Proposal—Project 1

*Standardized Testing and Its Impact in Academic Success*

The phrase “Academic Success” has a variety of meanings in higher education. Accordingly, there are many factors in quantifying “success.” Our project will utilize the *College Scorecard (CS)* dataset, which aims to shed light on relative college performance. *College Scorecard,[[1]](#footnote-1)* a website created by the *United States Department of Higher Education*, houses information about US colleges from the past 20 years. The dataset includes information on the institutional level, such as: SAT averages, completion rates, loan default rates, student ethnicity proportions, etc. Accordingly, the dataset is very large, with thousands of institutions, which then contain hundreds or thousands of data points for each subsequent year. As such, we will utilize the *College Scorecard “Data Dictionary”[[2]](#footnote-2)* as an aid for variable selection in conjunction with data analysis with Python in Jupyter Notebooks.

There are challenges with this dataset, which should be noted in this proposal. The primary challenge is the size of the dataset—as stated above, this dataset contains potentially hundreds of thousands of data points, with varying meanings, spanning a long time period. We may be concerned with some schools, for example, over others. As such, filtering and delineating various subsets of data from the whole dataset will be a key to answer our core questions. Accordingly, various components or pieces of the dataset may have changed over time, which may leave us with “null” values for different years, depending on the reporting or collection standards from the *Department of Education* over time.

We are primarily interested in standardized testing, specifically the *SAT*, through the lens of “academic success.” Our project with have two primary topics, each with a subset of questions. The first topic concerns the *SAT* averages of US colleges. How have SAT averages changed over time? Do those averages correlate with institutional completion rates? The second topic will be minority-centered and concerned with a broader definition of “academic success” to include loan default. Is there correlation with SAT averages and loan default rates for minority students or minority schools? To answer these questions, we will plot our selected data and plan to perform regression analysis. Additionally, we hope to utilize a map visualization of our data as well. We will subsequently summarize and submit our findings via presentation and Jupyter Notebook format.

1. U.S. Department of Education, *College Scorecard Data*, website, 2019, <https://collegescorecard.ed.gov/data/>. [↑](#footnote-ref-1)
2. U.S. Department of Education, *College Scorecard Data*, website, 2019, <https://collegescorecard.ed.gov/data/documentation/>. *College Scorecard* does contain an API, but due to the sheer size of the data involved, we will work with the entire dataset for each year, provided in *.csv* format. The *Data Dictionary* will help determine variable name and meaning (ie. ‘’CDR3” stands for for 3-year credit default rate). [↑](#footnote-ref-2)