IBM APPLIED DATA SCIENCE CAPSTONE COURSE

THE BATTLE OF NEIGHBORHOODS PROJECT – WEEK 4 REPORT

1. INTRODUCTION

This is a capstone project for IBM Applied Data Science and Data Science Professional Specializations. In this project, I will discover the opportunities to found "coworking space" in Toronto, since coworking spaces gain huge popularity between freelancers, small companies which cannot afford to have an own-office and entrepreneurs by offering lower costs and interaction with other employers. Therefore, the people who are planning to find suitable places for founding the coworking space are the main target audience for my project. At the end of this project, this target group can find the best alternative neighborhoods to open coworking space in Toronto, Canada.

2. DATA

To extract the best alternative neighborhoods in Toronto, it is needed to have list of neighborhoods in Toronto, coordinates of these and finally venue data related to existing coworking spaces in these neighborhoods.

To utilize these data, firstly, I will scrap the Toronto postal codes, boroughs and neighborhoods from related Wikipedia page via Python beautifulsoup package. Then, getting the coordinates of each of these postal codes from this link (http://cocl.us/Geospatial_data) in the csv format. And finally, by utilizing Foursquare API, I will get the coworking space venues data.

By merging the first 2 data sources, I will have the coordinates of each neighborhoods in Toronto. Then, with the help of Foursquare venue data, I will list all coworking spaces in these neighborhoods. Lastly, I will cluster the neighborhoods based on all venues and observe the clusters which offers the opportunity founding the new coworking space.