

COURSERA CAPSTONE

IBM Applied Data Science Capstone

Opening Restaurant in Toronto, Canada

By : Bernardo Andrey Panggabean

July 2020



Introduction

Restaurant is the a place where cooked food is sold to the public. Many people visit to Restaurant because hang out with friends, meet family or many more. Many businessman visit to Restaurant because want to client and establish cooperation. So, Location is an important aspects. We have to know if we have any competitor in the area, the crowd of the area, rent cost of the area etc. Of course, as with many business decision, opening a new restaurants requires serious consideration and a lot more complicated than it seems. Particularly, the location of the restaurants is one of the most important decisions that will determine whether the restaurants will be a success or a failure. If, restaurants is built in quiet place, so restaurants will be a failure.

Business Problem

The objective of this capstone project is to analyze and select the best location in Toronto, Canada to open a Restaurants. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question

Target Audience of This Project

This project is particularly useful to property developers and investors looking to open or invest in new restaurants in Toronto, Canada. This project is timely as the city is currently suffering from oversupply of Restaurants. Recently, there are a lot of restaurants are opened in Toronto. Chairman Specialty Restaurants Association of Canada predict the number of Restaurants in Toronto will increase to 20% by the end of the year. So this project will be useful for anyone who want to open Restaurants in Toronto.

Data

To solve the problem, we will need the following data:

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

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) contains a list of districts in Canada. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and BeautifulSoup packages. Then we will get the geographical coordinates of the districts using Python. After that, we will use Foursquare API to get the venue data for those districts. Foursquare has one of the largest database of 105+ million places and is used by over 125000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Restaurants category in order to help us solve the business problem described above. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), machine learning (K-means clustering) and map visualization (Folium).