IBM APPLIED DATA SCIENCE CAPSTONE

Where to Open a Restaurant in Toronto

Xinyao Jie

Introduction

This project is a capstone project for IBM applied data science course by coursera. The topic I choose is to explore where to open a new restaurant in Toronto. For many customers, going to restaurant is a great way for food after working for full day. It is also a great way to relax and enjoy themselves with their friends during weekends and holidays. For restaurant owners, the central location and the large crowd provides a great distribution channel for them. Property developers are also taking advantage of this trend to open a new restaurant to cater to the demand. As a result, there are many restaurants in the city of Toronto and many more are being built. Opening new restaurants allows owners to make money by selling foods. Of course, as with any business decision, opening a new restaurant requires serious consideration and is a lot more complicated than it seems. Particularly, the location of a restaurant is one of the most important decisions that will determine whether the restaurant will be a success or a failure.

Business Problem

The objective of this capstone project is to analyze and select the best locations in the city of Toronto, Canada to open a new restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Toronto, Canada, if someone is looking to open a new restaurant, where would you recommend that they open it?

Target Audience of Project

This project is particularly useful to someone who want to start their own restaurant and investors looking to open or invest in new restaurants in the city of Toronto. Also, it provides some guidance for tourists and customers about the density of restaurants in different neighborhoods.

Data

The following data is needed to solve this problem

- List of neighbourhoods in Toronto. This defines the scope of this project which is confined to the city of Toronto.
- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to different restaurants. We will use this data to perform clustering on the neighbourhoods.

Sources of data

This Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) contains a list of neighbourhoods in Toronto, with a total of 103 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python Pandas packages. Then we will get the geographical coordinates of the neighbourhoods using a csv. File I already downloaded. I use Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods to get that data. After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Restaurant category in order to help us to solve the business problem put forward. Notice there are mant kinds of restaurant like Asian Restaurant, African Restaurant, etc. We will consider all of them in one piece. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).