



# THE BATTLE OF THE NEIGHBORHOODS

# Introduction & Business Problem

## Introduction & Business Problem

---

Analysis of a locality based on the various venues available in that locality in give useful insights into the kind of Business thriving in that area. This profiling can be used to come up with the type of business which is likely to succeed in that locality.

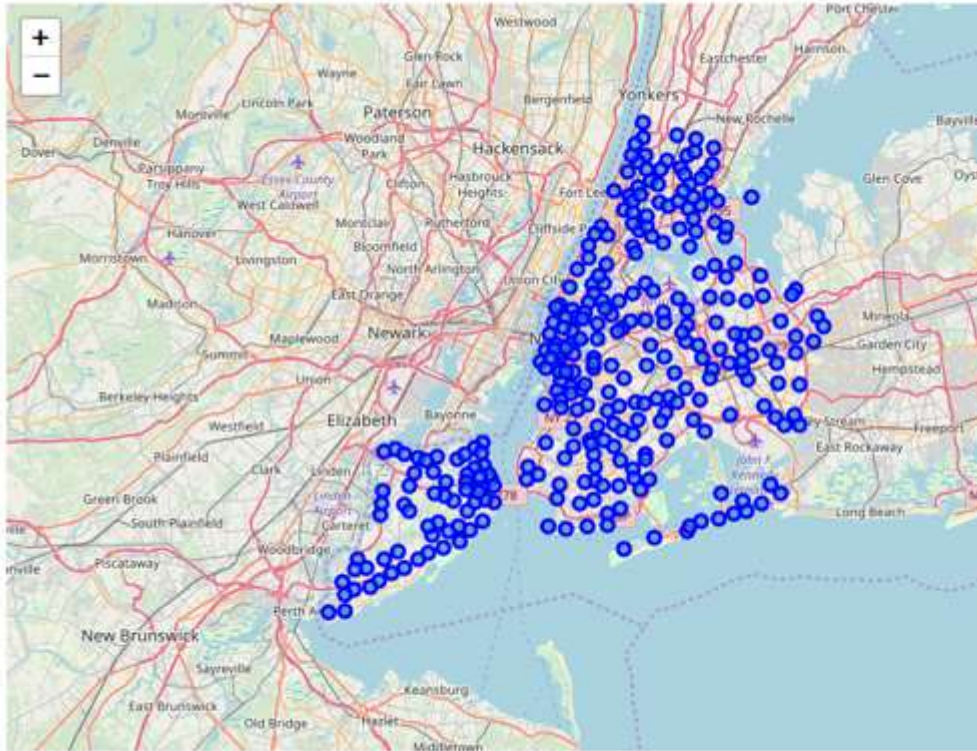
This project aims at profiling the neighborhoods to come up with the best location for starting a new restaurant. The project aims to analyze the localities of New York and Toronto. This will involve profiling these neighborhoods. The profiling will be based on the number and category of venues of various types present in an area.

## Data & Tools

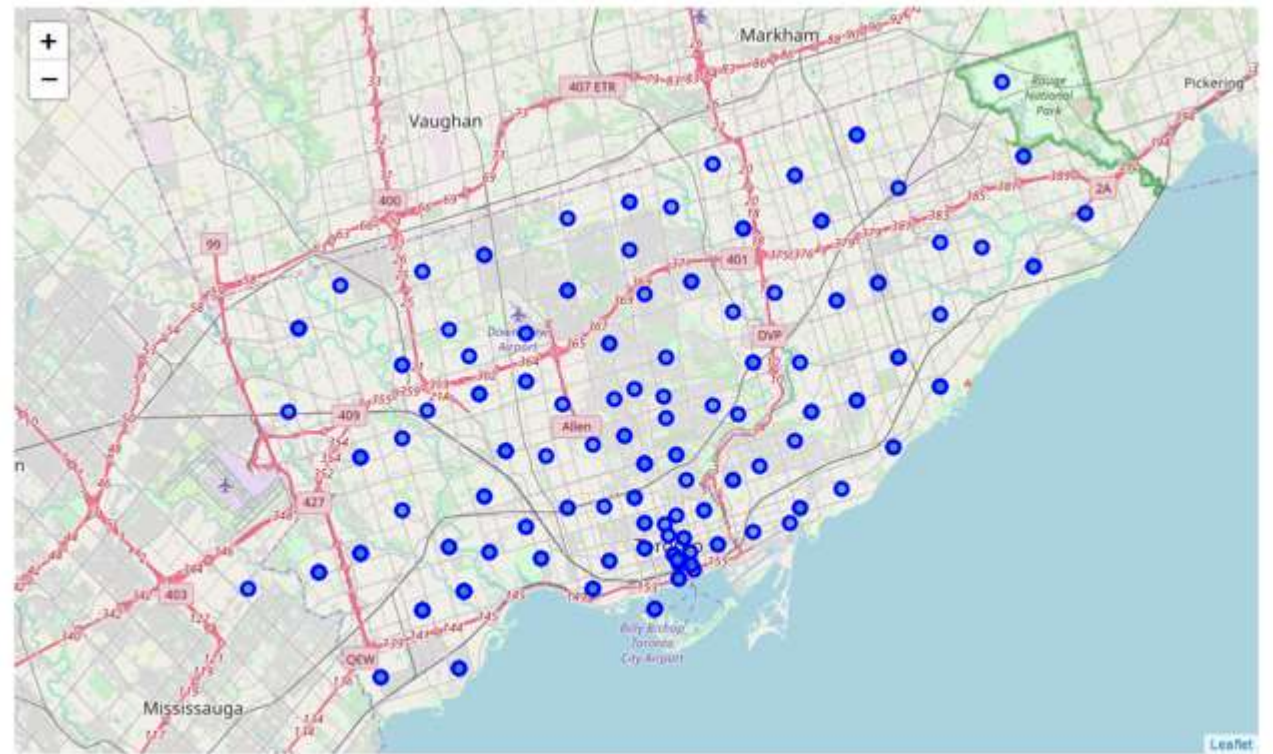
---

New York and Toronto are the two cities which are planned to be analyzed as part of this assignment. New York data was provided as part of the previous assignment in the course. Information of the neighborhoods names of Toronto will be extracted from Wikipedia article as was done in assignment in week 3. Coordinates will be extracted using the Geocoder API, which will then be used as input for Foursquare to obtain venue information and map generation.

# Neighborhood Visualizations



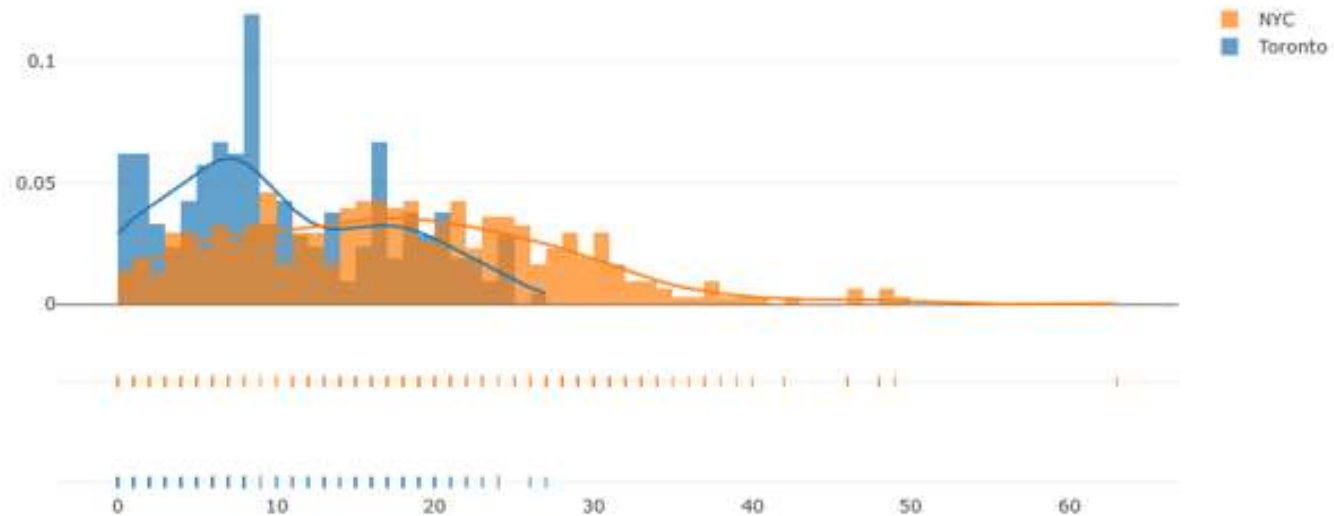
**New York Neighborhoods Visualization**



**Toronto Neighborhoods Visualization**



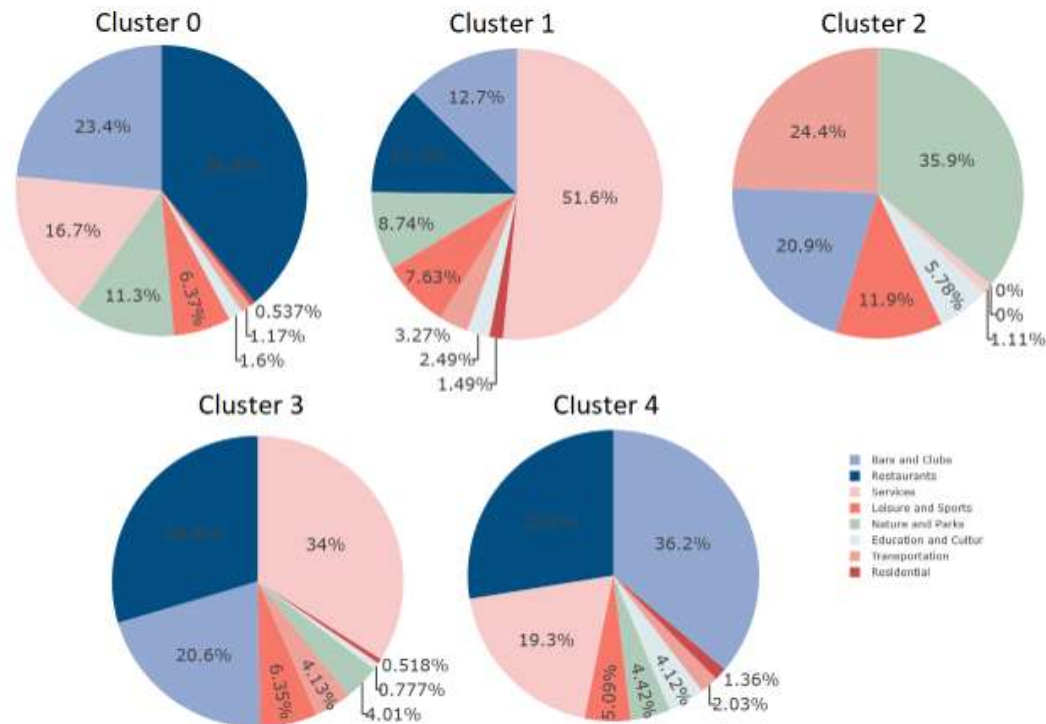
**General Services Distribution in Each City:** The general services distribution is somewhat similar between the two cities, but still following the same pattern than the previous ones - New York seems more saturated while in Toronto some neighborhoods have a deficiency of services.



# Results

**Neighborhood K-Means Clustering based on Mean Occurrence of Each Venue Category.** With the previously encoded data, we will now aim to cluster the neighborhoods into five clusters, each one with different major characteristics. K-means clustering is the algorithm that will be used - this algorithm aims to partition  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest mean. K-Means uses an iterative refinement technique, and it is also referred to as Lloyd's algorithm. In the next pictures we show how the five clusters are characterized, in terms of median percent share of each kind of neighborhood.

**Clusters in Toronto.** The Toronto city clusters are presented below.



Cluster 0 aggregates neighborhoods with a huge proportion of restaurants, followed by bars and then services. The parks and nature proportional share in this cluster is higher than similar clusters in New York, indicating few residential venues like closed residential buildings, and bigger parks and forested areas (as is expected in some parts of Toronto than in an urban sprawl like NYC).

Cluster 1 aggregates neighborhoods with a huge proportional share of services (more than 50%!) indicating that neighborhoods classified in this cluster probably are commercial districts or city centers.

Cluster 2 aggregates neighborhoods with a high prevalence of parks and nature, and also transportation infrastructure. There are services, restaurants and bars but they aren't a common occurrence in these areas. Here we also have the neighborhoods with the highest proportional share of leisure, education, cultural, and sports venues (probably indicating some kind of university presence).

Cluster 3 aggregates neighborhoods with balanced shares of restaurants, bars, clubs and leisure. This cluster also has the lowest parks and nature proportional share, indicating that neighborhoods in these areas are highly urbanized parts of the city, indicating high development.

Cluster 4 aggregates the highest proportional share of bars and clubs in Toronto, indicating that these regions for some reason are attracting business related to nightlife and food. This cluster will be further analyzed to answer our specified business problem.

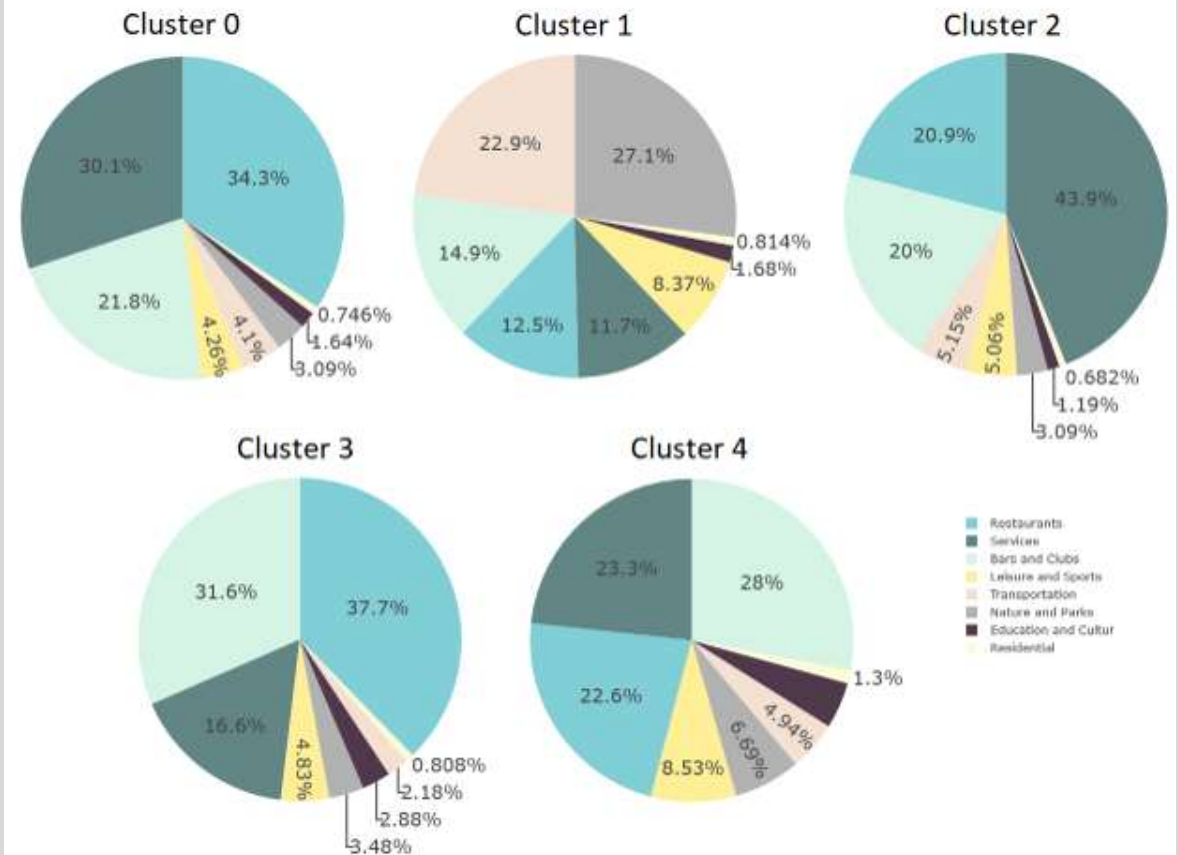
In the map below we can see the geographical visualization of the different types of clusters created using K-Means for Toronto. The colors indicate the biggest proportional share of venue category (using the same legend from the pie charts, except for the red one that indicate the balanced Cluster 3).

# Results

**Neighborhood K-Means Clustering based on Mean Occurrence of Each Venue Category.** With the previously encoded data, we will now aim to cluster the neighborhoods into five clusters, each one with different major characteristics. K-means clustering is the algorithm that will be used - this algorithm aims to partition  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest mean. K-Means uses an iterative refinement technique, and it is also referred to as Lloyd's algorithm. In the next pictures we show how the five clusters are characterized, in terms of median percent share of each kind of neighborhood.



**Clusters in New York City.** The New York City clusters are presented below.



# Analysis & Conclusion

**Toronto Analysis.** For Toronto, we selected the clusters 3 and 4. The cluster 4 aggregates neighborhoods with high numbers of bars and good demand and accessibility for the public (suburban areas), indicating places with lower rents and property prices - relative to city center neighborhoods, of the also selected cluster 3. Cluster 3 groups the more urbanized and developed parts of Toronto, with several services category venues - making these areas great neighborhoods with high demand for restaurants, bars, nightclubs, etc. The list of neighborhoods in this cluster is presented below, and they basically form a list of places with well-established business in the restaurant/bar/club's segment. The optimal location for a new business in the restaurant or bar category can be further studied with the granular data about the specific themes of restaurants and bars. High demand signals high offerings and also higher competitiveness, meaning that it's probably better to start a "new" kind of venue, in an untapped market in an underdeveloped or suburban area.

## **Good neighborhoods for establishing new restaurant venues in Toronto (possible untapped markets):**

Richview Gardens, Roselawn, Rouge, Royal York South West, Scarborough Town Centre, Silverstone, Parkview Hill, Mimico NW, Mount Olive, Oriole, Maryvale, South Steeles, Wexford Heights, Wilson Heights, Woodbine Gardens, York Mills West, Wexford, Thorncliffe Park, Thistletown, South of Bloor, St. Phillips, Steeles West, The Beaches West, The Queensway West, Martin Grove Gardens, Cliffside West, Downsview North, Downsview Northwest, Downsview West, East Birchmount Park, Dorset Park, Albion Gardens, Bathurst Manor, Beaumont Heights, Bedford Park, Birch Cliff, Caledonia-Fairbanks, Cedarbrae, Jamestown, Kennedy Park, Kingsview Village, Kingsway Park South West, L'Amoreaux West, Lawrence Heights, Lawrence Manor, Lawrence Manor East, Leaside, Malvern, Ionview, India Bazaar, Glencairn, Henry Farm, Hillcrest Village, Humbergate, Fairview

## **Good neighborhoods for restaurant venues in Toronto (but probably saturated markets):**

Railway Lands, Silver Hills, York Mills, South Niagara, Bathurst Quay, CN Tower, King and Spadina, Lawrence Park, Island airport, Harbourfront West



# Analysis & Conclusion

**New York City Analysis.** For NYC, we selected the clusters 0 and 4. The cluster 4 aggregates neighborhoods with high but balanced proportions of services, restaurants, and bars/clubs. Comparing cluster 4 with cluster 0, that aggregates city center neighborhoods with very high proportional share of restaurants and also relatively high proportion of services, we can notice that cluster 4 is behind cluster 0 in the gentrification, or urban development process. This information can be used to plan the best locations for a restaurant business based on the intentions of our business sponsor: does he want to open a restaurant in some place with high demand, but also high price of entry or he wants to bet in a place with less entrenched competitors and also good demand? We list the possible neighborhoods for each group in the next subsection.

## **Good neighborhoods for establishing new restaurant venues in NYC (not so much competition and good demand - "safe bets"):**

Oakland Gardens, North Side, North Riverdale, North Corona, Noho, Prospect Heights, Pelham Parkway, New Brighton, Murray Hill, Manhattanville, Manhattan Valley, Manhattan Beach, Lower East Side, Murray Hill, Morningside Heights, Midtown South, Ravenswood, Upper West Side, Turtle Bay, Tudor City, Tottenville, Throgs Neck, Sunnyside Gardens, Williamsburg, Whitestone, West Village, West Brighton, Sunnyside, Stuyvesant Town, Roosevelt Island, Rockaway Beach, Riverdale, Ridgewood, Schuylerville, Sheepshead Bay, Steinway, South Side, South Ozone Park, City Island, Chinatown, Central Harlem, Bushwick, Clifton, Clinton Hill, Bayside, Bay Ridge, Astoria, Annadale, Bedford Stuyvesant, Briarwood, Bergen Beach, Woodside, Hamilton Heights, Great Kills, Gramercy, Kingsbridge, Jackson Heights, Inwood, Hunters Point, Elmhurst, Edgewater Park, East Village, East Harlem, Fieldston, Fort Hamilton, Fort Greene, Forest Hills Gardens, Flushing, Flatbush, Yorkville

## **Good neighborhoods for establishing new clubs and bars in NYC (not so much competition and good demand - "safe bets"):**

Ozone Park, New Springville, New Dorp Beach, New Dorp, Queens Village, Pleasant Plains, Malba, Lindenwood, Marine Park, Mount Hope, Morrisania, Mill Basin, Middle Village, Maspeth, Westerleigh, Rosedale, Richmond Valley, Sea Gate, Starrett City, South Jamaica, Soundview, Lefrak City, Charleston, Castle Hill, Co-op City, Baychester, Bay Terrace, Arlington, Bellerose, Bloomfield, Heartland Village, Holliswood, Hunts Point, Glendale, Elm Park, Eastchester, East Flatbush, Erasmus, Georgetown, Floral Park

## **Good neighborhoods for restaurant venues in NYC (but probably saturated markets - entrenched business):**

Allerton, Olinville, Old Town, Ocean Parkway, Ocean Hill, Norwood, New Lots, Paerdegat Basin, Park Hill, Parkchester, Queensboro Hill, Prospect Park South, Prospect Lefferts Gardens, Prince's Bay, Port Richmond, Park Slope, Manhattan Terrace, Madison, Longwood, Little Neck, Mariner's Harbor, Melrose, Mount Eden, Mott Haven, Morris Park, Midwood, Rego Park, Utopia, University Heights, Unionport, Tribeca, Sutton Place, Sunset Park, Washington Heights, Woodhaven, Williamsbridge, Westchester Square, Weeksville, Wakefield, Rugby, Rosebank, Rochdale, Richmond Hill, Remsen Village, St. Albans, Spuyten Duyvil, Soho, Laurelton, City Line, Canarsie, Cambria Heights, Civic Center, Dongan Hills, Ditmas Park, Cypress Hills, Corona, Concord, College Point, Brownsville, Bay Terrace, Bath Beach, Auburndale, Baywater, Bronxdale, Borough Park, Douglaston, Bensonhurst, Belmont, Bellaire, Hillcrest, Highland Park, Greenwich Village, Greenridge, Grasmere, Grant City, Graniteville, Hollis, Howard Beach, Kew Gardens Hills, Kensington, Jamaica Hills, Jamaica Estates, Jamaica Center, Homecrest, Glen Oaks, Eltingville, Edenwald, East Tremont, East New York, Far Rockaway, Fresh Meadows, Fordham, Flatlands, Flatiron, Forest Hills