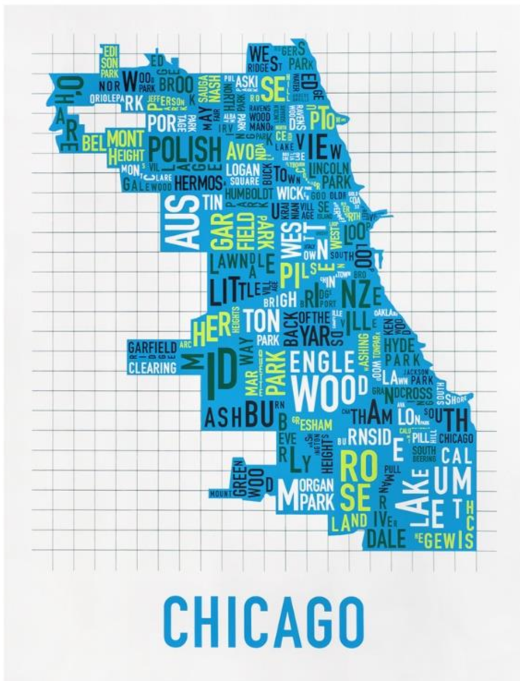
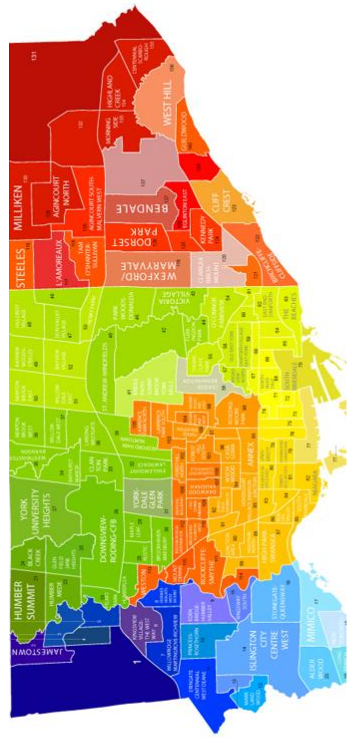
The Battle of Neighborhoods

Aditya Sinha

October 12, 2019





1. Introduction/Business Problem

1.1 Background

Chicago is often referred to as Toronto's closest sibling. They were both built in similar times, with similar sizes and populations. Both have experienced an influx of immigrants and have a diverse food culture. For this final capstone, I chose to examine the neighborhoods in Chicago (my hometown) and compare them to the ones in Toronto.

1.2 Problem Statement

The goal of this project will be to help a businessman choose which city to open their restaurant in, and in which neighborhood. If there is enough analysis available, we might even be able to isolate which *type* of restaurant we should open.

1.3 Process

To do so, we will gather data for Chicago and Toronto from two different datasets and analyze them. We will gather the coordinates for the neighborhoods in both these cities. We then use the Foursquare API to gather additional information such as venues for each neighborhood. Finally, we use the clustering approach to explore similar neighborhoods, segment them, and group them into clusters. This will allow us to compare the two cities and the individual neighborhoods since we'll be able to look at the most visited venues in each neighborhood. Based on this information, our businessman will be able to determine whether it would be better to open a restaurant in either city or in which neighborhood as well.

1.4 Tools Required

We will be using the Foursquare API as the main data gathering source since it has a wide database of millions of places/venues and it has the ability to perform location search and obtain details about a business. Since there is a limit on the number of http requests that can be placed, our neighborhoods will be limited to a reasonable 100 and the radius will be set to 500 meters.

We will be using the following libraries:

- Pandas - library for data analysis
- Numpy – library to handle data in a vectorized manner
- Geopy – retrieve location data
- Matplotlib – plotting module
- Sklearn – used for the clustering algorithm
- Folium – map rendering library
- Geocoder – retrieve the coordinates for a particular neighborhood
- JSON_normalize – transform json into pandas dataframe
- Nominatim – convert an address to latitude and longitude values

2. Data Extraction and processing

2.1 Data Source

For the initial data set, I will start by extracting the data from Open Data Soft (ODS) to get a list of the neighborhoods in Chicago. ODS has thousands of datasets available publicly that can be filtered according to state, keywords, etc. For Toronto, the data will be extracted from the Wiki link that was used in the previous exercise. Both links are provided below:

Chicago dataset:

https://public.opendatasoft.com/

	State	County	City	Name	RegionID
1	IL	Cook	Chicago	Beverly View	403368
2	IL	Cook	Chicago	Ranch Triangle	403305
3	IL	Cook	Chicago	Schorsch Forest View	403280
4	IL	Sangamon	Springfield	Vinegar Hill	761482
5	IL	Cook	Chicago	Lakewood - Balmoral	403260
6	IL	Cook	Evanston	West Village	764020
7	IL	Rock Island	Moline	Wheelock/Veile	761011
8	IL	Cook	Chicago	Jackson Park Highlands	403354
9	IL	Cook	Chicago	Hyde Park	269586
10	IL	Sangamon	Springfield	West Koke Mill	761475
11	IL	Cook	Chicago	East Beverly	403375
12	IL	Cook	Chicago	Montclare	274579
13	IL	Cook	Chicago	Cragin	28173
14	IL	Cook	Chicago	The Loop	269593
15	IL	Cook	Chicago	South Old Irving Park	403316
16	IL	Cook	Chicago	Wentworth Gardens	403341
17	IL	Vermilion	Danville	Center City	268056
18	IL	Cook	Chicago	Chrysler Village	403336
19	IL	Rock Island	Moline	Karsten's Park/City Line	761016
20	IL	Cook	Chicago	Portage Park	275083

Toronto dataset:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Postcode	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Harbourfront
M5A	Downtown Toronto	Regent Park
M6A	North York	Lawrence Heights
M6A	North York	Lawrence Manor
M7A	Queen's Park	Not assigned
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue
M1B	Scarborough	Rouge
M1B	Scarborough	Malvern
M2B	Not assigned	Not assigned
M3B	North York	Don Mills North
M4B	East York	Woodbine Gardens
M4B	East York	Parkview Hill
M5B	Downtown Toronto	Ryerson
M5B	Downtown Toronto	Garden District
M6B	North York	Glencairn
M7B	Not assigned	Not assigned
M8B	Not assigned	Not assigned
M9B	Etobicoke	Cloverdale
M9B	Etobicoke	Islington

2.2 Data cleaning and processing

Once these datasets are extracted and read into a pandas data frame, I will first filter for the respective cities and clean the data. This will ensure there are no duplicates or null values and that the data is consistent. To gather the latitude and longitude values, we will be using the **geocoder** package and create a function that will get the coordinates for each neighborhood. We will also need data about the different venues in different neighborhoods for both cities. In order to get this information, we will be using the **Foursquare API** to make calls. Foursquare provides data on location which includes multiple venues and areas of interest in a given radius. This can also include venue names, locations, and photos. After finding the list of neighborhoods and creating the respective data frames, we will connect with the Foursquare API to get information about the location within a 500-meter radius.

The data retrieved from the Foursquare API will contain information such as:

- Name of neighborhood
- Neighborhood latitude and longitude
- Name of venue
- Venue latitude and longitude
- Type of venue

2.3 Data driven goals

We will be using a clustering approach to explore and segment the neighborhoods. In the end, we can look at the top venues that were visited in each cluster and determine answers to the following questions:

- Comparing Chicago and Toronto, which city has more restaurants as the 1st most visited venue?
- In particular, are their certain neighborhoods that do not have restaurants in their "top 10 visited" list?
- Is there a specific type of restaurant that we can open that isn't already there?

Doing so can help our potential businessman in choosing the best city and location for opening his/her restaurant. In addition, this could also help future businessmen in determining the different types of business that are already largely available and what type of business they can invest in or open in these locations.

3. Exploratory Data Analysis

3.1 Processing the data

I first started with the Toronto dataset since this is something I had worked with previously and had a rough idea of what sort of results to expect. Both the cities have been explored in a similar fashion, only the data importing process varied slightly. In the end, the goal was to look at both the cities and be able to determine which city, and then perhaps the particular neighborhood within the city, would be an ideal venue to open a restaurant in. If there was enough data provided, we could also pinpoint which *type* of restaurant the businessman can open. For example, a Greek restaurant or an Italian restaurant – based on the list of venues that are visited most frequently.

After extracting the data from the Wikipedia page, I first created a dataframe that would include the Postal Code, Borough, and the Neighborhood. I then wrote a function that would look for the 'mw-parser-output' class and then iterate through each row and fill in the postal code, borough and neighborhood values into the dataframe. Taking a look at this dataframe, it has a size of (289, 3). This number will later reduce as we filter for the postal codes that are only related to Toronto. We clean up the data by setting some rules such as ignoring the cells with a borough that is not assigned and checking to see if a cell had a borough but a 'not assigned' neighbor. In this case, we would assign the value of the borough to the neighborhood. We also dropped an unassigned values in this stage of data cleaning and grouped the postal codes that contained multiple neighborhoods. Checking the size of the dataframe now, it was at (103, 3).

	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West
9	M1N	Scarborough	Birch Cliff, Cliffside West

Initially, the geocoder package would not work and I wasn't able to get the coordinates data for Toronto. Fortunately, there was a csv file provided in the previous lab and I simply used that create another dataframe that contained the postal codes along with their respective latitude and longitudes. I then performed an inner join on the postal codes column and received the following dataframe:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

3.2 Clustering and visualizing

To actually start visualizing the data, I first searched for all the boroughs that contained the word 'Toronto'. I then used the Nominatim tool to search for the geographical coordinates of Toronto. The next step was to visualize the neighborhoods of Toronto.

I then specified my Foursquare credentials and obtained the coordinates for the first neighborhood that was in the dataframe. The next step was to get the top 100 venues that were in that neighborhood within a radius of 500 meters. The GET request spits out data in a JSON format. I then used the function that was introduced in the Foursquare lab that extracts the category of the venue. This data was then structured into another pandas dataframe and filtered and cleaned to obtain the following dataframe:

	name	categories	lat	lng
0	Glen Manor Ravine	Trail	43.676821	-79.293942
1	The Big Carrot Natural Food Market	Health Food Store	43.678879	-79.297734
2	Grover Pub and Grub	Pub	43.679181	-79.297215
3	Upper Beaches	Neighborhood	43.680563	-79.292869
4	Seaspray Restaurant	Asian Restaurant	43.678888	-79.298167

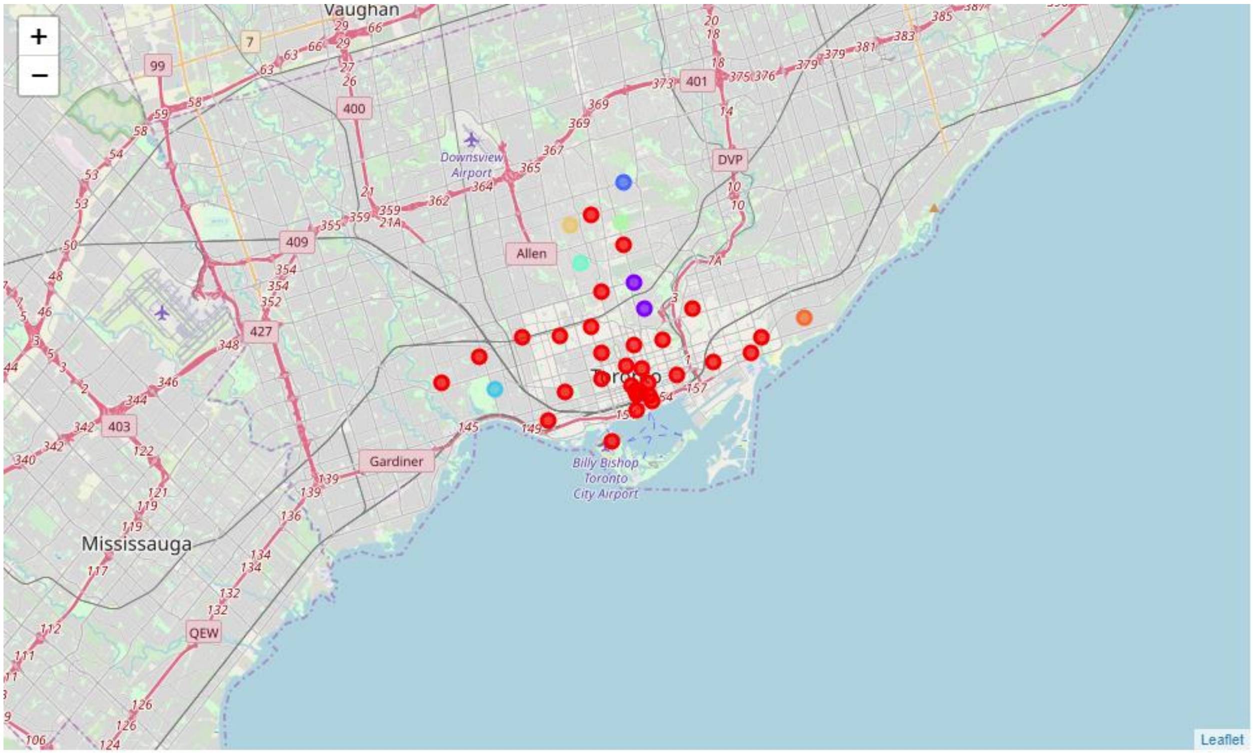
The next step was to create a function that would include the url that gets the information from the Foursquare API and returns relevant information for each nearby venue. This function was run and a new dataframe was created to store all nearby neighborhoods along with their coordinates, venues, venue coordinates and venue category.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Beaches	43.676357	-79.293031	Seaspray Restaurant	43.678888	-79.298167	Asian Restaurant
5	The Danforth West, Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant

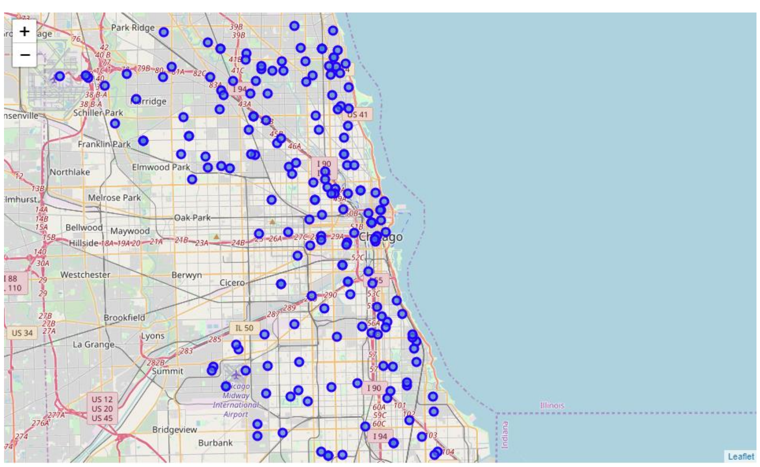
After taking a look at how many venues were returned for each neighborhood, we also took a look at how many unique categories there are. Before visualizing this information on a map, there were a few steps to complete. The first step was to perform a one-hot encode on the types of venues. The second was to group the rows by neighborhood and take the mean frequency of occurrence for each category. Next, I printed each neighborhood along with the topmost 5 common venues. I then wrote a function to sort the venues in descending order. We then create another dataframe that displays the top 10 for each neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Café	Bar	Steakhouse	Restaurant	Hotel	Burger Joint	Cosmetics Shop	Thai Restaurant	American Restaurant
1	Berczy Park	Coffee Shop	Cocktail Bar	Bakery	Café	Cheese Shop	Steakhouse	Farmers Market	Beer Bar	Italian Restaurant	Seafood Restaurant
2	Brockton, Exhibition Place, Parkdale Village	Coffee Shop	Breakfast Spot	Café	Yoga Studio	Stadium	Burrito Place	Caribbean Restaurant	Restaurant	Climbing Gym	Pet Store
3	Business Reply Mail Processing Centre 969 Eastern	Light Rail Station	Yoga Studio	Spa	Garden Center	Garden	Fast Food Restaurant	Farmers Market	Comic Shop	Park	Gym / Fitness Center
4	CN Tower, Bathurst Quay, Island airport, Harbo...	Airport Lounge	Airport Service	Airport Terminal	Plane	Harbor / Marina	Airport	Airport Food Court	Boutique	Bar	Boat or Ferry

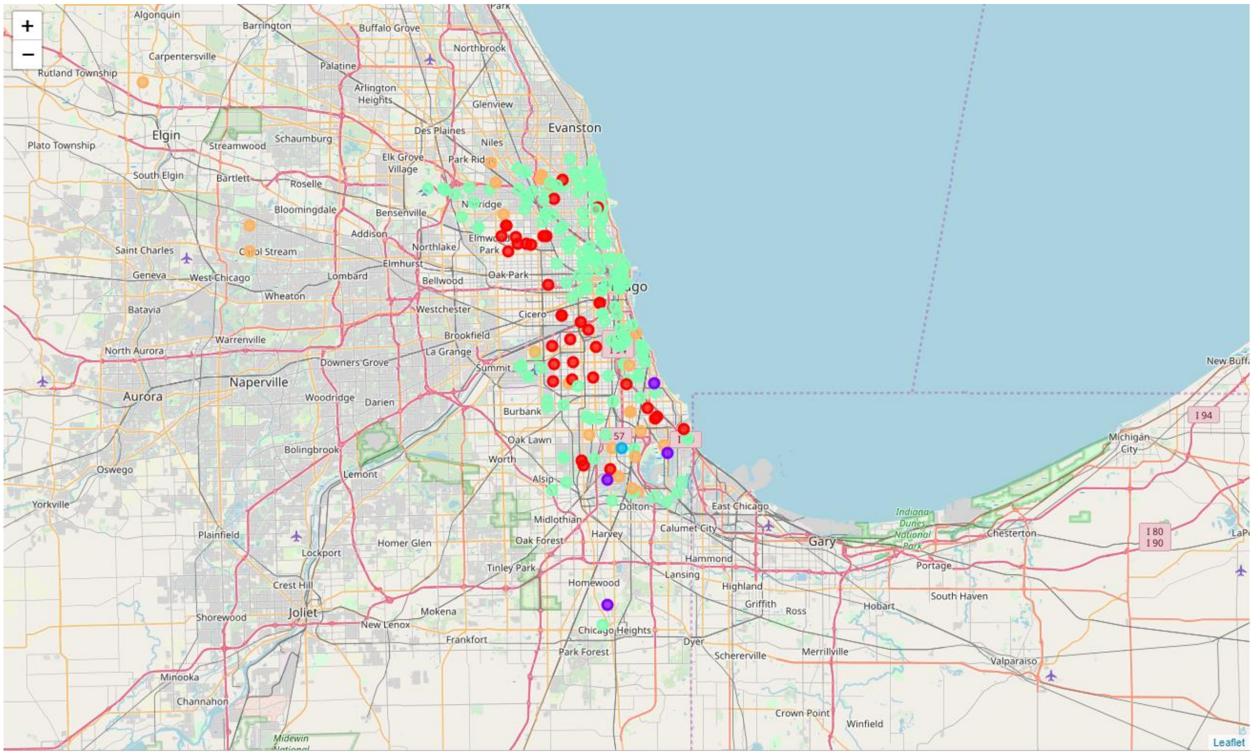
K-means clustering partitions n number of observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. I applied k-means clustering to this data and set the number of clusters to 5. I then created a new dataframe that includes the cluster as well as the top 10 venues for each of the neighborhoods. When visualizing the results, you can see in the map below that each cluster is color-coded. We can now examine each cluster and determine the discriminating venue categories.



The data for Chicago was processed in a similar way. However, instead of processing the data using beautifulsoup, I simply imported a csv file and read the data in using pandas. The data was cleaned up by removing any null values and filtering the City by ‘Chicago’ to only get all the neighborhoods that were in Chicago. Once cleaned up and filtered, the neighborhoods were displayed on a leaflet map just like the one used for Toronto.



After applying the k-means clustering to this set of data, we get 5 clusters where each cluster contains different neighborhoods in Chicago.



4. Results

4.1 Cumulative view

To look at both the data from Toronto and Chicago, we first looked at the data from each cluster which contained the 10 most common venues for each neighborhood. For example, this is what cluster one looks like for Chicago:

```
chicago_merged.loc[chicago_merged['Cluster Labels'] == 1, chicago_merged.columns[[1] + list(range(5, chicago_merged.shape[1]))]]
```

	County	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
66	Cook	-87.559510	1	Bar	Mexican Restaurant	Women's Store	Filipino Restaurant	Event Space	Exhibit	Eye Doctor	Factory	Falafel Restaurant	Farm
90	Cook	-90.064170	1	Bar	Women's Store	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Factory	Falafel Restaurant	Farm	Farmers Market
167	Cook	-87.646201	1	Food	Bar	Women's Store	Fast Food Restaurant	Event Space	Exhibit	Eye Doctor	Factory	Falafel Restaurant	Farm
172	Cook	-87.578940	1	Donut Shop	Bar	Women's Store	Fast Food Restaurant	Event Space	Exhibit	Eye Doctor	Factory	Falafel Restaurant	Farm
189	Cook	-87.646459	1	Bar	Cosmetics Shop	Bowling Alley	American Restaurant	Basketball Court	Women's Store	Filipino Restaurant	Eye Doctor	Factory	Falafel Restaurant

I went ahead and used the DataFrame.to_excel that is built into pandas and shifted the dataframes that contained the information regarding the clusters into Excel. This was so that I could easily analyze which neighborhoods or clusters were better for opening a restaurant.

After setting conditional statements to show any venues that are restaurants, we get something like this:

	County	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Cook	-87.69948	0	Mexican Restaurant	Grocery Store	Taco Place	Park	Bus Station	Burger Joint	American Restaurant	Check Cashing Service	Sandwich Place	Breakfast Spot
6	Cook	-87.76339	0	Pharmacy	Cuban Restaurant	Pet Store	Mexican Restaurant	Fast Food Restaurant	Sandwich Place	Chinese Restaurant	Restaurant	Diner	Coffee Shop
23	Cook	-87.72341	0	Mexican Restaurant	Ice Cream Shop	Gas Station	Bakery	Fried Chicken Joint	Bar	American Restaurant	Donut Shop	Factory	Falafel Restaurant
25	Cook	-87.65759	0	Mexican Restaurant	Bar	Bakery	Pizza Place	Thrift / Vintage Store	Latin American Restaurant	Vietnamese Restaurant	Liquor Store	Breakfast Spot	Art Gallery
27	Cook	-87.68093	0	Food	Playground	Breakfast Spot	Record Shop	Snack Place	Mexican Restaurant	Fast Food Restaurant	Skating Rink	Park	Women's Store
31	Cook	-87.79165	0	Polish Restaurant	Czech Restaurant	Gas Station	Discount Store	Mexican Restaurant	Pizza Place	Video Store	Pharmacy	Liquor Store	Home Service
37	Cook	-87.73424	0	Mexican Restaurant	Supermarket	Check Cashing Service	Park	Optical Shop	Discount Store	Diner	Recording Studio	Music Venue	Department Store
45	Cook	-87.73074	0	Cosmetics Shop	Grocery Store	Discount Store	Shoe Store	Gas Station	Fried Chicken Joint	Caribbean Restaurant	Fast Food Restaurant	Pizza Place	Intersection
58	Cook	-87.64189	0	Fast Food Restaurant	Cosmetics Shop	American Restaurant	Mobile Phone Shop	Bar	Bank	Sandwich Place	Electronics Store	Clothing Store	Discount Store
64	Cook	-87.61859	0	Fast Food Restaurant	Train Station	Farmers Market	Ethiopian Restaurant	Event Service	Event Space	Exhibit	Eye Doctor	Factory	Falafel Restaurant
67	Cook	-87.57449	0	Convenience Store	Mexican Restaurant	Construction & Landscaping	Dance Studio	Park	Sandwich Place	Bus Station	Fried Chicken Joint	Cosmetics Shop	Intersection
76	Cook	-87.67338	0	Diner	Seafood Restaurant	Grocery Store	Bike Rental / Bike Share	Thrift / Vintage Store	Video Store	Gas Station	Pharmacy	Mexican Restaurant	Discount Store
84	Cook	-87.69634	0	Pizza Place	Mexican Restaurant	Cosmetics Shop	Fast Food Restaurant	Intersection	American Restaurant	Food	Business Service	Event Space	Exhibit
87	Cook	-87.72338	0	Mexican Restaurant	Bank	Discount Store	Cosmetics Shop	Pet Store	Bus Station	Seafood Restaurant	Clothing Store	Coffee Shop	Playground
89	Cook	-87.69561	0	Convenience Store	Mexican Restaurant	Asian Restaurant	Currency Exchange	Sandwich Place	Sporting Goods Shop	Construction & Landscaping	Hockey Arena	Falafel Restaurant	Elementary School
94	Cook	-87.72445	0	Ice Cream Shop	Bowling Alley	Fast Food Restaurant	History Museum	Seafood Restaurant	Park	Discount Store	Diner	Mexican Restaurant	Sandwich Place
99	Cook	-87.66273	0	Mexican Restaurant	Grocery Store	Bank	Discount Store	Fried Chicken Joint	Boutique	Gas Station	Shoe Store	Farm	Pharmacy
110	Cook	-87.72556	0	Mobile Phone Shop	Mexican Restaurant	Gas Station	Grocery Store	Hot Dog Joint	Pharmacy	Sandwich Place	Bar	Gym / Fitness Center	Big Box Store
116	Cook	-87.71073	0	Eye Doctor	Mexican Restaurant	Sandwich Place	Coffee Shop	Clothing Store	Bakery	Fried Chicken Joint	Italian Restaurant	Pharmacy	Video Store
119	Cook	-87.68361	0	Bar	Cosmetics Shop	Skating Rink	Park	Gym / Fitness Center	Gym	Burger Joint	Mexican Restaurant	Fast Food Restaurant	Lawyer
129	Cook	-87.68455	0	Fast Food Restaurant	Intersection	Business Service	Bus Stop	Factory	Lake	Farmers Market	Event Space	Exhibit	Eye Doctor
150	Cook	-87.65759	0	Mexican Restaurant	Bar	Bakery	Pizza Place	Thrift / Vintage Store	Latin American Restaurant	Vietnamese Restaurant	Liquor Store	Breakfast Spot	Art Gallery
156	Cook	-87.71172	0	Mexican Restaurant	Pizza Place	Bank	Mobile Phone Shop	Bar	Grocery Store	Dessert Shop	Discount Store	Restaurant	Park
158	Cook	-87.775591	0	Mexican Restaurant	Convenience Store	Fast Food Restaurant	Grocery Store	Bakery	Chinese Restaurant	Gas Station	Automotive Shop	Food	Hot Dog Joint
161	Cook	-87.66726	0	Fast Food Restaurant	Cosmetics Shop	Train Station	Wings Joint	Currency Exchange	American Restaurant	Liquor Store	Discount Store	Event Space	Exhibit
166	Cook	-87.58816	0	Burger Joint	Boutique	Fast Food Restaurant	Sandwich Place	Diner	Cajun / Creole Restaurant	Grocery Store	Pizza Place	ATM	Flea Market
170	Cook	-87.79165	0	Polish Restaurant	Czech Restaurant	Gas Station	Discount Store	Mexican Restaurant	Pizza Place	Video Store	Pharmacy	Liquor Store	Home Service
175	Cook	-87.79847	0	Wings Joint	Home Service	Financial or Legal Service	Pizza Place	Mexican Restaurant	Flea Market	Farmers Market	Event Service	Event Space	Exhibit
176	Cook	-87.75601	0	Mexican Restaurant	Donut Shop	Ice Cream Shop	Sandwich Place	Bus Station	Restaurant	Coffee Shop	Discount Store	Financial or Legal Service	Flea Market
181	Cook	-87.65943	0	Coffee Shop	Bar	Mexican Restaurant	Yoga Studio	Bank	Sandwich Place	Asian Restaurant	Grocery Store	Dog Run	Park
193	Cook	-87.73772	0	Fast Food Restaurant	Grocery Store	Restaurant	Greek Restaurant	Check Cashing Service	Bakery	Mexican Restaurant	Discount Store	Diner	Department Store
202	Cook	-87.71172	0	Mexican Restaurant	Pizza Place	Bank	Mobile Phone Shop	Bar	Grocery Store	Dessert Shop	Discount Store	Restaurant	Park
213	Cook	-87.57742	0	Convenience Store	Fried Chicken Joint	Sandwich Place	Gym / Fitness Center	Bus Station	Park	Mexican Restaurant	Dance Studio	Cosmetics Shop	Fish & Chips Shop
221	Cook	-87.77734	0	Mexican Restaurant	Fast Food Restaurant	Liquor Store	Gas Station	Butcher	Grocery Store	Chinese Restaurant	Latin American Restaurant	Bus Stop	Donut Shop
223	Cook	-87.53625	0	Seafood Restaurant	Hot Dog Joint	Harbor / Marina	Mexican Restaurant	Women's Store	Farmers Market	Event Service	Event Space	Exhibit	Eye Doctor
227	Cook	-87.78917	0	Sporting Goods Shop	Sandwich Place	Pizza Place	Museum	Bank	Fried Chicken Joint	Mexican Restaurant	Chinese Restaurant	Farmers Market	Exhibit

As shown, the fifth column shows how many restaurants are in the '1st most visited' column in the dataframe. I performed this same step with each cluster for a total of 10 clusters.

This is one of the clusters for Toronto:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	East Toronto	0	Health Food Store	Trail	Pub	Wings Joint	Donut Shop	Diner	Discount Store	Dog Run	Doner Restaurant	Eastern European Restaurant
1	East Toronto	0	Greek Restaurant	Coffee Shop	Ice Cream Shop	Italian Restaurant	Furniture / Home Store	Bakery	Indian Restaurant	Sports Bar	Spa	Juice Bar
2	East Toronto	0	Sandwich Place	Pizza Place	Ice Cream Shop	Liquor Store	Brewery	Board Shop	Burger Joint	Fast Food Restaurant	Burrito Place	Fish & Chips Shop
3	East Toronto	0	Café	Coffee Shop	Bakery	Italian Restaurant	American Restaurant	Cheese Shop	Bank	Bar	Middle Eastern Restaurant	Diner
5	Central Toronto	0	Gym	Food & Drink Shop	Park	Clothing Store	Breakfast Spot	Sandwich Place	Hotel	Discount Store	Dog Run	Doner Restaurant
6	Central Toronto	0	Sporting Goods Shop	Clothing Store	Coffee Shop	Health & Beauty Service	Bagel Shop	Burger Joint	Café	Chinese Restaurant	Dessert Shop	Diner
7	Central Toronto	0	Dessert Shop	Sandwich Place	Sushi Restaurant	Gym	Coffee Shop	Italian Restaurant	Café	Pizza Place	Fried Chicken Joint	Seafood Restaurant
9	Central Toronto	0	Coffee Shop	Pub	Health & Beauty Service	Vietnamese Restaurant	Light Rail Station	Supermarket	Sushi Restaurant	Pizza Place	Liquor Store	American Restaurant
11	Downtown Toronto	0	Coffee Shop	Pet Store	Restaurant	Grocery Store	Pub	Italian Restaurant	Bakery	Café	Pizza Place	Sandwich Place
12	Downtown Toronto	0	Coffee Shop	Sushi Restaurant	Japanese Restaurant	Gay Bar	Restaurant	Café	Hotel	Mediterranean Restaurant	Men's Store	Gastropub
13	Downtown Toronto	0	Coffee Shop	Park	Café	Bakery	Theater	Breakfast Spot	Mexican Restaurant	Pub	Restaurant	Spa
14	Downtown Toronto	0	Coffee Shop	Clothing Store	Cosmetics Shop	Café	Fast Food Restaurant	Pizza Place	Bakery	Ramen Restaurant	Diner	Bookstore
15	Downtown Toronto	0	Café	Coffee Shop	Restaurant	Hotel	Italian Restaurant	Clothing Store	Cosmetics Shop	Gastropub	Breakfast Spot	Bakery
16	Downtown Toronto	0	Coffee Shop	Cocktail Bar	Café	Cheese Shop	Steakhouse	Seafood Restaurant	Farmers Market	Beer Bar	Bakery	Indian Restaurant
17	Downtown Toronto	0	Coffee Shop	Café	Italian Restaurant	Fried Chicken Joint	Burger Joint	Ice Cream Shop	Sandwich Place	Bubble Tea Shop	Gym / Fitness Center	Japanese Restaurant
18	Downtown Toronto	0	Coffee Shop	Café	Bar	Thai Restaurant	American Restaurant	Steakhouse	Hotel	Asian Restaurant	Restaurant	Burger Joint
19	Downtown Toronto	0	Coffee Shop	Hotel	Aquarium	Café	Fried Chicken Joint	Italian Restaurant	Brewery	Scenic Lookout	Plaza	History Museum
20	Downtown Toronto	0	Coffee Shop	Café	Hotel	Italian Restaurant	Restaurant	Bar	Seafood Restaurant	Steakhouse	Deli / Bodega	Gastropub
21	Downtown Toronto	0	Coffee Shop	Hotel	Café	Restaurant	American Restaurant	Seafood Restaurant	Steakhouse	Gym	Deli / Bodega	Italian Restaurant
23	Central Toronto	0	Trail	Mexican Restaurant	Jewelry Store	Sushi Restaurant	Wings Joint	Diner	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store
24	Central Toronto	0	Sandwich Place	Café	Pizza Place	Coffee Shop	Vegetarian / Vegan Restaurant	BBQ Joint	Cosmetics Shop	Pharmacy	Liquor Store	History Museum
25	Downtown Toronto	0	Café	Bar	Sandwich Place	Italian Restaurant	Japanese Restaurant	Bookstore	Bakery	Restaurant	Beer Bar	Beer Store
26	Downtown Toronto	0	Café	Vegetarian / Vegan Restaurant	Chinese Restaurant	Bar	Vietnamese Restaurant	Mexican Restaurant	Coffee Shop	Bakery	Dumpling Restaurant	Comfort Food Restaurant
27	Downtown Toronto	0	Airport Lounge	Airport Service	Boat or Ferry	Harbor / Marina	Boutique	Bar	Coffee Shop	Plane	Sculpture Garden	Airport
28	Downtown Toronto	0	Coffee Shop	Café	Restaurant	Beer Bar	Fast Food Restaurant	Seafood Restaurant	Hotel	Italian Restaurant	Cocktail Bar	Farmers Market
29	Downtown Toronto	0	Coffee Shop	Café	Hotel	Restaurant	Steakhouse	Gastropub	Bar	Burger Joint	Gym	Deli / Bodega
30	Downtown Toronto	0	Grocery Store	Café	Park	Nightclub	Diner	Restaurant	Italian Restaurant	Baby Store	Athletics & Sports	Convenience Store
31	West Toronto	0	Pharmacy	Supermarket	Bakery	Middle Eastern Restaurant	Music Venue	Park	Pizza Place	Café	Brewery	Bar
32	West Toronto	0	Bar	Men's Store	Coffee Shop	Asian Restaurant	Cocktail Bar	French Restaurant	Pizza Place	Restaurant	Café	New American Restaurant
33	West Toronto	0	Breakfast Spot	Café	Coffee Shop	Pet Store	Caribbean Restaurant	Stadium	Sandwich Place	Burrito Place	Restaurant	Climbing Gym
34	West Toronto	0	Café	Mexican Restaurant	Thai Restaurant	Grocery Store	Fried Chicken Joint	Music Venue	Diner	Discount Store	Cajun / Creole Restaurant	Bookstore
35	West Toronto	0	Breakfast Spot	Gift Shop	Bookstore	Dog Run	Italian Restaurant	Bank	Movie Theater	Bar	Dessert Shop	Eastern European Restaurant
36	West Toronto	0	Coffee Shop	Café	Bookstore	Italian Restaurant	Sushi Restaurant	Dessert Shop	Fish & Chips Shop	Food & Drink Shop	Food	Bar
37	East Toronto	0	Light Rail Station	Yoga Studio	Recording Studio	Butcher	Restaurant	Auto Workshop	Spa	Fast Food Restaurant	Farmers Market	Smoke Shop

As you may notice, there are far less restaurants in the '1st most visited' or '2nd most visited' columns. Looking at the other clusters that are in Toronto, there were even less venues that were categorized as restaurants. This is what the breakdown of the other clusters looked like:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
8	Central Toronto	1	Gym	Playground	Trail	Wings Joint	Dumpling Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
22	Central Toronto	2	Pool	Garden	Wings Joint	Dessert Shop	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	Central Toronto	3	Bus Line	Park	Swim School	Wings Joint	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
10	Downtown Toronto	4	Park	Playground	Trail	Building	Wings Joint	Donut Shop	Diner	Discount Store	Dog Run	Doner Restaurant

4.2 Results

Taking a look at both the clusters and their respective ‘most visited’ places, I believe it is easy to conclude that Toronto would be a better overall city to open a new restaurant in. It doesn’t have as many restaurants listed in some clusters – for example, cluster 4 only has a restaurant listed in the 10^{th} most visited place. This would arguable be the best neighborhood to open up a restaurant in since there isn’t much competition. On the contrary, all over Chicago’s clusters had multiple restaurants listed even in the top 5 most visited places. If someone is looking to open a restaurant, they would want a neighborhood where there isn’t a lot of competition. This is why, I believe downtown Toronto, near the Rosedale cluster, would be an ideal location for a new restaurant.

5. Discussion

When comparing both Chicago and Toronto and their neighborhoods, it is clear that downtown Chicago has more neighborhoods and has more venues as well. We know that Chicago downtown is congested, and the leaflet map reflects this. On the contrary, the clustering worked better for a city like Toronto where it is mildly congested and has just about enough venues to work with. Due to this, it was easier to pinpoint that Toronto was a better city to open a restaurant in. Comparatively, Chicago has multiple restaurants listed in each cluster and there are maybe only one of two neighborhoods where a restaurant can be opened. However, due to the varying congestion levels of both these cities, it can be said that any location that might look ideal might not be in a real-world scenario because chances are that there is another cluster nearby that contains a neighborhood with multiple restaurants that are listed in the top 5 most visited venues.

6. Conclusion

In this project, I compared the neighborhoods in Chicago and Toronto and determined which city and neighborhood would be the ideal place to open a restaurant. I grouped the neighborhoods and looked at the top-10 venues that were the most visited in each. I also used the k-means clustering algorithm and visualized it on a leaflet map to get a better idea as to how these neighborhoods and their respective most-visited venues compare. The end result was that Toronto was a clear winner amongst the two if someone is looking to open a new restaurant. This was because there were less restaurants that were in the clusters created in Toronto and this therefore results in less competition, which is something a new business owner would want.