

Data

I will be using data kindly provided in a CSV file from the Boston Planning & Development Agency (BPDA) which contains the latitude and longitude of the centroids of each Boston Neighborhood derived from the neighborhood shape files as provide in Analyze Boston (<https://data.boston.gov>). The shape files were run through arcgis online and the tools they provided were used to calculate the latitude and longitude of the neighborhood centroids. Each neighborhood was run through Foursquare to obtain a list of venues of in the immediate area. The resultant list was processed using Python and scikit-learn to cluster the data using k-means clustering. Health status indicators were obtained by manually populating a CSV file from data contained within the online document “Health of Boston 2016-2017” at http://www.bphc.org/healthdata/health-of-boston-report/Documents/_HOB_16-17_FINAL_SINGLE%20PAGES.pdf.