The Battle of Neighborhoods week

1. Introduction

An average man change his location about twelve times in his lifetime. This brings us to some questions which are itemized below:

- Why do people move from place to place until they finally find a place.
- Do they move to settle down where they truly feel happy,
- Does what we needs change over time, prompting us to eventually leave a town we once called home for a new area that will bring us satisfaction?
- Do we often move to a new area without knowing the exact environment we moved into, which might result into forcing us to run at any sign of discomfort?
- Does social life affects people of all age group?

To reduce these frequent change of location, a proper research should be done before deciding our next move in life. Considering the following factors when picking a new place to live, you won't end up wasting your valuable time and money making moves you'll end up regretting or get tired of easily. One of the first thing to consider is "Safety" which is important before moving to a new area. If you don't feel safe in your own home, you're not going to be able to enjoy living there.Does the area have a high crime rate? Sometimes, an area may seem perfect - low taxes, moderate climate, affordable housing. However, it's important to check the prevalence of violent crimes in the area before making your final decision. Exposure to a high crime rate leaves your home and your family susceptible to theft, assault and personal harm. Crime and property value are entwined, so find out if that's why houses are cheap. You shouldn't sacrifice safety to save money.

Does the area have a high cost of living? The cost of goods and services varies from city to city and tends to be highest in the most densely-populated cities. The essential expenses to consider when relocating are: Groceries, Healthcare, Gas, Utilities, Entertainment, etc. Cost of living is the amount of money spent to maintain the average standard of day-to-day life.

How will the climate affect energy costs? You may think that leaving a high-cost city such as San Diego in favor of the more affordable Dallas is a no-brainer, but the price of homes or rent is not the only factor to consider. The moderate climate of San Diego will keep your energy costs low. Your air-conditioning bills will hit the roof when attempting to keep your home comfortable in humid Texas weather. If the climate is extreme, your cooling (or heating) costs will increase dramatically.

1.2 Problem
The statistical dataset from BeautifulSoup found on IBM-COURSERA Professional Data Science Capstone course which is a web scraping package, has information of each neighborhoods in Toronto and a wikipedia page which has a list of postal codes. This project aims to select the safest borough based on the total crimes, exploring the neighborhoods of that borough to find the 10 most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering.

1.3 Interest
Anyone considering to relocate to Toronto will be interested to identify the best borough to live, exploring its neighborhoods and common venues around each neighborhood. This project will be

using the Foursquare API to explore neighborhoods in some selected cities of Toronto. The Foursquare explore function will be used to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. The k-means clustering algorithm will also be used for the analysis. Finally, using the Folium library to visualize the neighborhoods in Toronto and their emerging clusters. The target audience for this project is twofold. Firstly, any planning to visit Toronto, Canada can use the plots and maps from this project to quickly select places that suit their budget and rating preferences. Secondly, a company can use this information to create a website or a mobile application, which is updated on a regular basis, to allow individuals to the city or even expand same functionality to other places.

2. Data Acquisition and Cleaning

2.1 Data Acquisition

The data acquired for this project is a combination of data from two sources. The first data source of the project uses a BeautifulSoup Package that shows the Borough in Toronto.
The dataset contains the following columns:

- Postcode
- Borough: common names for Boroughs in Toronto
- Neighbourhood

The second source of data is scraped from a wikipedia page that contains the list of Canada postal codes. This page contains additional information about the boroughs.
And lastly, the Foursquare API gave the project the ability to locate each borough and provide an overview of the common venues with a view on the world map.
Using the Foursquare's explore API (which gives venues recommendations), I fetched venues from the centre of Toronto and collected their names, categories and locations (latitude and longitude).

2.2 Data Cleaning

Drop any 'None' row in the DataFrame.

Drop any row which contains a 'Not assigned' value.

And all "Not assigned" will be replace to 'NaN' using numpy for convenience.
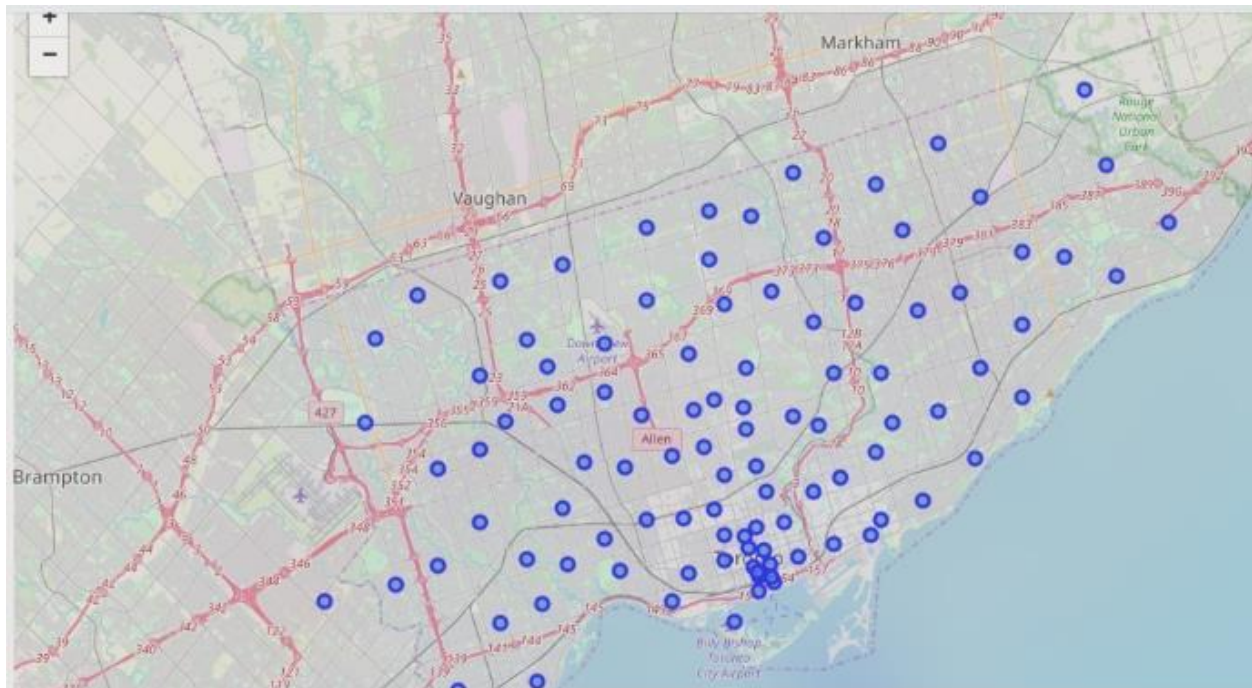
I checked that the latitude and longitude values of each corresponding venue match. After the careful analysis.
They can be categorized as follows:

1. There are venues that have specific restaurants/cafes inside them as provided by Foursquare API
2. Two locations are so close that they have practically same latitude and longitude values
3. Some venues have been replaced with new venues. Venues belonging to category 1 and 3 are perfect to keep. However, the venues that belong to category 2 should be dropped. After

careful inspection and removal, the final dataset had a total of 49 venues with which we can work. As a final dataset, we're left with 65 venues.
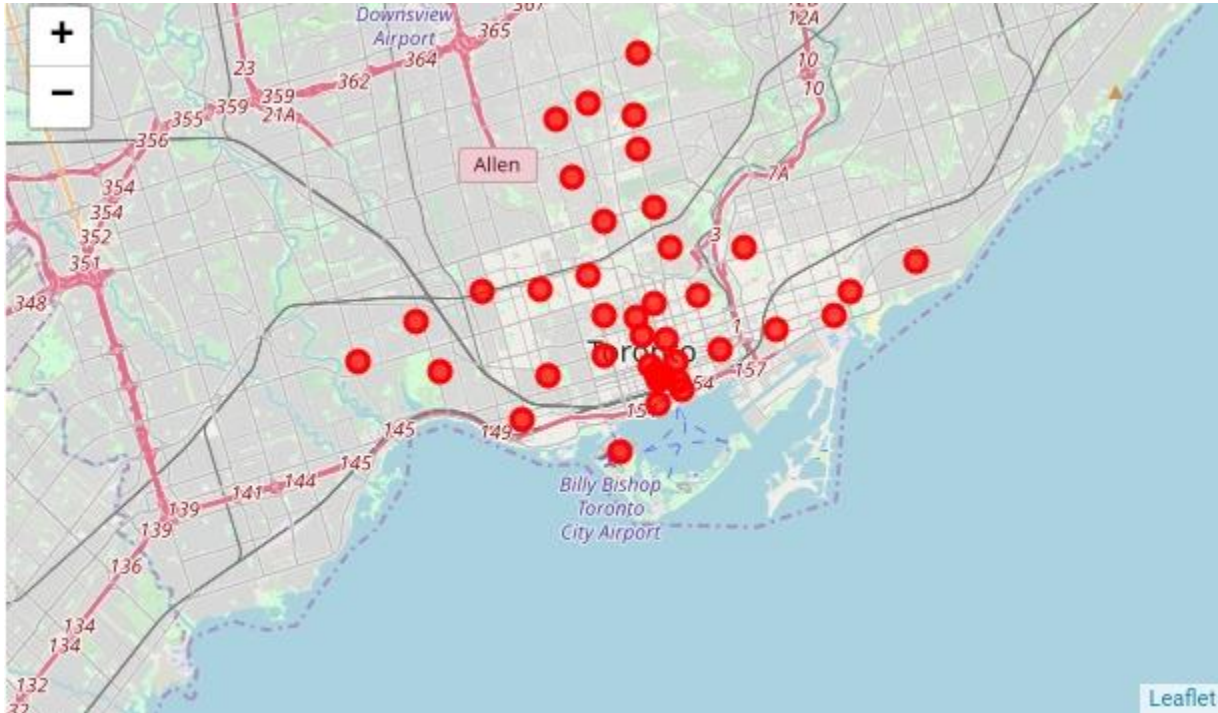
Using data cleaning, the dataset from the API will be used based on their venue names, latitude, and longitude values. One to one matching and careful data inspection would be used to remove any remaining outliers such as multiple venues at the same location from the dataset. The final data will include the venue name, category, address, latitude, longitude and rated based on the neighborhood.