# Loan Outcomes Among Pell Grant Recipients: A Comparison of HBCUs and Non-HBCUs

## 1 Introduction

Student financial aid is one of the primary programs that keeps the American higher education system intact. Pell Grant is a need-based federal grant for undergraduate students with no repayment, and is designed to help students from low-income families afford college. However, Pell Grant recipients apply for loans to cover the remaining costs (Ma et al., 2014). Students relying on borrowed funds for their undergraduate education bring about a concern in their long-term financial state, having noticed a relationship between high student loan debt and high default rates. Studies have shown that higher borrowing rates, together with insufficient income after graduation, lead to an increased risk of default (Looney & Yannelis, 2015). This research is relevant as it works to understand the disparities in the outcomes coming from education across different institutions. Understanding the differences between students from Historically Black Colleges and Universities (HBCUs) and non-HBCUs helps us determine a possible larger issue of economic inequality. However, we can see a gap in the literature regarding analyzing the disparities coming from Pell Grant Recipients of different types of institutions. This research works on a refined approach of comparing these values between HBCUs and non-HBCUs, aiming to understand if the disparities in loan default rates of Pell Grant recipients can be connected to the differences in institutions. By integrating different variables, our research also addresses this gap by shedding light on the intersection of borrowing behavior and the type of institution (HBCU vs non-HBCU) a student attends. Exploring the issue can provide awareness and ease the financial burden of the impacted students.

## 2 Motivation

Although Pell Grant provides a non-repayable grant, it is often insufficient to cover the entire cost of attendance at these institutions. This results in students trying to supplement their aid through student loans. This study seeks to examine how additional borrowing, called Pell Grant debt, is associated with loan default and repayment rates. One of the key aspects of this research is to find any discrepancies in these results between HBCUs and non-HBCUs. HBCUs primarily serve low-income students who are dependent on federal loans to finance their education (Lee & Keys, 2013). We aim to discover whether there is a relationship between borrowing amount and repayment success rates while trying to identify what institutional benefits reduce the risks of taking high loans.

Furthermore, we try to incorporate key characteristics that provide a full view of the factors influencing default and repayment rates. Through different analyses that include or remove these characteristics, we intend to find the direct relationship between student loan debt of Pell Grant recipients from HBCUs and non-HBCUs. Our goal is to gather information that can help create fair financial aid for both institutions equally. Having this analysis can help us understand and work towards eliminating it from our education system. This can help improve the financial stability and success in academics for low-income Pell Grant students, regardless of which institution they attend.

# 3 Research Questions

In this project, we aim to address the following research questions:

1. How does the loan default rate vary between Pell Grant students in HBCUs and non-HBCUs, or is it relatively the same across all universities?
2. How does the loan debt differ between students attending HBCUs and students attending non-HBCUs?
3. How is Pell Grant debt associated with loan default and loan repayment rates, and does this association differ for HBCUs versus non-HBCUs after controlling for other institutional characteristics such as net cost, admissions rate, completion rate, retention rate, and median earnings?
4. Can we predict loan default risk using student-level data?

# 4 Expected Outcomes

Working on top of the research that indicates borrowing profiles lead to higher default rates (Looney & Yannelis, 2015), and considering the financial context of HBCUs serving low-income students that rely on federal loans (Lee & Keys, 2013), we hypothesize that students attending HBCUs will generally have higher loan default rates due to higher levels of debt, exacerbated by institutional underfunding. Additionally, we expect to find that the amount of loan debt influences repayment success, with students facing higher debt likely to have more difficulty in repaying their loans. Furthermore, we anticipate that by controlling for other institutional factors, we will gain a deeper understanding of the relationship between Pell Grant debt and financial outcomes, which will inform the creation of more equitable financial aid policies for low-income students across different educational settings.

# 5 Data Description

The data we used in this project is from the U.S. Department of Education College Scorecard website. College Scorecard is a project designed to increase transparency about postsecondary institutions in the United States and enable students and families to compare how well individual postsecondary institutions are preparing their students for success. It also helps them compare college costs and outcomes while accounting for their needs and educational goals. The data in this project were provided through federal reporting from institutions, data on federal financial aid, and tax information (College Scorecard, 2025).

College Scorecard provides two data files: institutional-level data and field of study data. Due to the limited nature of the field of study data (most of the entries in the data were null values), we focused on institutional-level data for our analysis. To clean and prepare the data, we used Python data libraries (NumPy, Pandas, Matplotlib, and Seaborn). College Scorecard provides documentation for the dataset, and we used this documentation to learn about the variables in the data and the research questions that can be answered from the data.

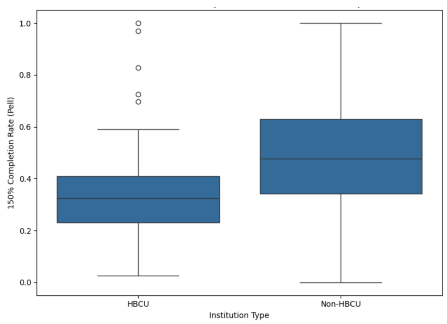
The institutional level data contains 3305 variables and 6484 observations.

The variables in the dataset ranged from basic information about the college (name, location, degree type, etc) to information about cost, financial aid, loan default rate and repayment, earnings of students, completion rate, etc. Since we were only interested in loan debt, default, and repayment rate for Pell Grant students in HBCUs and non-HBCUs, we removed variables that weren't useful for our research purposes. Hence, we reduced the number of entries to 45 during the exploratory data analysis stage. We further reduced the number of entries to 10 before we started our analysis. The full list of the variables we finally used for our analysis is listed at the end of this section.

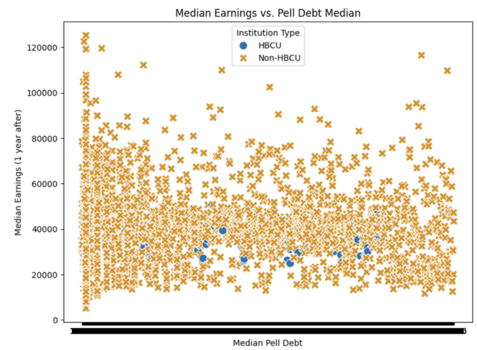
Some of the variables (e.g., PELL_DEBT_MDN) (Median Debt for Pell Student), ADM_RATE (Admission Rate), etc.) had to be converted into numerical variables so that we could make box and bar plots with them. Some of the variables had inconsistent observations. For example, the Default_Rate variable had some missing values, inequality symbols, extraneous characters, a range of figures, etc. To clean these variables, we defined a cleaning function that removed all of these inconsistencies. Moreover, we dropped all NaN rows from the dataset, and since we are only dealing with HBCUs and non-HBCUs, we set every other school which were not HBCU (schools that had an observation of zero under the HBCU column) to non-HBCU schools. We also changed the variables with vague names to easily understandable ones. For example, BBRR4_FED_PELL_DFLT and BBRR4_FED_PELL_PAIDINFULL variable names were changed to Default_Rate and Repayment_Rate, respectively.

In the discussion section, we wanted to touch on the difference in the endowment of HBCUs and non-HBCUs as a possible explanation for why students in HBCUs take more loans and have a higher default rate. Hence, we used the ENDOWBEGIN variable, which is the value of the school endowment at the beginning of the fiscal year, to calculate the average endowment of both HBCUs and non-HBCUs. This is the only variable we used that wasn't in the analysis section.

Figures 1 (a) and (b) visualize examples of some of the plots we made for during the Exploratory Data Analysis (EDA) stage.




(a) Boxplot of 150% Completion Rate for Pell students (C150_4_PELL) by Institution Type



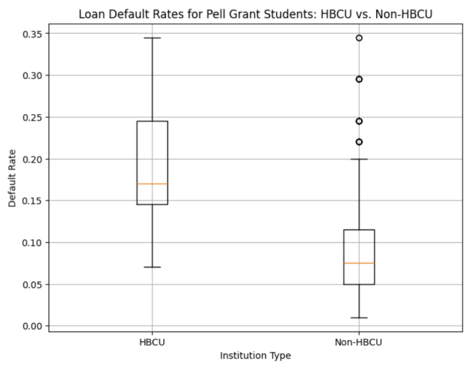

(b) Scatter plot of Median Earnings vs. Pell Debt Median for Pell Grant Recipients

Figure 1: EDA Box Plot and Scatter Plot

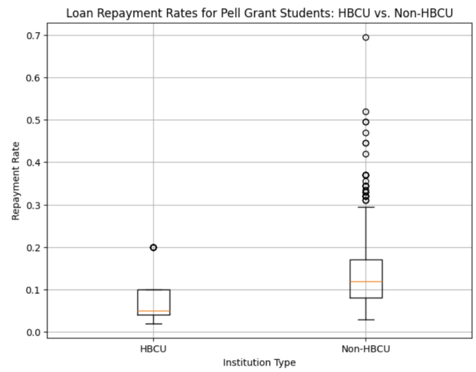
# 6 Methods

The first question we attempted to answer was how loan default and repayment rates vary between Pell Grant students in HBCUs and non-HBCUs. We started with some exploratory data analysis which included box plots for loan default rate, and repayment rate among Pell Grant recipients in HBCU vs non-HBCU institution. The exploratory analysis suggested a trend that both loan default rate and repayment rate differ significantly between HBCU and non-HBCU, so we wanted to test the statistical significance of the difference. Hence, we did a t-test to compare the mean of these two groups (HBCUs and non-HBCUs) for both the default and repayment rates. Next, we built a linear regression model to examine the relationship between HBCUs and default rates (as well as repayment rates). This helps us to know how these rates differ on average for HBCUs compared to non-HBCUs, just like the t-test. It also helps us to know, on average, the increase in these rates (in percentage points) between HBCUs and non-HBCUs. Unlike the t-test, we can draw insights from R-squared values and F-statistics values given from these models, and we can add more variables (e.g., median earnings, completion rate) to control more factors.

For our second question, we wanted to find out how the loan debt differs




(a) Loan Default Rate: HBCU vs Non-HBCU

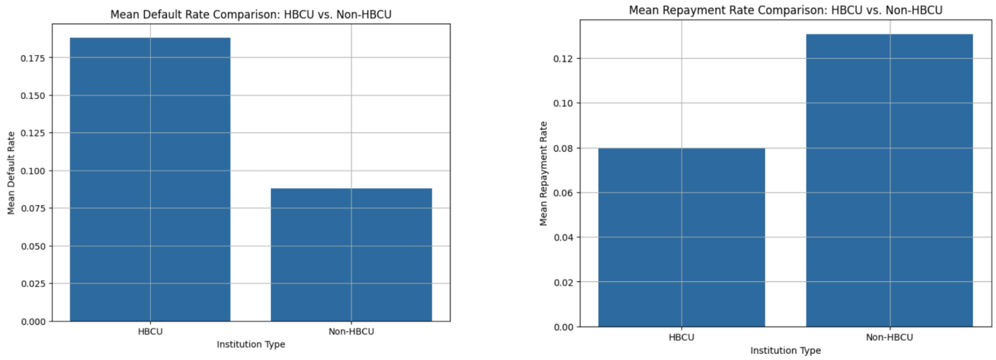



(b) Loan Repayment Rate: HBCU vs Non-HBCU

Figure 2: Loan Default and Repayment Rate EDA Box Plots

by HBCU status for Pell Grant students. To answer this question, we used the median debt for Pell Grant students (PELL_DEBT_MDN). Again, we used a box plot to show the distribution of loan debt across these institutions. We saw some differences in these distributions, but they were not as large as the default and repayment rates. To find out if these differences were statistically significant, we conducted a Welch's t-test and built a linear regression model with median debt as the dependent variable, and HBCU status as the independent variable.

Next, we explored how the loan debt we investigated in question two is associated with the loan default and repayment rates (which we explored in question one). The scatterplots in figure 5 (a) and (b) showed that. Next, we proceeded to building the multiple linear regression model. Before building this model, we had to control for other institutional variables, such as net cost, admission rate, completion rate, retention rate, and median earnings. Hence, our independent variable changed from just HBCU to the following: Pell median debt, HBCU, Pell median debt * HBCU, Cost of attendance, Admission Rate, Graduation within 150% of expected time, Retention Rate, Median Earnings Six years after graduation By using multiple regression, we were able to account for multiple factors that could affect the loan debt and default and repayment rates. More so, we were able to build an improved model that gave us a deeper understanding of the complex relationship be-




(a) Mean Default Rate: HBCU vs Non-HBCU

(b) Mean Repayment Rate: HBCU vs Non-HBCU

Figure 3: Loan Default and Repayment Rate EDA Bar Plots

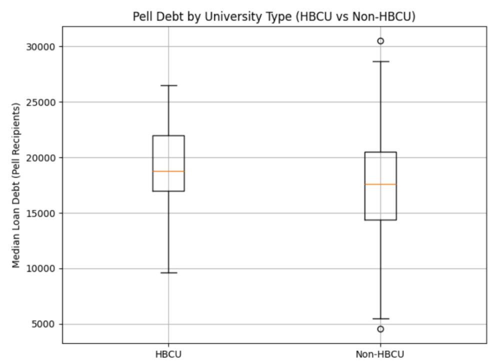
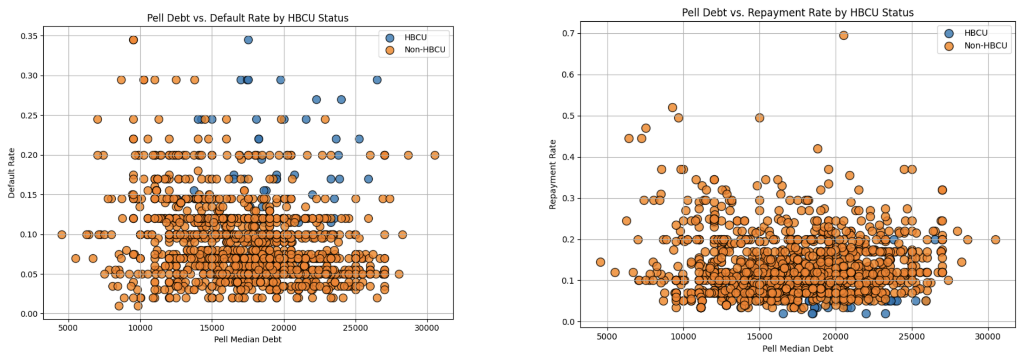



Figure 4: Pell Debt by University Type: HBCU vs Non-HBCU Box Plot

tween loan debt and default and repayment rates.

For the fourth research question, we wanted to build a predictive machine learning model that could predict the risk of loan default for students based on student level characteristics. We hypothesize that such a model would be used by loan providers in assessing the risk of a student loan default. To test the feasibility of such model, we intended to build a comprehensive model that would take in features such as student's academic performance, their family income, and the characteristics of the institutions they plan to attend. We include variables such as Pell Grant Status, and whether a student is attending a HBCU in our model, as it relates with our previous research




(a) Pell Debt vs Default Rate (b) Pell Debt vs Repayment Rate

Figure 5: Pell Debt vs Default & Repayment Rate Scatter Plots

questions on whether these factors impact student loan default rates. Since we do not have student level data to build the proposed model, we decided to use synthetic data that is simulated with assumptions that match the real world data.

We use the student's GPA, and SAT score as a proxy for academic performance. These variables are generated using a normal distribution of average GPA and SAT Scores of US students. We use the following income levels to classify students into different income groups as an indicator of their financial stability: (1) \$0–\$30,000 (2) \$30,001–\$48,000 (3) \$48,001–\$75,000 (4) \$75,001–\$110,000 (5) \$110,000+. The students are randomly assigned into one of these groups with each group's probabilities being 30%, 25%, 20%, 15% and 10% respectively. A binary variable to indicate whether a student is a Pell Grant recipient. This variable is categorized as 1 with a 90% probability if the student's family income level is \$0–\$30,000, otherwise, it is categorized as 1 with a 20% probability. Next, we use institutional-level variables for the institution the student is attending, such as net cost, admission rate, completion rate, and median earnings after graduation. Each student is randomly assigned to either an HBCU or a non-HBCU institution using weighted probabilities that reflect the real-world proportion of such schools. After classification, we generate institution-level variables using normal distributions. These distributions are set up differently for HBCU and non-HBCU schools, based on statistics from the College Scorecard dataset for each group. Once all the predictor variables are generated, we define a function to calculate the log odds and probabilities of student default based on the predictor variables

$$log\_odds = \begin{pmatrix} -3 + (2.50 - GPA) + 0.00015 \times Cost\ of\ attendance \\ +0.25 \times \begin{pmatrix} 4 - Income\ Level\ Category \\ +0.40 \times Pell\ Receipeint\ Indicator \\ +0.40 \times HBCU\ Indicator \end{pmatrix} \end{pmatrix}$$

Then, the outcome variable, Loan Default, is simulated as a binary variable based on each student's risk probability of default. With a synthetic data ready, it is split into train and test the dataset and a logistic model is fit on the data. The logistic regression model is used to predict the risk of loan default risk for students based on the predictor variables. While we recognize the limitations of using synthetic data, this approach allowed us to explore the feasibility of predictive modeling in this context and simulate how such a model could be applied in real-world scenarios.

# 7 Results

Pell Grant recipients at HBCUs undergo substantially more repayment strain than their counterparts at non-HBCUs. In our simple OLS regressions, attending an HBCU was associated with a ten-percentage-point higher four-year default rate and a 5.1-point lower full-repayment rate reflecting raw averages of 18.3 percent default and 61.4 percent repayment at HBCUs versus 8.3 percent default and 66.5 percent repayment at non-HBCUs. These unadjusted models make clear that, before accounting for any other factors, HBCU Pell recipients default at more than twice the rate of their peers and are significantly less likely to clear their loans in full.

Regarding borrowing behavior, we found that HBCU students have a median Pell debt of $12,211 compared with $10,551 at non-HBCUs, a difference of $1,660 or 15.7 percent. In the corresponding OLS regression, the HBCU indicator predicted this same $1,660 increase with a p-value of 0.006. While statistically significant, this debt difference is relatively small compared to the steep default gap, suggesting that higher borrowing is part of, but not the sole driver behind, the poorer outcomes observed at HBCUs.

We then estimated multivariate models that include median debt alongside net cost of attendance (mean \$28,450), admission rate (mean 64 percent), six-year Pell completion (mean 42 percent), first-year retention (mean 72 percent), and median earnings six years post-enrollment (mean \$36,800). In the default model ($R^2 = 0.495$), each additional \$1,000 of median debt raised default probability by 0.24 percentage points, with a p-value less than 0.001. Crucially, once these controls are in place, the HBCU coefficient falls to +0.023 ($p = 0.38$), and the interaction term (debt $\times$ HBCU) is marginally non-significant ($2.4 \times 10$ per dollar, $p = 0.078$). This indicates that higher debt and institutional conditions, rather than HBCU status itself, largely explain default risk, though there is a hint that each dollar of debt may carry slightly more risk at HBCUs. In the parallel repayment model ($R^2 = 0.497$), each \$1,000 of extra debt reduced repayment probability by 0.29 points (p less than 0.001), and neither the main HBCU effect nor its interaction with debt reached significance, demonstrating that the negative impact of debt on repayment is consistent across institution types when we account for broader context.

Our synthetic student-level logistic regression (20,000 simulated observations) further illustrates how individual and institutional factors combine to shape default risk. Achieving 90 percent classification accuracy, the model shows that a one-point increase in GPA reduces default odds by about 12 percent, while Pell eligibility and HBCU attendance independently raise odds by factors of 1.5 and 1.4, respectively. Low family income (below \$30,000) increases default odds by roughly 25 percent compared to the next bracket, and each \$1,000 of institutional cost adds 1.5 percent to the log-odds of default. These coefficients confirm that, although debt and college context are dominant, student-level characteristics like academic preparedness and financial background are vital to understanding and forecasting repayment outcomes.

Across all models, rising debt emerges as the single strongest, most consistent predictor of loan distress: higher borrowing correlates with higher default and lower repayment, regardless of institution type once controls are applied. The simple regressions highlight stark disparities by HBCU status, the multivariate analyses attribute these gaps to debt burdens and institutional resource differences, and the synthetic-data exercise demonstrates the power of personalized risk profiling. Together, these findings argue that policy and institutional reforms should prioritize reducing overall debt, through mechanisms such as expanded grant aid, cost subsidies, and completion supports, and leveraging individual-level risk assessments to deliver targeted counseling and tailored aid. By addressing both the structural under-resourcing of colleges that serve low-income students and the specific vulnerabilities of individual borrowers, stakeholders can work toward equitable financial outcomes for all Pell Grant recipients.

All the statistics for the regression models are displayed in Appendix B.

# 8 Discussion

The research found that Pell Grant Recipients at HBCUs had greater loan and debt amounts, as well as default rates, compared to their counterparts in non-HBCUs. The research found an interesting discrepancy: the average median debt for Pell Grant Recipients at HBCUs was only slightly higher than that of students at non-HBCUs, but HBCU student default rates were almost twice as high.

The differences between the default rate and the repayment rate among Pell Grant recipients in HBCUs versus non-HBCUs were both statistically significant. Although the AIC values for the chosen models were satisfactory, the adjusted $R^2$ values are of concern. The median amount of debt for Pell Grant recipients in HBCUs was also found to be statistically significant; however, the constructed regression model may require further investigation, as both the AIC values and the adjusted $R^2$ values suggest possible overfitting. When controlling for other variables such as net cost, admissions rate, completion rate, retention rate, and median earnings, model fits improved significantly. Nonetheless, a notable concern is that in both models the HBCU indicator was not statistically significant, which potentially points to an area for further research. Synthetic student-level data were used to predict the odds of defaulting, and the model yielded favorable results. However, given that many confounding characteristics may affect students at HBCUs versus non-HBCUs, the predicted statistic may not be accurate.

As expected, the schools with more financially needy students had more students struggle to repay their debt amounts and often received higher amounts of loans. Although there are clear discrepancies in debt amounts and default rates between Pell Grant Recipients in HBCUs and non-HBCUs, this analysis could not pinpoint a causal relationship between HBCU enrollment and signs of post-graduate financial struggles. It does, however, suggest that the lack of resources at HBCUs may result in lower amounts of financial aid awards to students already receiving a Pell Grant. For example, the median debt amount for a Pell Grant Recipient at Vassar College is $12,531, significantly less than the average median debt amount for Pell Grant Recipients at HBCUs, even though the average cost to attend an HBCU is nearly $60,000 less than that of Vassar College.

Ideally, student-level information regarding debt amounts and postgraduate income is available. However, given the proprietary nature of an income dataset, only school-level information was available. Even if student-level data were available, debt repayment is a matter of financial intelligence and awareness, which cannot be quantified without highly proprietary information. As we also do not have intimate knowledge regarding how schools give aid and how their resources are used, it is difficult to pin responsibility on the school or the endowments that they receive. Based on the data available, however, we discovered that there was a significant discrepancy in endowment amounts between HBCUs and PWIs, as they averaged approximately $58 million and $385 million, respectively.

Ultimately, further research would have to delve into student-level data to find a causal relationship between a student's enrollment status at an HBCU and their ability to repay their loans. That does, however, face additional difficulties regarding proprietary information laws and maintaining data privacy. However, if research can find a reason why students at HBCUs struggle to repay their loans, it could potentially improve the lives of millions of Americans in helping them find financial security and decreasing financial burden post graduation.

# 9 Limitations

We were faced with several limitations while working on this project. Some limitations were profound enough to change the course of our research plan. We had previously planned to compare the loan default rate between black students in HBCUs and non-HBCUs and examine how the choice of academic major influences loan repayment for these students. This research had never been done before, and we were excited for the novel insights that we would discover in the research. Unfortunately, we realized that we didn't have enough data to make our findings. The data we planned to use didn't have the loan default rate disaggregated by race. Moreover, we weren't able to find any other public datasets that have this information. In light of this, we decided to pivot to using Pell Grant students for our study.

While looking at data for Pell Grant students, one of the research questions we wanted to explore was the relationship between the choice of major of Pell Grant students and the earnings of these students, and if there is a difference in this relationship between HBCUs and non-HBCUs. We thought we had the field of study data (which disaggregates by course major) to answer this question, but we soon found out that most of the entries in the variables are NaN values. We decided to stop exploring this question because using a dataset with many NaN values can lead to statistical errors. Getting the right data to find answers to our research proposals was difficult, and most of the time, we had to tweak our proposals to fit the data we had.

Furthermore, this project uses Ordinary Least Squares (OLS) regression, which assumes that the relationship between the dependent and independent variables are linear in nature. Although this is a standard approach in OLS, we wanted to state this assumption because real-world relationships (such as relationships between Cummulative Median Debt and Average Cost of tuition) may be more complex. Future research could explore nonlinear models or transformations to check potential curvilinear effects. Also, some independent variables, such as institutional cost, student debt and earnings, may be correlated. This could lead to multicollinearity, which we didn't check for in our project. We understand that while this doesn't bias OLS estimates, it can inflate statistical errors and complicate interpretation of individual predictor effects.

As discussed in the Discussion section, we were only able to get school-level data on the variables in our dataset given the proprietary nature of income datasets. This limited us from having a deeper understanding of how students from various socioeconomic backgrounds pay back their debts. Moreover, other factors could be in play which we might not necessarily be able to get data for, such as the school's philosophy or goals regarding financial aid, students' financial literacy, social bias in terms of schools that students attend which can affect their career path, resources available to students in school to aid them enhance their career and their path to financial freedom, socioeconomic discrimination of low income students, etc. We understand that this factors could also interfere in loan default and repayment rates in complex ways that might not be easily quantifiable on paper.

# 10 Ethical Considerations

Based on our analysis, when we controlled for factors, such as net cost, completion rate, median post-graduation earnings, and admission rate, we realized that it becomes insignificant whether a student went to an HBCU school or not. This implies that the differences in loan default and repayment rates among students is not necessarily tied towards HBCU status, but rather as a result of underlying structural inequities. There is likely no bias in loan servicing and repayment but rather an amplification of preexisting inequalities that disadvantage institutions that serve historically marginalized populations. However, we do realize that there may be some vulnerabilities relating to limited institutional support, especially when it comes to the relationship between loan debt and HBCU status. Hence, the marginally insignificant p value of 0.078, which we got from loan debt and HBCU status. But these vulnerabilities could also stem from long standing structural inequalities. There is a great need for reforms that address the root causes of educational and economic inequality and not to only focus on ensuring that borrowers have equal conditions for repayment.

Due to the sensitive nature of income and loan default datasets, we couldn't get student-level information on completion rate, repayment rate, default rate, income level, etc. Therefore, we made a synthetic dataset that allowed us to explore student loan default on an individualized basis. Although, the model which trained on this synthetic dataset had high predictive accuracy and aligned with our earlier findings, care must be taken to treat these results as conclusive because the data relies on assumptions that need to be validated against real-world outcomes. The results should be interpreted as suggestive and not to be blindly used to draw policy inferences.

# 11 Future Works

There are four potential paths following the conclusions of this research to further rule out any correlative relationships that may be dictating our observations. The first two paths are relatively more feasible than the latter two, but do not provide a clear and concise causal conclusion as the latter may achieve.

First, as aforementioned, student-level data can be incorporated to examine whether there is an underlying variable that may be dictating the debt default rate. Although higher average earnings could suggest higher repayment rates and thus lower default rates, that is not always the case. Financial literacy is a greater issue in identifying whether a person will be able to effectively partition their income and savings to repay their debt. However, overcoming data privacy regulations and laws can be a challenge, as income and debt information for individuals is hard to find, especially if Personally Identifying Information (PII) such as alma mater is included.

Second, an in-depth review of a university's expenditures can be helpful to understand whether schools are providing less financial aid due to funding or due to their spending philosophy. This would be a much easier approach to understanding a potentially underlying cause of the disparities we observed, but there is no guarantee that it tells the whole story.

Third, much more expensive than the earlier two but significantly more informative, a longitudinal study examining how students fare at and after college. Although the added costs may be significant, this could be an easier way to understand a student's financial literacy and debt accumulation at the student level than acquiring student-level data. The downside is that there is always a possibility of losing research candidates along the way, and although a lot easier while they attend the school, once they graduate, following up can become a very challenging process.

Finally, a matched randomized controlled experiment would be the best way to understand a potentially underlying causal relationship while accounting for confounding variables. It would be the most sound approach to account for the tremendous amount of confounding variables that may be causing the discrepancies that we observed. This is extremely unfeasible and unethical, however, as it consists of randomly assigning students to HBCUs or non-HBCUs.

Ultimately, these four paths could be useful in understanding why there were such differing default rates among the observed universities. HBCUs are an important and foundational part of the United States' educational history, and it has long supported the Black academics of America. If there is an underlying issue with them that puts the alumni of those schools under financial distress, it must be researched and understood before another obstacle attempts to hinder the success of Black Americans.

# 12 If we could start over

Looking back, if we could restart this project, we'd definitely spend more time at the beginning really digging into the dataset. With around 3000 variables, we could have tried to bring in more details or even combined this data with other source. We would also make sure to go through the data description carefully right away. This would have helped us know exactly what the data could tell us and shape our research questions to be more targeted and answerable with the data we had. Moreover, we would try to identify causal relations by exploring if there were any specific events, like policy changes, that might have caused shifts in loan outcomes. In our regression models, we would also make sure to check for multicollinearity to ensure the validity of the models and our interpretation of them. Lastly, for we could try to strengthen the synthetic data. We would aim to make the simulations even more realistic by incorporating more real-world complexities and dependencies between variables. We would also try to find some real-world student level loan outcome data, even if it's just a small sample or aggregated in some way, to validate how well our synthetic data and the model trained on it actually reflect reality.

# References

[1] Looney, A., & Yannelis, C. (2015). A crisis in student loans?: How changes in the characteristics of borrowers and in the institutions they attended contributed to rising loan defaults. *Brookings Institution.*

[2] Lee, J. M., & Keys, S. W. (2013). Repositioning HBCUs For the Future: Access, Success, Research & Innovation. *APLU Office of Access and Success Discussion Paper 2013-01*. Washington, DC: Association of Public and Land-grant Universities.

[3] Ma, J., Baum, S., Pender, M., & Bell, D. (2014, November 30). Trends in college pricing, 2015. *Trends in Higher Education Series*. College Board. Retrieved from `https://eric.ed.gov/?id=ED572540`

[4] United States Department of Education College Scorecard. (2025, April). Technical Documentation: College Scorecard Institution-Level Data. *College Scorecard*. Retrieved from `https://collegescorecard.ed.gov/assets/InstitutionDataDocumentation.pdf`

# A Variable Definitions

The following is a list of the variables that we used in our analysis and their definitions:

HBCU - Historically Black Colleges and Universities

INSTNM - Institution's name

BBRR4_FED_PELL_DFLT - Default rate for Pell Grant Students on BBRR (Borrower Based Repayment Rate) on Federal Loans after four years of entering repayment

BBRR4_FED_PELL_PAIDINFULL - Rate of Pell Grant Students who paid their loans in full for BBRR on Federal Loans after four years of entering repayment

PELL_DEBT_MDN - Cumulative Median Student Debt for Pell Grant students

COSTT4_A - Average cost of Attendance, Tuition, and Fees for academic year institutions

ADM_RATE - Admission Rate

C150_4_PELL - Completion rate for Pell Grant students who complete within 150 percent of their expected completion time (6 years for 4-year institutions)

RET_FT4 - Retention rate for full-time students in a 4-year institution

MD_EARN_WNE_P6 - Median entry cohort earnings 6 years after a student enrolls in college

# B Regression Model Statistics

```
==============================================================================
OLS Regression: Repayment_Rate ~ HBCU
==============================================================================
                            OLS Regression Results
==============================================================================
Dep. Variable:         Repayment_Rate   R-squared:                       0.019
Model:                            OLS   Adj. R-squared:                  0.019
Method:                 Least Squares   F-statistic:                     28.41
Date:                Fri, 11 Apr 2025   Prob (F-statistic):           1.14e-07
Time:                        18:32:28   Log-Likelihood:                 1832.9
No. Observations:                1445   AIC:                            -3662.
Df Residuals:                    1443   BIC:                            -3651.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.1308      0.002     71.657      0.000       0.127       0.134
HBCU          -0.0508      0.010     -5.330      0.000      -0.070      -0.032
==============================================================================
Omnibus:                      595.832   Durbin-Watson:                   1.721
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3317.486
Skew:                           1.852   Prob(JB):                         0.00
Kurtosis:                       9.433   Cond. No.                         5.33
==============================================================================
```

Figure 6: OLS Regression for HBCU on Repayment Rate

```
==============================================================================
OLS Regression: Default_Rate ~ HBCU
==============================================================================
                            OLS Regression Results
==============================================================================
Dep. Variable:           Default_Rate   R-squared:                       0.136
Model:                            OLS   Adj. R-squared:                  0.135
Method:                 Least Squares   F-statistic:                     227.1
Date:                Fri, 11 Apr 2025   Prob (F-statistic):           9.05e-48
Time:                        18:32:28   Log-Likelihood:                 2355.8
No. Observations:                1445   AIC:                            -4708.
Df Residuals:                    1443   BIC:                            -4697.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0881      0.001     69.292      0.000       0.086       0.091
HBCU           0.1000      0.007     15.070      0.000       0.087       0.113
==============================================================================
Omnibus:                      346.243   Durbin-Watson:                   1.779
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              812.701
Skew:                           1.305   Prob(JB):                    3.34e-177
Kurtosis:                       5.586   Cond. No.                         5.33
==============================================================================
```

Figure 7: OLS Regression for HBCU on Default Rate

```
=======================================================
OLS Regression: PELL_DEBT_MDN ~ HBCU
=======================================================
                            OLS Regression Results
==============================================================================
Dep. Variable:          PELL_DEBT_MDN   R-squared:                       0.005
Model:                            OLS   Adj. R-squared:                  0.005
Method:                 Least Squares   F-statistic:                     7.574
Date:                Fri, 11 Apr 2025   Prob (F-statistic):            0.00600
Time:                        18:32:28   Log-Likelihood:                -14143.
No. Observations:                1445   AIC:                         2.829e+04
Df Residuals:                    1443   BIC:                         2.830e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   1.746e+04    115.570    151.050      0.000    1.72e+04    1.77e+04
HBCU        1660.7555    603.452      2.752      0.006     477.019    2844.492
==============================================================================
Omnibus:                       15.838   Durbin-Watson:                   1.573
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               10.451
Skew:                          -0.045   Prob(JB):                      0.00538
Kurtosis:                       2.593   Cond. No.                         5.33
==============================================================================
```

Figure 8: OLS Regression for HBCU on Pell Debt Median

```
=======================================================
OLS Regression: Default_Rate
=======================================================
                            OLS Regression Results
==============================================================================
Dep. Variable:           Default_Rate   R-squared:                       0.495
Model:                            OLS   Adj. R-squared:                  0.492
Method:                 Least Squares   F-statistic:                     175.9
Date:                Fri, 11 Apr 2025   Prob (F-statistic):         7.58e-207
Time:                        18:32:29   Log-Likelihood:                 2743.7
No. Observations:                1445   AIC:                            -5469.
Df Residuals:                    1436   BIC:                            -5422.
Df Model:                           8
Covariance Type:            nonrobust
======================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
Intercept              0.2907      0.009     31.964      0.000       0.273       0.309
PELL_DEBT_MDN      -1.786e-06   2.63e-07     -6.782      0.000    -2.3e-06   -1.27e-06
HBCU                   0.0234      0.027      0.881      0.378      -0.029       0.076
PELL_DEBT_MDN:HBCU  2.421e-06   1.37e-06      1.763      0.078   -2.72e-07    5.11e-06
COSTT4_A            4.051e-07   7.12e-08      5.691      0.000    2.65e-07    5.45e-07
ADM_RATE              -0.0346      0.006     -6.233      0.000      -0.045      -0.024
C150_4_PELL           -0.0693      0.008     -9.058      0.000      -0.084      -0.054
RET_FT4               -0.0693      0.011     -6.246      0.000      -0.091      -0.048
MD_EARN_WNE_P6      -1.53e-06   9.29e-08    -16.464      0.000   -1.71e-06   -1.35e-06
==============================================================================
Omnibus:                      501.836   Durbin-Watson:                   1.838
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2094.440
Skew:                           1.624   Prob(JB):                         0.00
Kurtosis:                       7.923   Cond. No.                     1.88e+06
==============================================================================
```

Figure 9: OLS Regresion for Default Rate

```
========================================================
OLS Regression: Repayment_Rate
========================================================
                            OLS Regression Results
==============================================================================
Dep. Variable:         Repayment_Rate   R-squared:                       0.497
Model:                            OLS   Adj. R-squared:                  0.494
Method:                 Least Squares   F-statistic:                     177.1
Date:                Fri, 11 Apr 2025   Prob (F-statistic):          6.41e-208
Time:                        18:32:29   Log-Likelihood:                 2314.8
No. Observations:                1445   AIC:                            -4612.
Df Residuals:                    1436   BIC:                            -4564.
Df Model:                           8
Covariance Type:            nonrobust
======================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
Intercept             -0.0130      0.012     -1.063      0.288      -0.037       0.011
PELL_DEBT_MDN      -2.866e-06   3.54e-07     -8.089      0.000   -3.56e-06   -2.17e-06
HBCU                   0.0337      0.036      0.944      0.346      -0.036       0.104
PELL_DEBT_MDN:HBCU -1.649e-06   1.85e-06     -0.893      0.372   -5.27e-06    1.97e-06
COSTT4_A            3.424e-07   9.58e-08      3.575      0.000    1.54e-07     5.3e-07
ADM_RATE              -0.0397      0.007     -5.309      0.000      -0.054      -0.025
C150_4_PELL            0.1088      0.010     10.570      0.000       0.089       0.129
RET_FT4                0.1110      0.015      7.438      0.000       0.082       0.140
MD_EARN_WNE_P6       1.38e-06   1.25e-07     11.035      0.000    1.13e-06    1.63e-06
==============================================================================
Omnibus:                      448.302   Durbin-Watson:                   1.856
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1984.279
Skew:                           1.412   Prob(JB):                         0.00
Kurtosis:                       7.998   Cond. No.                     1.88e+06
==============================================================================
```

Figure 10: OLS Regression for Repayment Rate