Selecting location for a new health cafe in North York, Canada

1. Introduction:

When it comes to deciding a neighbourhood to position a new cafe, the important aspect to consider is who is the target audience and which neighbourhood would provide the maximum accessibility to that audience.

Let's say that the new cafe is a health cafe serving high protein, low calorie drinks and snacks. The best neighbourhood for such a cafe would be one with maximum number of health and training centres, gyms and other workout places and minimum number of other cafes.

It is not feasible to suggest a single neighbourhood to the entrepreneur since we are not aware of her budget and the property prices in that specific neighbourhood. So, instead we will suggest the neighbourhoods to be preferred based on number of fitness centres and neighbourhoods to be avoided based on number of cafes/restaurants. We will also create clusters of neighbourhoods based on their similarity in terms of number of cafes, types of restaurants, number of workout studios, etc. This will give entrepreneur the flexibility to select a neighbourhood based on her requirements. In the end we will visualize these clusters on the map.

2. Data Description and Collection:

Data for this project is collected and merged from three different sources – Wikipedia, Geospatial Coordinates csv and Foursquare API. We perform the following steps in order and generate the required dataset for analysis

   a. Scrape list of postal codes of Canada from Wikipedia to get Postal Code, Borough, and Neighbourhood details
   b. Read the Geospatial Coordinates csv to get latitude and longitude details corresponding to each postal code
   c. Merge the datasets obtained in a) and b)
   d. Fetch top 100 venues that are within a radius of 500 meters to all the neighbourhoods in North York using Foursquare API based on latitude and longitude details collected in c)

The data collected from Foursquare API contains Neighbourhood, Neighbourhood Latitude, Neighbourhood Longitude, Venue, Venue Latitude, Venue Longitude, and Venue Category details. Once we have all this information, we will fetch the neighbourhoods which have fitness related venues, and which have the largest number of eateries. This will give a good estimation of preferred and undesirable locations for a new health café in North York. To provide flexibility to the entrepreneur in terms of budget and the property prices, we performed clustering on the neighbourhoods to provide a list of locations similar to the preferred (undesirable) ones.