

From Campus to Career: Predicting Earnings Through Institutional Traits

Anaya Tention

ABSTRACT

This study examines the relationship between institutional characteristics and the median earnings of college graduates ten years after enrollment, focusing on factors such as family income, faculty salary, completion rates, and socioeconomic diversity. Using data from the U.S. Department of Education's College Scorecard, individual institutions were analyzed through Multiple Linear Regression (MLR) for interpretability and Random Forest (RF) for predictive accuracy. Key findings indicate that higher family income, faculty salaries, and completion rates are associated with greater graduate earnings, with STEM program availability included as a control variable to account for earning potential linked to in-demand fields. Institutions with a higher proportion of Pell Grant recipients tend to show lower graduate earnings, suggesting socioeconomic disparities in post-graduation financial outcomes. These insights provide actionable recommendations for institutions and students to enhance support systems and close earning gaps linked to socioeconomic backgrounds.

INTRODUCTION

As the cost of higher education rises, understanding the economic returns of a college degree has become a critical concern for students and institutions alike. For students, choosing an institution that promises a positive return on investment (ROI) can be essential to their long-term financial success. For institutions, identifying factors that enhance graduates' earnings can help in developing programs that improve student outcomes and boost institutional reputation. Analyzing the institutional characteristics that impact post-graduation earnings provides a clearer picture of how colleges and universities can optimize their resources to support student success.

This study aims to investigate the influence of specific institutional characteristics on the median earnings of graduates a decade after enrollment. Key characteristics explored include family income, faculty salary, completion rate, and percentage of Pell Grant recipients. The STEM Index is included as a control variable to account for the higher earnings associated with STEM fields, allowing us to isolate the effects of other institutional characteristics. By examining these factors, the study seeks to provide valuable insights that inform institutional strategies and support students' decision-making processes. This report is structured as follows: Data and Methodology provides a detailed description of the data and preprocessing steps; Experiments and Results summarize the modeling approaches and presents findings; Discussion includes an analysis of key insights and recommendations for institutions and students; and Conclusion offers a brief summary and suggestions for future research.

DATA AND METHODOLOGY

4.1 Dataset Description

The dataset used in this analysis is sourced from the U.S. Department of Education's College Scorecard (cohort 2014-15), which aggregates data at the institutional level, including enrollment statistics, student demographics, financial aid, and graduate outcomes. Each row represents an individual higher education institution in the United States. Primary variables examined in this study are Family Income (MD_FAMINC), Average Faculty Salary (AVGFAC SAL), Percentage of Pell Grant Recipients (PCTPELL), Completion Rate, and Average Cost of Attendance (AVG_ATTENDANCE). The target variable, Median Earnings Ten Years After Enrollment (MD_EARN_WNE_P10), represents the financial success of graduates a decade post-enrollment.

4.2 Data Preprocessing

To prepare the dataset for analysis, several preprocessing steps were undertaken. Several financial variables that contained privacy suppressed values were converted to NA values and dropped, and the columns were converted to numeric. Any missing values were dropped, since imputations might cause incorrect results due to the volume. Variables with significant skewness, such as Family Income, Average Faculty Salary, and Median Earnings, were log-transformed to normalize distributions for Multiple Linear Regression. For the Random Forest model, no transformations were applied. I converted institution type to binary, and STEM Index was created as a control to account for higher earning potential linked to STEM fields.

4.3 Methodological Approach

Two predictive models were applied to the data. The Multiple Linear Regression (MLR) model was selected for its interpretability, allowing for a clear understanding of the relationships between institutional characteristics and graduate earnings. The Random Forest (RF) model, a non-linear ensemble method, was used to improve predictive accuracy and manage complex interactions among variables. Hyperparameter tuning and five-fold cross-validation were employed for the Random Forest model, optimizing it for robust performance.

EXPERIMENTS AND RESULTS

5.1 Experimental Setup

The dataset was divided into a training and testing split of 70/30 for both models. For the Random Forest model, five-fold cross-validation was conducted, yielding an average R^2 score of 0.7062, while the final test set R^2 was 0.7631. These evaluations were conducted in Python, utilizing Pandas for data management, Seaborn for visualization, and Scikit-Learn for model training and tuning.

5.2 Results

The Random Forest model showed that Family Income, Faculty Salary, and Completion Rate are among the strongest predictors of graduate earnings, with Family Income accounting for 44% of the importance. Other key predictors included Pell Grant Percentage and Average Cost of Attendance, while the STEM Index was maintained as a control. The Multiple Linear Regression model achieved an R^2 of 0.703, with

Family Income, Faculty Salary, Completion Rate, and Pell Grant Percentage emerging as statistically significant predictors, indicating strong influences on graduate earnings.

5.3 Analysis

The analysis identified several institutional characteristics that significantly influence graduate earnings. Higher family income, measured as the median income of enrolled students, was the strongest predictor of post-graduation earnings, highlighting the lasting impact of socioeconomic background on financial outcomes. Faculty salary was also positively correlated with earnings, suggesting that investment in faculty quality may contribute to better long-term success for students. Pell Grant Percentage showed a negative association with earnings, which may reflect systemic financial challenges faced by students from lower-income backgrounds. Additionally, higher completion rates were associated with improved financial outcomes, underscoring the importance of student retention and support.

DISCUSSION

Key Insights and Implications

The findings from this analysis offer actionable insights for both educational institutions and prospective students. The strong correlation between family income and graduate earnings highlights a persistent socioeconomic gap that institutions could address through enhanced support services for low-income students. Expanding financial literacy programs, mentorship, and career services may help bridge this gap, particularly for students from lower-income families. Faculty quality, indicated by faculty salary, appears to have a substantial impact on graduate earnings, suggesting that institutions may benefit from investing in faculty resources and development. High completion rates also correlate with better financial outcomes, reinforcing the importance of retention programs. For students, evaluating institutional support systems and resources, particularly if from lower-income backgrounds, can be beneficial in selecting a college. Additionally, considering the presence of STEM programs, as controlled for in this study, can provide insights into earning potential.

Recommendations for Stakeholders

To bridge the socioeconomic gap, institutions should strengthen support systems that cater to low-income students, offering resources such as financial literacy workshops and career counseling. Investing in faculty quality is another area of focus, as it appears to correlate with better student outcomes. Retention programs that promote high completion rates are essential for improving graduate financial success. For students, prioritizing institutions with strong support for low-income individuals and evaluating faculty resources can inform better college selection, leading to improved financial prospects post-graduation.

CONCLUSION AND FUTURE WORK

This study highlights the institutional factors that influence the financial success of college graduates, with family income, faculty quality, and completion rates emerging as major predictors. These findings point to the critical role that socioeconomic background plays in determining post-graduation earnings,

as well as the importance of institutional resources and support. Future research could examine additional institutional factors, explore demographic variables beyond socioeconomic background, or investigate longitudinal outcomes across different fields. Overall, improving access to quality support and resources for all students is essential for enhancing the long-term economic benefits of higher education.