Capstone 1 Proposal

Fred Berendse

Essential Questions

1. Is there a significant difference in post-secondary graduation rates, outcomes, and degrees awarded for Pell Grant/Subsidized Stafford Loan (SSL) recipients and/or minority students compared to students outside these groups?

2. Is there a state-by-state variance in these metrics?

3. Does this disparity persist when one accounts for variable levels in college preparedness between these groups?

Data

The primary data set I intend to use is the Integrated Post-secondary Education Data Set (IPEDS) from the National Center for Education Statistics. (https://nces.ed.gov/ipeds/use-the-data) The set is comprised of annual surveys of 7153 public and private 2-year and 4-year post-secondary institutions nationwide.

The set is organized into several survey tables that can be joined by institution ID. The tables most likely to be of interest to my study include Head Directory, Admissions and Test Scores, Completions, Graduation Rates, and Outcome Measures. Most if not all measures are disaggregated by race and Pell Grant/SSL status.

Challenges I will face processing this dataset include interpreting and extracting the appropriate column headings from each table. Between the tables mentioned above, there are 191 columns of data available. Each table is accompanied by a descriptive file that I will use to determine which columns will be needed for my analysis. Each column has a supplementary column containing a code on how the data were imputed from the surveys. Codes like "Do not know," " Value not derived – data not usable," and "Institution left item blank" will need to be scrubbed to make a clean data set.

Viable Products

A minimum viable product for this project is a clean database containing data vital to answering the essential questions of the project, along with exploratory plots of the data set. A secondary viable product would be the results of A/B hypothesis testing to determine if a disparity exists between members of a group and members outside the group for selected metrics.