Battle of the neighbourhoods (Week 1)

Using Machine Learning to find locations to open up a Sports Nutrition shop

Robert Ashton-Jones (https://github.com/ RobEAJ123)

1.Introduction

1.1 Background

For this Capstone project, I am creating a hypothetical scenario, where a Canadian entrepreneur is looking to open a new sports nutrition store. This project aims to clarify where the ideal location to open a store in Toronto would be. This will help the store to gain reputation quickly with the idea of expanding quickly. The starting location for a store like this is important, as it will define how profitable the store can be (and will ultimately affect how quickly the next store will be opened as a result). Finding the location to open such a restaurant is one of the most important decisions for this entrepreneur and I am designing this project to help him find the most suitable location.

1.2 Business Problem

The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new sports nutrition store in Toronto. Using data science, including machine learning methods (e.g. clustering), this project aims to answer the question: “Where to start?”

1.3 Target Audience

The entrepreneur looking to begin his sports nutrition start-up.

2.Data

To examine this problem, I will use data from the sources below:

● List of neighbourhoods in Toronto.

● Latitude and Longitude of said neighbourhoods.

● Venue data related to the neighbourhoods in Toronto. This will help us to maximise the footfall for people who may visit the store.

I will use the data above to determine the frequency of different types of venue. This will help us to infer whether a neighbourhood is more interested in restaurants (in which case, a nutrition store will see less footfall) or other venues including: Gyms, parks, or health food stores.

3. Extracting Data

● Scraping of Toronto neighbourhoods (from Wikipedia)

● Latitude and Longitude data of the above neighbourhoods (from the Geocoder package).

● Venue data related to these neighbourhoods (from Foursquare API)