On Most Common and Trendy Venues in Toronto and New York

# Introduction to the Business Problem

Cross-border investment is becoming ever more important for economic prosperity. For the United States and Canada, a significant number of normal barriers to trade are non-existent due to the similarities in these two countries’ culture, language, and legal frameworks. Without these barriers, entrepreneurs from either country looking to invest in the other can dedicate more of their resources on identifying profitable opportunities rather than overcoming trade barriers.

The goal of the project was to generate useful analytical information for entrepreneurs to identify profitable investment opportunities in Toronto and New York. The project aimed to identify the most popular venues in the two cities, and if significant overlap existed in the types of venues that are popular in both places. By identifying this information, the project could assist entrepreneurs to identify business opportunities for investment in these cities, as well as providing insight into whether the targets of their investment are profitable or not.

# Data Used

The data used in the project came from three sources: the New York University, the City of Toronto, and Foursquare. The data from these sources were used to obtain the lists of neighborhoods in Toronto and New York, as well as the coordinates of the neighborhoods and trending venues.

The data on neighborhoods and GIS coordinates of these neighborhoods were provided by the City of Toronto and New York University, respectively. The data for Toronto comes from the Neighborhood Profiles dataset provided by the City of Toronto, which can be found at the address <https://open.toronto.ca/dataset/neighbourhood-profiles/>. The New York Neighborhood Names dataset comes from the NYU Spatial Data Repository, which can be found at <https://geo.nyu.edu/catalog/nyu_2451_34572> .

The data on trending venues in each of the cities is provided by Foursquare API. The GIS coordinates of each neighborhood are passed through the API to obtain the most trending venues in each neighborhood. With this data, statistical analysis can be conducted on the number of different types of venues in each neighborhood.

# Analytical Methods

To identify business insights that are useful for entrepreneurs, both traditional statistics techniques, as well as machine learning algorithms, were employed in the analysis. The traditional statistics techniques, such as summary statistics on the number and type of venues, were used to obtain high-level insights such as the types of popular venues, and to approximate their relative popularity in each city. Machine learning algorithms were also employed to group neighborhoods in Toronto and New York with similar characteristics together therefore potential opportunities for investors to open similar venues.

To identify the higher-level similarities, summary statistics were used to identify the most popular venues in both cities. Techniques used include ranking venues by their frequency both on the municipal level as well as on the neighborhood levels. For example, by comparing the frequency of cafes in the two cities, it was possible to determine if cafes were popular and therefore provided good investment opportunities in both cities. More broadly, the information could be used to identify the general areas that are favorable for investors.

For more in depths analysis at the neighborhood level, clustering algorithms were used to cluster neighborhoods with similar characteristics together. Two machine learning algorithms, K-Means and DBSCAN, were used. By identifying neighborhoods that share the most common characteristics, entrepreneurs that are successful in one neighborhood can identify similar business opportunities in the other neighborhoods in the same cluster. For instance, if the boroughs of Downtown Toronto and Manhattan were clustered together, the information could be further analyzed to determine if investment opportunities exist that can leverage the similarities of both boroughs.

# Analysis Results

## Most Numerous Venues in Toronto

Table 1: most common venues in Toronto and New York

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Venue Category** | **Toronto Venues** | **Toronto Percentages** | **New York Venues** | **New York Percentages** | **Venue Total** |
| Pizza Place | 76 | 7.55% | 443 | 2.97% | 519 |
| Coffee Shop | 155 | 4.14% | 291 | 1.81% | 446 |
| Italian Restaurant | 57 | 4.04% | 298 | 1.69% | 355 |
| Park | 85 | 3.70% | 177 | 4.53% | 262 |
| Bakery | 38 | 2.97% | 216 | 1.87% | 254 |
| cafes© | 83 | 2.78% | 165 | 3.04% | 248 |
| Bar | 42 | 2.19% | 203 | 2.00% | 245 |
| Sandwich Place | 61 | 2.05% | 183 | 2.07% | 244 |
| Grocery Store | 45 | 1.85% | 196 | 2.21% | 241 |
| Chinese Restaurant | 27 | 1.80% | 202 | 1.54% | 229 |
| Pharmacy | 35 | 1.71% | 171 | 1.75% | 206 |
| Bank | 37 | 1.32% | 151 | 2.06% | 188 |

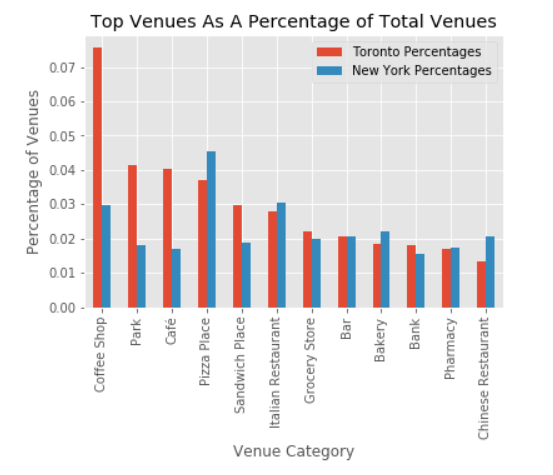
The summary results show the topo 20 most numerous venues that can be found in both cities. Percentages for each venue were calculated using the number of venues out of the total number of venues in each city. Percentages represent the proportion of the venues out of the total number of venues present in each of the cities. As New York was a much larger city with much higher number of venues, percentages were better indicators of the popularity of the types of restaurant in each city than the raw number of venues. The data shows that fast dining establishments are consistently the most numerous. Except for parks, grocery stores, pharmacies, and banks; all the top venue types present are eateries.

Coffee shops and cafes were consistently the most numerous of the establishments, coming 1st and 3rd in terms of numbers in Toronto; as well as 3rd and 11th in New York. Percentagewise, coffee-oriented eateries were much more popular in Toronto. Coffee shops and cafes together comprise 10.33 percent of all venues in Toronto but are only 4.85 percent of all venues in New York.

In addition to coffee shops, fast eateries were also among the numerous venues in Toronto and New York. Pizza places were the most numerous venues in terms of total number of venues. Bakeries and sandwich places were also popular, coming 5th and 8th places respectively in the total number of venues. Together with a large number of cafes and coffee houses present, it could be reasonably concluded that fast eateries were the most numerous types of venues in the two North American cities.

For other dining services, Italian and Chinese restaurants proved to be the most numerous types of higher-class eateries that were in the top 20 most common venues in both cities. Overall, Italian restaurants were more popular than Chinese restaurants, both in terms of the number of venues as well as a percentage of the total venues.

Other than eateries, the most numerous venues in both cities were grocery stores and pharmacies. This was not surprising, as groceries and pharmacies supply essential goods that are required by the populace.



Comparing the percentages of the top venues that were present in the two cities, it could be seen that compared to Toronto, pizza places, Italian and Chinese restaurants, and bakeries were significantly more numerous as a proportion of venues in New York. Pharmacies also had a slightly higher proportion in New York than in Toronto. It could, therefore, be concluded that compared to Toronto, investment in New York should be more oriented towards providing dining services such as restaurants and other eat-in venues.

On the other hand, the proportion of coffee shops and cafes in Toronto was significantly higher than in New York. In addition, sandwich places were proportionally more numerous in Toronto as well. From the existing data, it could be inferred that social venues that provide beverages and side snacks were the more popular choice in Toronto, and therefore investors should target these areas.

## K-Means Clustering

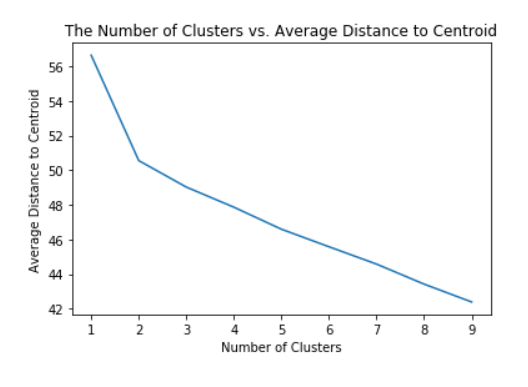


Figure 1: Average distance to the centroid in different clusters

K-means clustering is a powerful machine learning algorithm provided by Scikit-Learn for the Python language. K-means clustering works by creating a centroid, measure the average distance of data points in the cluster to the centroid, and shift the centroid towards the cluster’s “center” with each iteration. Because the average distance between data points to the centroid is always decreasing as the number of clusters increase, it is recognized as best practice to choose the number of clusters that shows a dramatic change in the slope of decrease for the average distance. This is commonly known as the “elbow method”. In the case of neighborhood clustering, this “elbow” occurred when there were only 2 clusters. Therefore, the k-means algorithm was set with n\_clusters=2.

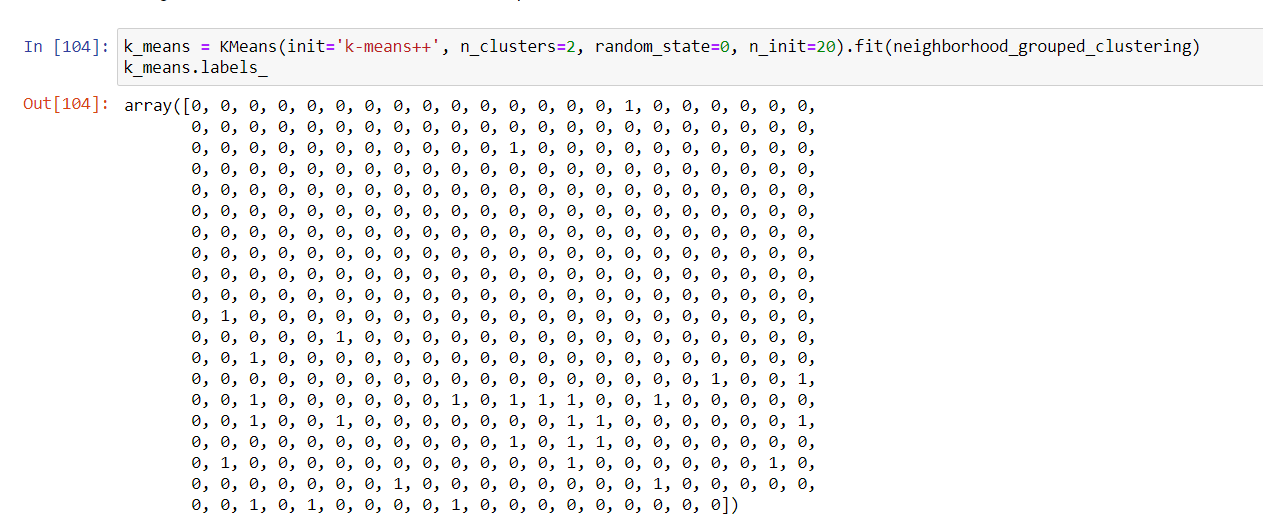


Figure 2: clustering labels from k-means clustering

The result of the k-means clustering algorithm shows a significant number of “0”s and very few “1”s. As an algorithm, K-Means lacks the ability to identify outliers, and therefore places all data into one of the clusters it creates. Because of this, in this case, the “0” cluster was not analyzed, as it was believed this cluster was the equivalent of a “miscellaneous” cluster where unrelated data points are grouped. The cluster with the label “1” was analyzed, as it was believed to be a more the more cohesive cluster.

## Top Venues in K-Means Cluster 1

|  |  |
| --- | --- |
| Venue Category | Venue |
| Park | 41 |
| Playground | 5 |
| Pool | 4 |
| Moving Target | 3 |
| Tennis Court | 3 |
| Bus Stop | 3 |
| History Museum | 3 |
| cafes© | 3 |
| Construction & Landscaping | 3 |
| Dog Run | 2 |

Table 2: most common venues in Cluster 1

Only one coherent cluster could be determined by the algorithm from a table of all venues and their frequencies in different neighborhoods. This cluster comprised of the New York neighborhoods Bayswater, Clason Point, Randall Manor, Somerville, Todt Hill; and the Toronto neighborhoods Bathurst Manor, Bayview Woods-Steeles, Bendale, Brookhaven-Amesbury, Caledonia-Fairbank, Casa Loma, Centennial Scarborough, Clanton Park, Downsview-Roding-CFB, Edenbridge-Humber Valley, Forest Hill South, Glenfield-Jane Heights, Hillcrest Village, Kingsway South, Lambton Baby Point, Lansing-Westgate, Milliken, North Riverdale, Parkwoods-Donalda, Rosedale-Moore Park, Taylor-Massey, West Hill, Westminster-Branson, and Willowdale West. Therefore, this cluster is significantly Toronto biased.

From the result of the top 10 most common venues, it was clear that recreation facilities were by far the most numerous of the venues in these neighborhoods. Parks were the most numerous types of venues in these neighborhoods. Playgrounds, pools, and other sports facilities follow parks in terms of numbers. It could, therefore, be assumed that the residents of these neighborhoods had comparatively more spare time, which they could spend at these recreational facilities. Therefore, investment in these neighborhoods can follow a similar pattern of providing recreational facilities to the residents.

## DBSCAN

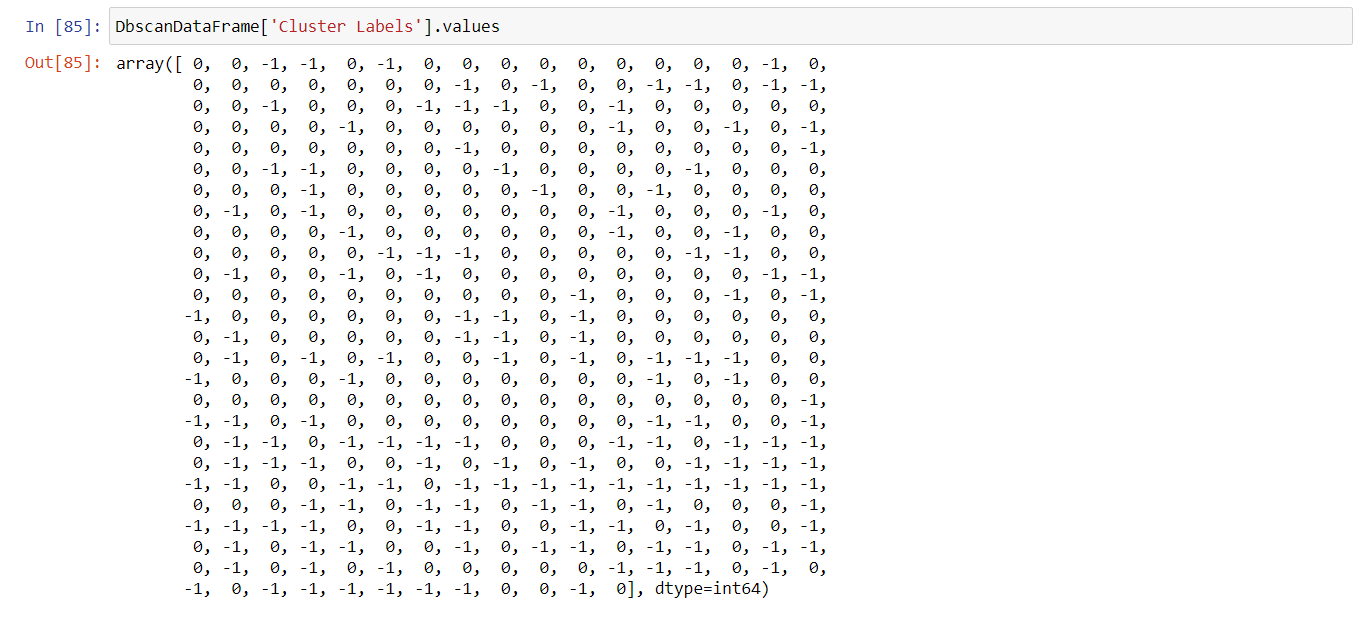


Figure : DBSCAN results

DBSCAN is another powerful machine learning algorithm offered through Scikit-Learn. DBSCAN works by scanning the radius around each data point to determine if a sufficient number of data points are around the central data point for it to be considered to be part of a cluster. For this project, the radius of the scan was set at 0.3, and the threshold for clustering was set at 7 data points.

Compared to K-Means clustering, DBSCAN has the advantage of accounting for irregular shaped clusters and outliers that are often missed by clustering through the K-Means method. Datapoints that were determined by DBSCAN to not be part of any cluster is given the label of “-1”. In this case, DBSCAN determined that there is only one cohesive group (Cluster 0), with the remaining datapoints being outliers.

### Cluster 0

|  |  |
| --- | --- |
| Venue Category | Venue |
| Pizza Place | 519 |
| Coffee Shop | 446 |
| Italian Restaurant | 355 |
| Deli / Bodega | 275 |
| Bakery | 254 |
| Bar | 245 |
| cafes© | 245 |
| Sandwich Place | 243 |
| Grocery Store | 240 |
| Chinese Restaurant | 229 |
| Park | 221 |
| Pharmacy | 206 |
| Bank | 188 |
| Mexican Restaurant | 173 |
| Donut Shop | 171 |
| American Restaurant | 160 |
| Gym / Fitness Center | 143 |
| Ice Cream Shop | 142 |
| Restaurant | 134 |
| Gym | 132 |

Table 3: DBSCAN cluster 0 most common venues

The most common venues in cluster 0 reflect the most common venues overall in the cities of Toronto and New York. Similar to the list of top venues, the top venues in Cluster 0 were all eateries such as pizza and sandwich places, delis and snack places, and different types of restaurants serving a variety of food from different cultures. This was not surprising, as Cluster 0 encompassed the majority of neighborhoods in the datasets, and their venue preferences would be reflected in the higher-level results.

# Discussion

The project demonstrated that the most popular venues in Toronto and New York related to eating and socializing. The most numerous venues in both cities were venues that offered food and beverages, not least the various coffee shops, snack places, and restaurants. It was found there were slight variations in the preferences between Toronto and New York. In Toronto, coffee houses and sandwich places were, as a proportion of the total venues, more popular than they are in New York. Conversely, New Yorkers have more preferences for pizza places and restaurants than Torontonians.

The machine learning algorithms employed found the neighborhoods in the two cities could be broadly divided into two grouped. The first group, comprised of the majority of the neighborhoods, had food-related venues as the most numerous. This was followed by grocery shops and pharmacies. The second cluster comprised of neighborhoods where recreational and sports venues, such as parks and sports facilities, were the main type of venues in these neighborhoods. The second cluster was heavily dominated by Toronto neighborhoods, but the New York neighborhoods of Bayswater, Clason Point, Randall Manor, Somerville, Todt Hill were also in this group.

For investors who want to open shop in either of the cities; opening fast-food, restaurant or beverage places are popular choices that are frequented by residents in both cities. The data showed whether they offer quick snacks or sit-in dining, eateries are the undisputed leading type of establishments. As both Toronto and New York are the financial centers of their respective countries, high demand for ready-made food for a fast-paced working population would be expected. Therefore, catering is an investment opportunity for those looking to set up shop in either of the cities.

Those who are interested in investing in neighborhoods grouped under Cluster 1 may also consider recreational facilities that cater to various types of sports or fitness. These neighborhoods are more likely to be more residential with lower levels of commercial activities. Therefore, investing in facilities that cater to family activities is another area of potentially good investments.

# Conclusion

The study showed food-related venues were the most numerous types of venues in both Toronto and New York. However, in certain neighborhoods, recreational facilities were the most numerous types of venues. An investor looking at what type of venue to invest in may consider these two broad areas when deciding.

While the macro-view supports investments in the catering and customer service industry, investors should keep in mind several factors when making decisions. While food and recreation were the most numerous venues in both cities, care should be taken to study individual neighborhoods in more detail, so investors are not entering an oversaturated market. Furthermore, as the differentiation in the type of restaurants in the data can attest, different neighborhoods have different preferences for certain types of restaurants and these tastes need to be considered.

A further area of study can be to study the flow of people at different venues to further determine what is popular in each city. For example, heatmaps can be used that show which venues people tend to congregate. This approach would give much better insight into what is popular in each city, beyond counting the number of venues that are on offer. Such studies would, however, would raise privacy concerns and the data needed may be out of reach for commercial organizations.

# Appendices

## Top 20 Venues

Table 5: Top 20 venues in Toronto by number of venues



Table 6: top 20 venues in New York based on the number of venues



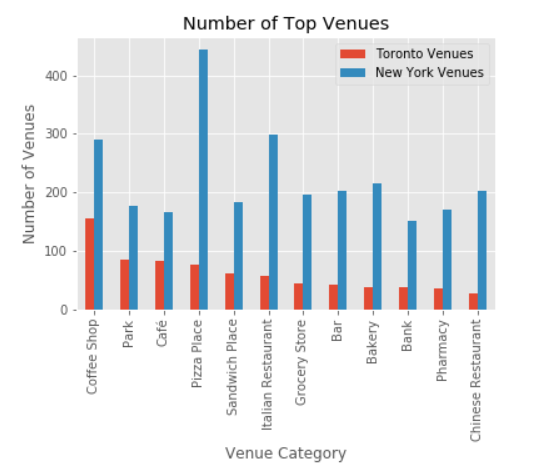


Figure 4: Comparing Toronto and New York top venues by number of venues

## Clustering

Table 7: k-means clustering cluster 1 neighborhoods and common venues

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Neighborhood** | **1st Most Common Venue** | **2nd Most Common Venue** | **3rd Most Common Venue** | **4th Most Common Venue** | **5th Most Common Venue** |
| Bayswater | Playground | Park | Tennis Court | Filipino Restaurant | Event Space |
| Clason Point | Park | Pool | Moving Target | South American Restaurant | Boat or Ferry |
| Randall Manor | Park | Bus Stop | Playground | Yoga Studio | Event Service |
| Somerville | Park | Yoga Studio | Ethiopian Restaurant | Event Space | Exhibit |
| Todt Hill | Park | Yoga Studio | Ethiopian Restaurant | Event Space | Exhibit |
| Bathurst Manor (34) | Playground | Park | Baseball Field | Convenience Store | Fish Market |
| Bayview Woods-Steeles (49) | Park | Basketball Court | Dog Run | Zoo Exhibit | Egyptian Restaurant |
| Bendale (127) | Park | History Museum | Tennis Court | Greek Restaurant | Zoo Exhibit |
| Brookhaven-Amesbury (30) | Park | Zoo Exhibit | Dumpling Restaurant | Egyptian Restaurant | Electronics Store |
| Caledonia-Fairbank (109) | Women's Store | Park | Zoo Exhibit | Falafel Restaurant | Egyptian Restaurant |
| Casa Loma (96) | Park | History Museum | Historic Site | Modern European Restaurant | Museum |
| Centennial Scarborough (133) | Construction & Landscaping | Park | Zoo Exhibit | Falafel Restaurant | Electronics Store |
| Clanton Park (33) | Park | IT Services | Zoo Exhibit | Falafel Restaurant | Electronics Store |
| Downsview-Roding-CFB (26) | Business Service | Park | Moving Target | Eastern European Restaurant | Electronics Store |
| Edenbridge-Humber Valley (9) | Indian Restaurant | Fast Food Restaurant | Park | Zoo Exhibit | Electronics Store |
| Forest Hill South (101) | Playground | Park | Dumpling Restaurant | Egyptian Restaurant | Electronics Store |
| Glenfield-Jane Heights (25) | Park | Pool | Tea Room | Moving Target | Zoo Exhibit |
| Hillcrest Village (48) | Park | Residential Building (Apartment / Condo) | Zoo Exhibit | Dumpling Restaurant | Egyptian Restaurant |
| Kingsway South (15) | Pool | Lounge | Park | Zoo Exhibit | Event Space |
| Lambton Baby Point (114) | Garden | River | Park | Zoo Exhibit | Event Service |
| Lansing-Westgate (38) | Park | Zoo Exhibit | Dumpling Restaurant | Egyptian Restaurant | Electronics Store |
| Milliken (130) | Sandwich Place | Park | Sushi Restaurant | Egyptian Restaurant | Electronics Store |
| North Riverdale (68) | Cafes© | Park | Pool | Dog Run | Zoo Exhibit |
| Parkwoods-Donalda (45) | Food & Drink Shop | Park | Zoo Exhibit | Event Space | Egyptian Restaurant |
| Rosedale-Moore Park (98) | Playground | Candy Store | Tennis Court | Park | Eastern European Restaurant |
| Taylor-Massey (61) | Theater | Poutine Place | Park | Zoo Exhibit | Event Space |
| West Hill (136) | Park | Gym / Fitness Center | Construction & Landscaping | Zoo Exhibit | Egyptian Restaurant |
| Westminster-Branson (35) | Gym / Fitness Center | Construction & Landscaping | Baby Store | Park | Event Space |
| Willowdale West (37) | Mobile Phone Shop | Park | Zoo Exhibit | Falafel Restaurant | Electronics Store |

Table 8: DBSCAN cluster 1 neighborhoods and most common venues

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Neighborhood** | **1st Most Common Venue** | **2nd Most Common Venue** | **3rd Most Common Venue** | **4th Most Common Venue** | **5th Most Common Venue** |
| Bayswater | Playground | Park | Tennis Court | Filipino Restaurant | Event Space |
| Clason Point | Park | Pool | Moving Target | South American Restaurant | Boat or Ferry |
| Randall Manor | Park | Bus Stop | Playground | Yoga Studio | Event Service |
| Somerville | Park | Yoga Studio | Ethiopian Restaurant | Event Space | Exhibit |
| Todt Hill | Park | Yoga Studio | Ethiopian Restaurant | Event Space | Exhibit |
| Bathurst Manor (34) | Playground | Park | Baseball Field | Convenience Store | Fish Market |
| Bayview Woods-Steeles (49) | Park | Basketball Court | Dog Run | Zoo Exhibit | Egyptian Restaurant |
| Bendale (127) | Park | History Museum | Tennis Court | Greek Restaurant | Zoo Exhibit |
| Brookhaven-Amesbury (30) | Park | Zoo Exhibit | Dumpling Restaurant | Egyptian Restaurant | Electronics Store |
| Caledonia-Fairbank (109) | Women's Store | Park | Zoo Exhibit | Falafel Restaurant | Egyptian Restaurant |
| Casa Loma (96) | Park | History Museum | Historic Site | Modern European Restaurant | Museum |
| Centennial Scarborough (133) | Construction & Landscaping | Park | Zoo Exhibit | Falafel Restaurant | Electronics Store |
| Clanton Park (33) | Park | IT Services | Zoo Exhibit | Falafel Restaurant | Electronics Store |
| Downsview-Roding-CFB (26) | Business Service | Park | Moving Target | Eastern European Restaurant | Electronics Store |
| Edenbridge-Humber Valley (9) | Indian Restaurant | Fast Food Restaurant | Park | Zoo Exhibit | Electronics Store |
| Forest Hill South (101) | Playground | Park | Dumpling Restaurant | Egyptian Restaurant | Electronics Store |
| Glenfield-Jane Heights (25) | Park | Pool | Tea Room | Moving Target | Zoo Exhibit |
| Hillcrest Village (48) | Park | Residential Building (Apartment / Condo) | Zoo Exhibit | Dumpling Restaurant | Egyptian Restaurant |
| Kingsway South (15) | Pool | Lounge | Park | Zoo Exhibit | Event Space |
| Lambton Baby Point (114) | Garden | River | Park | Zoo Exhibit | Event Service |
| Lansing-Westgate (38) | Park | Zoo Exhibit | Dumpling Restaurant | Egyptian Restaurant | Electronics Store |
| Milliken (130) | Sandwich Place | Park | Sushi Restaurant | Egyptian Restaurant | Electronics Store |
| North Riverdale (68) | Caffe© | Park | Pool | Dog Run | Zoo Exhibit |
| Parkwoods-Donalda (45) | Food & Drink Shop | Park | Zoo Exhibit | Event Space | Egyptian Restaurant |
| Rosedale-Moore Park (98) | Playground | Candy Store | Tennis Court | Park | Eastern European Restaurant |
| Taylor-Massey (61) | Theater | Poutine Place | Park | Zoo Exhibit | Event Space |
| West Hill (136) | Park | Gym / Fitness Center | Construction & Landscaping | Zoo Exhibit | Egyptian Restaurant |
| Westminster-Branson (35) | Gym / Fitness Center | Construction & Landscaping | Baby Store | Park | Event Space |
| Willowdale West (37) | Mobile Phone Shop | Park | Zoo Exhibit | Falafel Restaurant | Electronics Store |