Ranking Neighborhoods based on Pizza Restaurants

IBM Data Science Capstone Project

Written by: Muhammad Saleh Anwar

Introduction:

The goal of this project is to cluster neighborhoods in Toronto based on the Pizza restaurants located in the neighborhood. If you are a pizza lover in Toronto or any city in the world, you would like to know which neighborhoods in your city have the best options available for Pizza restaurants.

This project will answer the question: Can we rank the neighborhoods in Toronto based on the Pizza restaurant options available? In order to answer this question, two main factors are going to be analyzed. One is the number of Pizza restaurant options available in a neighborhood. The more options are available in a neighborhood, the more choices its residents have. Therefore, the number of options available should affect the ranking of the neighborhood. The other factor is the popularity of the Pizza restaurant options in the neighborhood. There are different metrics available that can be utilized for gauging popularity of any restaurant. Most people rely on ratings provided by platforms such as Foursquare, Yelp and Google to decide on a particular restaurant. Customers can log on to these platforms and submit reviews, ratings or likes for a particular restaurant. For this project, the number of likes associated with each Pizza restaurant is extracted using the Foursquare API. These two factors are used to cluster the neighborhoods in the Toronto based on Pizza restaurants.

Data:

In order to cluster the neighborhoods in Toronto, the list of all neighborhoods in Toronto is required. This data is extracted from the Wikipedia page titled 'List of Postal Codes of Canada' (https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=942655364). This page only contains the postal codes starting with 'M' which is the convention used for the neighborhoods in Toronto. Along with the postal codes, it also contains the name of Borough and the neighborhood name.
In addition to the list of all neighborhoods in Toronto, the geospatial data (latitude and longitude) for the neighborhoods in Toronto is also required. This is needed to perform

Foursquare API queries as well as to plot the neighborhood on a map using Folium for visualization purposes. This dataset was provided during this course and it can be downloaded from http://cocl.us/Geospatial_data. This dataset lists all the postal codes in Toronto with their latitude and longitude.

List of all restaurants in a particular neighborhood was required for this project. This was extracted through a Foursquare API call. This API call provided a list of all venues for the neighborhoods in Toronto. Using pattern matching, the list of restaurants and pizza restaurants were extracted from the venue's dataset. The number of likes submitted for each pizza restaurant was also obtained through a Foursquare API call.

In summary, the following data was collected and used for the project:
- Postal Codes in Toronto (Wikipedia)
- Geospatial data for Toronto neighborhoods (provided in previous assignment)
- List of all venues for the neighborhoods in Toronto (Foursquare API call)
- Number of likes submitted for each Pizza restaurant in Toronto (Foursquare API call)