Stat 159 Final Project - Providing Credit to Students

Team W.O.S: Yoon Jung Rho, Young Hoon Kim, Ellen Hwang, Joseph Simonian

December 3, 2016

Abstract

Our client profile is a credit institution that provides financial aid to students. The managers are interested in expanding their customer base but they would also like that most of the loans be payed back. Our team's purpose is to perform exploratory data analysis and create predictive models to find which schools and what kind schools credit institutions should provide credit.

1 Introduction

Our client, a credit institution that provides financial aid to students, want to expand their customer base but would also like that most of the loans be payed back. Our role as the analyst is to use the publicly available [College Scoreboard Datasets](https://collegescorecard.ed.gov/data/) to figure out what features of a college make it more reliable for credit. Our team will be using the 3 year repayment rate as am indicator of the school's overall reliability rate. Using exploratory data analysis we will examine the relationships between repayment rate and other features of the school. In our analysis, we will use ridge, lasso, partial least squares, and principle component regression to find significant features that influence a school's overall repayment rate.

2 Data

The data from College Scorecard provide insights into the performance of schools eligible to receive federal financial aid, and offer a look at the outcomes of students at those schools. The Data that appear on the College Scorecard provides data on student completion, debt and repayment, earnings, and more. The files include data from 1996 through 2016 for all undergraduate degree-granting institutions of higher education. This data was last updated on September 13th, 2016. The data is available at: https://collegescorecard.ed.gov/data/

For our project, besides the main data, our team also used featured downloads provided by College Scorecard. These data downloads provide quick access to some of the data in which users may be most interested, including a file that offers the most current data for each element. Among variety of data, we used [Post-School earnings](https://ed-public-download.apps.cloud.gov/downloads/Most-Recent-Cohorts-Treasury-Elements.csv) to narrow down the analytical component.

There is also a documentation that provides more on how to use the data, including: Sources of the data, The construction of metrics, and Data considerations and limitations available at: https://collegescorecard.ed.gov/data/documentation/

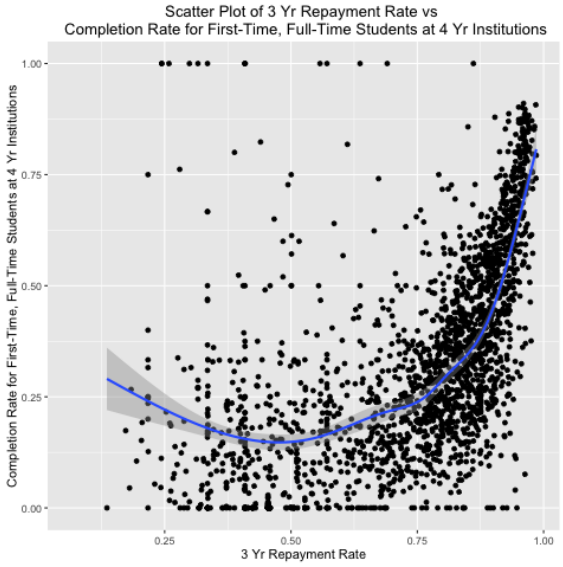
3 Methods

4 Results

WRITE SOMETHING ABOUT USING REGRESSIONS TO SELECT VARIABLES

After using correlation and regression to select features related to repayment, we were able to perform some exploratory data analysis.

Figure 1: Figure 1: Scatter Plot of Completion Rate vs. 3 Year Repayment Rate



5 Analysis

6 Conclusions