IBM Data Science Capstone Project - Week 2

Keivan Mokhtarpour
Coursera

August 17, 2020

1 Introduction

This is a capstone project for IBM Data Science Professional Certificate. In this project, a data-analyzed approach to find the best possible locations for opening up a coffee house in the greater Toronto area (GTA) is considerable. As you may know, many immigrants enter Canada every year using investment and entrepreneurship programs. Most of them end up buying local businesses or simply opening up a franchised coffee house like Tim Hortons or Starbucks in large cities where the turnover of their investment is somehow guaranteed. Therefore, many of them are concerned about the location as a key factor in their annual total sales. GTA has been host to many immigrants over the past few decades and they have enlarged the city in terms of size by occupying new neighbourhoods and territories. The development of the Toronto city has extended its boundaries all the way to the northern regions like York where many immigrants specially Iranians, Koreans and the Chinese have chosen North York as their new home. This is a challenging decision for many of them whether to buy their first business in the north or prefer Downtown Toronto over it. This project is aimed at helping them to narrow down the search area they are looking for opening up a new coffee house that is a $6.2 billion industry in Canada.

2 Business Problem

The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new coffee house in the following boroughs of the greater Toronto area (GTA):

- Central Toronto
- Downtown Toronto
- East Toronto
- East York
- North York
- West Toronto

By using data science methods and tools along with machine learning algorithms such as clustering, this project aims to provide solutions to answer the business question: In Toronto, if an entrepreneur wants to open a coffee house, where should they consider opening it?

3 Client

The entrepreneur who wants to find the location to open a coffee house.

4 Data

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

- List of neighborhoods in Toronto, Canada. (Web Scrapping through Wikipedia)
- Latitude and Longitude of these neighborhoods (Geocoder Package)
- Venue data related to coffee houses (Foursquare API). This will help us find neighborhoods that are more suitable to open a coffee shop.