Kensho College Equity Data

Data Source

Kaggle user Kensho Impact combined data from four sources: Opportunity Insights, Brookings Institution, College Scorecard, and Integrated Postsecondary Education Data System (IPEDS). The third and fourth datasets are sourced from the US federal government. Brookings Institution is a nonprofit public policy organization in D.C. with a ".edu" domain. Opportunity Insights is also a nonprofit based out of Harvard. All of these are trustworthy data sources.

Contents Overview

Each observation has 117 characteristics. These include basic information about the college—such as name, location, admission rate, and enrollment number—undergraduate demographic stats, parent income stats, SAT score stats, tuition rates, financial aid stats, student debt stats, degree completion demographics, graduate income stats, and major choice breakdown. See full documentation on column names here.

Limitations

The documentation on the data integration is not very robust. For all intents and purposes, it only tells us the meaning of the column names. Through this, we don't know which source each column came from. There also is not information on what year the data is from. It is safe to assume that it is recent, though, as Kensho Impact's dataset was published four months ago. I reached out to them on Kaggle, asking what year the data is sourced from.

Relevancy

While searching for a set of data, I aimed to take on a project that will evaluate equity in postsecondary education. With this dataset, I can do so by determining whether degree completion demographics are proportional to undergraduate demographics for each school. Disproportionalities in any demographic categories would indicate possible inequities. Then, I can see if there are equity trends by state or general areas of the country. It would also be interesting to compare public, nonprofit, and for-profit universities.

Additional angles would be to examine the differences across school type (public v. private) and region of graduate's income, parent income, student debt across school types and regions, and financial aid.

Questions to Explore

Between writing the previous sections and this section, I performed data cleaning steps. Because I had to throw out the degree completion demographic data (too many schools had null values in those variables), I must shift my focus for this analysis toward the additional angles I mentioned in the relevancy section.

Adjoining Questions:
- Are there geospatial trends with any of the quantitative variables?
- How do the quantitative variables of different school types (public, nonprofit private, for profit private) compare?

Elevating Questions:
- How proportionate are the college enrollment demographics to state-wide population demographics, including those who are not in college?
- How does graduate income data compare with state-wide income data, including those who did not attend college?

Clarifying and Funneling Questions:
- What are differences in graduate income?
    - Across regions
    - Between public/private
- What are differences in parent income?
    - Across regions
    - Between public/private
    - Is there a correlation between parent income and graduate income?
- What are differences in student debt?
    - Across regions
    - Between public/private
- What are differences in financial aid?
    - Across regions
    - Between public/private
    - Is there a correlation between the amount of financial aid and student debt?