STATS 131: Final Project

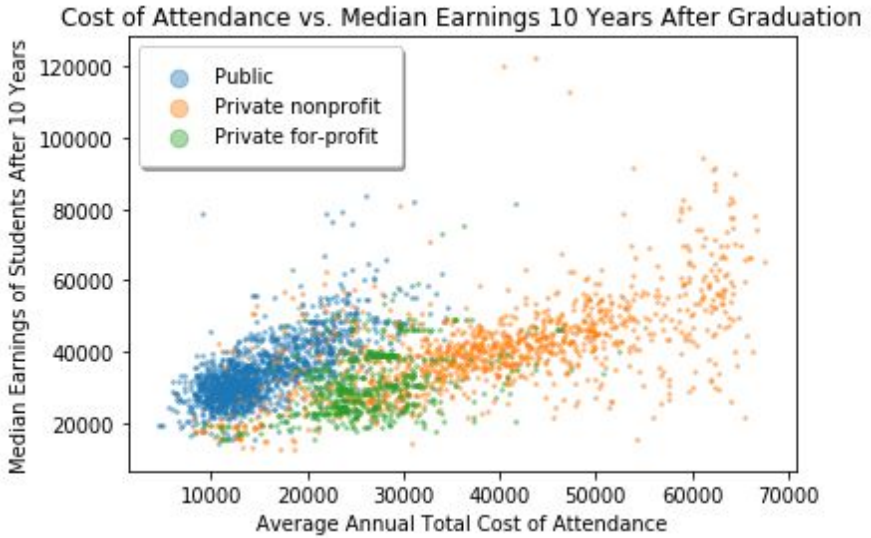
Analyzing College Scorecard Data

by

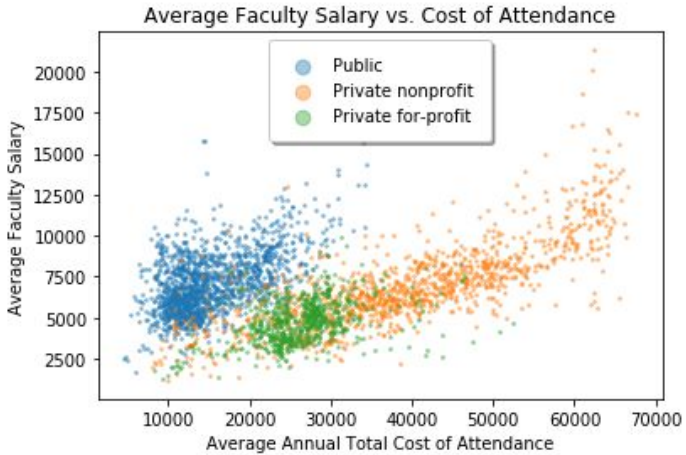
Brian Lin, Aida Ylanan and Ignat Kulinka

Analysis: Institution Type

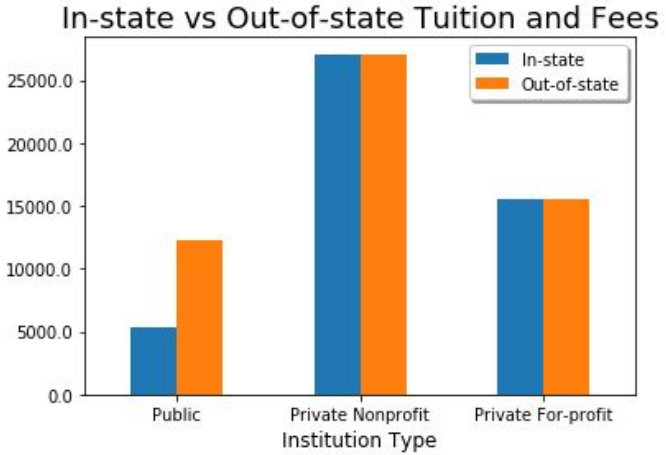
- Comparing public, private for-profit and private nonprofit institutions
  - Based on median earnings of students after 10 years and average annual total cost of attendance
- Outcomes:
  - Public schools appear to be a better return on investment, followed by private nonprofit and private for-profit last
  - Best option: elite private nonprofit schools



Analysis: Salaries & Tuition



- Public schools offer on average better salaries compared to other schools
- Private schools only offer faculty good salaries only at the most expensive colleges
- No difference in pay for private for-profits vs nonprofit



- Public school is the cheapest option even for out-of-state students.
- Private for-profit schools are still more expensive than public schools
- Best option: local public college

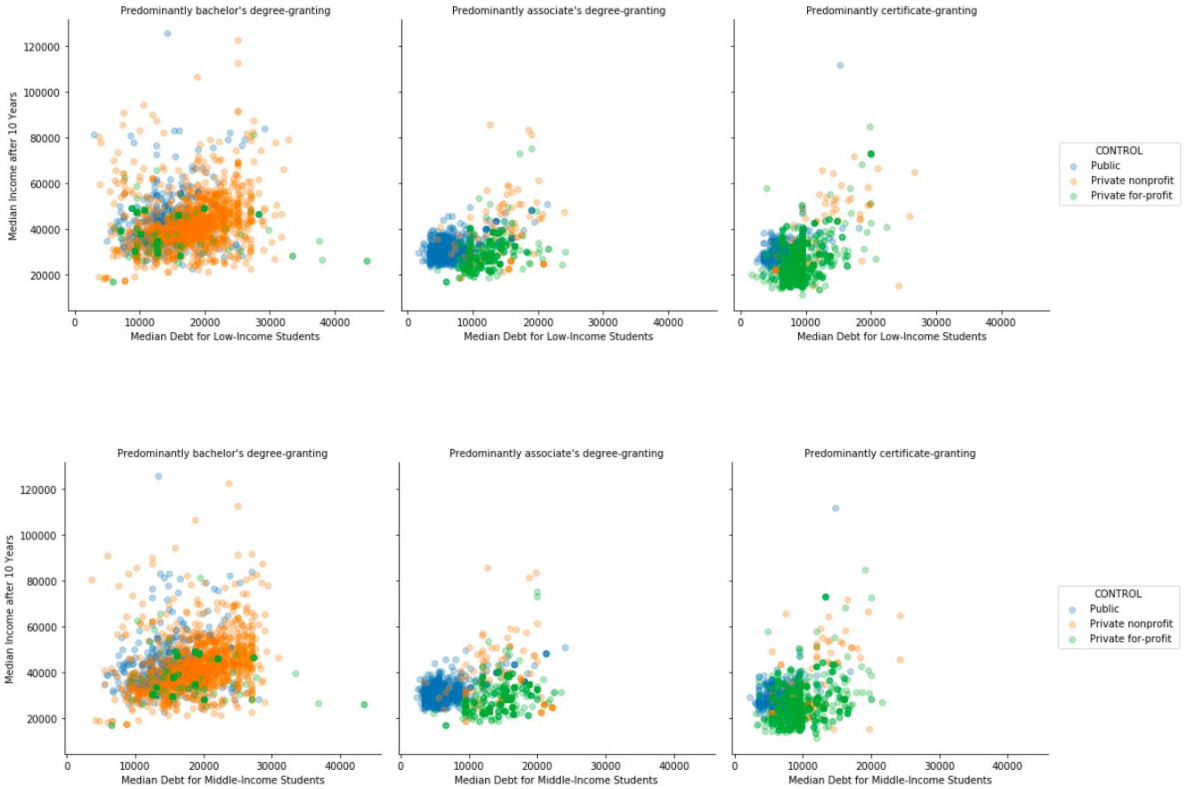
Analysis: Median Debt vs Median Income

Plots

- Top plot: Low-Income Students
- Bottom Plot: Middle-Income Students

Analysis

- Similar trends for both income groups
- Greatest variation for predominantly Bachelor's Degree-granting schools



Analysis: Debt Ratio vs Median Income

Plots

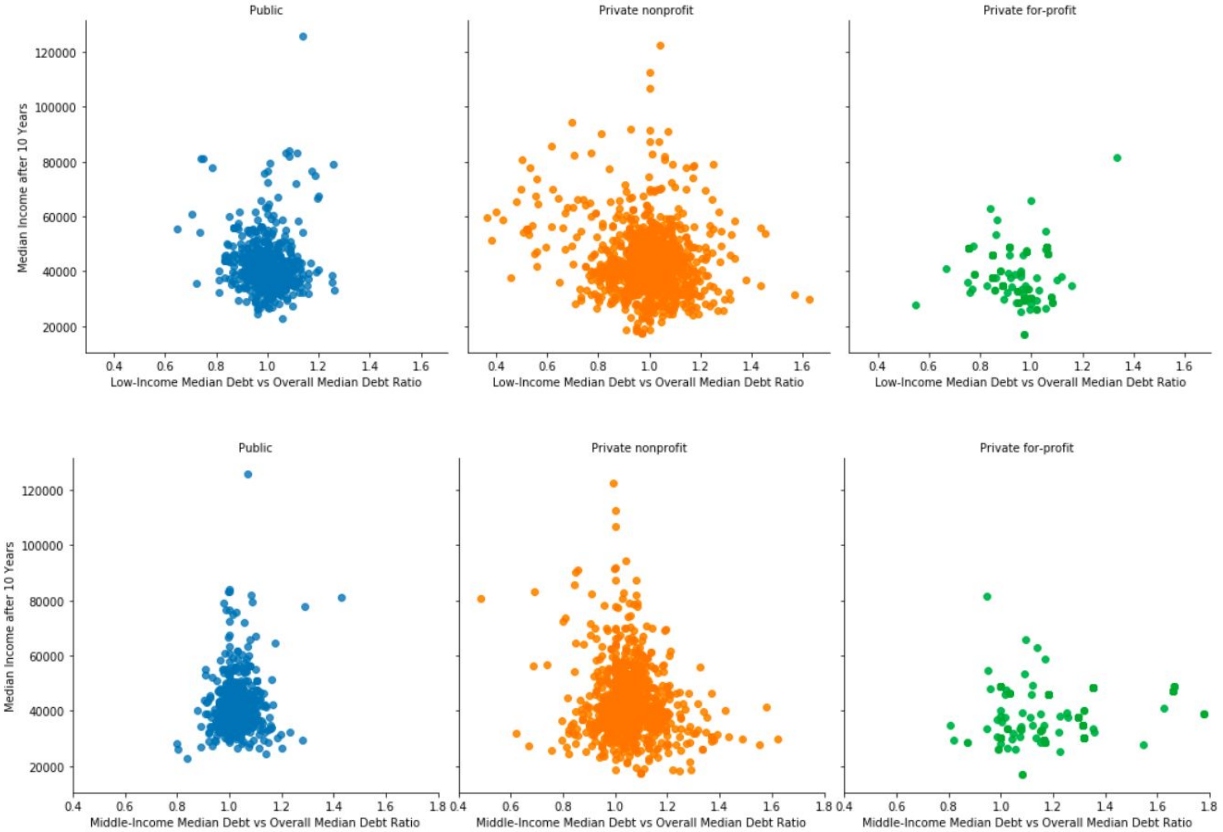
- Top plot: Low-Income Students
- Bottom Plot: Middle-Income Students

Analysis

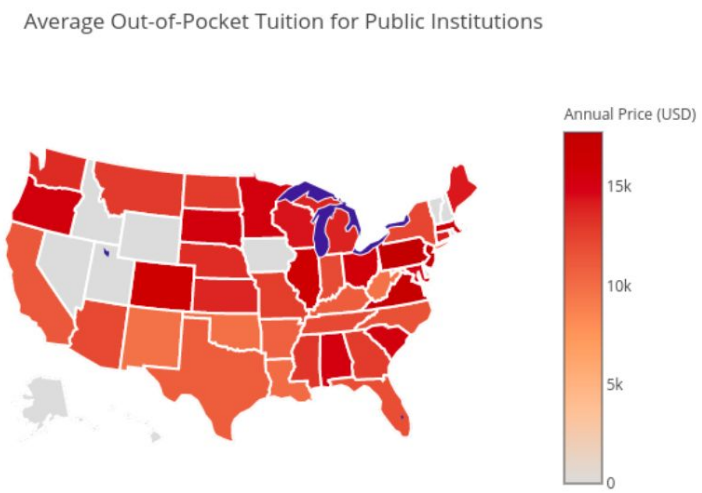
- Middle-income students pay a slightly greater share of the debt than low-income students
- Greatest variation within public nonprofit schools

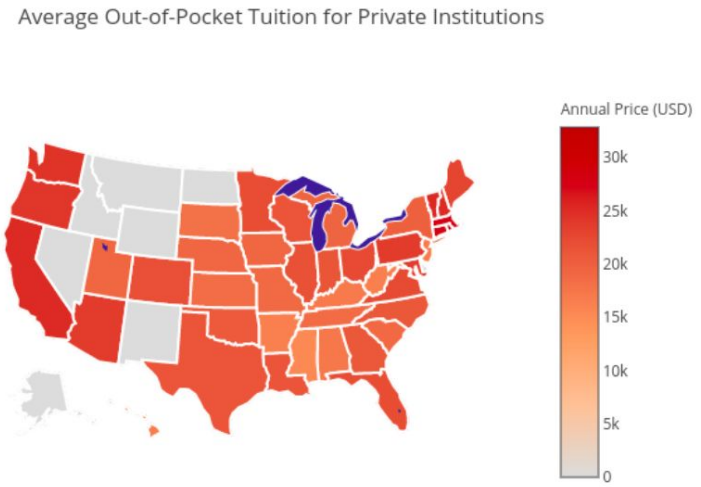
$$debt\ ratio = \frac{median\ debt\ owed\ by\ low{-}income\ students}{median\ debt\ for\ all\ students}$$

- Ratios calculated per institution



Analysis: Out-of-Pocket Cost Across States



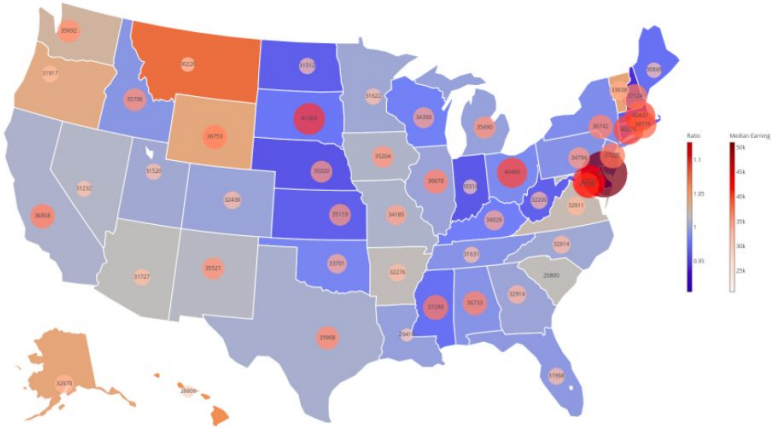


Analysis

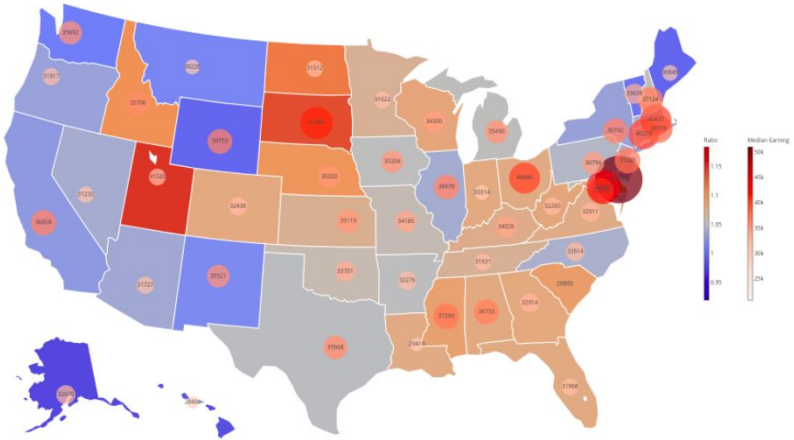
- Out-of-pocket cost generally higher for private schools than for public schools
- Schools in South consistently lower in out-of-pocket cost compared to rest of the country

Analysis: Debt Ratios Across States

Low-Income Students



Middle-Income Students



Plots

- State colors: debt ratio
- Circle size/color: median income after ten years

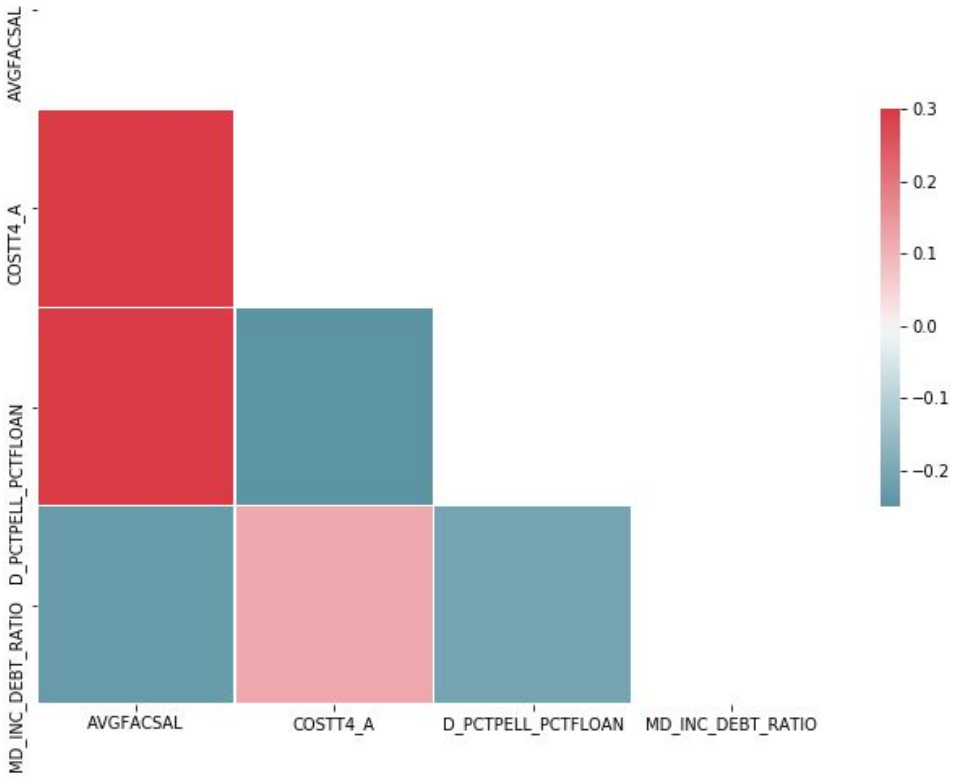
Analysis

- Middle-income students generally shouldering a debt ratio of one or higher
- East Coast a favorable location for low-income students because of low debt ratios and high median income

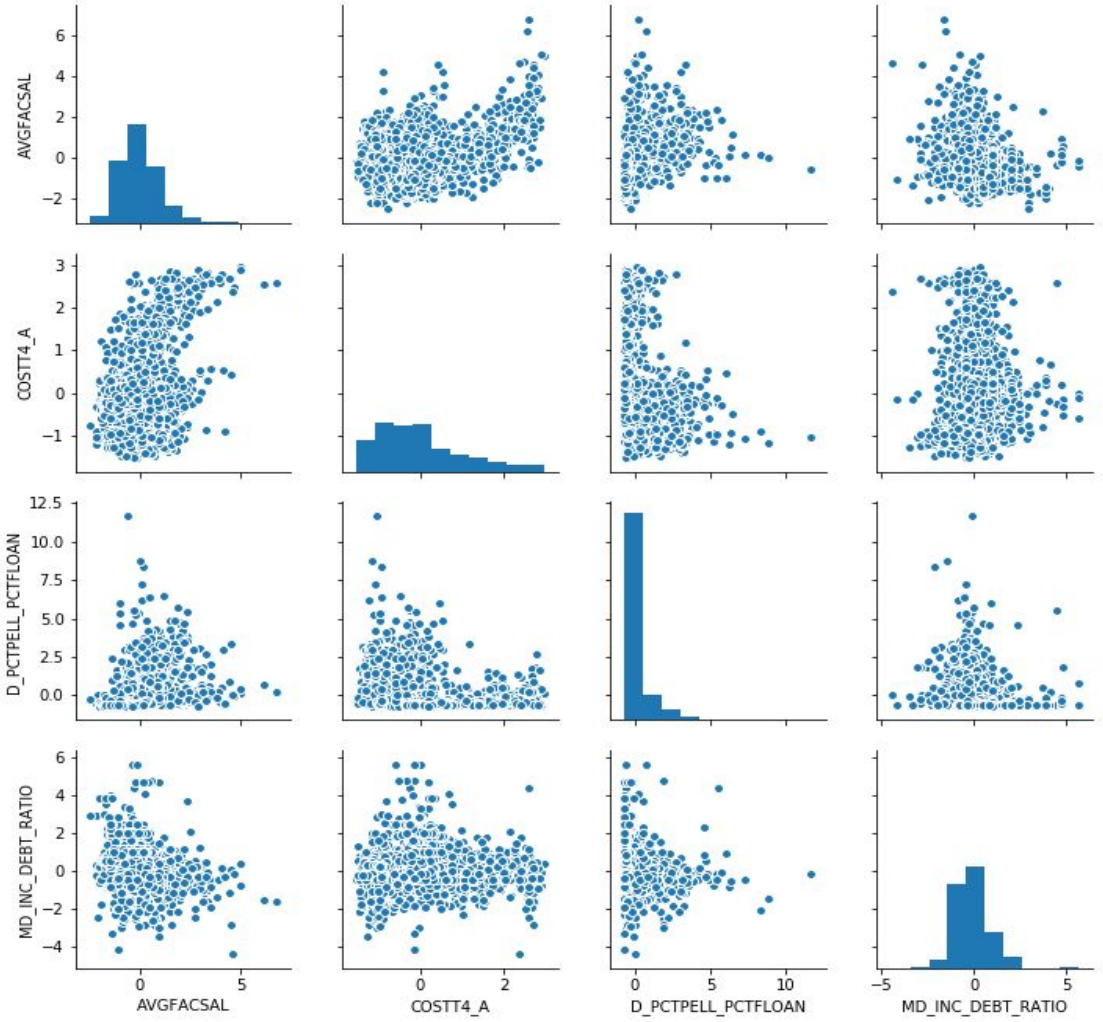
Data Modeling: Logistic Regression

- From the result we gathered from our exploratory analysis, we are interesting in predicting the school type.
- Dependent variable: Control (public or private school)
- Independent variables: AVGFACSAL (average faculty salary), COSTT4_A (average cost of attendance) , D_PCTPELL_PCTFLOAN (% of students receving a Pell Grant or federal loan), MD_INC_DEBT_RATIO (middle-income debt ratio)
- Private for-profit and private nonprofit merged into the same group.
- Model Preprocessing: Standardization

Predictor Behavior Correlation Matrix

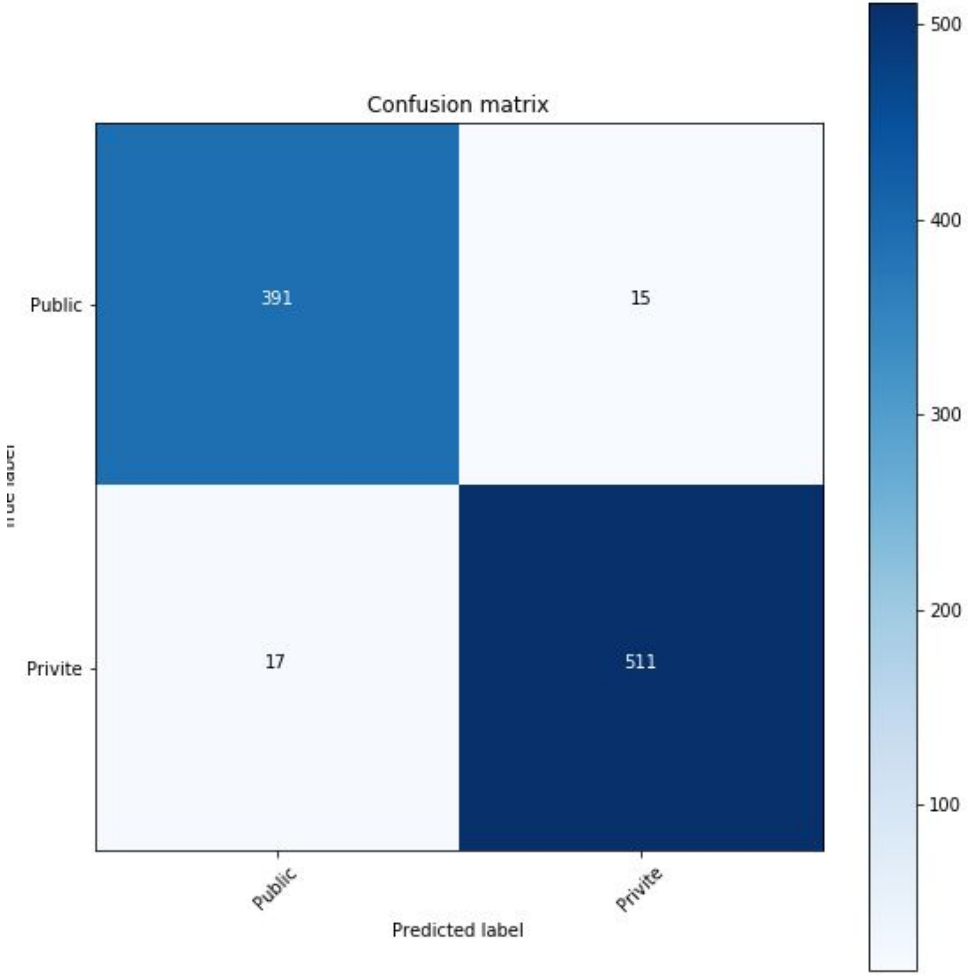


Predictor Behavior Pair Plot



Model Results

- 10 fold cross validation on logistic model
- Model has a pretty strong predicting power.
- Variation explained by the model: 83%
- Average Model accuracy: 97.43%
- Standard deviation: 0.00800



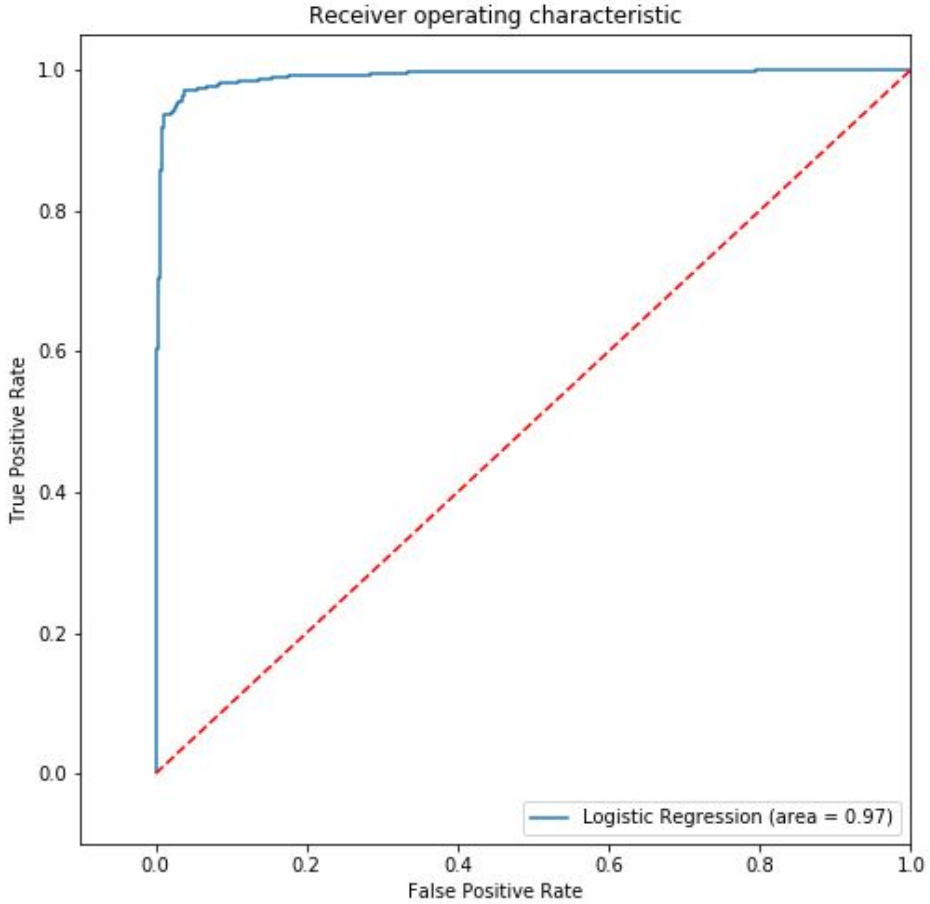
Model Accuracy Stats

- Sensitivity (True Positive Rate): 96%
- False Positive Rate: 3.2%
- Specificity (True Negative Rate): 96.7%
- Precision: 95.8%
- Number of public schools: 1354
- Number of private schools: 1759

Model Result: Odds Ratio

- A standard deviation increase in the *average faculty salary* has 0.92 times less likely to be in private school.
- A standard deviation increase in *average cost of attendance* for academic year institutions has 169 times more likely to be in private school.
- A standard deviation increase in *number of undergraduate students* (denominator is % receiving a Pell Grant or federal student loan) has 0.78 times less likely to be in private school.
- A standard deviation increase in *median income debt ratio* has 2.64 times more likely to be in private school.

ROC Graph


Receiver operating characteristic
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Logistic Regression (area = 0.97)