

Introduction/Business Problem

I would like to study the Neighborhoods in Knoxville, TN, and try to discover what is the diversity of Food in Knoxville. If someone would like to open an Asian restaurant, or a Pizza place etc, utilizing the data from FourSquare and Unsupervised learning clusters we would like to see if we can discover which neighborhoods would be better options for this placement. Perhaps there are neighborhoods which already have too many pizza places, and some that have so few that they are very rare. This information will allow the business user to see where it is better to spend their resources.

Data

The first data set that we would need is the list of Neighborhoods in Knoxville, TN. Once we have that, we will need to get the latitude and longitude locations for each of the neighborhoods. After the creation of a visual map, just to make sure that the data is validated, we are going to start using FourSquare API to look for the top 100 venues in a 500 meter radius of each neighborhood. These calls will be specific for the categoryId of Food. Once we have the data for each neighborhood, and their top food venues by category we are going to put this data through k-means unsupervised learning in order to create clusters. Post-visualization of these clusters, we will study what are the main differences between the clusters, and if we can discover a pattern which allows us to help answer our business problem.