**The final report**

Determining a feature vector for classification the food serving venue according to risk for a public health

Introduction/Business problem

Sanitary inspection of various cities in United States publish the results of checking the sanitary and other relevant conditions of venues that serve a food in the form of dataset with list of venues’ metadata and the result of the inspection on the healthdata.gov web site (http://www.healthdata.gov).

Idea of the project is to correlate this dataset with Foursquare database for one city, cluster a data and find a similarity with the clusters in another city. Based on those data or eventually incorporating a data from another datasets if available (for example: environmental health indicators – quality of drinking water, air pollution indicators etc.), determine a set of features relevant to classify the restaurant in another city as risky or another class, according to available list of classes in the results of inspections (multivalue classification). As evaluation dataset will be used the exact results of inspection control for that city, acquired from healthdata.gov.