Coursera Applied Data Science Capstone Project

Exploring the opportunity to open a new restaurant in Toronto, Canada

By Kiet Phung

May 2020

Introduction

The city of Toronto is very diverse. Different ethnic neighborhoods throughout the city focus on specific cuisines, for example Chinese and Vietnamese cuisines found in the city's Chinatowns, Korean food in Koreatown, Italian cuisine in Little Italy, etc. In addition to ethnic cuisines, Toronto is also home to many fine dining establishments and chain restaurants ranging from fast food to casual or upscale dining.

One purpose of the analysis is to help people to open a new restaurant in Toronto, Canada based on the provided data on income, neighborhood population, and other existing restaurants within the same region. Another purpose is to help travelers to select a correct to dine based on the location of the restaurant, what type of food, the cost of the meals, etc.

In order to solve these above questions, we need to create a system to help travelers and restaurant owners to find the recommended restaurants as well as to understand the current food / restaurant market conditions in Toronto.

Data Source

In order to provide the necessary information to people who want to open new restaurants in Toronto, Canada, I will need to obtain information from the city's most recent census regarding population and average income from different neighborhoods in Toronto. I also need to use Foursquare API to obtain information regarding restaurants that already exists in the same area.

The neighborhood data of Toronto is taken from the Wikipedia page:

https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Toronto

This website provides public information on Toronto Census:

https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/#8c732154-5012-9afe-d0cd-ba3ffc813d5a

Beautiful soup library is used to extract data from the Wikipedia page. . The geocode library to be used is geopy. Folium library will be used to represent the data on maps. And scikit-learn will be used to utilize machine learning. And Foursquare API will be used to gather neighborhood data.

Importing data is divided into 3 stages. The first stage is getting list of neighbourhoods of the four cities from the above Wikipedia links. The second stage is getting location of neighbourhoods. The final stage is getting the venues in the neighbourhoods from Foursquare.