Classification of Chicago Schools

Using Foursquare Data and Machine Learning to Identify Patterns in Chicago Schools

By Zac Yauney

1. Introduction and Statement of Problem

Not all schools are the same. That much seems like it should be common sense, and yet when we try to help schools improve performance, we tend to offer generalized solutions. While it is true that there are guiding principles for what makes a school better, the neighborhood in which a school is embedded makes a difference in what interventions are most successful.

While it is outside the scope of this exercise to prescribe specific interventions to improve school performance, it should be possible to classify the different types of neighborhoods in which the schools are located. This can be done by combining Chicago Census data and Chicago Public School data with geographic venue data from the Foursquare API. We can list all the different types of venues near a school and then use an unsupervised machine learning algorithm (kmeans clustering) to segment the schools.

This segmentation data should allow policy makers to develop individualized approaches to providing aid to schools in need.