Analysis of Undergraduate Student Debt

Abstract

In this project, we looked at data of undergraduate students to analyze the cost incurred by looking at student debt and their earnings after 6 years. Utilizing data science tools and working with data from the Department of Education and Department of Labor Statistics, a visualization using Tableau was created.
Then a supervised machine learning algorithm was recommended to build a regression model to predict a student's future earnings. This model would help incoming university students to make better informed decisions about taking on student loan debt.

Design

According to the National Center for Education Statistics, the United States had 3.4 million high school students graduated the school year 2020-2021. Of the 3.4 million, it is estimated 2.2 million of the high school graduates will attend some form of higher education in the form of a 2-year college or 4-year university.
One of the hardest hurdles for most students is the financial burden of 4 years. It is often easy to take loans but hard to pay back. Therefore, this EDA and regression model will bring some attention to the need to plan about future earnings as well.

Data

Website - https://collegescorecard.ed.gov/data/

- College Scorecard - is a dataset provided by the U.S. Department of Education. It contains institution-level data for all accredited institutions in the United States offering undergraduate degrees. It contains institutional admissions and academics data, student loan and student earnings data.
- Use of Python and API to retrieve the data

Website - https://www.bls.gov/bls/blswage.htm

- U.S. Bureau Of Labor Statistics - provides occupational wage data in many formats such as by state or job characteristics with degree requirements.

Algorithm & Tools

Perform a thorough exploratory analysis in Excel and Tableau; clean, explore, aggregate, and visualize the data as appropriate to address the client's problem

- Exploratory data analysis in Excel
- Visualization in Tableau
  - Acquisition tools use of APIs
  - Programming Language use of Python