IBM Data Science Capstone: Week 4

August 8, 2019

Overview

This document was made for the Capstone project of Coursera's IBM Data Science Specialization. Using the Foursquare Developer API, we were asked this week to formulate the beginnings of a Data Science project, starting with the overall problem and background, followed by the description of the data an how it will be used.

The main machine learning tactic I will use is K-Means Clustering to find similarities in different Canadian cities, both on diversity and immigration-related data, as well as data from Foursquare about the different kind of food places available.

Goals

The analysis can be split into two scenarios:

1. **If you'd like to open an restaurant**

Say you'd like to open a new restaurant and you're looking for areas to set up your new shop. Maybe you'd like to open an Italian restaurant, or a Sandwich shop, or a Vietnamese restaurant. You're thinking of a certain price point -- maybe you're a stickler for fine dining or maybe you want to open a family-owned shop with cheap eats. Based on your requirements, the following analysis will pinpoint the cities in Canada that have proven to be the most popular foods at that price point.

2. **If you're a new immigrant or looking to move within Canada**

Though Canada is known to be diverse, some places are just more diverse than others. For example, almost half of Toronto's population is made up of immigrants. Maybe you want somewhere to feel at home, and you want somewhere with a wide variety of different cultural

foods. The clustering will help you do that, and also find an intersection between kinds of food, as well as the diversity of the population.

Data

A few datasets will be used, as follows:

For diversity measures amongst population:

- From Statistics Canada (2016 Census)
    - Immigration Data
        - This details the number of immigrants in Canada, in each of its major cities, distribution of immigratnts in Canada, and proportion compared to the rest of the population. It also reports these numbers for immigrants between 2011 and 2016 only.
    - Population Data
        - Only one column is needed: the population of each city. This will be used with the visible minorities data to figure out the proportion of Visible Minorities in each city
    - Visible Minorities
        - Statistics Canada defines a visible minority as: 'Visible minority' refers to whether a person belongs to a visible minority group as defined by the Employment Equity Act and, if so, the visible minority group to which the person belongs. The Employment Equity Act defines visible minorities as 'persons, other than Aboriginal peoples, who are non-Caucasian in race or non-white in colour'. The visible minority population consists mainly of the following groups: South Asian, Chinese, Black, Filipino, Latin American, Arab, Southeast Asian, West Asian, Korean and Japanese.

For coordinates I used a dataset from Simplemaps.com: Coordinates of major cities in Canada

For venue information, I used the Foursquare Places API. In particular, I used the explor endpoint for each city to find that 50 closest “food places”. I excluded venues that are categorized as Cafes, since I would consider them a different kind of venue all on its own, and I

excluded ‘Restaurants’ since this was too broad a term and probably arises from bad venue tags.