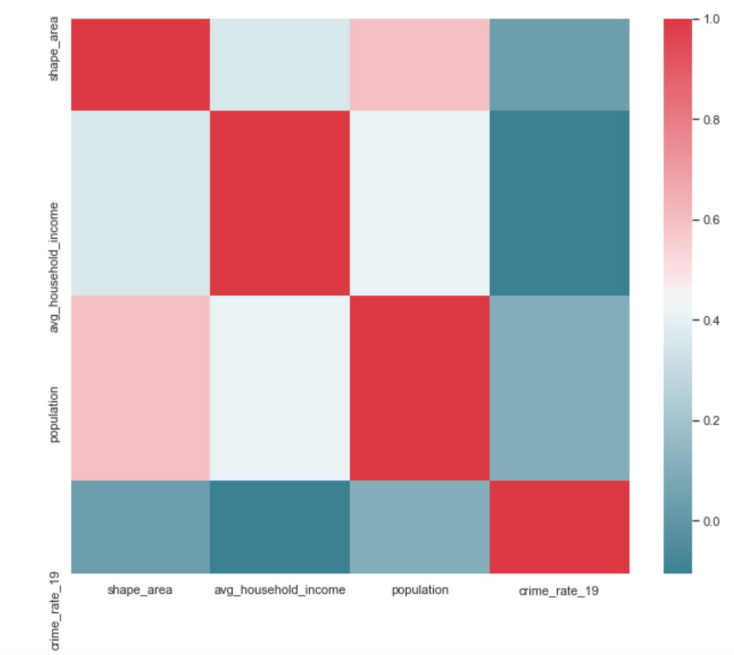


Figure 2

	VIF	features
0	3.086059	shape_area
1	3.631153	avg_household_income
2	6.896971	population
3	2.758644	crime_rate_19

Table 2

3.2. Mapping Neighborhoods

The 140 observations from the combined data represent a neighborhood, and two of the variables are the geographical latitude and longitude coordinates. To visualize the neighborhoods, I plot the coordinates in a map using the folium package. **Figure 3** shows the locations of all the neighborhoods within the Toronto area.

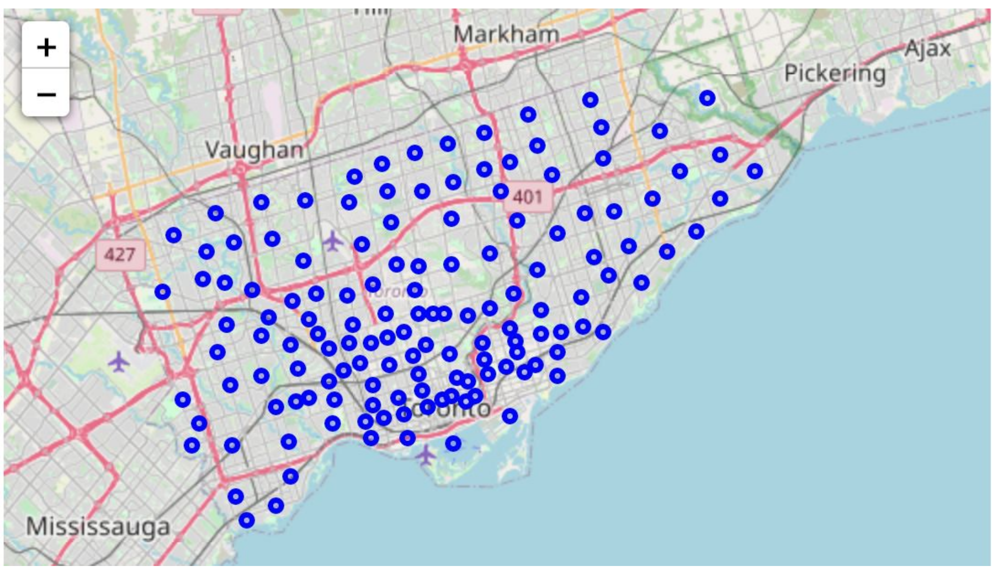


Figure 3

3.2.1. Neighborhoods by Population

To explore the population in Toronto neighborhoods, I found it useful to plot the neighborhoods by 10 highest and lowest each variable. For example, **Figure 4** shows the map of the 10 most populated neighborhoods in red and the 10 least populated. This map shows five of the least populated and two of the most populated around the downtown area. Additionally, most of the neighborhoods with large populations are located on the northeastern side of the city.

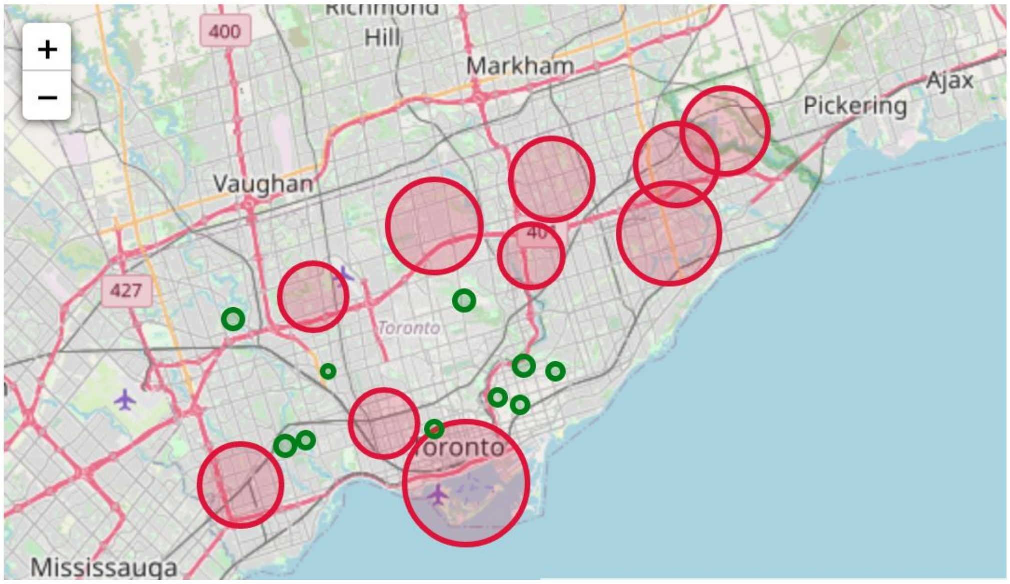


Figure 4

According to **figure 5a,** The most populated neighborhood, Waterfront Communities-The Island, is located by the lake and has a population of 65,913. In **figure 5b**, on the other hand, the neighborhood with the least population, Beechborough-Greenbrook, is located in the borough of York and has a population of 6,577.

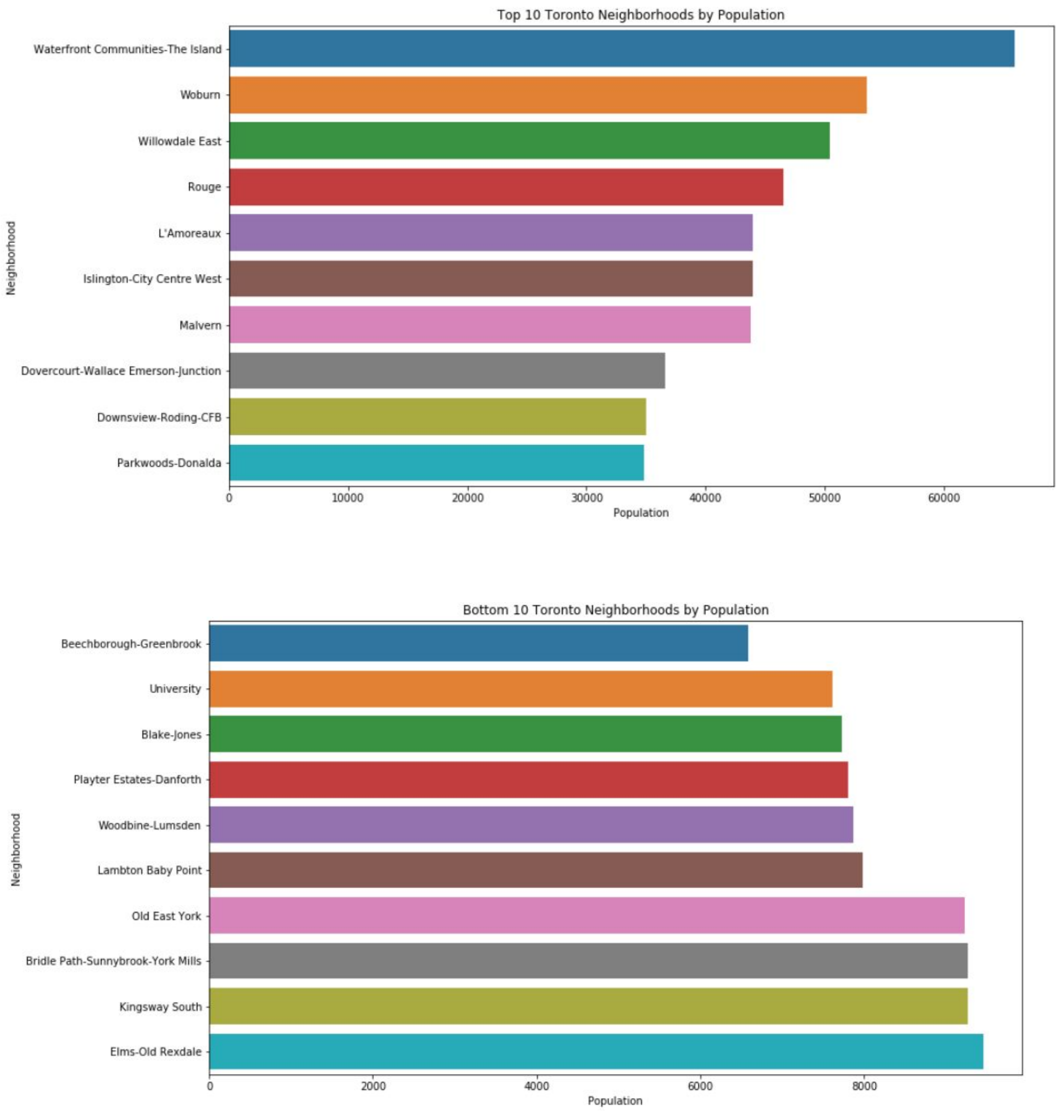


Figure 5

3.2.2. Neighborhoods by Average Household Income

Neighborhood income level is one target demographic that helps determine a location for a restaurant business. In addition to site surveys that show the population base of a neighborhood, a contrast between household income and the number of competing restaurants within the area will help determine the best type of restaurant to open. Like the neighborhood population, I explored the average income variable by plotting the areas with the highest and lowest household income. **Figure 6** displays the location of the neighborhoods with the highest and lowest average household income. The neighborhoods with the largest income are displayed in red and the lowest in green. Because the averages were too high for the high-income neighborhoods, the coefficient used to adjust the bubble sizes resulted in tiny dots for the low-income neighborhoods on the map.

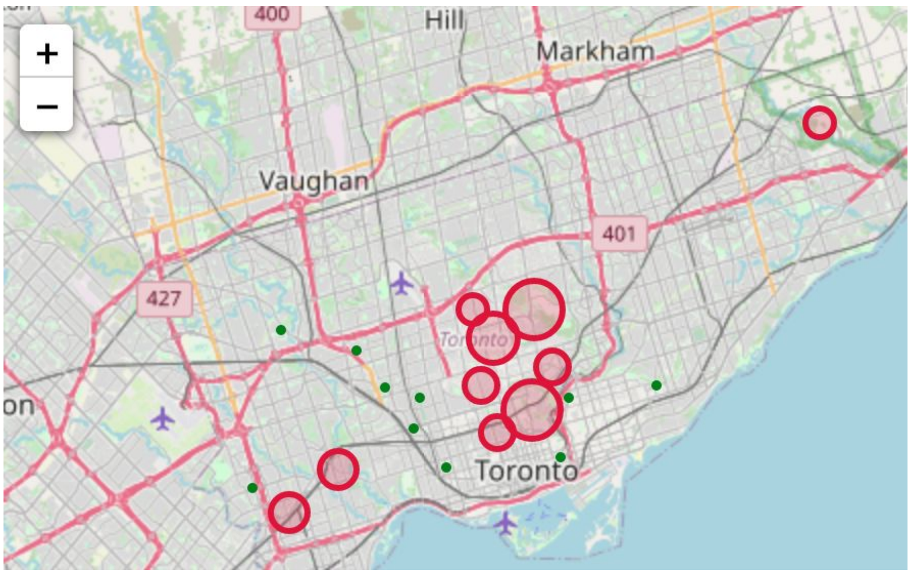
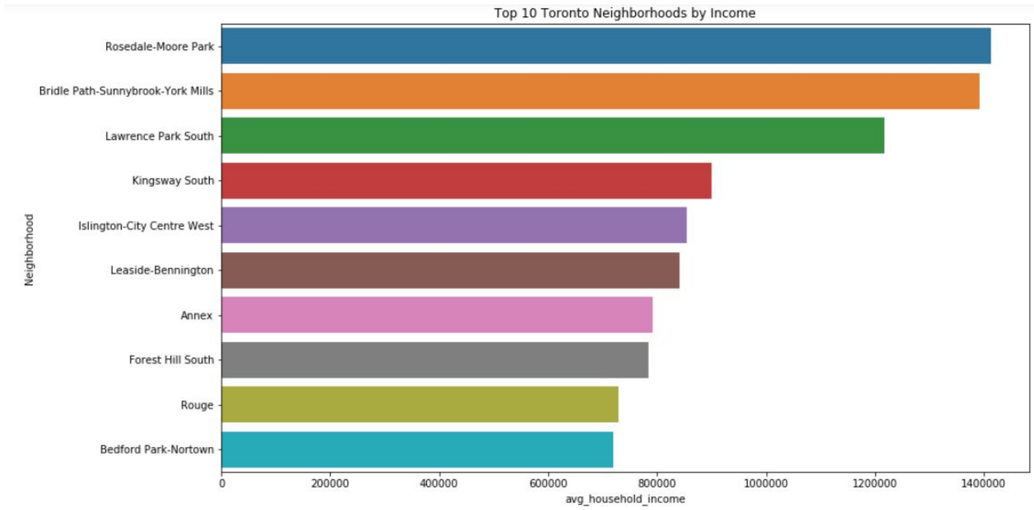


Figure 6

Fortunately, the bar graphs for the featured neighborhoods in **figure 7** helps with the low-income part of this visualization. Most of the high-income neighborhoods in the map are clustered in the downtown Toronto area surrounding some low-income neighborhoods. The neighborhood with the largest average household income, Rosedale-Moore Park, is in the downtown area. On the other hand, the neighborhood with the lowest average income, Taylor-Massey, is the easternmost point in the map below.



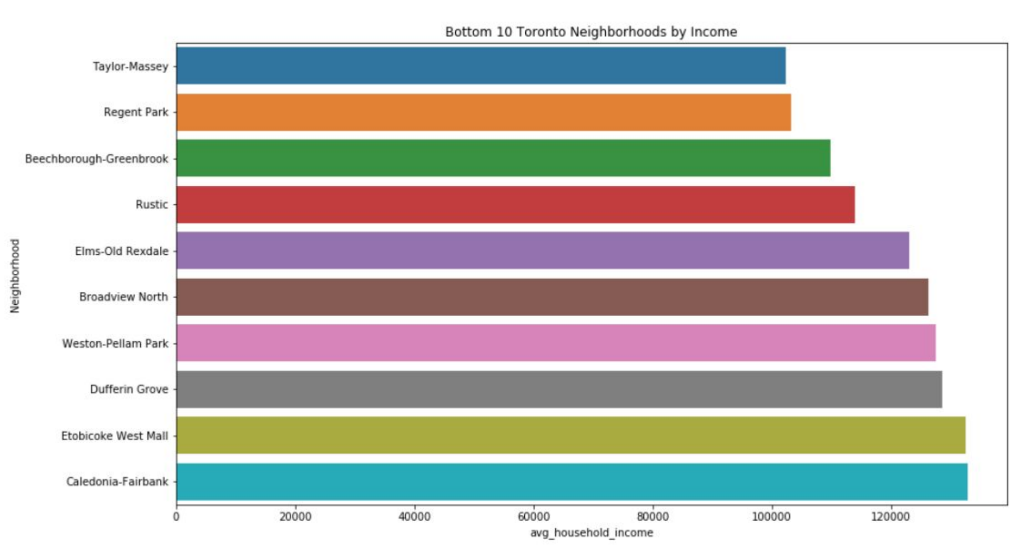


Figure 7

3.2.3. Neighborhoods by 2019 Crime Rate

Another factor in considering locations for a Colombian restaurant is crime statistics. While the original crime dataset consisted of many columns for crime incidents since 2014, the focus for the crime data is better if narrowed down to 2017 and after. For this analysis, I calculated the crime rate for each year since most of the data consisted of incident frequencies by type of crime. Moreover, it became clear to use the overall crime rate by year and then drop the variables used to calculate it. Since the crime rates from the past three years are highly correlated, I explored the 2019 crime rate in this section.

The map in **figure 8** shows that many neighborhoods with high crime rates are clustered close to each other, along with others, in downtown Toronto. This implies that opening a restaurant in this area needs to be handled with care, considering city crime. The low-crime neighborhoods, shown in blue, are more scattered, but most are located in the east part of the city.

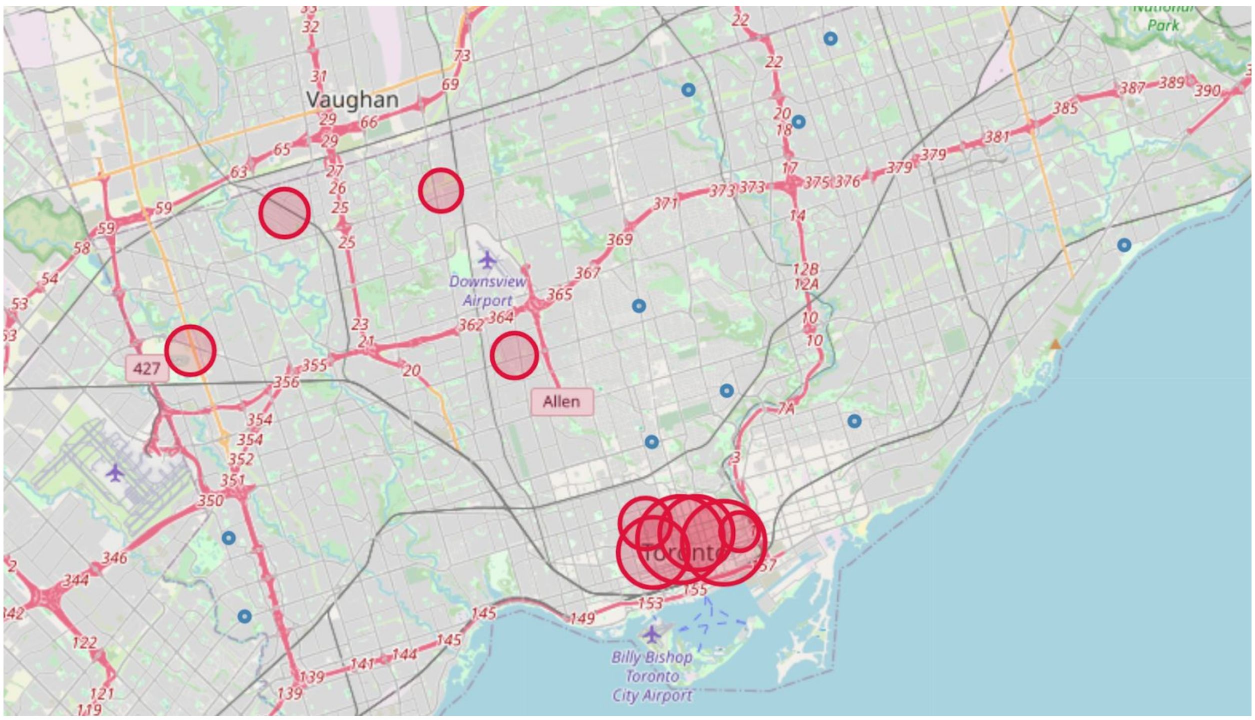
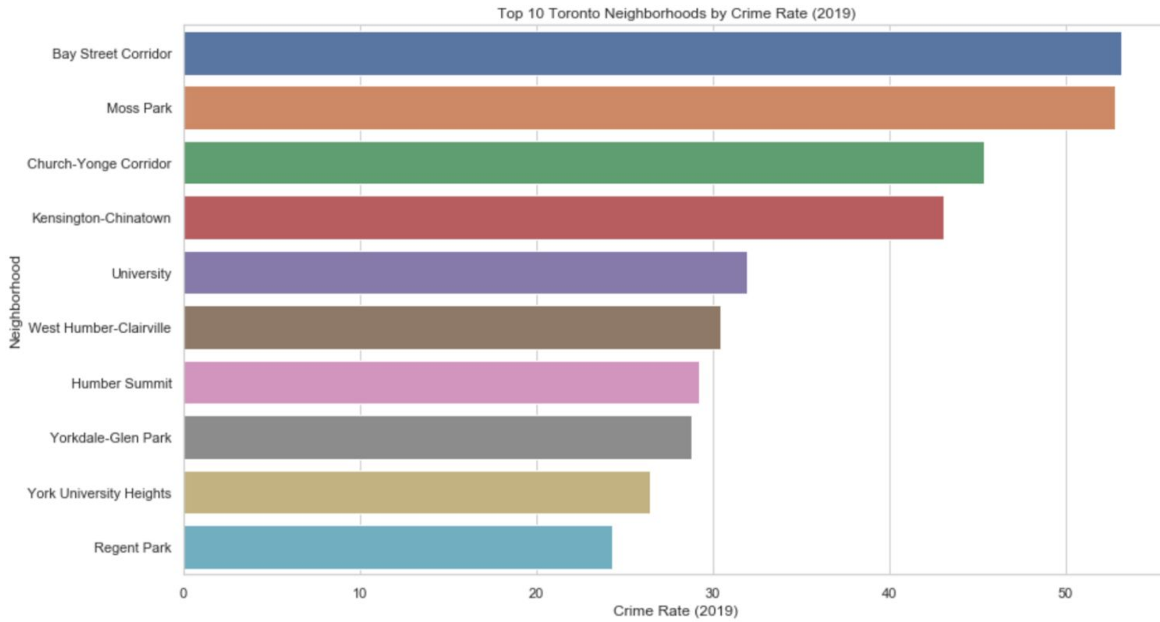


Figure 8

According to **figure 9a**, the neighborhoods with the largest crime rate are Bay Street Corridor, Moss Park, and Church-Yonge Corridor. The top five neighborhoods are basically next to each other in the downtown area. In **figure 9b,** the neighborhoods with the lowest crime rates are Guildwood, Lawrence Park North, and Pleasant View. In contrast to their high-crime counterparts, Guildwood and Lawrence Park North are far from each other.



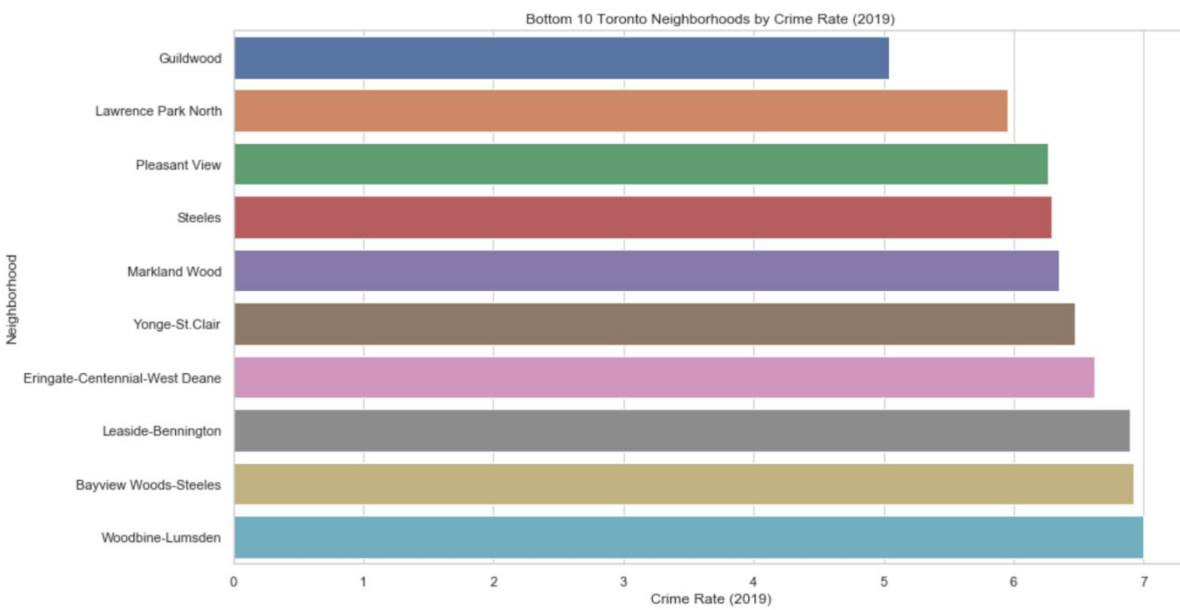


Figure 9

3.2.3. Toronto Venue Data

With 5700 venues with 348 different categories gathered from the Foursquare Places API, it is enough to make a simple map plot quite messy. Alternatively, in **figure 10,** I made a heatmap that shows the density of all the venues over the city. The venues form the API show that downtown Toronto has the most venue density while in areas farther away from downtown have low venue density. Downtown areas have more people than anywhere else in the city, therefore, it is not surprising that downtown neighborhoods have the highest number of venues.

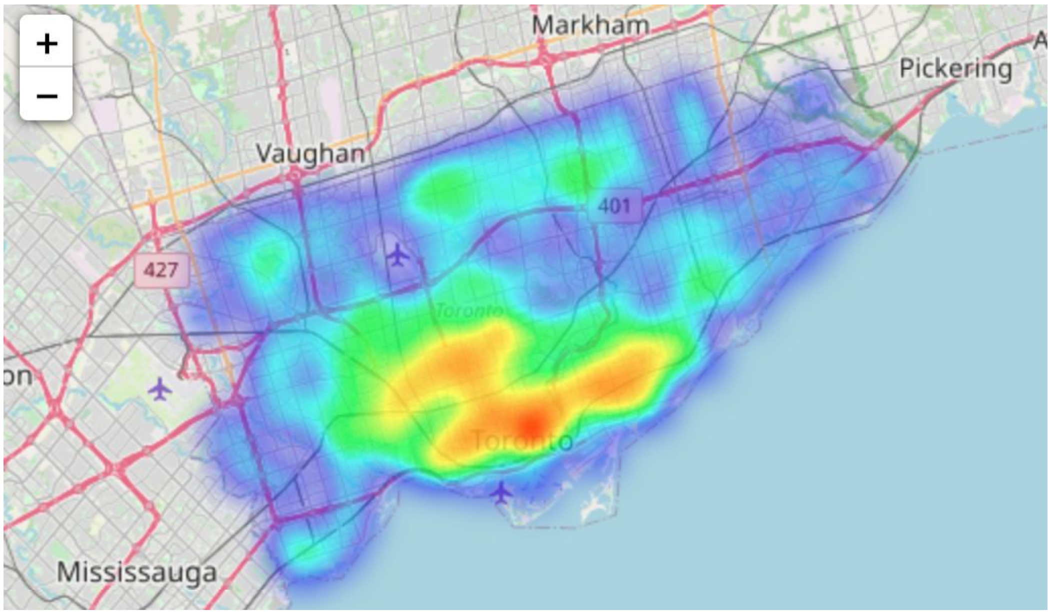


Figure 10

Another look at the Foursquare Toronto venue data is the most common venue categories. Like most cities, the most common venues in Toronto are places that attract locals and tourists alike. Specifically, the most common venues in the data are coffee shops, cafes, parks, pizza places, bakeries, Italian restaurants, etc. **Figure 11** displays the quantity of said venues. The data has 407 coffee shops, 227 cafes, 227 parks, 129 Italian restaurants, and so on. Although there are less than 100 entries for each ethnic restaurant, the clustering algorithm in this project will still help in segmenting neighborhoods to choose the location for a Colombian restaurant.

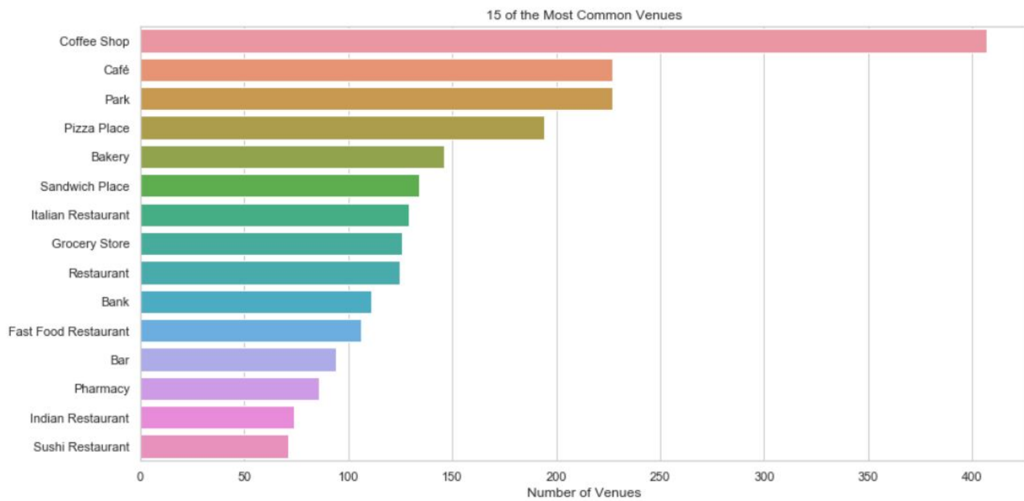


Figure 11

4. Modeling

For each neighborhood, I spread the venue category variable by converting each into dummy variables. This results into an expanded dataset with 349 columns, Then, I grouped the rows by neighborhood by calculating the mean frequency for each category. Since the frequencies are labeled as 1 or 0, the means are actually proportions of venue categories by neighborhood. Now that the neighborhoods are grouped, I merged this dataset with dataset 3, the combined Toronto data. The modeling for this data will account for not only all venue categories but also annual crime rates, income, area and population.

4.1 K-Means Clustering

The modeling of this data consists of implementing an unsupervised learning algorithm named K-means clustering. K-means is a simple and elegant approach for partitioning data into K distinct, non-overlapping segments. To perform this algorithm, we must specify a desired number of clusters *K* so that the algorithm assigns each observation to exactly one of the *K* clusters. Identifying the optimal value of *K* is a fundamental step for partitioning clustering. While there are several ways to do this, a good method is the elbow method. This method looks at the within-cluster sums of squares (WSS) as a function of the number of clusters. In other words, the user chooses a number of clusters such that adding another cluster does not improve the total WSS. The plot of WSS and number of clusters generally displays a bend at the location of the appropriate number of clusters.

Before implementing the clustering algorithm on the grouped Toronto data, I standardized the neighborhood shape area, income, population, and crime rate since they have different scales. The rest of the columns are the 351 venue dummy variables, which do not need to be standardized. If these variables are otherwise standardized, then **figure 12** would not be an elbow plot, but rather a different kind of line indicating non clusterable data. The plot indicates that the within-cluster sum of squares decreases significantly until the number of clusters hits 5. After that, the decreases are less significant, indicating five as the optimal number of clusters for the K-means algorithm on the neighborhood data.

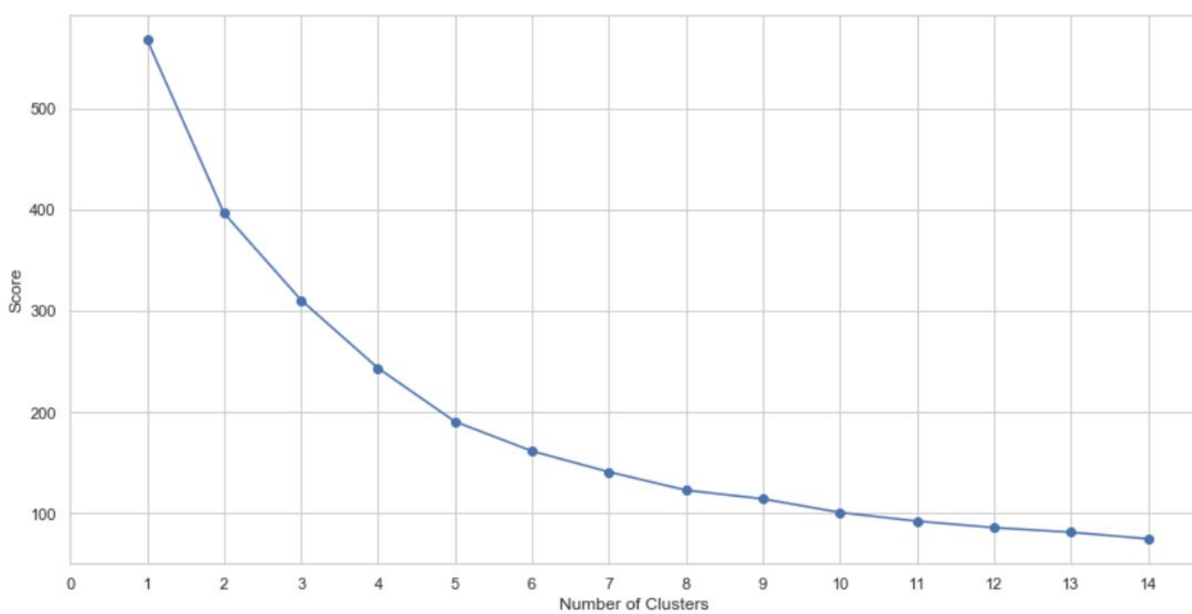


Figure 12

After creating the five clusters on the neighborhood data, I attached the cluster column to the grouped dataset with the venue proportions, geographical coordinates, and the additional variables in their original scales. For this analysis, I added 1 to the cluster indices for identification purposes. For instance, cluster 0 would be cluster 1, cluster 1 would be cluster 2, and so on. To understand the clusters, I plotted the number of neighborhoods within each cluster in **figure 13.** This bar plot shows that cluster 1 is the largest, with 81 neighborhoods. The second largest is cluster 4, with 39 neighborhoods, and the other three have less than 10.

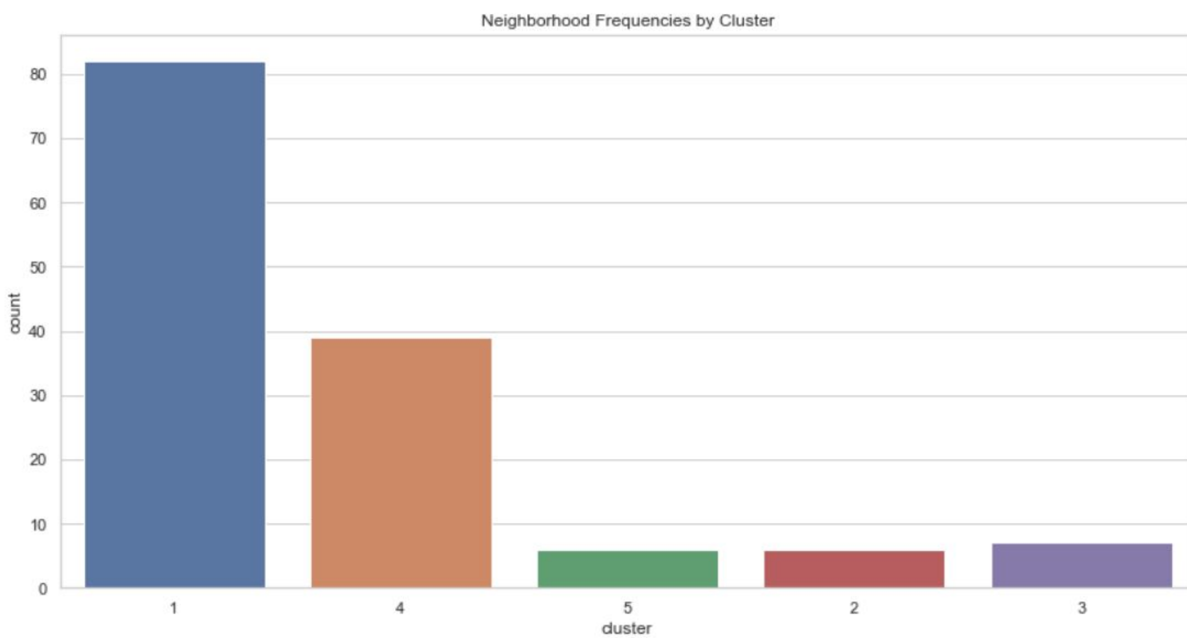


Figure 13 - Number of neighborhoods by cluster

With respect to venues, the cluster sizes may be similar compared to neighborhood size. Indeed, in **figure 14,** the bar plots show that cluster 1 has the most venues. Like **figure 13**, cluster 4 has the second largest number of venues. As cluster 3 is the third largest and cluster 5 is the smallest.

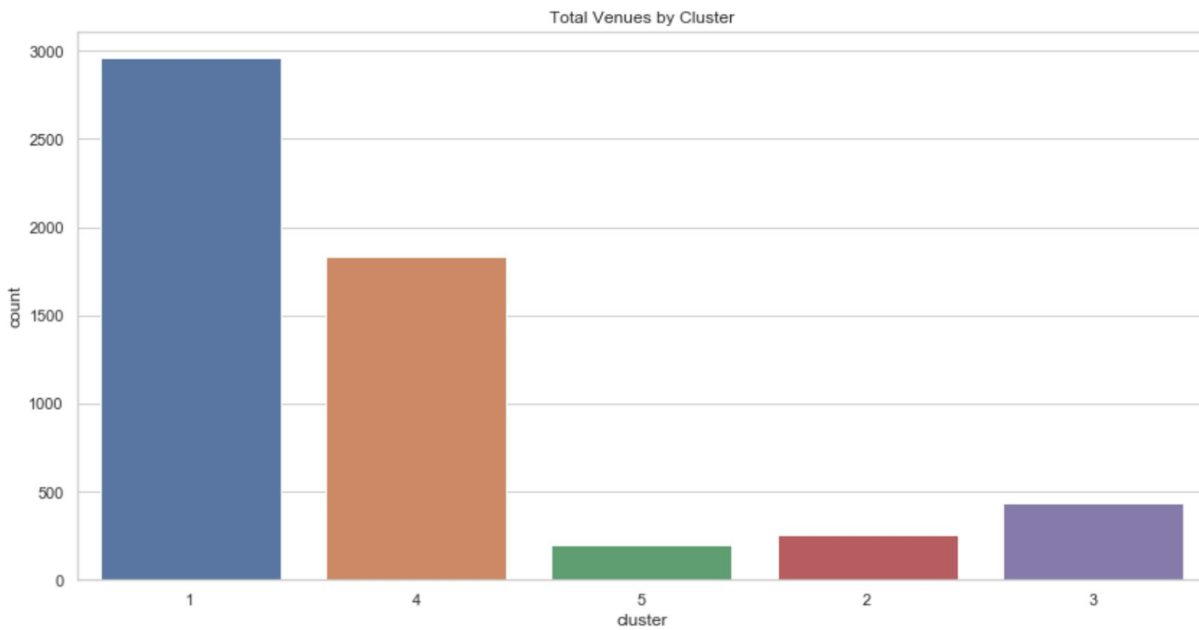


Figure 14 - Number of venues by cluster

4.2 Clustering Results

For each neighborhood, I wrote a function to sort venues in descending order. This is a technique from the capstone class that would bring a better picture for each neighborhood. In other words, this function would output a dataframe with a neighborhood and 10 of their most common venues. The first 5 neighborhoods and 5 most common places are in **table 3.**

Neighborhood	**1st most common**	**2nd most common**	**3rd most common**	**4th most common**	**5th most common**
Agincourt North	Chinese Restaurant	Bakery	Indian Restaurant	Vietnamese Restaurant	Pharmacy
Agincourt South - Malvern West	Chinese Restaurant	Cantonese Restaurant	Asian Restaurant	Restaurant	Pool Hall
Alderwood	Gas Station	Convenience Store	Pizza Place	Park	Pharmacy
Annex	Italian Restaurant	Coffee Shop	Cafe	Restaurant	Bakery

Banbury - Don Mills	Restaurant	Coffee Shop	Cafe	Women's Store	Japanese Restaurant

Table 3 - The first five rows of the most common venues by neighborhood

The map in **figure 15** shows the five clusters across the Toronto area. The cluster colors in this map are the same as those in **figure 13.** In other words, cluster 1 is in blue, cluster 2 in red, cluster 3 in purple, 4 in orange, and 5 in green. While the neighborhoods in clusters 1 and 4 are abundant, the neighborhoods from the other three clusters are very much scattered across the city. We could assume at this point that the best neighborhood for a Colombian restaurant in Toronto is more likely from clusters 1 and 4 than any of the other clusters.

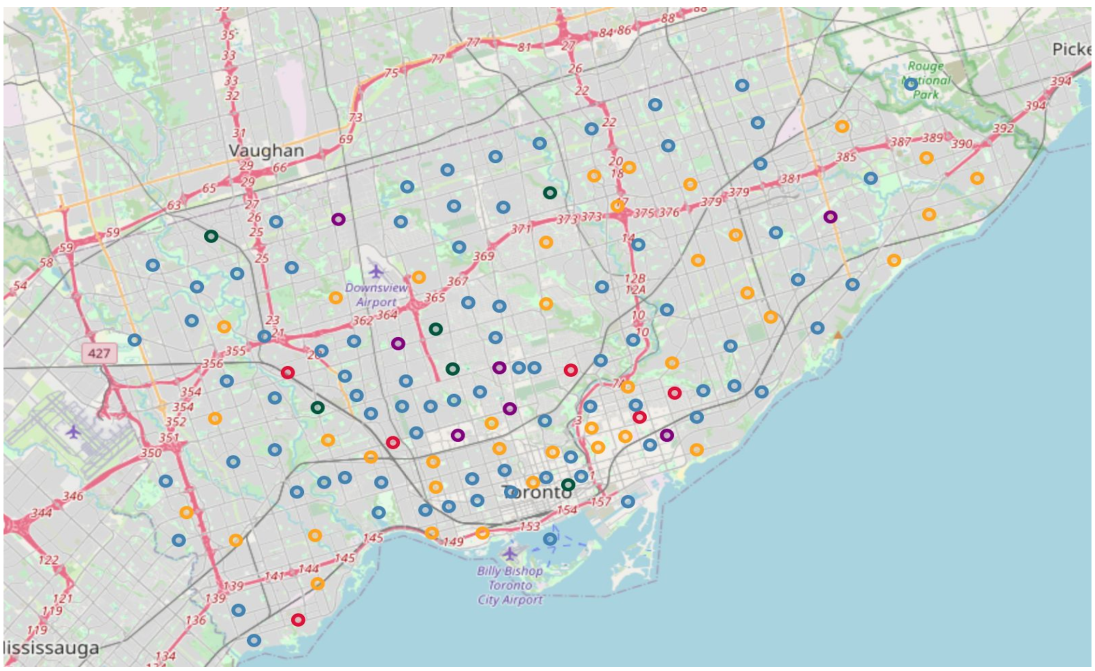


Figure 15 - Map of all neighborhoods by cluster

To proceed with cluster analysis, I attached each neighborhood's corresponding cluster to the data frame representing **table 3**. Then, I split it by cluster label to get ethic restaurant insights in each. The clusters are portrayed in the following tables.

4.2.1 Cluster 1

Neighborhood	1st most common	2nd most common	3rd most common	4th most common	5th most common
Agincourt North	Chinese Restaurant	Bakery	Indian Restaurant	Vietnamese Restaurant	Pharmacy
Agincourt South - Malvern West	Chinese Restaurant	Cantonese Restaurant	Asian Restaurant	Restaurant	Pool Hall
Alderwood	Gas Station	Convenience Store	Pizza Place	Park	Pharmacy
Banbury - Don Mills	Restaurant	Coffee Shop	Cafe	Women’s Store	Japanese Restaurant
Bathurst Manor	Park	Sports Bar	Playground	Hardware Store	Baseball Field

Table 4 - Cluster 1 (5 of 82 neighborhoods)

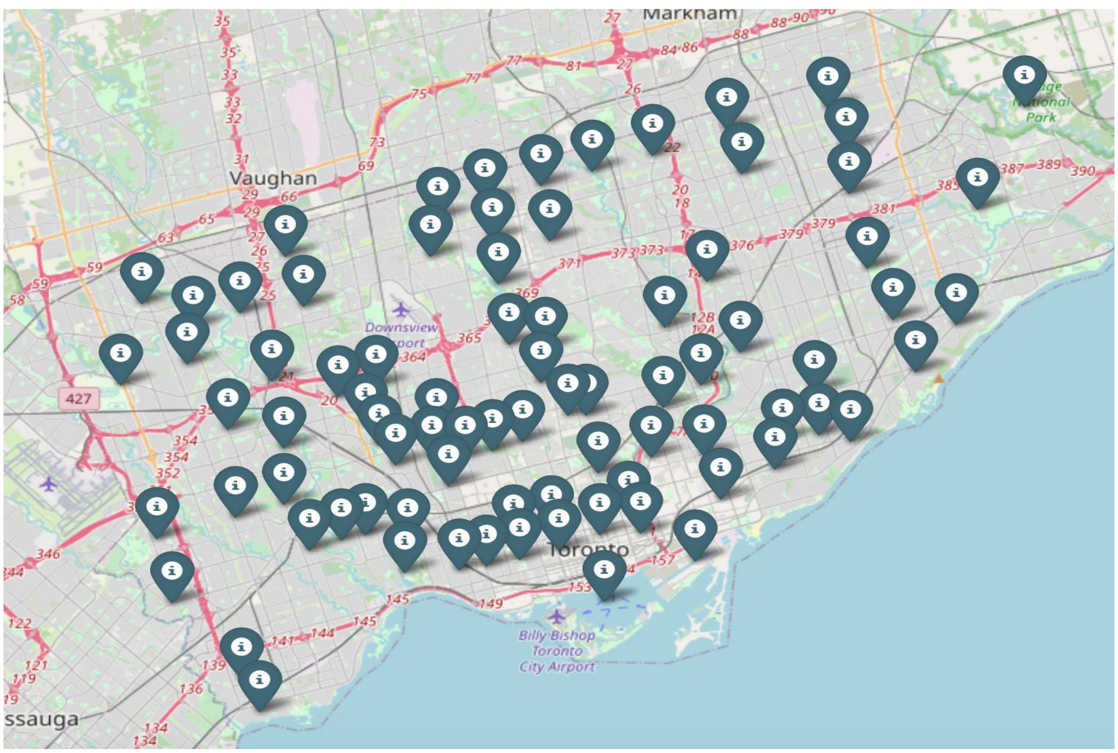


Figure 16 - Neighborhood locations from cluster 1

4.2.2 Cluster 2

Neighborhood	1st most common	2nd most common	3rd most common	4th most common	5th most common
Danforth	Coffee Shop	Cafe	Gastropub	Ethiopian Restaurant	Pizza Place
Leaside - Bennington	Bakery	Indian Restaurant	Burger Joint	Sushi Restaurant	Sporting Goods Shop
New Toronto	Park	Coffee Shop	Sushi Restaurant	Bakery	Skating Rink
Weston	Train Station	Pizza Place	Coffee Shop	Diner	Laundromat
Weston-Pellam Park	Italian Restaurant	Burger Joint	Cafe	Mexican Restaurant	Restaurant
Woodbine - Lumsden	Grocery Store	Park	Pizza Place	Pharmacy	Coffee Shop

Table 5

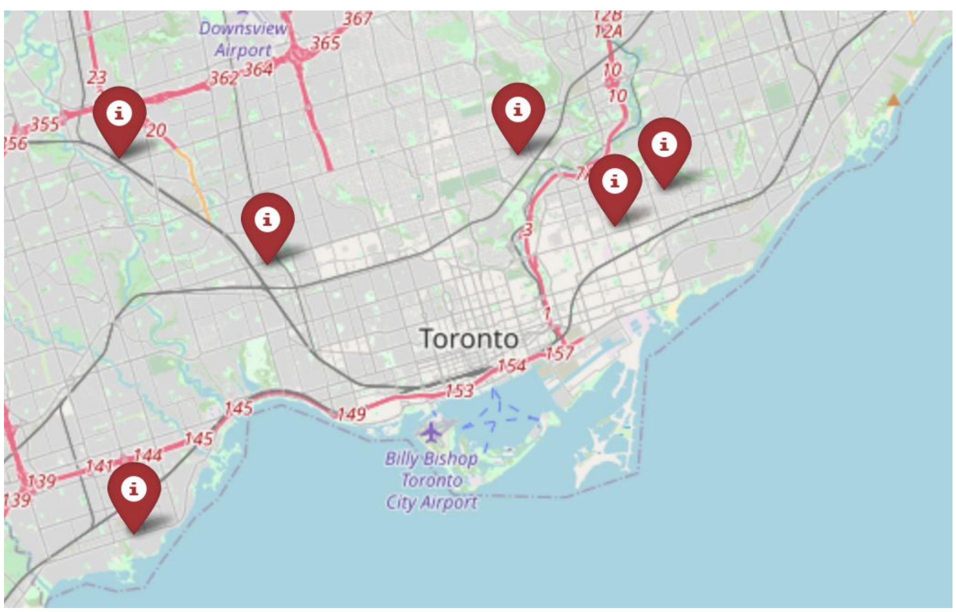


Figure 17 - Neighborhood locations in cluster 2

From the map, the neighborhoods from cluster 3 are located outside the downtown Toronto area.

4.2.3 Cluster 3

Neighborhood	**1st most common**	**2nd most common**	**3rd most common**	**4th most common**	**5th most common**
Woburn	Coffee Shop	Fast Food Restaurant	Bank	Pharmacy	Discount Store
Woodbine Corridor	Indian Restaurant	Cafe	Park	Grocery Store	Coffee Shop
Wychwood	Coffee Shop	Cafe	Italian Restaurant	Grocery Store	Indian Restaurant
Yonge-Eglinton	Coffee Shop	Italian Restaurant	Sushi Restaurant	Cafe	Bakery
Yonge - St. Clair	Coffee Shop	Sushi Restaurant	Italian Restaurant	Thai Restaurant	Park
York University Heights	Pizza Place	Furniture Store	Coffee Shop	Sports Bar	Shopping Mall
Yorkdale - Glen Park	Fast Food Restaurant	Restaurant	Grocery Store	Coffee Shop	Dessert Shop

Table 6

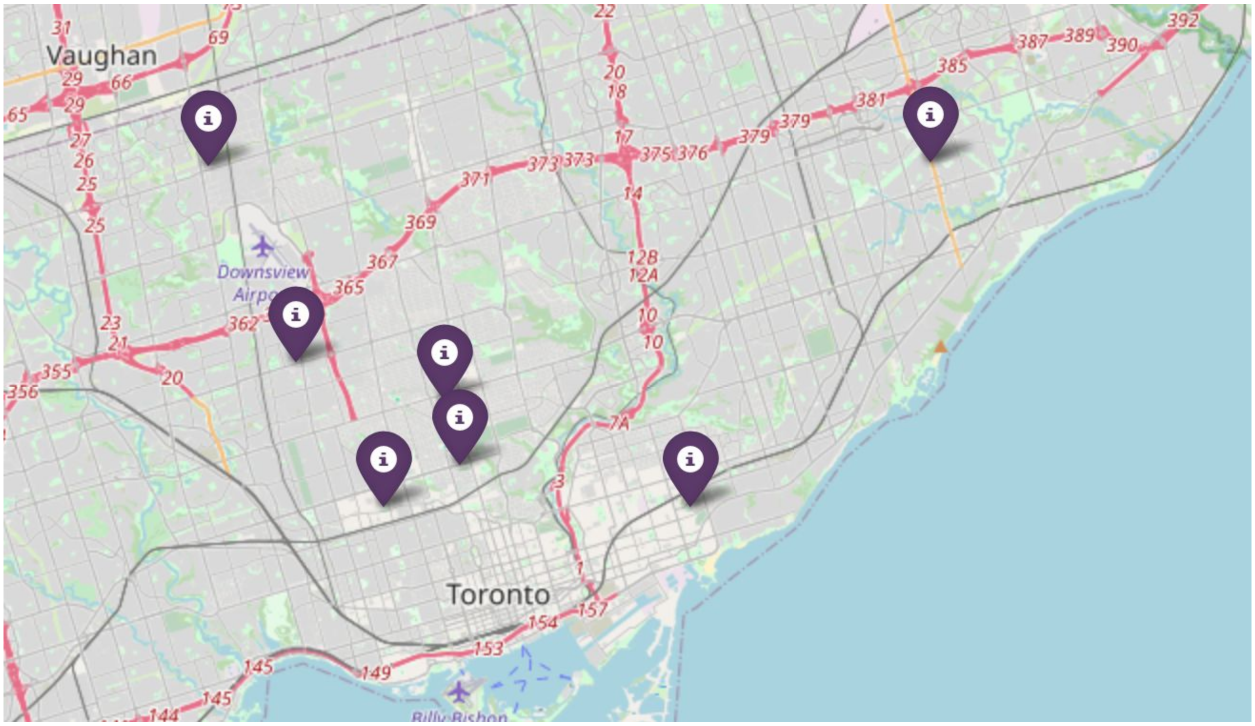


Figure 18 - Neighborhood locations in cluster 3

4.2.4 Cluster 4

Neighborhood	1st most common	2nd most common	3rd most common	4th most common	5th most common
Annex	Italian Restaurant	Coffee Shop	Cafe	Restaurant	Bakery
Bay Street Corridor	Coffee Shop	Cafe	Clothing Store	Sushi Restaurant	Park
Blake-Jones	Cafe	Coffee Shop	Greek Restaurant	Fast Food Restaurant	Pizza Place
Bridle Path-Sunnybrook	Cafe	Restaurant	Bookstore	Coffee Shop	Trail
Casa Loma	Coffee Shop	Sandwich Place	Cafe	Park	Mexican Restaurant

Table 7 - The first five out of 39 neighborhoods in cluster 4

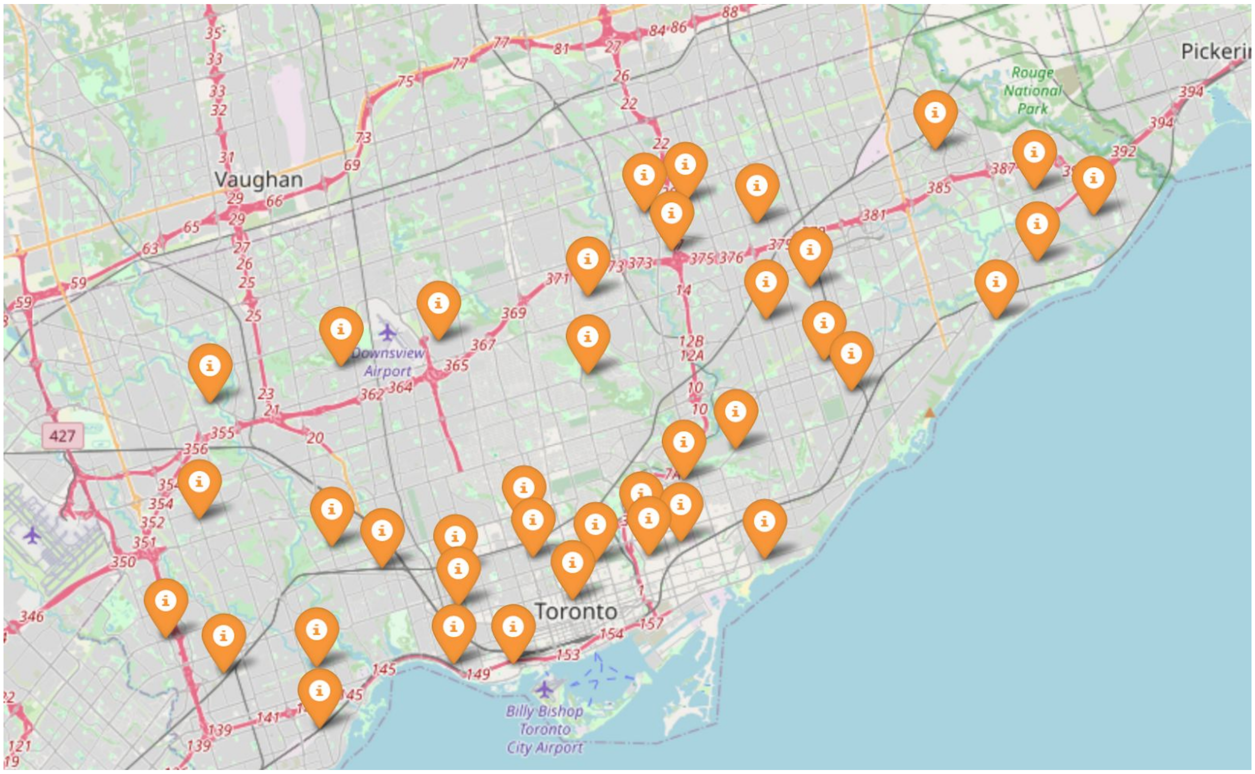


Figure 19 - Neighborhood Locations for cluster 4

4.2.5 Cluster 5

Neighborhood	1st most common	2nd most common	3rd most common	4th most common	5th most common
Bayview Village	Park	Sporting Goods Shop	Outdoor Supply Store	Fish Market	Metro Station
Englemount - Lawrence	Coffee Shop	Fast Food Restaurant	Flower Shop	Gas Station	Sushi Restaurant
Forest Hill North	Bank	Pizza Place	Coffee Shop	Trail	Middle Eastern Restaurant
Humber Summit	Gym	Arts & Crafts Store	Latin American Restaurant	Skating Rink	Pizza Place
Moss Park	Coffee Shop	Cafe	Restaurant	Gastropub	Theater
Mount Dennis	Furniture Store	Coffee Shop	Pizza Place	Hockey Arena	Golf Course

Table 8

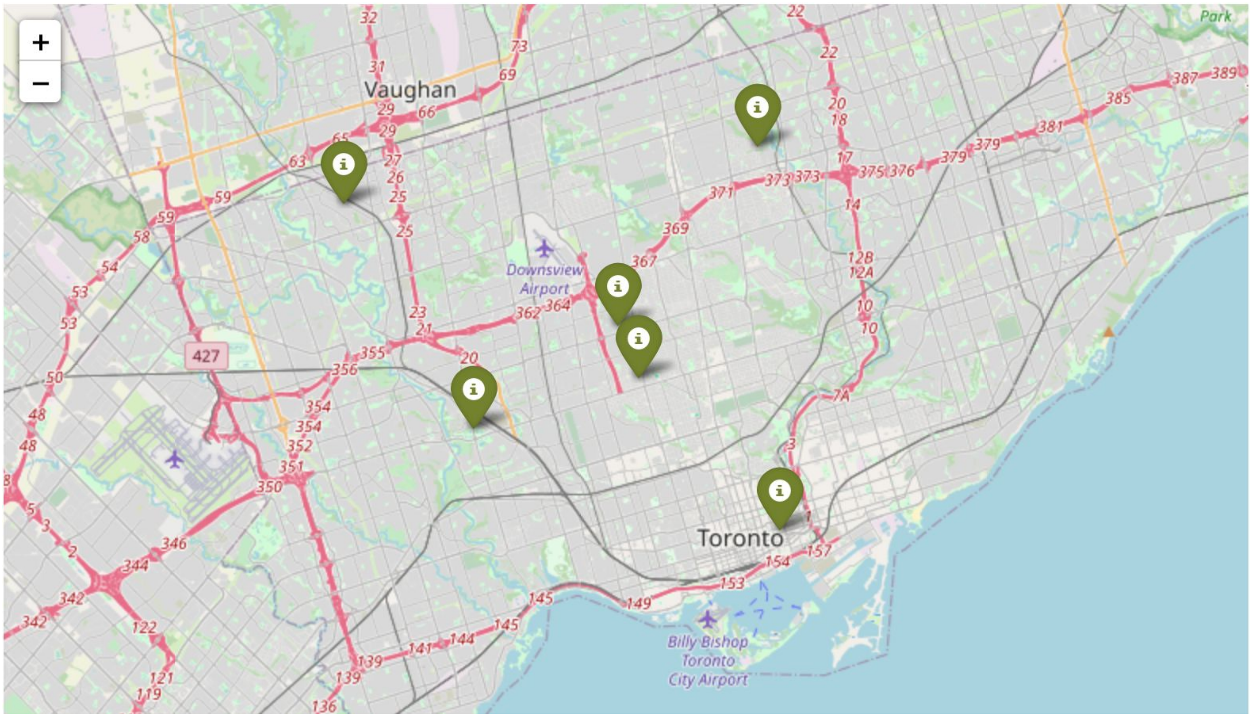


Figure 20 - Neighborhood locations for cluster 5

4.3 Restaurant Search

Since the five clusters represent the neighborhoods, the next step is to look inside the clusters for our restaurant location. My method to determine the best neighborhoods for a Colombian restaurant is to focus mainly on ethnic restaurants, especially Latin American Restaurants. While locations near similar restaurants can help with business marketing and customer attraction, it may also be a challenge to gain a foothold in a community. Since Toronto is one of the most diverse cities in the world, the target market is for everyone regardless of origin and identity. Therefore, the criteria for neighborhood location for our restaurant are not only areas where Spanish is the second most common language after English but also areas with restaurants serving other international cuisines such as Vietnmaese, Japanese, Indian, Mediterranean, etc. This way, the restaurant would have a lot of potential in customer growth with tourists and locals.

4.3.1 Latin American Restaurants in Clusters

Before looking at other restaurant venues, let's look at the locations of existing Colombian restaurants. Although there is no 'Colombian Restaurant' venue category in the API dataset, both Latin American and South American restaurants could sell Colombian cuisine as well as from other countries such as Chile, El Salvador, and Peru. The map in **figure 21** shows location pins on restaurants labelled as 'Latin American' or 'South American' as well as all international restaurants across Toronto. The clusters are colored according to the previous maps in this report. According to the picture, cluster 1 has 3 Latin American restaurants, cluster 3 has 4, cluster 4 and cluster 5 have one. To narrow this down, let's take a look at the neighborhoods that have these restaurants among their top 10 venue categories in the five clusters of the previous section.

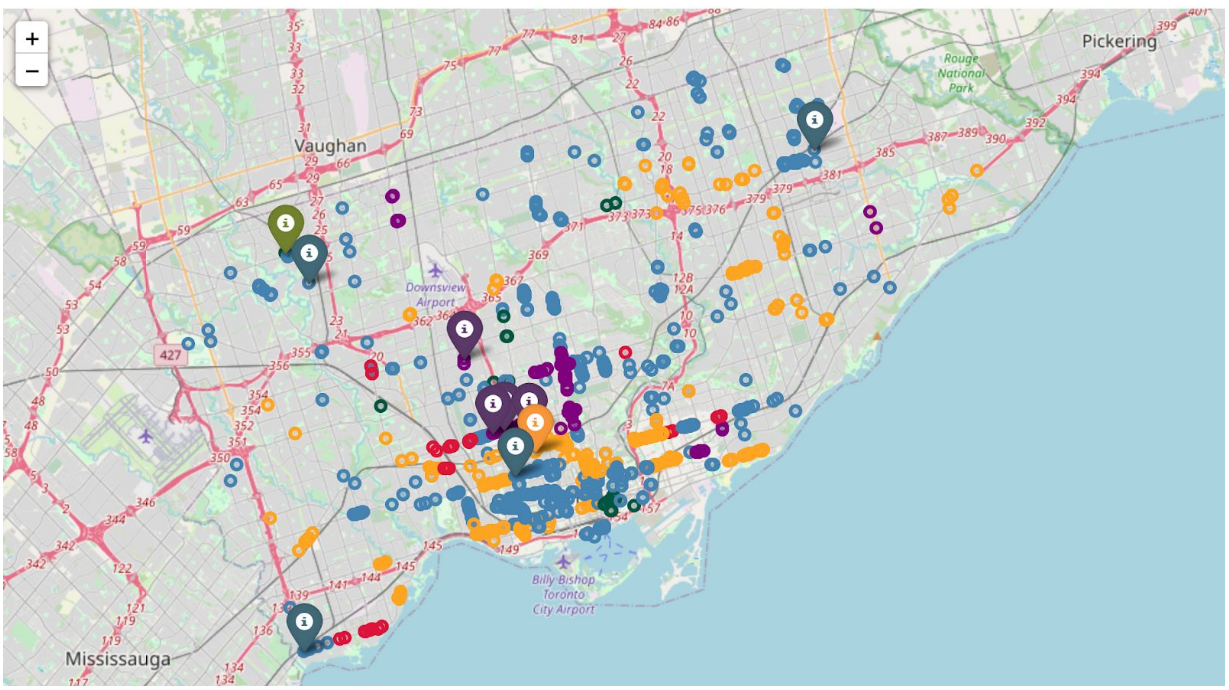


Figure 21

Of the five clusters, only 1 and 5 have at least one neighborhood where Latin or South American restaurants rank among the 10 most common venue categories by neighborhood. In cluster 1, Latin American restaurants rank as the 7th most common venue in the neighborhood of Humbermede. Other international cuisines among the top 10 venue categories in this neighborhood are Vietnamese and Caribbean restaurants.

Table 9 - Top 10 venue categories in Humbermede in cluster 1

Neighborhood	Humbermede
1st Most Common Venue	Vietnamese Restaurant
2nd Most Common Venue	Fast Food Restaurant
3rd Most Common Venue	Discount Store
4th Most Common Venue	Pharmacy
5th Most Common Venue	Sandwich Place
6th Most Common Venue	Grocery Store
7th Most Common Venue	Latin American Restaurant
8th Most Common Venue	Nightclub
9th Most Common Venue	Coffee Shop
10th Most Common Venue	Caribbean Restaurant

In cluster 5, Latin American restaurants rank as the 3rd most common venue in Humber Summit. In other words, Latin American is the most common international restaurant venue category in this area. Another international restaurant among the top 10 in Humber Summit is Ethiopian. **Table 10** lists the other most common venue categories in this neighborhood.

Table 10 - Top 10 venue categories in Humber Summit in cluster 5

Neighborhood	Humber Summit

1st Most Common Venue	Gym/Fitness Center
2nd Most Common Venue	Arts & Crafts Store
3rd Most Common Venue	Latin American Restaurant
4th Most Common Venue	Skating Rink
5th Most Common Venue	Pizza Place
6th Most Common Venue	General Entertainment
7th Most Common Venue	Electronics Store
8th Most Common Venue	Escape Room
9th Most Common Venue	Ethiopian Restaurant
10th Most Common Venue	Event Space

After identifying the neighborhoods where Latin American restaurants rank among the top 10 most common venues, I plotted their locations in **figure 22.** The markers represent the restaurant locations in Humbermede and Hummer Summit, even though both have only one restaurant of this category. The circle markers, on the other hand, represent other Latin American restaurants in their respective clusters. The blue markers represent cluster 1 while the green ones represent cluster 5.

The restaurant in Humbermede is named Mi Pueblo, which serves cuisine from multiple Latin American countries including Colombia, such as Colombian Empanadas. The Humber Summit restaurant, on the other hand, is named Plaza Latina. This venue is a mini-mall with a great food court serving all types of authentic Latin American cuisine. In other words, there are many stalls serving Colombian as well as Cuban, Peruvian, Ecuadorian, and Salvadorian food.

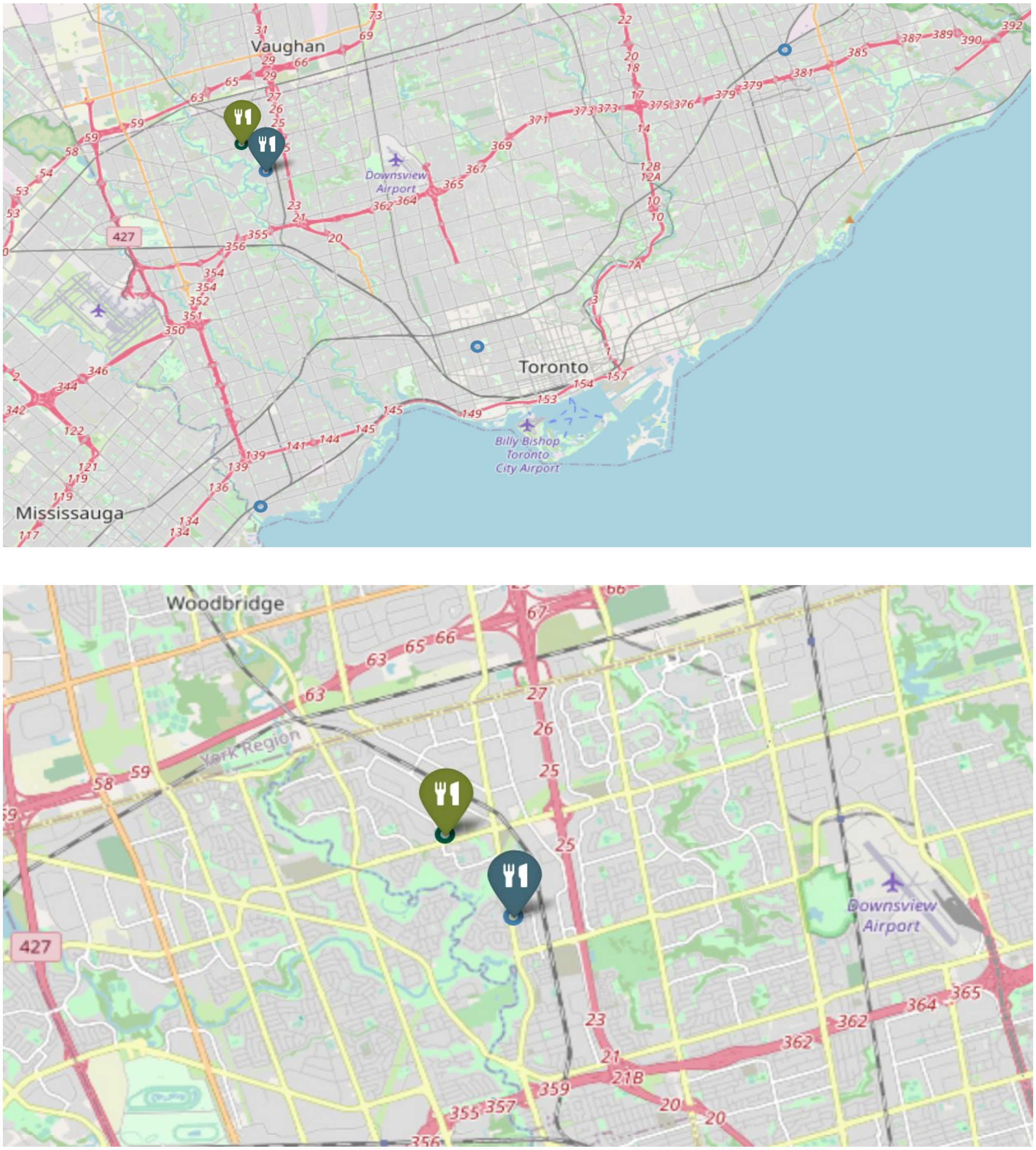


Figure 22

4.3.2 Spanish-Speaking Neighborhoods

In addition to looking at the neighborhoods that have Latin American restaurants among their 10 most common venues, I looked into neighborhoods where Spanish is the most spoken language after English. According to a Toronto website there are 20 such neighborhoods such as Humbermede, Weston, Casa Loma, Glenfield-Jane Heights, etc. For these neighborhoods, I looked at what other international restaurants operate, where are Latin American restaurants located, if any, as well as their average household income and crime rate statistics. These factors would help decide up to five locations for a Colombian restaurant in addition to the two neighborhoods from the last section.

Not all of the 20 neighborhoods have an international restaurant, therefore it is better to look at those that have at least one such restaurant. The bar plot in **figure 23** shows 14 out of 20 neighborhoods where Spanish is mostly spoken has at least one international restaurant. The neighborhood with the most international restaurants in Toronto is Annex. They include Italian, Hungarian, Greek, Japanese, and Thai. This neighborhood has a Latin American restaurant named Gordo Ex. The neighborhood of Mount Dennis has one restaurant named Golden Crisp, an American restaurant.

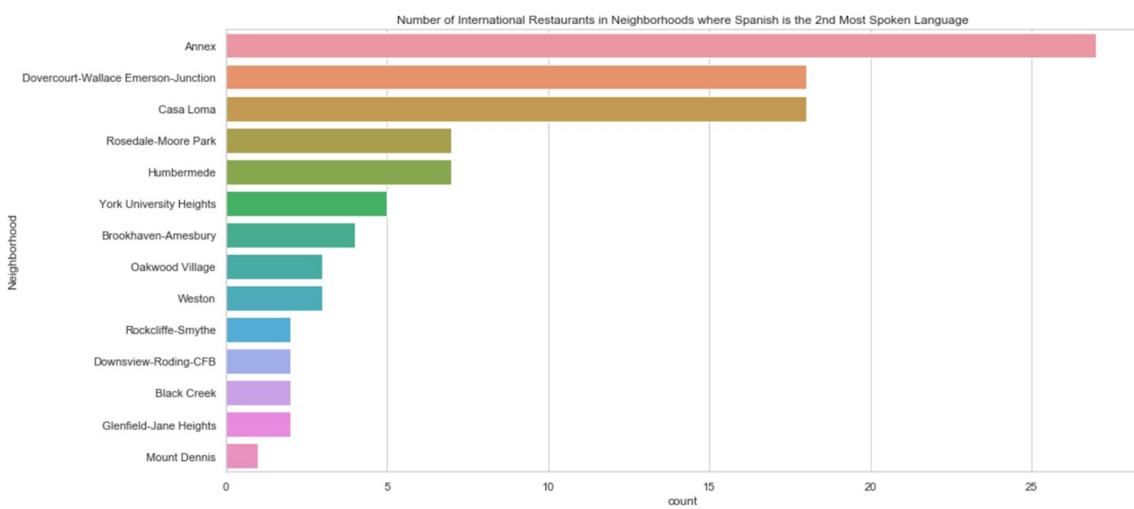


Figure 23

Figure 24 shows the locations of the above neighborhoods as well as restaurant locations around them. Unlike all of the other maps in this analysis, the pins in the map represent the neighborhoods from **figure 23** while the circle markers represent restaurants. All points are colored with respect to clusters. From this map, the top four neighborhoods with the most international restaurants are located in the downtown Toronto area.

Figure 24

Up next, let's look at a few demographics on the remaining neighborhoods. **Figure 25** shows the 2019 crime rate, population, and average household income. According to **figure 25a,** the neighborhood of Dovercourt-Wallace Emerson-Junction is the most populated in the Spanish-speaking group while Elms-Old Rexdale is the least populated. In **figure 25b,** average household income is the highest in Rosedale-Moore Park. This neighborhood's income of $1,400,000 is significantly larger than that of the second largest income of $800,000 in Annex. The neighborhood with the lowest average income is Elms-old Rexdale with $123,119. Finally, the 2019 crime rate is highest in York University Heights with over 2500 crimes per 100,000 people. The lowest, on the other hand, is in Casa Loma with less than 1000 incidents per 100,000 people.

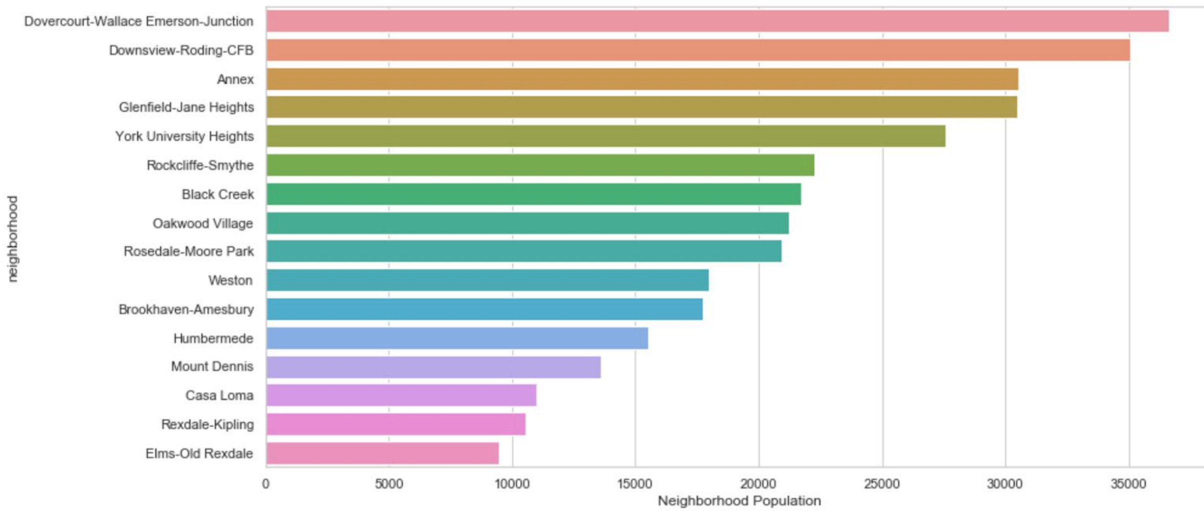


Figure 25a - Population in Spanish-speaking neighborhoods

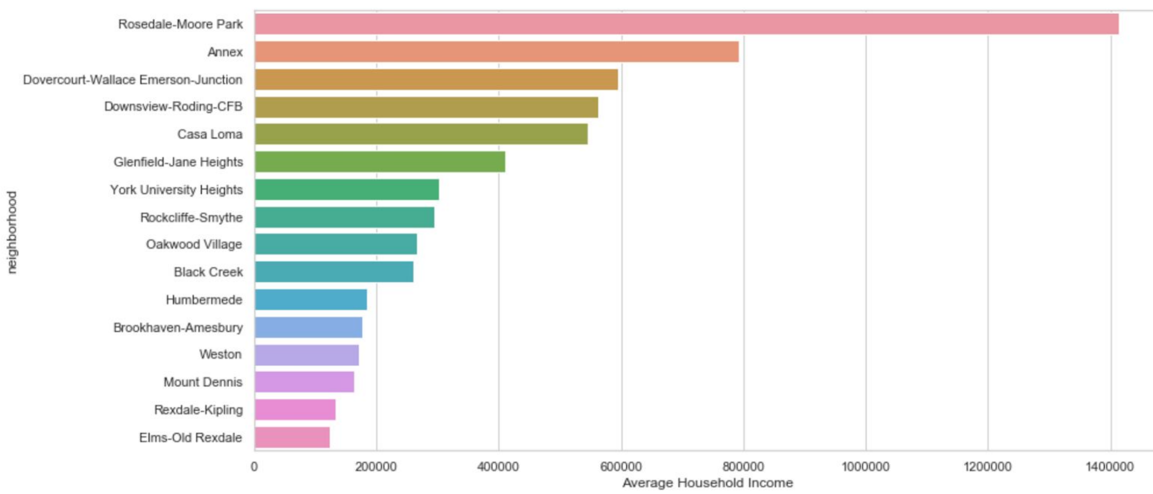


Figure 25a - Average household income in Spanish-speanking neighborhoods

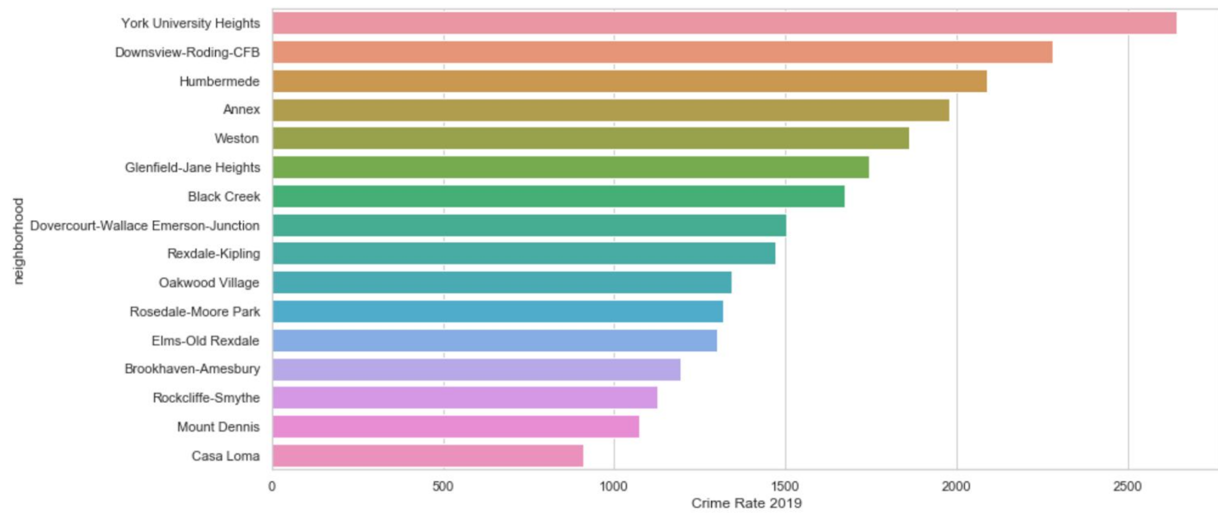


Figure 25b - 2019 crime rate by 1,000 people in Spanish-speaking neighborhoods

4.4 Final Results

With the figures and maps from **section 4.3** I managed to narrow down the list of potential neighborhoods to five. The neighborhoods include Annex, Dovercourt-Wallace Emerson-Junction, Humbermede, Casa Loma, and Rosedale-Moore Park. These are listed for the following reasons: Annex has the most international cuisine venues, Dovercourt-Wallace Emerson-Junction is the most populated neighborhood in the group, Humbermede is the Spanish-speaking neighborhood featured in the cluster in **section 4.3.1,** Casa Loma has the lowest crime rate, and Glenfield-Jane Heights has the most highest population of Spanish speakers in Toronto. The five neighborhoods are organized in **table 11** with their most common venues and cluster label.

Neighborhood	Annex	Glenfield-Jane Heights	Humbermede	Casa Loma	Dovercourt-Wallace Emerson-Junction
1st Most Common	Italian Restaurant	Vietnamese Restaurant	Vietnamese Restaurant	Coffee Shop	Bar

Venue					
2nd Most Common Venue	Coffee Shop	Tea Room	Fast Food Restaurant	Sandwich Place	Coffee Shop
3rd Most Common Venue	Cafe	Coffee Shop	Discount Store	Cafe	Bakery
4th Most Common Venue	Restaurant	Bank	Pharmacy	Park	Cafe
5th Most Common Venue	Bakery	Grocery Store	Sandwich Place	Mexican Restaurant	Pharmacy
6th Most Common Venue	Japanese Restaurant	Plaza	Grocery Store	Liquor Store	Park
7th Most Common Venue	French Restaurant	Moving Target	Latin American Restaurant	Pizza Place	Mexican Restaurant
8th Most Common Venue	Grocery Store	Food & Drink Shop	Nightclub	Bank	Restaurant
9th Most Common Venue	Vegetarian Restaurant	Pizza Place	Coffee Shop	Grocery Store	Portuguese Restaurant
10th Most Common Venue	Mexican Restaurant	Gym	Caribbean Restaurant	Ice Cream Shop	Italian Restaurant
Number of Int. Restaurants	27	2	7	18	18

Table 11

Looking at **table 11** as well as the graphs in **figure 25**, the three neighborhoods with the most international restaurants are Annex, Casa Loma, and Dovercourt-Wallace Emerson-Junction. These three belong to cluster 4 and are among the top 5 neighborhoods with largest average household income for Spanish-speaking neighborhoods. **Table 12** shows that only Annex and Casa Loma have at least one Latin American restaurant in the area. Although Annex has more restaurants as well as higher income level and population, Casa Loma has a lower crime rate. According to **figure 26**, Casa Loma has significantly less assault and property crime incidents. Therefore, since it is important to acknowledge safety issues and crime rates when choosing a restaurant location, the first choice for a Colombian restaurant is Casa Loma.

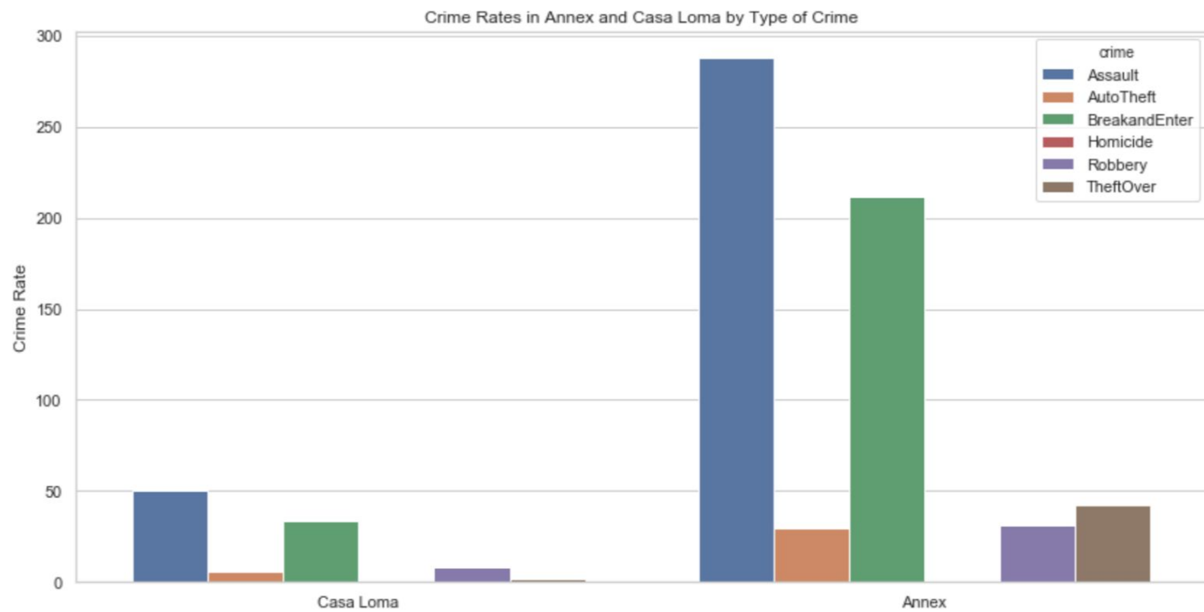


Figure 26 - Specific crime rates in Annex and Casa Loma

5. Discussion and Conclusion

In this analysis, I gathered four Toronto datasets from different sources to compare and find a neighborhood to open a Colombian restaurant in the city. In addition to neighborhood venue data, I used neighborhood population, crime rate, and average household income to cluster the neighborhoods as a method to narrow down locations. After wrangling the venue data, I joined it with standardized versions of the additional demographics variables to cluster the data in addition to finding the optimal number of clusters. While sourcing through the clusters, an important part was to list the neighborhoods where Spanish is the second most spoken language. Although two neighborhoods from two different clusters stood out by including Latin American restaurants among their 10 most common venues, one was in the list of Spanish-speaking neighborhoods. However this neighborhood did not stand out in terms of demographics, so I included four other neighborhoods that did for each variable. Humbermede stood as fourth place in the top five neighborhoods for the restaurant due its number of international cuisine venues. Although Casa Loma does not include Latin American restaurants among its 10 most common venues, it has one named Gourmet Gringos. Moreover, this neighborhood's venues include Mexican, Indian, French, Japanese, and Middle Eastern restaurants. This makes Casa Loma a great location in terms of diversity, where customers from different origins can eat. Even though the K-means clustering algorithm helped narrow down at least one neighborhood for the top 5 locations, this proved to be quite a useful method in narrowing restaurant locations down with data.

References:

1 -
https://torontoglobal.ca/TG-Blog/March-2019/How-multicultural-is-Toronto-Let-us-count-the-way

2 - https://www.visaplace.com/canadian-immigration/country/colombia/

3 - https://open.toronto.ca/dataset/neighbourhoods/

4 - https://open.toronto.ca/dataset/neighbourhood-profiles/

5 -
https://www.thebalancesmb.com/how-to-find-out-the-population-base-of-an-area-2888660

6 -

https://www.datanovia.com/en/lessons/determining-the-optimal-number-of-clusters-3-must-know-methods/

7 - https://www.webstaurantstore.com/article/81/restaurant-environmental-analysis.html

8 - https://www.zomato.com/toronto/mi-pueblo-north-york-gta

9 - https://www.blogto.com/eat_drink/2011/10/inside_plaza_latina/