IBM Applied Data Science Capstone –Wk4

Introduction

In the present, the young generations are increasingly becoming more conscious of their diet and activity. In an effort to reduce their risk of health care problems everyone is indulging in going to gym to fit their body fit. The instructions people receive at the gym makes them want to eat a healthy diet. This has led to the increase in people adopting plant based diet. Adopting a whole-foods, plant-based diet not only benefits your waistline, but it can also lower your risk and reduce symptoms of certain chronic diseases. Perhaps one of the most well-known benefits of plant-based diets is that they are heart-healthy. So we are going to figure out the neighbourhoods to in Canada if you are into plant based eating and gyms.

Target Audience

Hypothetically let's assume that a young man decides to move to Toronto for a prolonged stay due to work and is looking for a neighbourhood to rent an apartment. Some of his expectations for the neighbourhood are,

1. The neighbourhood should have gyms since our man is into the healthy lifestyle.
2. The neighbourhood should have plant based restaurants since our man doesn't know how to cook and orders in very frequently.

Business Problem

A neighbourhood that houses a sufficient number of gyms and a plant based or vegan restaurants needs to be identified in Toronto. Machine learning algorithms such as clustering needs to be performed to group neighbourhoods to identify which will be suitable for our hypothetical man here.

Data

To solve the problem, we will need the following data:

- List of neighbourhoods in Toronto: This is scraped from a Wikipedia page that contains a list of postal codes, boroughs, and neighbourhoods. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data. This is acquired through the Foursquare API.
- Venue data, particularly data around each neighbourhoods. We will use this data to perform clustering on the neighbourhoods.