

UNTIL DEBT  
TEAR US APART

# The Price of Education

Predicting the Level of Debt for Programs & Majors

# Presentation Agenda

- 
- Business Problem 01
  - Methodology 02
  - Results 03
  - Challenges 04
  - Next Steps 05
  - Limitations 06

# Student Debt is National Problem

Getting an education today in the US requires a much bigger sacrifice than it did for previous generations. In the past 30 years, public universities have tripled their attendance costs, and have private university costs have more than doubled.

**11%**

Of student debt is past due/in default

**54%**

Of College Attendees Take on Debt

**\$1.4 Trillion**

Outstanding Student Debt in 2019

**\$35,359**

Average Debt per Borrower



## Can Debt Forgiveness Save the Economy?

Moody's Investor Service predicts that eliminating the student debt owed could stimulate the economy as much as proposed tax cuts.



# Our Data

We used the public College Scorecard database provided by the US Department of Education. This provides information on university costs of attendance as well as the financing needed to cover those costs.

Because of the age of the dataset, we were able to find the debt to income ratio of an average graduate of each university by dividing the average debt upon graduation by the average income a student made after 10 years, since most loans are 10-years on average.

## Business Case:

Create more transparency  
on expected debt-to-income  
ratio at graduation before  
you attend.



# Technical Methodology



## Three Data Divisions

1. For-Profit
2. Non-Profit/Private
3. Non-Profit/Public

## Dropping Nulls

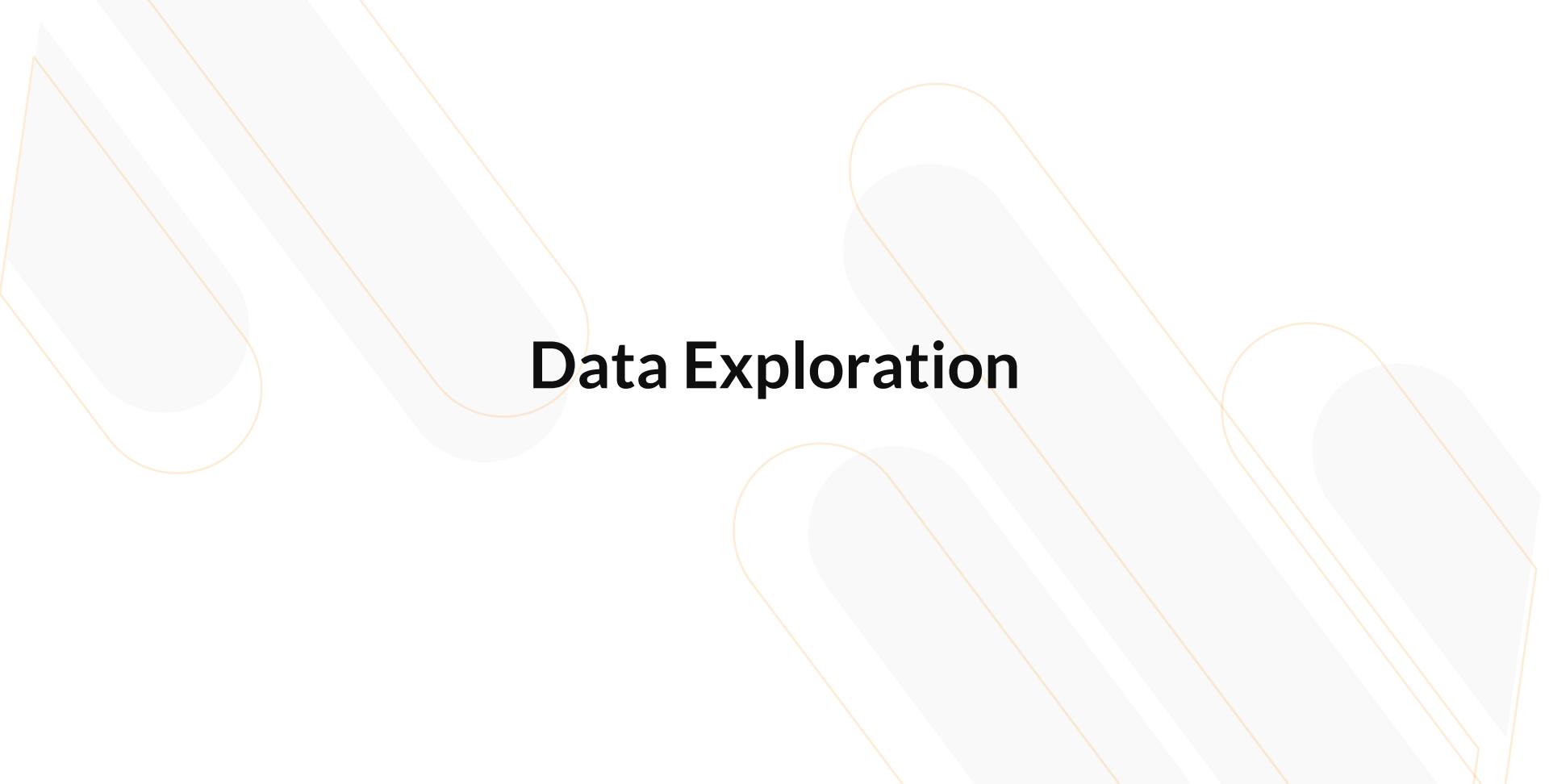
We dropped NaN values to scrub the one-hot-encoded dataset

## Criteria

- Debt-to-Income Ratio
- Demographic Breakdown
- Tuition and Fees

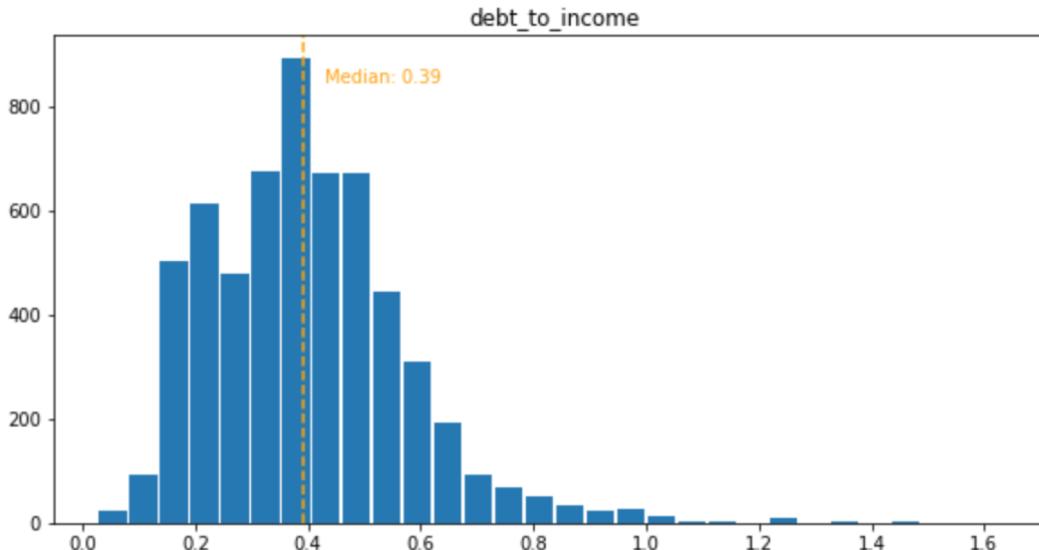
## Predictions

- Features of High Debt Institutions
- DTI with GDP

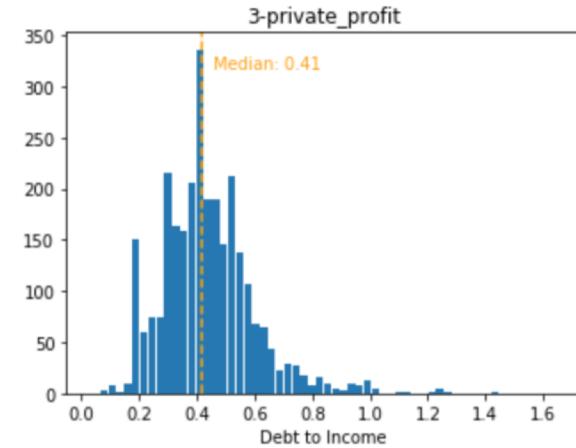
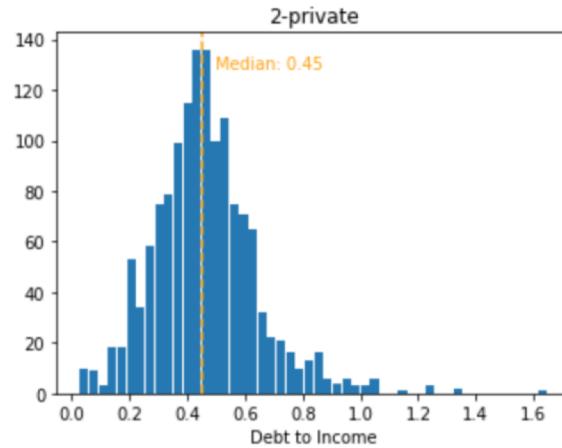
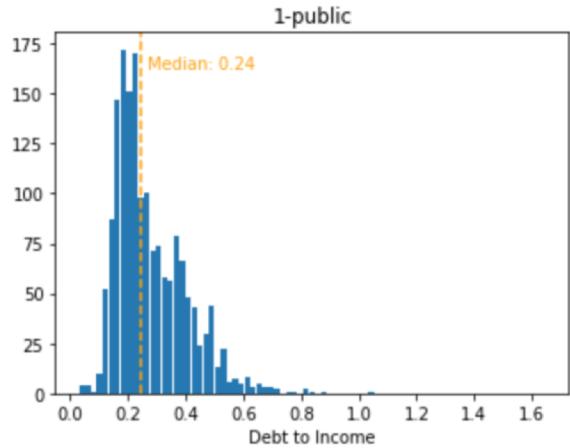


# Data Exploration

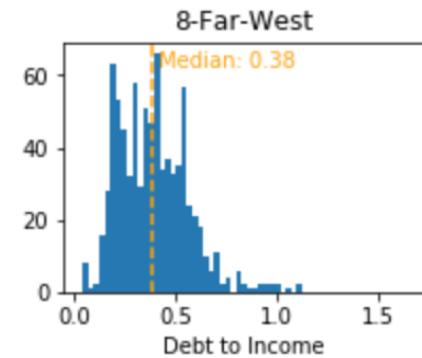
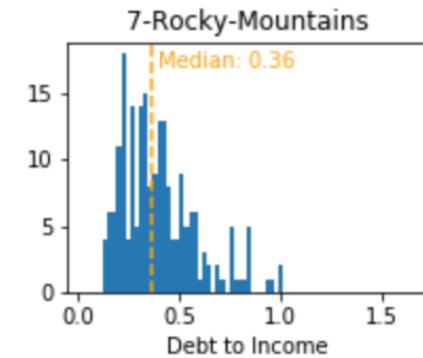
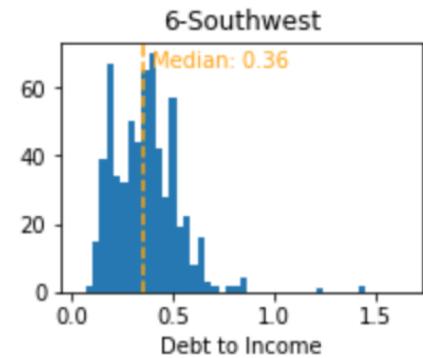
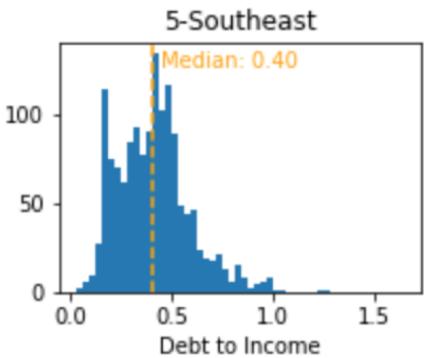
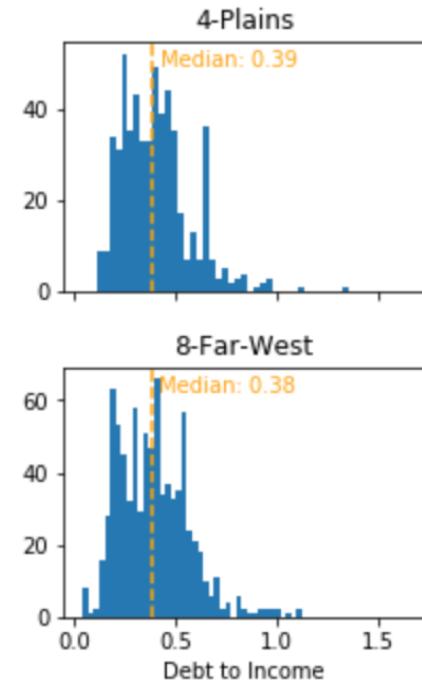
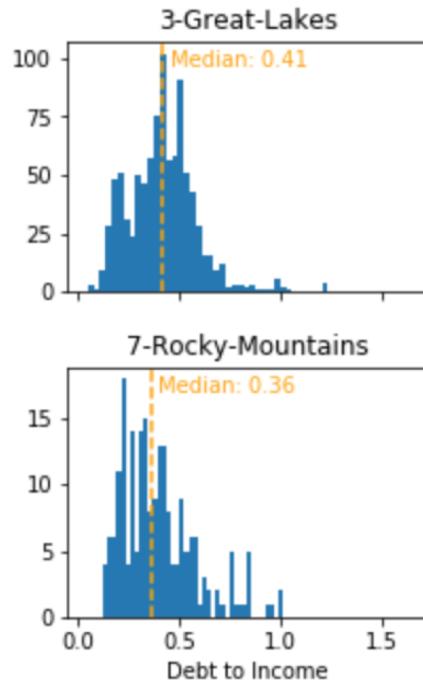
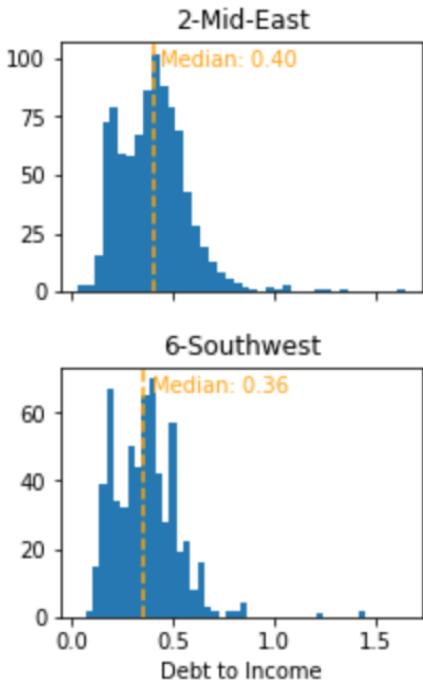
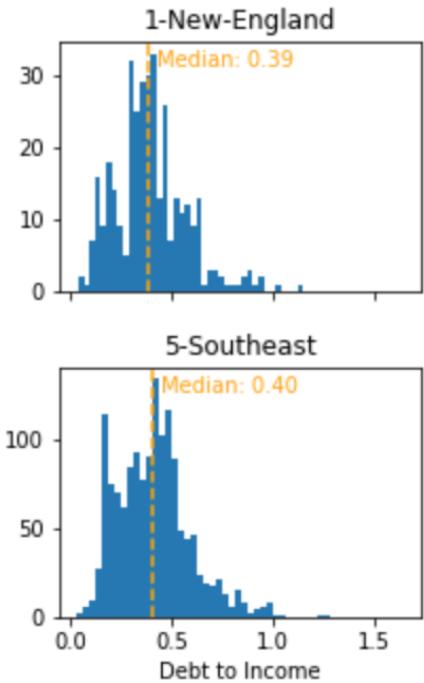
# Debt-to-Income Ratio: All



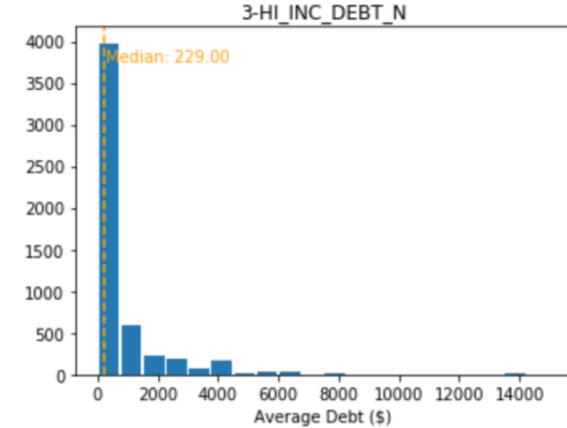
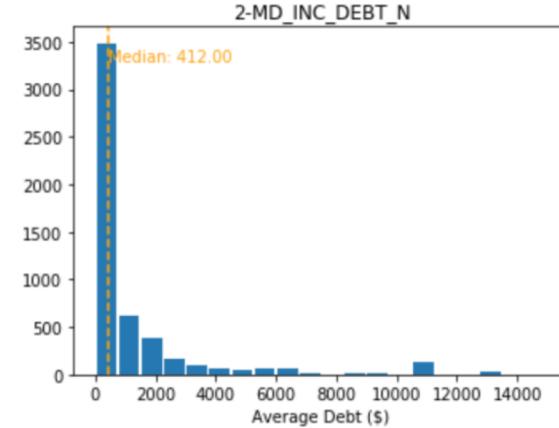
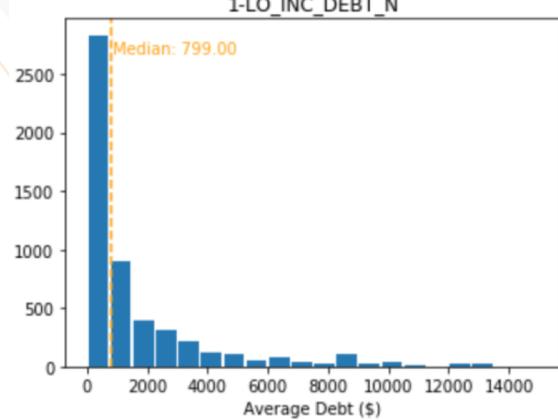
# Debt-to-Income Ratio: School Type



# Debt-to-Income Ratio: Geographical Breakdown

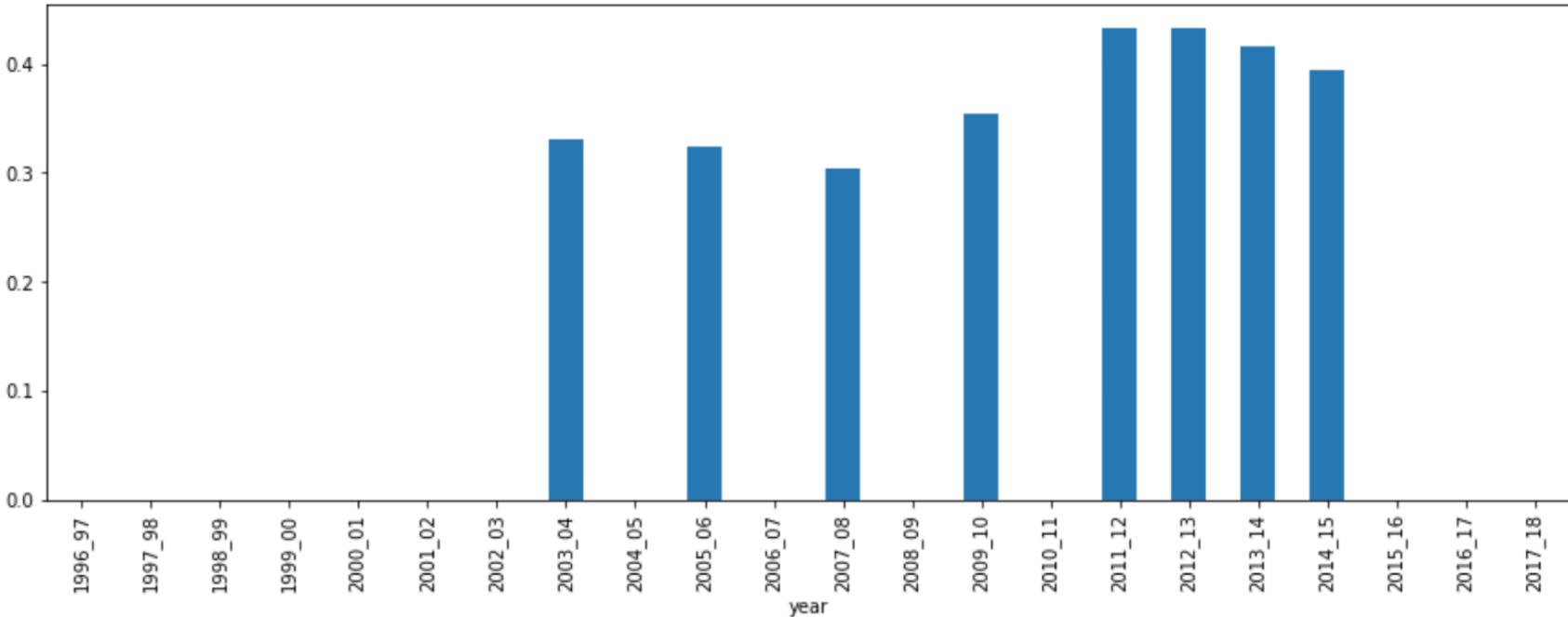


# Debt-to-Income Ratio: Household Income



# Debt-to-Income NYU

NYU Debt-To-Income Prediction



# Prediction Model

# Modeling Techniques

## K-Folds

Selected training set of data to hold against smaller test set for decision tree model

## Clustering

Clustered debt-to-income ratio into two KMeans to find a guideline for establishing binary criteria

## Box Plots

Generated box plots to visualize any effect top features had on D-T-I

## Decision Tree

Examine different features and create binary nodes from comparing different entropy values across the features

## Random Forest

Iterates over features to combine decision trees for D-T-I ratio and reveal top 20 predictive features.

# Random Forest Accuracy

	<b>n=10</b> (avg ~8s)	<b>n=100</b> (avg ~8s)	<b>n=1000</b> (avg ~63s)
Public	0.5985	0.7642	0.7736
Private	0.5417	0.6814	0.6869
Private for Profit	0.4058	0.6221	0.6349

\*OOB Score uses out-of-bag samples to estimate the R^2 on unseen data

# Random Forest Important Features

## Public

Average SAT equivalent score of students admitted for all campuses rolled up to the 6-digit OPE ID

First-time, full-time student retention rate at four-year institutions

Admission rate for all campuses rolled up to the 6-digit OPE ID

Total share of enrollment of undergraduate degree-seeking students who are black

Share of undergraduate, degree-/certificate-seeking students who are part-time

## Private

Percentage of degrees awarded in Visual And Performing Arts.

In-state tuition and fees

Admission rate for all campuses rolled up to the 6-digit OPE ID

Percentage of degrees awarded in English Language And Literature/Letters.

First-time, full-time student retention rate at four-year institutions

## Private for Profit

In-state tuition and fees

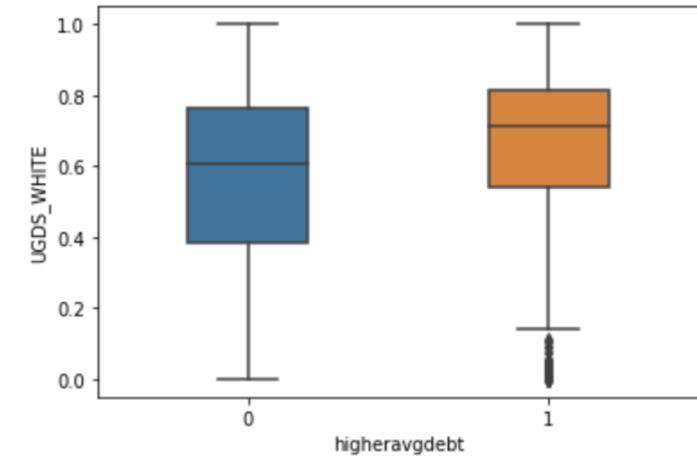
Percentage of degrees awarded in Visual And Performing Arts.

Share of undergraduate, degree-/certificate-seeking students who are part-time

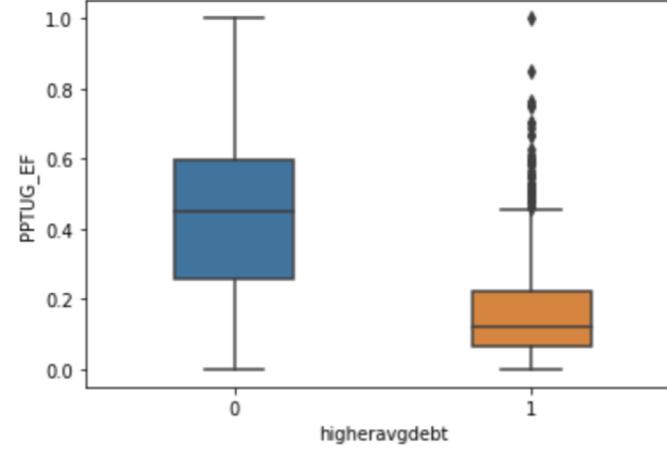
Tuition, fees, books, and supply charges for largest program (full program)

Number of graduate students

# Box Plots



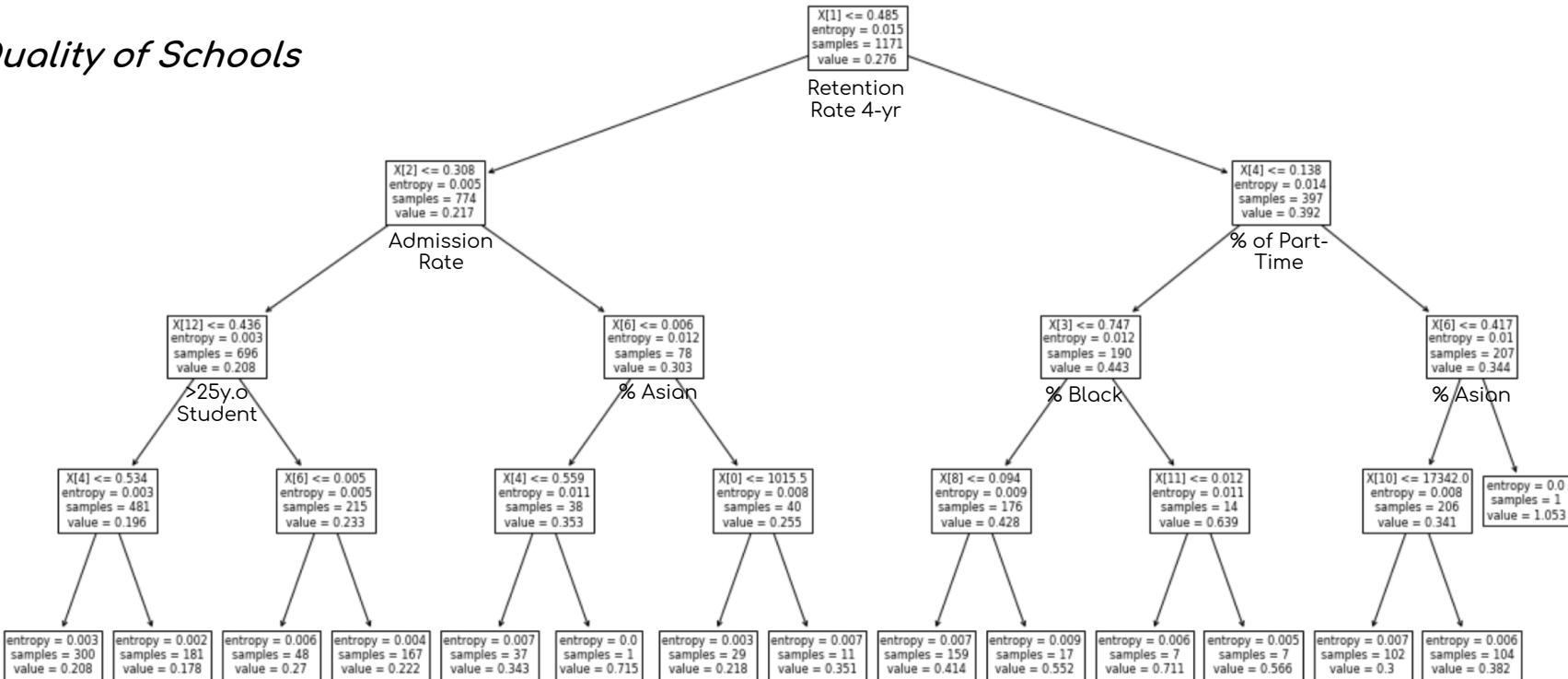
The percentage of **white students** seems to indicate that less diversity results in a **slightly higher average distribution** for debt.



Schools with a **higher percentage of part time students** show that they have a **lower average debt** than schools with 20% or less part time students

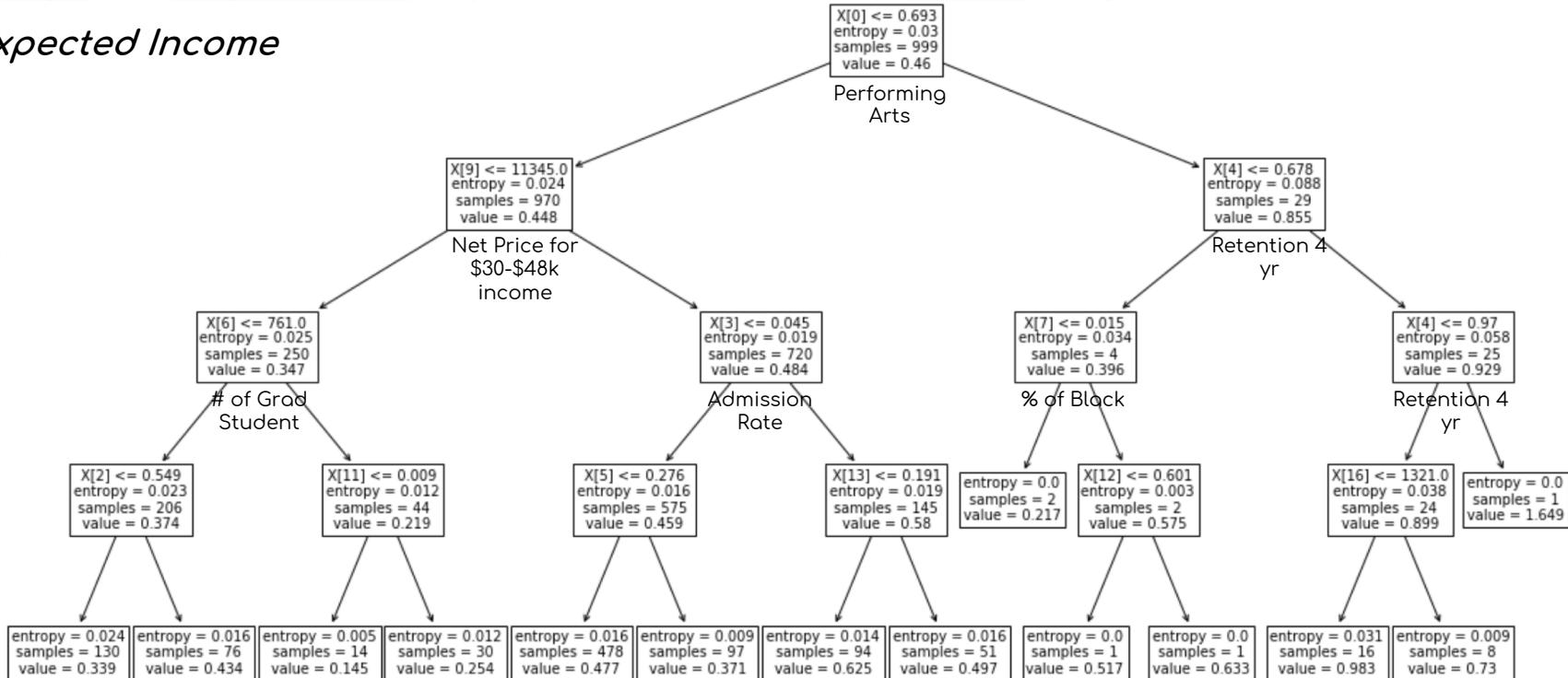
# Decision Tree: Public School

*Quality of Schools*

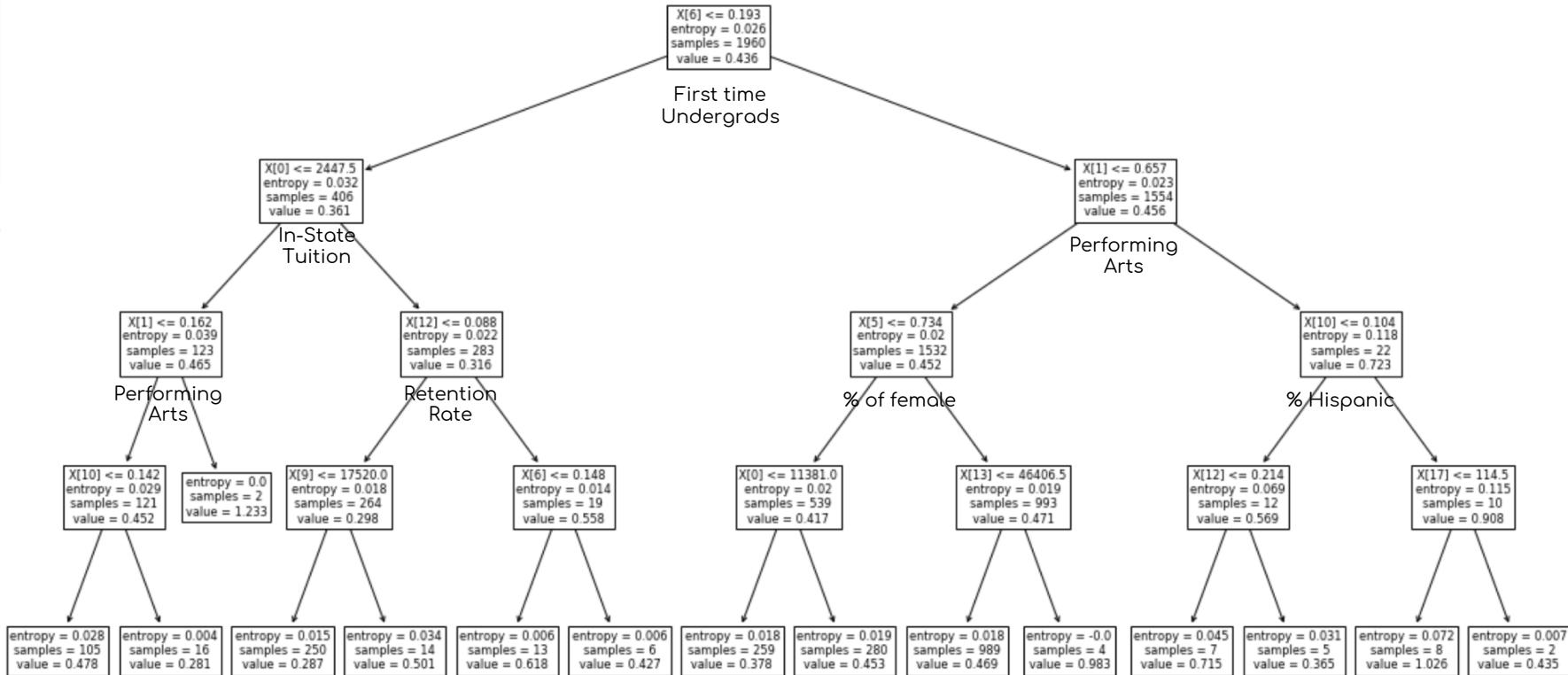


# Decision Tree: Private School

*Expected Income*



# Decision Tree: Private School - For Profit



# Our Challenges

When dealing with such a large dataset with so many features, we had to determine which incomplete features to keep or delete for analysis and reference the dictionary to understand the insights our work revealed.



Technical

- Debt-to-income ratio was difficult to designate as a binary value for the box plots.
- Decision Tree Rendering on a Windows Device was difficult to create.



Business

- Analyzing the significance of box plot and decision tree results amongst a diverse set of features.
- Creating a narrative to fit into our business case was not a clear path.

# Results

Consideration of Debt is Important when Deciding your Future Institution



Each School Type had differing features that better predicted the expected debt-to-income ratio.

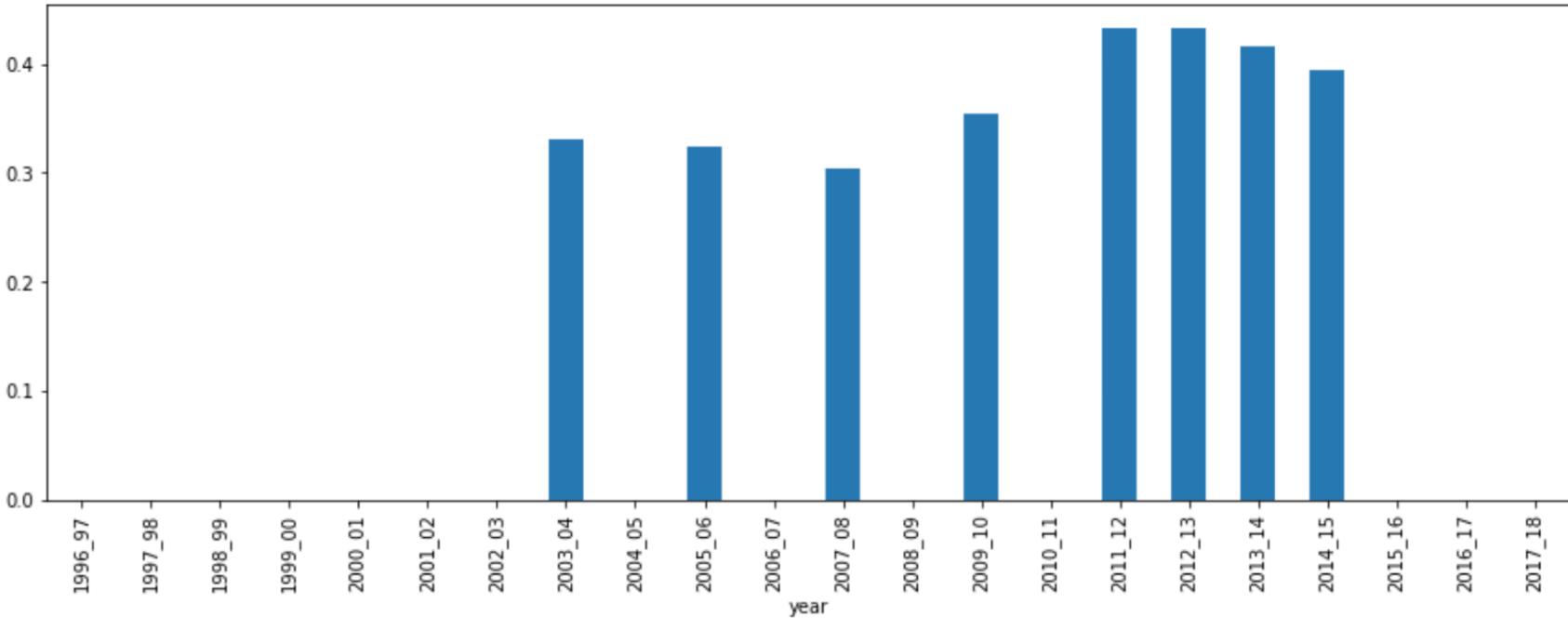
Public schools' D-T-I is heavily influenced by the program type, such as part-time and admission rate.

Private schools, both for- and non-profit schools, are influenced by expected income that a major would have, represented by % of Performing Arts majors.



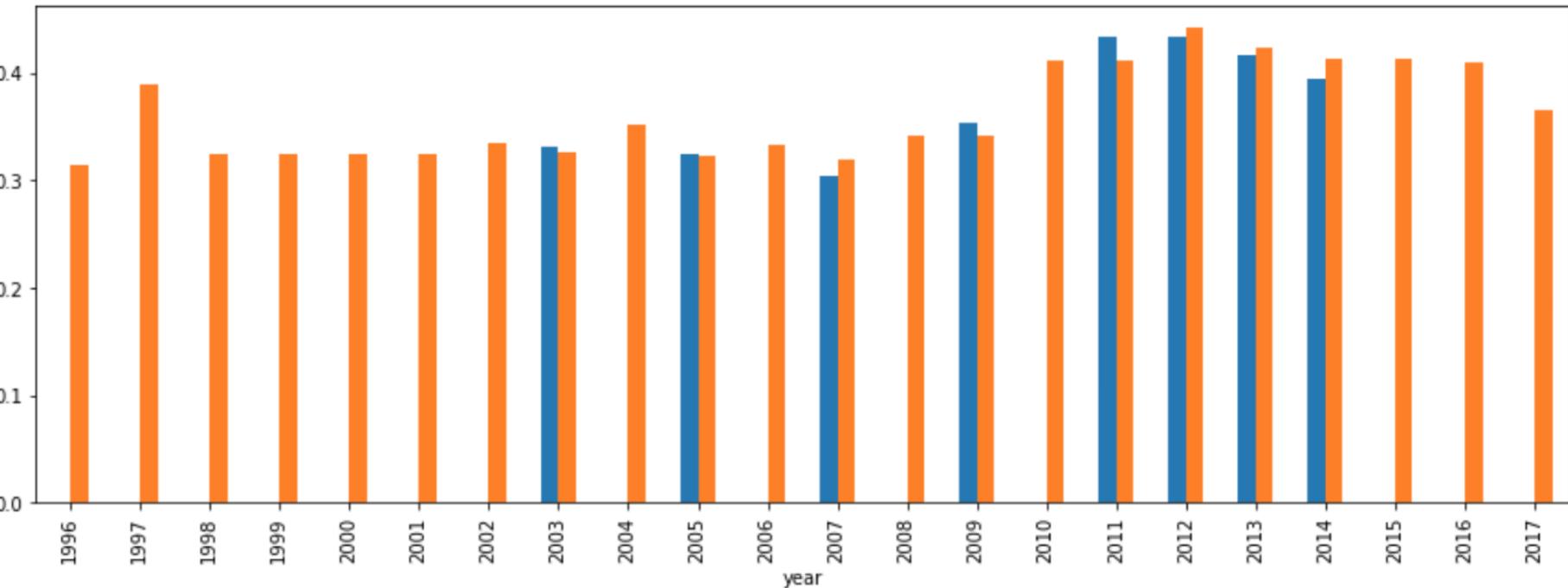
# Debt-to-Income NYU

NYU Debt-To-Income Prediction



# NYU: Predicting Debt-to-Income (Full Data)

NYU Debt-To-Income Prediction



# Predicting Debt-to-Income: Full Year Adding External Data (GDP)

## U.S. GDP Data

Source:

Gross domestic product (GDP) of the United States from 1990 to 2018 (in billion current U.S. dollars )

year	GDP	
0	2018	20500.6
1	2017	19485.4
2	2016	18707.2
3	2015	18219.3
4	2014	17521.7

# Predicting Debt-to-Income: Full Year Adding External Data (GDP)

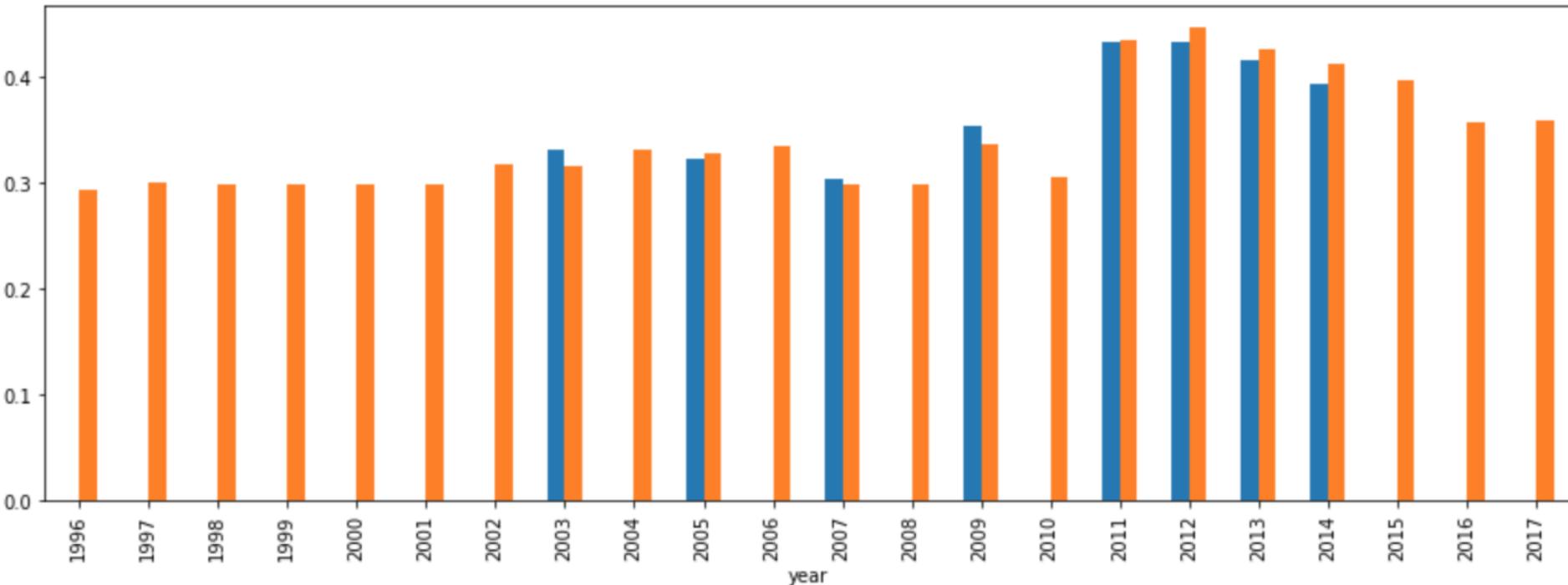
OOB Score	Before External Data	After External Data	Percent Change
Public	0.7642	0.8779	14.88%
Private	0.6814	0.8097	18.83%
Private for Profit	0.6221	0.7780	25.06%

\*OOB Score uses out-of-bag samples to estimate the R^2 on unseen data

\*n=100 trees used

# NYU: Predicting Debt-to-Income (Full Data + GDP)

NYU Debt-To-Income Prediction





# When choosing a university, consider.....

## Debt-to-Income Predictors

1. Economic Health of the US
2. Competitiveness of the Program
3. Demographic Breakdown of the Program
4. Potential salary of your major

# Limitations

- Granularity of Data
  - By University vs. Individual
  - By University vs. Schools/Program within a University
- Size of Data



# Thank You

Any Questions?