

US STUDENT DEBT CRISIS

COHORT B TEAM 8

Kyle Blackburn, Fernanda Lin, Hongyang Liu, Lyufan Pan & Mansi Tolia



Objective

TO

Student Debt

45 Million

Borrowers

\$1.5 Trillion

Existing debt

~ 40 %

Estimated Default
By 2023



Reduce Student Loan Default Rates

Identify Factors

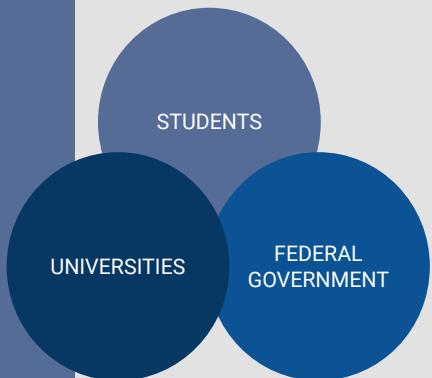
Determine variables that contribute most to students defaulting

Inform Policymakers

Enable policymakers to make more informed decisions

Help Students

Help students make better decisions when choosing their university



Data

20

Our Data



College Scorecard

Tool developed by the Department of Education to help students find information about universities, career paths, and more

Integrated Postsecondary Education Data



Statistics, research, and evaluation arm of Department of Education

Executive Summary



30

Executive Summary

Imputation

Lack of standardized reporting caused high amounts of missing values

Machine Learning

Evaluate imputations with ML and determine the critical attributes when defaulting



Females/caucasians, universities with high salaried faculty, and private non-profit colleges see lower default rates

Key Findings

Fewer student loans, increased student-faculty ratios, predicted default considered rates in university selection

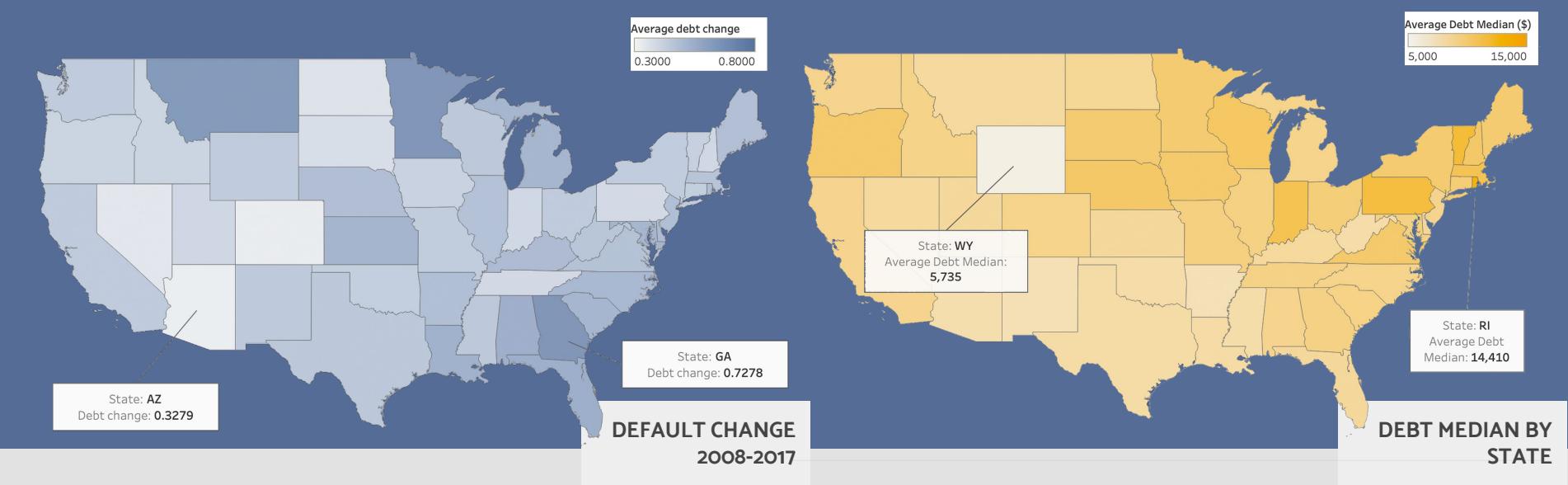
Policy advice

Exploratory Data Analysis

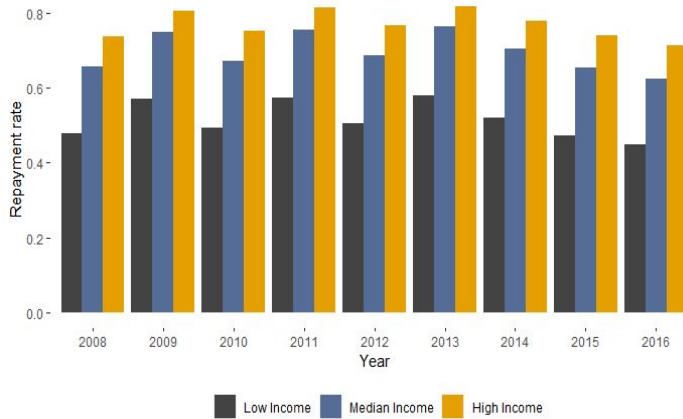


04

EDA



Repayment rate at different income level



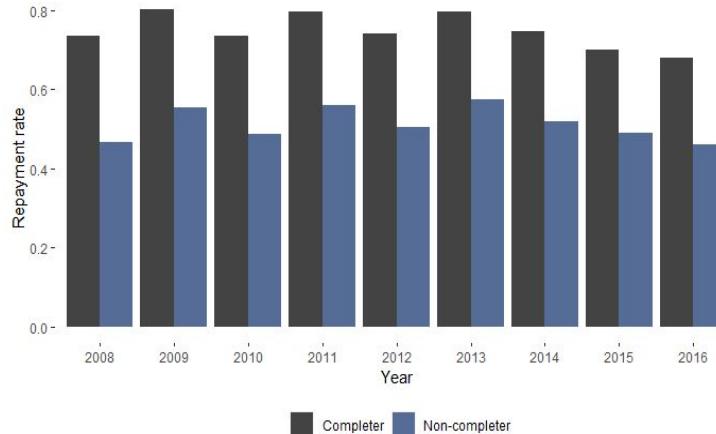
Income Level

Students from high-income families tend to have higher repayment rates, while students with low-income see the opposite

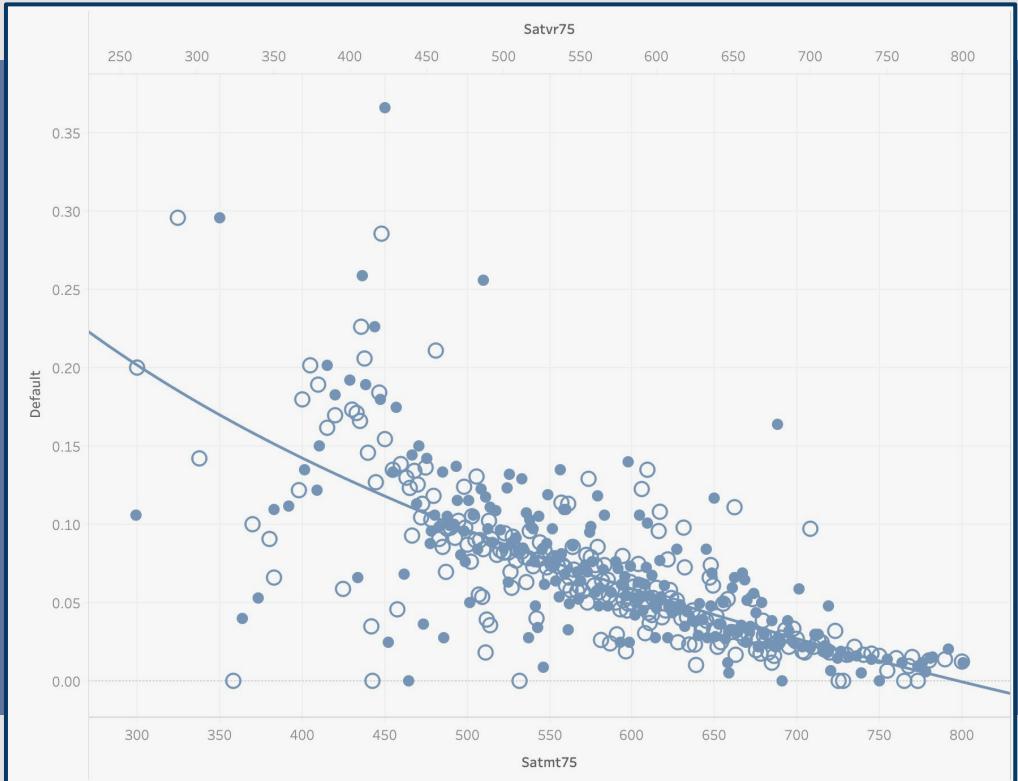
Completion Status

Students who complete their college study have much higher repayment rates compared with students who dropout

Completers VS non-completers repayment rate



SAT Scores



Higher SAT
scores
suggest lower
default rates

An aerial photograph of a dense urban skyline, likely Boston, featuring numerous skyscrapers of varying heights and architectural styles. In the foreground, a large, dark body of water is visible, with several boats and piers extending into it. The city extends far into the background under a clear blue sky.

Methodology

SO

Scope

Original

10%

- Determine if US education is over- or under-valued based on federal loan defaults

Result

20%

- Determine most relevant factors of student loan default rates, predict default rates, advise policy

40%

- Individual-level loan data
- Data cleaning

50%

- Cohort-level loan data
- 11 imputation methods

50%

- Risk score
- Attribute individual defaults to student characteristics (universities, loan amounts, lenders)

30%

- Prediction, clustering
- Defaulting rates based on university attributes

Methodology

Preprocessing
Variable selection, EDA,
dimension reduction,
[52,284 x 60]

Ensemble Imputation

11 different versions of data
with various imputation
methods

Random Forest

100 trees for each, tested
out-of-sample MSE to
identify best imputation

Segmentation

Analyze performance of 3
clusters

Regression

Random split 80/20,
forward-selection model to
identify top variables in
predicting default rates

Findings



60

Significant Variables

	<h2>MAJOR</h2> <p>(+) Education, Religious studies, Business, Theology (-) Law Enforcement, Liberal arts, Repair Technicians, Construction</p>
---	--

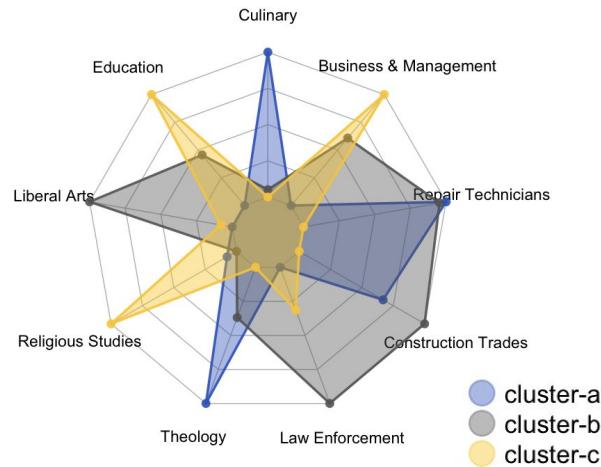
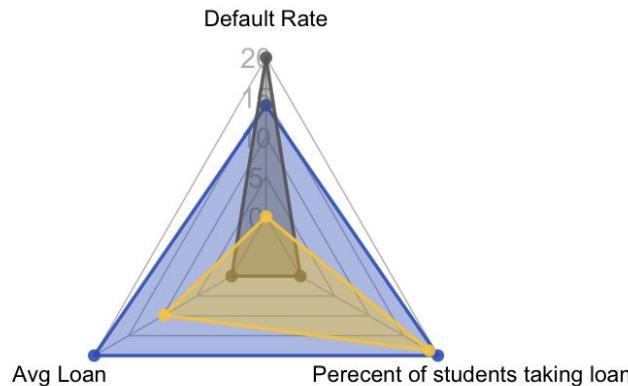
	<h2>CONTROL</h2> <p>Default rate from high to low 1) Private for-profit 2) Public 3) Private not-for-profit</p>
---	---

	<h2>YEAR</h2> <p>Default Rates continue to increase through years</p>
	<h2>GRADUATION RATES</h2> <p>Higher Caucasian and Women graduation rates within the cohort is correlated with lower default rate</p>

	<h2>DEBT & LOANS</h2> <p>Higher Federal loans taken is correlated with higher default rates</p>
---	--

	<h2>FACULTY</h2> <p>Higher Faculty salaries and Faculty-to-student ratio is correlated with lower default rates</p>
---	---

Clusters on Fields of Study

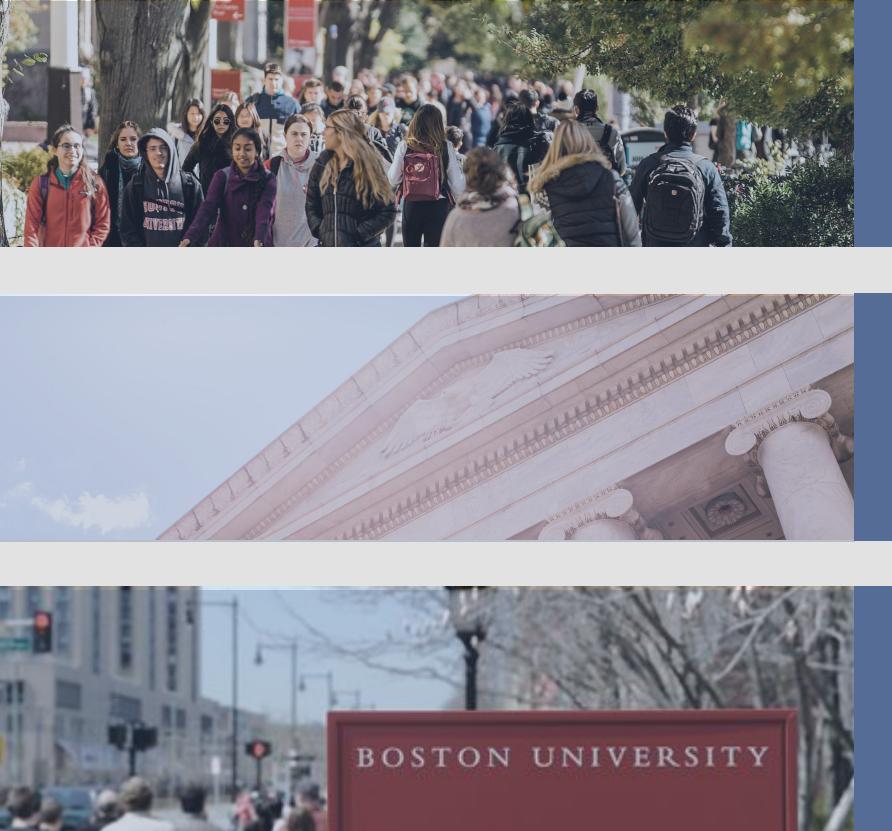


A	<p>Default Rates - Relatively High Average Loan Amount - Highest Percent of Loans Awarded - High</p> <ul style="list-style-type: none"> Personal and Culinary Services Theology and Religious Vocations Mechanic and Repair Technologies
B	<p>Default Rates - Highest Average Loan Amount - Low Percent of Loans Awarded - Low</p> <ul style="list-style-type: none"> Liberal Arts Law Enforcement and Homeland Security Construction Trades
C	<p>Default Rates - Lowest Average Loan Amount - High Percent of Loans Awarded - High</p> <ul style="list-style-type: none"> Business and Management Philosophy and Religious Studies Education

Conclusion



ZO



Recommendations

STUDENTS

- Evaluate predicted default rates when choosing university
- Focus on fields like Education, Religious Studies, Business, Management, Marketing and Theology
- Avoid fields like Law Enforcement, Liberal Arts, Repair Technicians and Construction

FEDERAL GOVERNMENT

- Increase the requirement of SAT and ACT scores
- Reduce the overall amount of federal loans awarded

UNIVERSITIES

- Provide higher incentives to students to finish school
- Reduce class sizes and focus on teaching quality
- Charge lower fees or reduce student costs

Discussion

INFLUENTIAL FACTORS

Cohort's demographic composition,
Cohort's completion rates &
Field of study

RECOMMENDATIONS

Provided for students, universities and
the federal government

LIMITATIONS

Recommendations might be biased
and discriminatory
Limited access to data

FUTURE WORK

Analyze private loans and student-specific
data for each school
Use more variables with fewer missing values
Impute missing values more accurately

THANK YOU

COHORT B TEAM 8



CREDITS

- Presentation template by Slidesgo
- Icons by Flaticon
- Infographics by Freepik
- Images created by Freepik, Unsplash & Boston University