

Applied Data Science Capstone

Opening a new Bookstore in Toronto, Canada

By: Riyadh Bin Rafiq

June 2019

Introduction

A bookstore is a place where different kinds of textbooks or reference books are sold. It is a very important place for all educated persons. Most of the customers of the bookstore are book lovers of all ages. It is mostly familiar to the students. In a bookstore, there are different shelves which are nicely arranged. From a book shop, we can buy our textbooks, story books, novels, dictionaries, science fiction books etc. In fact, a bookstore always plays an important role for the readers and mostly the students. There are many bookstores in the city of Toronto. Serious considerations are required to open a new bookstore than it seems. So the location of the bookstore is one of the most important decisions that will determine whether the bookstore will be a success or failure.

Business Problem

The objective of this project is to analyse and select the best locations in Toronto city to open a new bookstore. For this project, data science methodology and machine learning techniques like clustering are used. The provided solution gives answer to the question: If an investor wants to open a new bookstore, what would be the best place that you recommend to open? The project is useful to investors or book lovers who want to open a new bookstore in the Toronto city.

Data

To solve the problem we need the following data:

- List of neighborhoods in Toronto.
- Latitude and longitude coordinates of those neighborhoods. We will use this data to plot the map and also get the venue data.
- Venue data specially bookstores are used to perform clustering.

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, this

Wikipedia page contains our required data including post code, borough and neighborhoods. I will extract the data from the wikipedia page using web scraping techniques. I will also get latitude and longitude of our specified neighborhoods. Then I will use Foursquare API to get the venue data of those neighborhoods. Foursquare API usually provides various categories of the venue data. In this project, I am particularly interested in bookstore data which will help to solve the business problem. Data science methodologies are used in this project such as data scraping, data cleaning, data wrangling, machine learning i.e. K-means clustering, map visualization.