Applied Data Science Capstone – Capstone Project

The Battle of Neighborhoods
In Toronto

By: Manuela de Oliveira

October 2020

Contents

A brief description of the Problem

In this final project we were invited to explore a hypothetical situation. We live in Toronto and love our neighborhood mainly because have everything we need to feel like “home”. Great amenities and other types of venues such as gourmet fast food joints, pharmacies, parks, grad schools, we have all.
But “sadly” we receive a job offer from our dream company on the other side of
the city with great career prospects and is not feasible to keep living on our current place, and we need to move if we decide to accept the offer.

So, the challenge is to discover similar neighborhoods on the other side of the city! And the objective of this Capstone project is to analyze and select the most likely neighborhoods in the city of Toronto to move in if we accept changing job.

Toronto in Wikipedia words

«The strength and vitality of the many neighbourhoods that make up Toronto, Ontario, Canada has earned the city its unofficial nickname of "the city of neighbourhoods." There are 140 neighbourhoods officially recognized by the City of Toronto and upwards of 240 official and unofficial neighbourhoods within the city's boundaries. Before 1998, Toronto was a much smaller municipality and formed the core of Metropolitan Toronto. When the city amalgamated that year, Toronto grew to encompass the former municipalities of York, East York, North York, Etobicoke, and Scarborough. Each of these former municipalities still maintains, to a certain degree, its own distinct identity, and the names of these municipalities are still used by their residents, sometimes for disambiguation purposes as amalgamation resulted in duplicated street names. The area known as Toronto before the amalgamation is sometimes called the "old" City of Toronto, the Central District or simply "Downtown".»

https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Toronto

Data

To solve the problem, we will need the Toronto city data containing the neighborhoods and boroughs (postal codes, latitude and longitude coordinates). With this data it will be plot the map with the venue data to perform further analysis of the neighborhoods in order to better decide.

Data Source and methods to extract them

Toronto City data containing the neighborhoods and boroughs will be obtained from the open data source: http://cocl.us/Geospatial_data. After it, we will get the geographical coordinates of the neighborhoods (postal code, latitude and longitude).

After that, we will use Foursquare API to get the venue data for the neighborhoods defined at the previous step. Foursquare has one of the largest databases providing many categories of the venue data, that will help us to better decide.

Additionally, this project will require to apply many data science skills, such as web scrapping (open source dataset), working with API (Foursquare), data cleaning, data wrangling, and to map visualization (Folium).