Syracuse University

Memo

To: Dr. Landowski

From: Elizabeth Axworthy and Samuel Deery-Schmitt

Date: 8/13/2021

Re: Project Proposal

Topic: Compare college tuition and salary outcome information against economic information to see changes in their relationships over time.

Data Description:

Our main dataset is the College Tuition, Diversity, and Pay dataset, found on Kaggle. It has 5 CVSs containing information in tuition costs, historical tuition rates, salary potentials, student income brackets, and diversity of student populations. There are a total of a total of 33 columns and 263,845 rows; we are particularly interested in the following variables: minority enrollment, tuition cost by year, total in-state and out-of-state cost, in-state and out-of-state tuition, income level, and mid-career pay.

We are supplementing this dataset with additional information on tuition costs adjusted for inflation, general inflation data, and income by state. To examine the tuition costs and inflation, we will use Tuition Costs of Colleges and Universities from the U.S. Department of Education, National Center for Education Statistics Tuition Costs (62 observations of 6 variables) which is itself derived from the much more extensive Average undergraduate tuition and fees and room and board rates charged for full-time students in degree-granting postsecondary institutions, by level and control of institution: Selected years, 1963-64 through 2018-19 (200 observations of 25 variables), also from the U.S. Department of Education, National Center for Education Statistics Tuition Costs. We may or may not incorporate data from the second table.

For income data, we will use U.S. Census data from the following source: Information on income at a state level from the U.S. census: Median Household Income by State: 1984 - 2018 (102 observations of 68 variables) and supplement with additional data from the U.S. Census as necessary. The results of this search will provide any additional datasets we need.

We will also add inflation data to our analysis, as the previous dataset only captured tuition adjusted for inflation, not inflation itself. The Consumer Price Index Data from 1913 to 2021 (108 observations of 15 variables) is derived from the Bureau of Labor Statistics dataset (108 observations of 15 variables) and we will incorporate the source that we deem the most user-friendly.

Scope
This project will be larger in scope than the homework assignments as we will be using multiple datasets from different sources. We will also be investigating several open-ended questions and looking for patterns and trends that we will examine in more detail as warranted. Our project will remain exploratory in nature and will not veer into the realm of prediction by making forecasting models. We are going to look only at undergraduate programs at accredited US schools.

Research Questions:

- How has the rate of tuition increase changed compared to the rate of inflation?
    - Is there a difference in tuition increases among 2-year and 4-year, private and public institutions?
    - Which type of schools, and which schools specifically, have increased their costs compared to the rate of inflation, and which have remained the same?
- How does the rate of change in salary potential compare to the rate of change in tuition increase?
    - Is there a significant difference among 2-year and 4-year, private and public institutions?
- How as the "affordability" of tuition changed over time – how do changes in median income, at the national and state-level, compare to changes in tuition?
    - Which type of schools, and what schools specifically, are the most affordable?
    - Which are the least affordable?
- How has economic diversity at institutions changed over time – what are the changes in student income bracket distribution over time?
- How has racial diversity at institutions changed over time?
- What are the specific trends for tuition increases, salary potential, affordability, economic diversity, and racial diversity at:
    - the most selective schools (with a significantly lower acceptance rate than other institutions)?
    - the schools that have historically earned top-25 rankings in annual analyses, such as those conducted by the US News and World Report?

Data Preparation Plan:

1) Load all data into Python.
2) Create a single dataset and save in a reloadable format.
3) Check for and resolve NAs and other data inconsistencies.
4) Convert data to appropriate datatypes as necessary.
5) Determine a timeframe to subset that data by to help focus analysis.
6) Reduce redundant data, if necessary, to focus on timeframe of interest.
7) Create calculated fields as necessary.

Anticipated Challenges

- Efficiently merging data from multiple sources to make meaningful analysis possible.
- Scope-creep – trying to answer too many research questions.