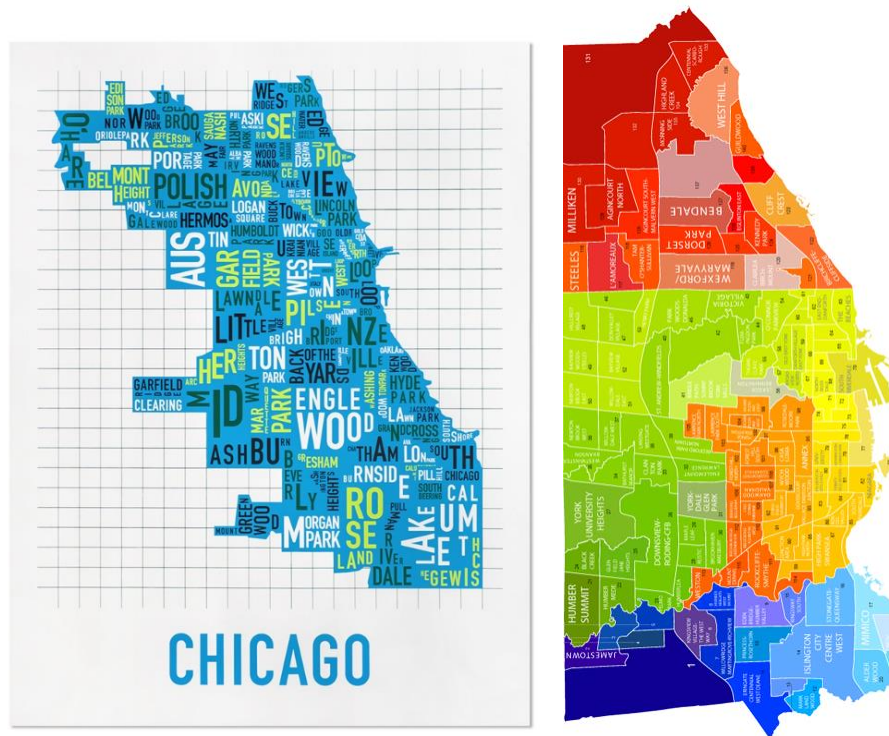


Capstone Project – The Battle of Neighborhoods

Introduction/Business Problem



Project description:

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.

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.

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.

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