

Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Coffee Shop in the city of Toronto

By: Jaekisen Agarwal

January 2020

Introduction

For many customers, visiting coffee shop is a great way to relax and enjoy themselves during weekends and holidays. It provides a friendly, comfortable atmosphere where the customer can receive quality food, service and entertainment at a reasonable price. Opening a successful coffee shop can be a rewarding experience. Because of this, hundreds of friends will have great conversations. Because of this, mornings will be brighter and afternoons will seem less stressful. Opening a cafe takes a big investment in both time and money. It's essential to spend time reaching out to coffee business owners and learning from their experience; finding out what works, and what doesn't. And here's the fun part — it also means visiting lots of cafés to get an insight into what you want your business to be like. Consider what you will take from other businesses and what will make you different. Learn about your customer base. Who will they be? What are their needs? What time of the day will be busiest? Knowing your customers well will assist with planning, creating a menu, price points. Particularly, the location of the coffee shop is one of the most important decisions that will determine whether it will be a success or a failure.

Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Toronto to open a new coffee shop. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: Recommend a good place to open a new coffee shop in the city of Toronto.

Data

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

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

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 neighborhoods in the city of Toronto, with a total of 103 neighborhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and BeautifulSoup package. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods. 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 coffee restaurant category in order to help us to solve the business problem put forward. 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). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.