Emily Van Auken

Erica Cordatos

Project Phase 3

Analysis Plan

After scrolling through the nearly 8,000 observations in the College Scorecard dataset, we decided that we would like to investigate first natural clusters of colleges that appear when we use specific variables of interest that could provide information to these institutions about what schools they most resemble based on our factors. Secondly, we hope to take these groupings and move over into Tableau to create interactive maps based on these clusters that applicant students could use during their college search.

Our first goal is to provide information to the institutions themselves for better understanding how they fit into the country’s college system. Knowing who they are similar to can help to aid them in future advertising and recruitment schemes using this information. This could be as simple as modifying pamphlets to resemble those of other schools, or mimicking other institution’s open houses and admissions events, or as drastic as purchasing ads for themselves on the website of a university that falls within their cluster.

We will be experimenting with the k-nearest neighbors clustering and k-means clustering to uncover natural clusters of schools that exist around the country. Once we determine which method of clustering creates the better-fitting clusters, we will hopefully generate plots in R using ggplot to represent this information. At this stage, the plots will be simple: color coded and labeled institutions. We hope to come up with a title and description of each cluster that appears. For example, given the results, we could have a cluster named “Small, Private, Northeast schools.” Next, we will take these clusters into Tableau and hopefully present them as an option on an interactive map.

Our second goal is to utilize Tableau to aid high school seniors in finding a large list of colleges tailored to their needs based on meaningful clustering and data analysis. Therefore, we will input our initial data table along with our newly discovered clusters, and make interactive maps for students. We will then add several factors for students to select as important to their college search. This could be, one of our clusters, retention rate, or even religious affiliation. Whatever is important to the individual student will hopefully be able to be inputted into the Tableau map, and provide the student with a list of colleges that might suit them, giving them ideas of where to apply.