**IBM Data Science Final Capstone Project**

# The Battle of Neighbourhoods: Saskatoon to Toronto

By:

Louis Cho

# Table of Contents

# 1. Introduction

## 1.1. Background

Given my recent move from Saskatoon, Saskatchewan to Toronto, Ontario , I wanted to look at various neighbourhoods in the city that were near my new workplace office. Some of the parameters that were key considerations on where to live were: nearby amenities such as coffee shops, shopping, nice restaurants, parks and safety (low crime). Given the vast amounts of location-based data that is available and accessible through tools like geo-coded data that maps various cities, bureaus and neighbourhoods, and apps like Foursquare that contains data on venues, category of venue types, reviews, trends and location, this project is for individuals considering relocating to Saskatoon and looking for an ideal place to live. This project will leverage data science methods of data wrangling, machine learning clustering algorithms and data visualization that will identify the best neighbourhoods to live in Saskatoon.

## 1.2 Business Problem

The major purpose of this project is to recommend the best neighbourhood to live in for a new comer relocating to Toronto. This project aims to provide insights and solutions to answer the business problem: What is the best neighbourhood to live in Toronto?

## 1.3 Approach

This project will use scraping methodologies of Saskatoono neighborhoods via Wikipedia. Extract Latitude and Longitude data of these neighborhoods via Geocoder package. Use Foursquare API as its prime data gathering source. And to compare the similarities between multiple locations, we decided to explore neighborhoods, segment them, and group them into clusters based on a set of attributes nearby amenities such as coffee shops, shopping, nice restaurants, parks and safety (low crime). To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

# 2. Data

## 2.1 Description of Data

Based on definition of our problem, we will primarily use Foursquare data to pull venues data in the Toronto area and use data attributes such as Name, Category, location and trending venues (based on foot traffic) to determine which locations have positively trending venues in a given neighbourhood as well as the mix venue types (categories) in the vicinity as well.

- Part 1: Build a dataframe that consists of postal code of each neighbourhood along with the borough name and neighbourhood name, including geocoding the longitudinal and lattidudinal coordinates of each neighbourhood in Saskatoon using the wikipedia data source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- Part 2: Scraping additional information of the different Boroughs in Toronto using Foursquare data. More information regarding the boroughs of Toronto is scraped using the Beautifulsoup library
- Part 3: Creating a new dataset of the Neighborhoods of the boroughs in Toronto with the venue data from Foursquare with their coordinates mapped to the neighbourhoods geocoded in Part 1.

## 2.2 Python Libraries Used in this Project

- Pandas: For creating and manipulating dataframes.
- Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
- Scikit Learn: For importing k-means clustering.
- JSON: Library to handle JSON files.
- XML: To separate data from presentation and XML stores data in plain text format.
- Geocoder: To retrieve Location Data.
- Beautiful Soup and Requests: To scrap and library to handle http requests.
- Matplotlib: Python Plotting Module.

# 3. Exploratory Data Analysis

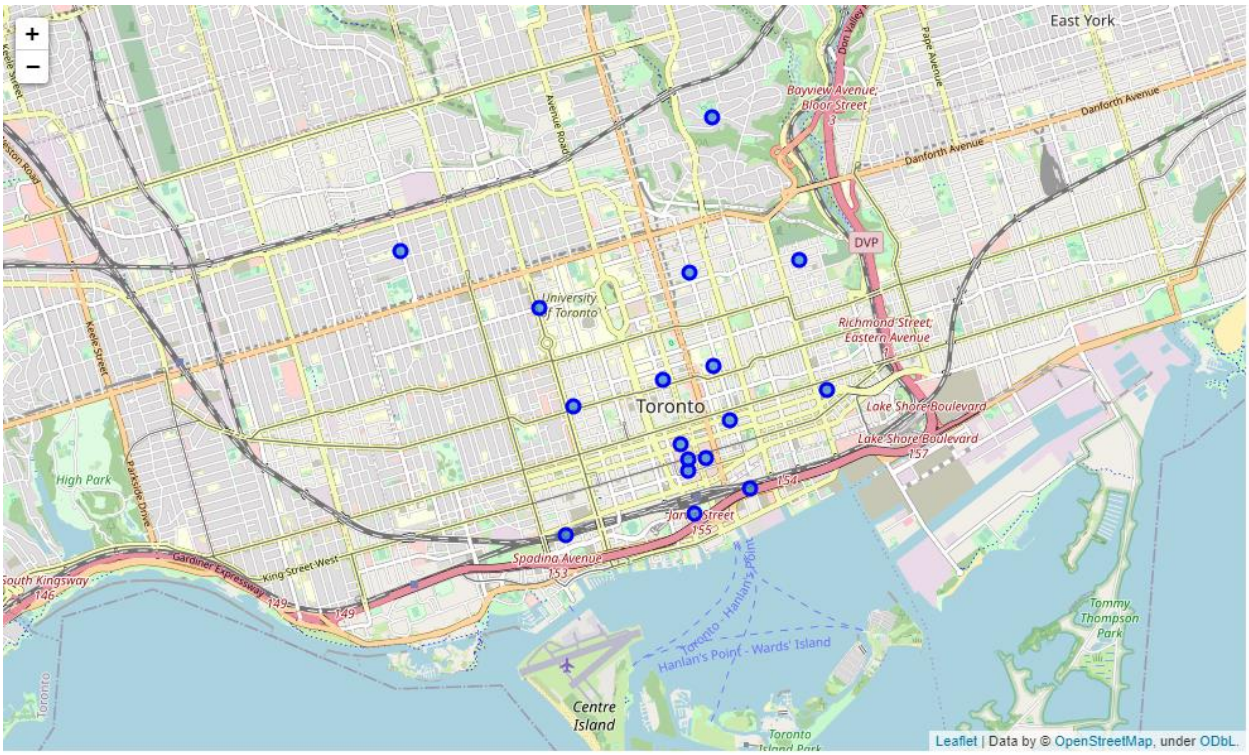
## 3.1 Data Preparation

Data was downloaded or scraped from multiple sources and combined into one table for data analysis. Wikipedia was scraped to pull FSA-level data for the postal codes in the City of Toronto (i.e. postal codes starting with “M”) to capture postal code, borough and neighbourhood descriptions. A dataframe was created using python library pandas.

Then using the postal code information scaped from Wikipedia, a geocoder was used to append geographical longitudinal and latitudinal coordinates to the dataframe for each of the neighbourhoods.

Using the geographical coordinates, the neighbourhoods were mapped using the folium mapping library in Python to visualize the locations of the neighbourhoods:

**Figure 3.1 – Map of Downtown Toronto Neighbourhoods**




## 3.2 Profiling the Neighbourhoods

Based on the neighbourhoods selected for analysis, API connection was used to pull location-based information on venues surrounding the neighbourhoods from Foursquare in order to score the neighbourhoods using this information.

Using the venue data from Foursquare, we can see the number venues that are within each neighbourhood.

**Figure 3.2.1 – Number of Venues by Neighbourhood**

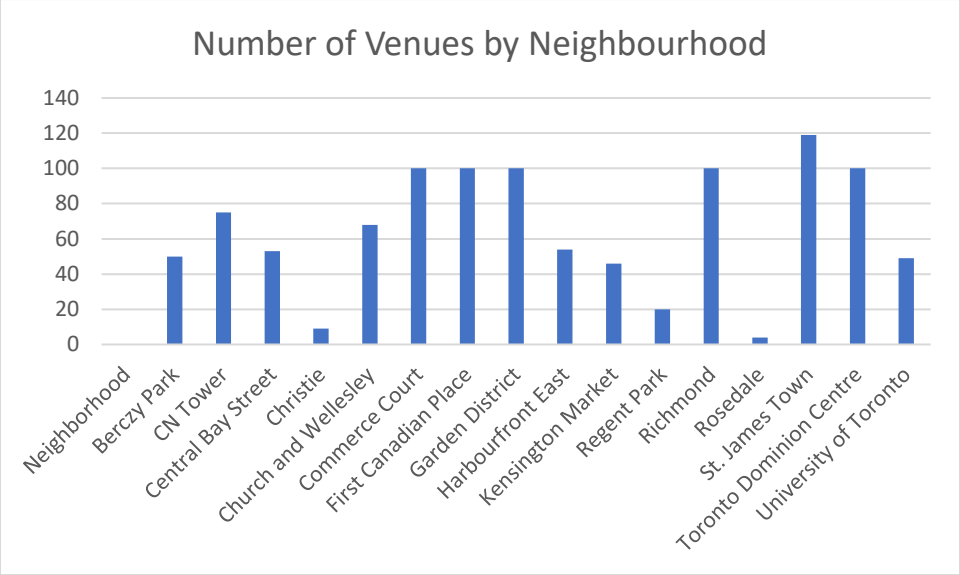
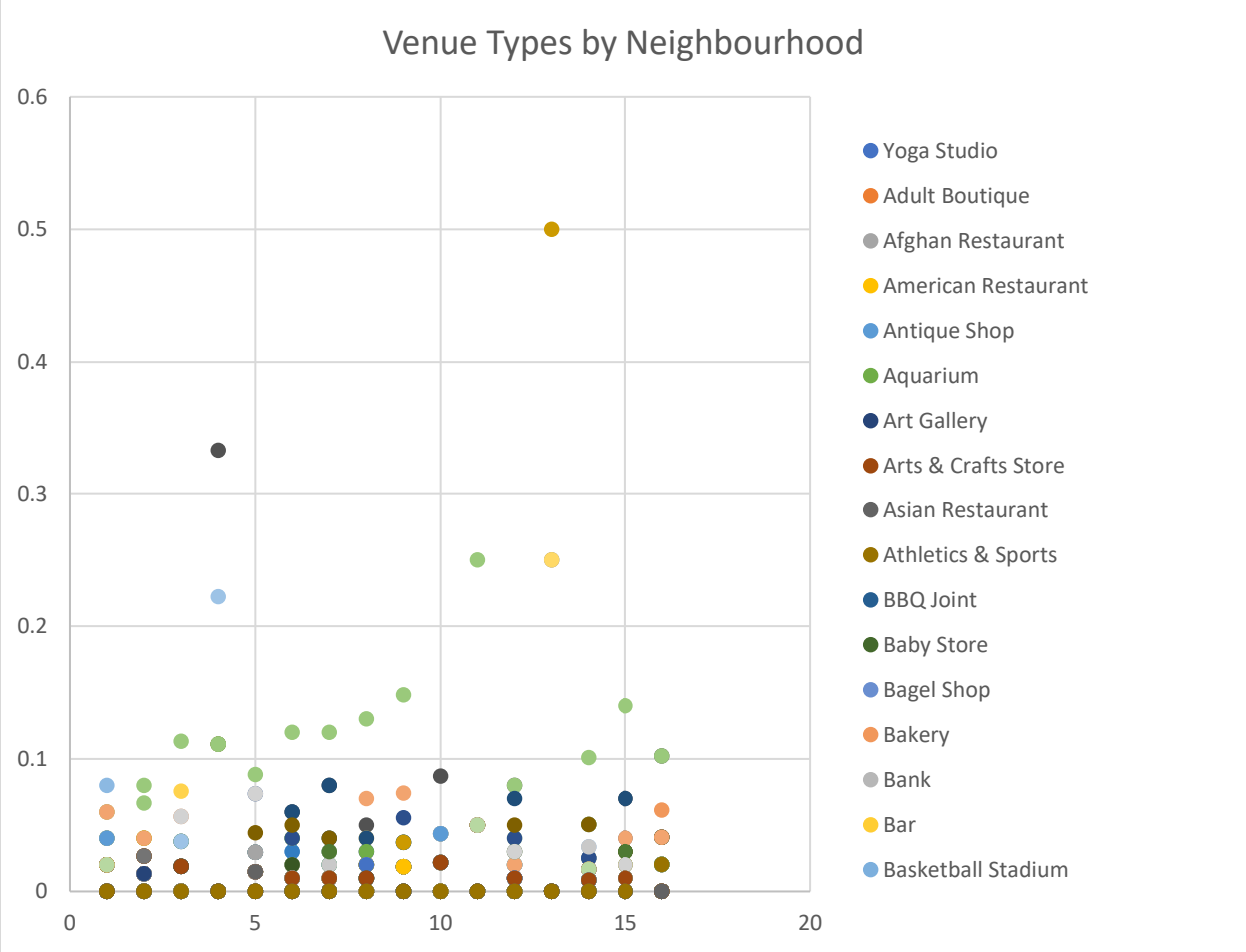



**Figure 3.2.2 – Venue Types by Neighbourhood**



Since there are 172 venue types, looking more closely at the top venues by neighbourhood would give more insightful information for the user to asses which neighbourhoods has the amenities they prefer when deciding to choose a neighbourhood to live.

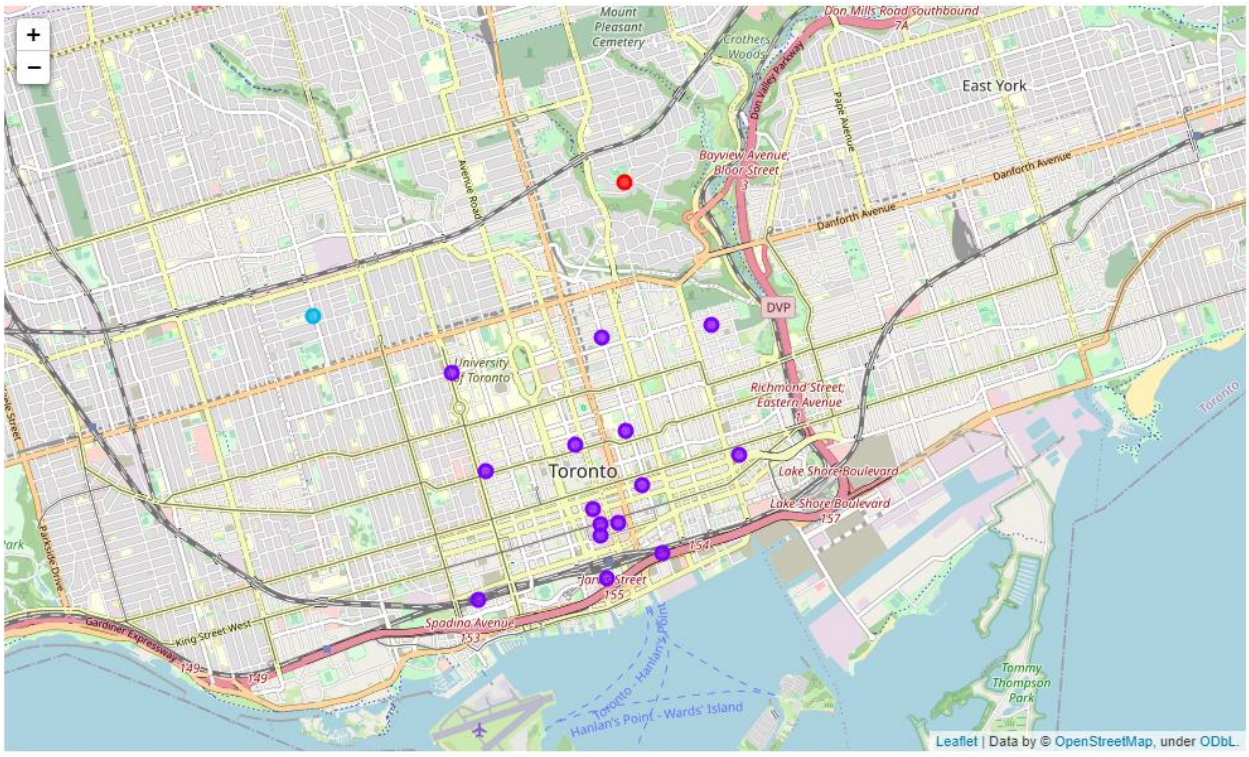
**Figure 3.2.3 – Top 10 Venue Types by Neighbourhood**

| | 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 | Berczy Park | Cocktail Bar | Sandwich Place | Coffee Shop | Bakery | Beer Bar | Farmers Market | Vegetarian / Vegan Restaurant | Seafood Restaurant | Butcher | Bistro |
| 1 | CN Tower | Italian Restaurant | Coffee Shop | Café | Park | Restaurant | Sandwich Place | Bar | French Restaurant | Gym / Fitness Center | Japanese Restaurant |
| 2 | Central Bay Street | Coffee Shop | Clothing Store | Pizza Place | Sushi Restaurant | Sandwich Place | Japanese Restaurant | Cosmetics Shop | Middle Eastern Restaurant | Restaurant | Plaza |
| 3 | Christie | Café | Grocery Store | Baby Store | Coffee Shop | Italian Restaurant | Department Store | Ethiopian Restaurant | Escape Room | Electronics Store | Distribution Center |
| 4 | Church and Wellesley | Coffee Shop | Japanese Restaurant | Sushi Restaurant | Restaurant | Dance Studio | Gym | Grocery Store | Indian Restaurant | Sandwich Place | Gay Bar |
| 5 | Commerce Court | Coffee Shop | Café | Hotel | Restaurant | Italian Restaurant | Beer Bar | Gym | Japanese Restaurant | Gastropub | Cocktail Bar |
| 6 | First Canadian Place | Coffee Shop | Hotel | Café | Gym | Japanese Restaurant | Restaurant | Concert Hall | Deli / Bodega | Steakhouse | Asian Restaurant |
| 7 | Garden District | Coffee Shop | Sandwich Place | Clothing Store | Café | Hotel | Middle Eastern Restaurant | Japanese Restaurant | Cosmetics Shop | Italian Restaurant | Pizza Place |
| 8 | Harbourfront East | Coffee Shop | Sandwich Place | Japanese Restaurant | Bank | Park | Aquarium | Boat or Ferry | Hotel | History Museum | Beer Bar |
| 9 | Kensington Market | Café | Art Gallery | Gaming Cafe | Vegetarian / Vegan Restaurant | Pizza Place | Burger Joint | Noodle House | Farmers Market | Comfort Food Restaurant | Restaurant |
| 10 | Regent Park | Coffee Shop | Italian Restaurant | Greek Restaurant | Pub | Discount Store | Restaurant | Breakfast Spot | Distribution Center | Electronics Store | Sandwich Place |
| 11 | Richmond | Café | Coffee Shop | Hotel | Restaurant | Japanese Restaurant | Gym | Breakfast Spot | Steakhouse | Asian Restaurant | Sushi Restaurant |
| 12 | Rosedale | Park | Playground | Bike Trail | College Rec Center | Department Store | Ethiopian Restaurant | Escape Room | Electronics Store | Distribution Center | Discount Store |
| 13 | St. James Town | Coffee Shop | Café | Italian Restaurant | Restaurant | Cocktail Bar | Pizza Place | Clothing Store | Bakery | Japanese Restaurant | Cosmetics Shop |
| 14 | Toronto Dominion Centre | Coffee Shop | Hotel | Café | Sandwich Place | Asian Restaurant | Japanese Restaurant | Pharmacy | Steakhouse | Restaurant | Deli / Bodega |
| 15 | University of Toronto | Café | Coffee Shop | Bakery | Sandwich Place | Bar | Pub | Japanese Restaurant | Gym | Restaurant | French Restaurant |

## 3.3 Predictive Modeling: K-Means Clustering

In order to assess the clusters of neighbourhoods based on the venue type surrounding the neighbourhoods, we used k-means unsupervised machine learning to determine the distinct clusters that exist between neighbourhoods.

**Figure 3.3 – Neighbourhood clustering using K-Means**



## 4. Results & Discussion

The objective of the project was to help those considering moving to the City of Toronto to identify the best neighbourhood in Toronto by identifying a number of relevant amenities/venues within the neighbourhood to make a decision of where best to live. This has been achieved by first making use of Toronto data from Wikipedia to identify all the neighbourhood within the borough for resident to be viable. After selecting the borough it was imperative to choose the right neighbourhood where coffee shops, parks and grocery stores were among the venues located in close proximity to each other. We achieved this by grouping the neighbourhoods into clusters to assist prospective relocators by providing them with relevant data about venues and safety of a given neighbourhood.

## 5. Conclusion

We have explored the amenities located within various neighbourhoods to profile the different types of neighbourhoods that Toronto has to offer. Based on a newcomers preferences on what amenities are important to them, such as being close to coffee shops, restaurants, pubs, or parks, this tool would help to identify what each neighbourhood has to offer and help them determine the best neighbourhood to live in the City of Toronto.