Title: Michigan Neighbourhood Segmentation

Md Asif Shahjalal

Introduction:

Suppose, a person has been living in Dearborn, Michigan for 15 sweet years of his/her life. All on a sudden, he may have to leave Dearborn and relocate to Grand Rapids, MI for a change in his job location or some other event. Now, he has been used to a particular lifestyle for a longtime. Maybe he likes to go to Mexican restaurants for breakfast, maybe he loves to visit some kind of park in the weekends or maybe he is a party guy. Now, he would more like to choose a neighborhood in Grand Rapids which has all the amenities he was used to in a close proximity. To come to his rescue, we will work on a Neighborhood Segmentation on the state of Michigan.

Objective:

1. Applying different clustering algorithms to cluster the neighbourhood based on their similarities in different amenities and venues.
2. For defining success we will try to figure out the optimal cluster size by doing some exploratory data analysis on different clusters and trying to observe their similarities.

Initial approach:

We will try to cluster our dataset using different clustering techniques like K-means algorithm, Agglomerative Clustering or DBSCAN. We will do exploratory analysis on different segments and will try to come to a conclusion which algorithm is doing better. We will do the data collection process, which is a tedious process in our case as it involves web scraping and data cleaning. Second member will analyze the collected data. We will try to visualize the data using leaflet MAP API and try to figure out what features can be extracted from the dataset and do the necessary preprocessing. We will apply different clustering algorithms and try to conclude what the segmentation means. The literature survey and background study was also distributed among three members. Based on our discussion, we may face challenge in determining the success of our approach.

Data sources:

We will be using FourSquare API to get data of a particular area. For getting all the areas from a State we will first use google api to extract all the zipcodes and its associated area name in a given state which consists all the postcodes and areas of Michigan. We will use geopy library of Python to collect the latitudes and longitudes of these locations. Finally, using FourSquare API we can collect different venues on that location. These venues can be chinese restaurant, spa centre, parks within 1 mile of radius. We will cluster different neighborhoods based on the presence of

these venues. If two neighborhood have similar kind of venues then they are more closely related and hence assigned to same cluster.