Capstone Project;
The Battle of the Neighborhoods;

Report prepared by Cristiano Venanzoni

Milan, April 2020

Summary

Disclaimer

The following report has been prepared as part of the IBM Data Science Certification course. The assignment was requested as final assignment of the Applied Data Science Capstone module.

The report contains consideration from the author, which is a beginner in the data science field.

It should be used solely for learning or practice purposes.

Introduction/Business problem

- The objective of the assignment is to leverage on Foursquare location data to compare neighbourhoods.
- The idea is to build a model that can compare neighbourhoods of 2 cities (A and B) based on the level of similarity of their most popular venues.

- The model will offer insights among neighbourhood of different cities and can be used in case of relocation but also for tourism purposes: when visiting city A, tourists will be able to pick a neighbourhood similar to one they know of their hometown (city B).
- The scope of the project is to build a model that can be easily understood and customized by users, adding data from any city they would like to test the model with.

Data

- The model will combine data from cities and datasets of venues with datasets of cities based on neighbourhood geographical codes.
  - Venues:
    - Venues data will be retrieved using Foursquare (www.foursquare.com).
    - In particular the venue/explore? function will be used to retrieve data of popular venues based per neighbourhood.
  - Cities:
    - The model will use 2 given cities data (for simplicity we used Toronto and New York), but as said before, the model can be easily customized and any city data can be inserted by the user.
    - Data of Toronto will be retrieved from: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M , while the geo data will be obtained by the following csv file: http://cocl.us/Geospatial_data
    - Data of NYC will be retrieved from the following csv file: https://cocl.us/new_york_dataset
    - All datasets are publicly available; user credentials (client_ID, client_secret_ version) are needed to access Foursquare data.