

# Analyzing the Neighborhoods of Toronto

Sarah Beth Sanchini

July 2021

## Introduction / Business Problem "The Battle of the Neighborhoods"

**1.1 Background:** Foursquare is a web service that has geographical data, and this combined with data from Wikipedia can be utilized to analyze the different area codes of Toronto Canada. Python libraries are also used to help clean, organize and visualize the data to relay useful information to the consumer. The top 5 venues in each neighborhood will be shown so people can see what is in each one.

**1.2 Problem:** Different neighborhoods vary from each other in size, population and what services they offer. When a new person moves to a location, it is helpful to know details on each neighborhood so they can choose where to live according to their needs. This project clusters neighborhoods and then shows the top 5 venues in each neighborhood.

**Problem Summary** - Need information based on neighborhoods

**1.3 Interest: Why it is important to solve** - People want to know what services are in the areas close to where they live, people rely on services for their daily lives to function.

**1.4 Target Audience:** People that are new to the area of Toronto, or people interested in discovering new information based on the different area codes. This information can be used for new people in the area deciding where they want to live, or also can be used for business owners to get an idea of what competing venues are in the vicinity.

## 2. Data

### 2.1 Data Sources

1) **Foursquare API Data:** The Foursquare data provide locational data - this includes venues, interesting locations, schools etc. given a specific location.

Information retrieved per Area Code:

- Neighborhood
- Neighborhood Latitude
- Neighborhood Longitude
- Venue
- Name of Venue
- Venue Latitude
- Venue Longitude
- Venue Category

2) **Wikipedia**: Information about different area codes in Toronto scraped from Wikipedia  
Information retrieved: Latitude, Longitude, Area/Zip Codes Data Link:  
[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

3) **Geospatial Data**: Postal Codes and their corresponding Latitude and Longitude values  
Data Link: [https://cocl.us/Geospatial\\_data](https://cocl.us/Geospatial_data)

## 2.2 Data Cleaning

Data downloading and data scraping from multiple sources to supply the data necessary for analysis. To get the data on Toronto, we first had to webscrape from Wikipedia. This data was cleaned so the columns would be in Postal Code, Borough and Neighborhood.

The postal code was the key to converting Latitude/Longitude values into Boroughs and Neighborhoods.

Out[8]:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Queen's Park	Ontario Provincial Government	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242

## 3. Exploratory Analysis

### 3.1 Parkwoods

To get an idea of what specifics to extract, we start with just one neighborhood = Parkwoods.

Out[36]:

	name	categories	lat	lng
0	Brookbanks Park	Park	43.751976	-79.332140
1	KFC	Fast Food Restaurant	43.754387	-79.333021
2	Variety Store	Food & Drink Shop	43.751974	-79.333114

### 3.2 Number of Venues per Neighborhood

Then we can use Foursquare to show which venues and categories are in that neighborhood. This process is repeated for all of the neighborhoods in Toronto so we can compare and explore what results returned.

A total of 4260 venues have been returned  
Grouped by neighborhood:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Agincourt	8	8	8	8	8	8
Alderwood, Long Branch	18	18	18	18	18	18
Bathurst Manor, Wilson Heights, Downsview North	46	46	46	46	46	46
Bayview Village	8	8	8	8	8	8
Bedford Park, Lawrence Manor East	50	50	50	50	50	50

However, just knowing the count of how many venues in a neighborhood is not enough information, so we can continue the analysis.

### 3.3 Top 5 Venues for each neighborhood

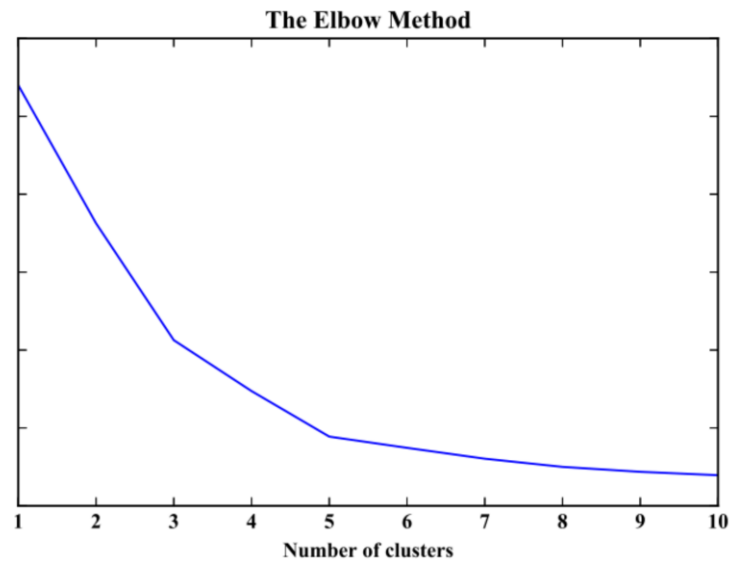
To have a more detailed look at what venues are in each neighborhood, we return the top 5 values which show the frequency of that particular category. Here are a few examples:

```
----Agincourt----
      venue  freq
0      Breakfast Spot  0.25
1              Lounge  0.25
2  Latin American Restaurant  0.25
3          Clothing Store  0.25
4      Accessories Store  0.00
```

```
----Alderwood, Long Branch----
      venue  freq
0      Pizza Place  0.22
1  Athletics & Sports  0.11
2      Coffee Shop  0.11
3              Pub  0.11
4      Sandwich Place  0.11
```

## 4. Clustering

In order to find the most common venue type per neighborhood, clustering using k-means was performed. The number of clusters was set to 5 because that is where the error – elbow curve showed where the best value would be.



The Cluster Labels were set to Latitude, Longitude and Neighborhood and were used to merge the dataframes together and perform the clustering.

Here are some examples of the output:

Out[49]:

	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	North York	1.0	Food & Drink Shop	Fast Food Restaurant	Park	Dumpling Restaurant	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Drugstore
21	York	1.0	Park	Women's Store	Pool	Yoga Studio	Drugstore	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant

## 5. Discussion

Observation: The number of venues in each neighborhood vary as different neighborhood have different sizes. Also, the size of the venue would be an important thing to note if a business is looking/deciding what the competition may look like.

### 5.1 First neighborhood analysis - Parkwoods

The first neighborhood we looked at was Parkwoods – and the 3 venues nearby were a Park, Fast Food Restaurant and Food & Drink Shop.

We then continue to analyze all the neighborhoods to compare with Parkwoods. The top 5 venues were listed for each neighborhood.

## 5.2 Finding the most common venue types

Clustering was performed to find the top 10 venue categories for each neighborhood.

For example, here is the result of the Parkwoods common venue analysis:

*Neighborhood = Parkwoods:*

1st Most Common Venue > Food & Drink Shop  
2nd Most Common Venue > Fast Food Restaurant  
3rd Most Common Venue > Park  
4th Most Common Venue > Dumpling Restaurant  
5th Most Common Venue > Discount Store  
6th Most Common Venue > Distribution Center  
7th Most Common Venue > Dog Run  
8th Most Common Venue > Doner Restaurant  
9th Most Common Venue > Donut Shop  
10th Most Common Venue > Drugstore

## 5. Conclusion

Analyzing the neighborhoods of Toronto using only open data sources was surprisingly insightful. The grouping of Toronto into neighborhoods (using Latitude and Longitude values) transforms the data into information that is easy to digest for users.

By listing the venues in each neighborhood, a person can understand what is available to them if they decide to move or work there.

Furthermore, by analyzing what are the most common types of venue categories in each neighborhood, a person is able to get a deeper understanding of each area. This information can be used for various purposes, such as what is lacking in a neighborhood so you can open a business there or even if you really enjoy fast food restaurant varieties you can move to a place that has a lot of them.

Utilizing Fourquare and Wikipedia data lead to detailed information on the city in Toronto. **The battle of the neighborhoods is concluded!**