II. Data source

This project made use of the following data sources:

1. **Postal codes of different neighborhoods in Seattle**

Postal codes in Seattle was retrieved from this websitehttp://seattlearea.com/zip-codes/

2. **Neighborhood location data retrieved using Google maps API**

Location coordinates of different neighborhoods is returned by using Google API

3. **Finding popular venues in different neighborhoods from FourSquare API**

I have used the FourSquare API to explore different neighborhoods in Seattle. The Foursquare explore function will be used to get the most popular venues in each neighborhood. The resultant data set will be shown in ascending order.

III. Methodology:

First of all, the postal codes of different neighborhoods in Seattle will be collected by Web scraping(HTTP request) in this web page "http://seattlearea.com/zip-codes/". Data will be cleaned and selected by Beautiful Soup library. After that, the geographical coordinates, latitude, and longitude of different neighborhood will be retrieved by using the Google API. The neighborhood with name, postal code, geographical coordinates will be regrouped and cleaned by the functions in Pandas library. Then FourSquare API will be called to retrieve all the venues in different neighborhoods with the radius of 1000m. Finally, the machine learning algorithm K-means clustering would be applied to form the clusters of different venues categories for all the neighborhoods.