Does the pay post graduation outweigh the amount of loans generated?

Felipe Rodriguez

2023-02-12

Introduction

In the fall of 2020, there were 15.85 million undergrad students registered for the fall semester (Hanson (2023)). To many, a bachelor's degree is considered to be an initial milestone to begin a long term career and create financial stability. A typical four year degree might not seem like the right path to some, but a big factor influencing this decision is the cost of attendance and the necessity to take out student loans. In this study, the pay of an undergraduate student, post graduation, will be analyzed and compared to the amount of average debt acquired by student to understand the correlations between the two.

A few research question that are worth investigating:

Which institutions have the highest paid undergraduates?

Out of those institutions, what is the average student debt?

Which institutions have the highest debt?

Does highest debt and highest pay have a correlation?

Which majors pay the most?

Is there a correlation between student debt and diversity in the institution?

Approach

The approach taken will be to identify data that contains salaries after graduation from a four year university. There will also be a need to have information regarding student loans, the pay each major is most likely to have, and demographic information. Using this data, we can join income and debt per institution to analyze the amounts. With the demographic information, an analysis can be done to see if there is a correlation in amount of minorities and total debt.

Data (Minimum of 3 Datasets - but no requirement on number of fields or rows).

The main sources for my data will be as follows:

- Brown, M. (2022, January 19). Student Loan Debt by School by State Report. LendEDU. Retrieved February 9, 2023, from https://lendedu.com/blog/student-loan-debt-by-school-by-state/
- Devastator, T. (2022, November 23). The schools that create the most student debt. Kaggle. Retrieved February 9, 2023, from https://www.kaggle.com/datasets/thedevastator/the-schools-that-create-the-most-student-debt
- Mostipak, J. (2020, March 9). College tuition, diversity, and pay. Kaggle. Retrieved February 10, 2023, from https://www.kaggle.com/datasets/jessemostipak/college-tuition-diversity-and-pay
- Payscale. (2021). Best universities and Colleges. The Best Universities For a Bachelor's Degree. Retrieved February 9, 2023, from https://www.payscale.com/college-salary-report/bachelors?search=nebraska

The source LendEDU contains loan information on student loans throughout the United States. This data set contains student loan information for states, private schools and public schools. The main columns that make this data a notable source will be Institution, State, and Average Student Loan debt per borrower. The data set provided by Payscale contains average salaries post graduation. The columns in this data set that will be used are School Name and Early Career pay. The last two data sets are in Kaggle. One contains data on schools that create the most debt and the other contains demographic information of school. These can be used to compare the percentage of minorities in relation to the amount of debt the school has to understand if there is a correlation of schools where there are more minorities and debt.

Required Packages

The anticipated packages that will be needed will be as follows:

- readxl
- dplyr
- purrr
- lm.beta
- ggplot2

The packages listed above will aid in plotting, visualizing, and creating calculations. These packages are anticipated are subject to change depending on the need of the study and the results uncovered. Any changes will be documented.

Plots and Table Needs

The plots that are anticipated are a bar graph and a scatter plot. Using a bar graph, the total amount of races can be visualized to see demographic information. With a scatter plot, an analysis can be done to view the amount of average debt per student and average early career salary to understand if there is a relationship between the two.

A table will need to be created containing all the data joined on University Name or Institution. There will need to be the following columns in the joined data set:

- School
- Average Student Debt
- Early Career Pay
- Total Students

- Minority Students

The stated plots and columns are anticipated in the study, but more can be created or added as necessary to the analysis being conducted.

Future steps

The next step will be to merge the data into one data set. One item that will need to be studied is renaming columns to have one consistent column to join on. The data sets have the University name as the following columns: Institution, Name, and School Name. These will need to be changed to University Name and then joined on the new name created. Another step that will be needed is to extract total students and total minority from the demographic data set. Total minority is a row field as a opposed to a column, this will need to be transformed.

References

Hanson, &amp., M. (2023). *College enrollment statistics [2023]: Total + by demographic.* Education Data Initiative. https://educationdata.org/college-enrollment-statistics