

Introduction

The COVID-19 epidemic has led to many changes. One of them is an increase in people working remotely from their homes. This change has led many of these types of workers to move to residences in new cities. Once they've moved, people will need to change their existing routines. For example, they will need to find new grocery stores at which to buy their food. Depending on their routines, it might not be easy for them to adapt to their new neighborhoods.

Some people follow customized health plans. These plans describe the types and amounts of food that people should eat and the types and frequency of exercise they should do. A fictional online company, XYZLifestyle.biz, has been creating such plans for their customers for a few years.

Over the past few months, increasing numbers of the company's customers have been asking for a specific type of help. These people have moved and are difficulty finding businesses in their new neighborhoods that can provide them with the types of food and exercise listed in the health plans they've purchased from the company.

The company wants to provide a new service. It will gather information to determine which neighborhoods in cities have the most health-related venues, such as parks and restaurants with healthier menus. Customers will be able to search through this information, to assist them when they're thinking about moving and want to know which places in a city would be a good fit for them. This new service will help existing customers and may encourage new people to become customers.

To test this new service, a prototype will be created. It will focus on neighborhoods in the city of Toronto.

Data

Wikipedia and the geocoder API will be used to determine the names and locations (latitudes and longitudes) of the various neighborhoods in the prototype's city, respectively. The postal codes of the neighborhoods (from Wikipedia) will be used by geocoder to determine their locations.

Foursquare will be used as the source for data about the venues in the various neighborhoods of the prototype's city. The data lists the venues' names, locations, and categories.

The data from Wikipedia and the geocoder API has been gathered and combined with the data from Foursquare. Next, the category names of the venues will be used to filter the health-related venues from the other ones. The frequencies of the health-related venues will be calculated in each neighborhood. With this dataset, k-means cluster modeling will be used to find neighborhoods with higher concentrations of health-related venues.