



1) Problem Description

The stakeholder is a successful restaurant located in uptown Manhattan, New York City: “Amy Ruth’s”. “Successful” is defined in this context as being in the top venues list of the foursquare data of that particular neighborhood. The restaurant owner wants to branch out and open a second restaurant in another larger metropolitan city or even in New York itself if appropriate. The goal of the project is to identify another neighborhood in an US city, which is similar to the one where “Amy Ruth’s” currently resides (Zip Code 10026).

2) Data and Processing

The project requires three data sources:

- List of US cities: this information is extracted from Wikipedia with BeautifulSoup: https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population
- Neighborhood definition: postal zip codes of the individual cities are used as a surrogate for the classical neighborhood definition (i.e. neighborhood names). This is done for ease of use as obtaining neighborhood lists of several US cities is less consistent. Zip codes of US cities are available as a python library and database “uszipcode” <https://pypi.org/project/uszipcode/>
- The data characterizing a neighborhood is obtained through the Foursquare API <https://developer.foursquare.com/>

Using the three data sources above a master table will be generated. First the zip codes are retrieved for the most populous cities in the US (≥ 1 M inhabitants) extracted from Wikipedia. Then venue data is extracted using the Foursquare API for each zip code. This information is added to a master table as one row per zip code and columns corresponding to venue types. A clustering will be performed on the one-hot encoded data. The cluster containing the neighborhood in which “Amy Ruth’s” is located is identified. This means that “Similarity” of neighborhoods is defined as neighborhoods belonging to the same cluster after a Kmeans clustering of all the collected neighborhood data. These neighborhoods of interest (zip codes) are then analyzed to ensure that there is no restaurant of the same type in the area to avoid direct competition. This then finally yields a list of zip codes in the US which could be of interest to open a new location of “Amy Ruth’s”.