Clustering Toronto neighborhood areas by surrounding restaurant type

Advay Janardhan
June 24, 2020

Data

The first dataset was scraped from the [Wikipedia page on Toronto postal codes](). This page has a list of all postal codes that start with M, giving a list of all Boroughs and Neighborhoods in Toronto. This was imported in table format with the Postal code, Borough, and Neighborhoods as columns. The second dataset came from a .csv file. This has the geographical coordinates for each postal code, which was used in conjunction with the first set. This was also imported in table format with postal code, latitude, and longitude as the columns. The third dataset was found by connecting to the FourSquare API. It yielded a list of venues within a 1000 m radius from each Neighborhood location used. This was imported by analyzing the .json file extracted from FourSquare, ultimately into a table with Neighborhoods, Neighborhood Latitude, Neighborhood Longitude, Venue, Venue Latitude, Venue Longitude, and Venue Category as columns.