Alex Kusek, Luke Baumer, Arianna Smith, Patrick Dunleavy

MATH 3570

Project Proposal

Project Title: Datadawgs Take on College

During your Junior year of High school, it is one of the most stressful times in your life. You do not know where you are going to attend school yet, what you want to major in, the cost of university, and worrying about all of this while also trying to be academically as successful as possible. Our goal is to analyze colleges in different demographics all around the United States of America and give prospective students important information related to degree earning outcomes, debt, cost, success rate, and other institutional characteristics. By giving prospective students this information, they will have a better idea of what to expect for their futures. Our goal is to answer some of the most important questions related to college selection: How much debt am I expected to be in after college? How long will it take to pay off based on what I study? What are my career earning potentials based on my major? What is the student success rate at each University?

We plan on using Python for data cleaning and analysis, packages related to this will be Numpy, Pandas, and the Matplot library. For data visualization, we plan on using R packages like ggplot2, Tidyverse, all in posit clout with GitHub collaboration. We plan on having multiple graphs of data for the multiple different variables that we are analyzing. For prediction models, we plan on using linear regression and logistic regression. To keep the information clean and simple for the end user, we do not plan on creating any models that utilize more than 4 variables to keep it simple for prospective students.

We will have a variety of datasets from the U.S. Department of Education. This data set is massive, and includes all of the previously mentioned variables that we want to analyze.