IBM Applied Data Science Capstone

Opening a New Restaurant in Toronto, Canada

By Suyash Singh

June 2020

Introduction

With the rising economic growth in the Greater Toronto Area (GTA), people have higher spending power, and this can be seen in the growing number of hotels, restaurants and entertainment centers. The diversity in this region has allowed the rise of a variety of cuisines being offered. We have been approached by a property dealer to identify potential areas for a new restaurant business. We are tasked with the goal of identifying what cuisines would work in each area.

Business Problem

The investors are not picky about the cuisine and want to choose the cuisine that will give them the highest profits. They are looking to open a restaurant in the Toronto area. We are required to analyze the different areas of Toronto and identify the more popular cuisines in these areas. For this purpose, we will implement a k-means algorithm to cluster different areas of Toronto based on the popular restaurants in the region. At the end of the analysis, we will be able to identify regions with potential for a new restaurant and will recommend the cuisine to offer.

Data

The data used in this analysis come from:

1. Wikipedia
2. Foursquare API

Wikipedia(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) provides a table containing postal codes and the associated neighborhoods of Toronto. We use the BeautifulSoup package in Python to scrape the table from the website and modify it for our use.

We will then extract the latitude and longitude values for these neighborhoods using the Geocoder package in Python.

List of restaurants will be extracted using the Foursquare API. We will use the latitude and longitude values of the neighborhoods from previous step and find restaurants within 500 metres radius.

Once we have cleaned the data after extraction, we will apply k-means clustering algorithm to identify the identical clusters.