Data

The geojson data of the city of Cleveland is used. It consists of the names and the GPS coordinates of 36 neighborhoods in Cleveland. Based on the coordinates, venues within the radius of 1km of each neighborhood are extracted from Foursquare API. A total of 1064 venues are returned under 224 unique categories. The top 10 venues for each neighborhood are sorted out and then converted from categorical variables to dummy variables for the convenience of clustering. These venues will help us determine whether there is an opportunity for a certain business in a neighborhood.