Capstone Project - Report

“The Battle of Neighborhoods”

in Old Toronto (Week 4)

for the Applied Data Science course (IBM)

by Coursera

Introduction, The business problem

The current study will try to provide an optimum location for opening a new business in a specific city, based on:

- A neighborhoods property, e.g. the second most common language spoken (after English) in the neighborhood,
- The number of competitors in the neighborhood,
- The population density in the neighborhood.

Let's investigate in the ***old City of Toronto*** (Canada), and propose the best possible place for opening ***a new restaurant with ethnic cuisine***. Assume that preferably we would like the new restaurant to be located in a neighborhood with a high degree of the same ethnic characteristics, i.e. assume the languages spoken in that neighborhood, so to make advantage of the cultural element of the area. In order to sustain the new business, there should be a lot of population and the less number of competitors possible.

The results could be highly usable ***for people having ethnic cooking skills or restaurant-businessmen, who want to open an ethnic restaurant in a neighborhood having some degree of the same ethnic culture in Old Toronto***

Data description

The datasets that will be used are for retrieving information about city's neighborhoods and their characteristics are taken from Wikipedia website for the Toronto demographic information:

- https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods

The geographic coordinates of the neighborhoods of Old Toronto are taken from:

- File 'oldToronto.csv' which the latitude and longitude per neighborhood in Old Toronto area.

Combining the above data sets, we get ***demographic information, focused on Old Toronto's neighborhoods*** and the exact coordinates per neighborhood. Then by using the Foursquare API, we can retrieve further information for venues, venue categories and venue coordinates for every area. ***The Foursquare data set combined with the neighborhood's data set with demographic information will be the main data set*** that we will be used for the analysis. Visualization of the results via maps and graphs, where possible, will help to explain the data.

Based on the language spoken (second language spoken after 'English'), the neighborhood's population and the number of ethnic restaurants (restaurants with ethnicity common with the language spoken) in a neighborhood *the best possible set of candidate neighborhoods can be retrieved*. Then by using ***k-means algorithm*** the candidate neighborhoods will be further analyzed. The final results, via tables and maps will conclude on finding ***the best neighborhood to start an ethnic restaurant in an ethnic-cultural neighborhood,*** show any existing patterns and similarities between ethnic restaurants and ethnic populated neighborhoods in Old Toronto area.