**Team Name**: Poor College Students

**Team Members**: Bill Su (rs2bd), Alex Struck (avs3ev), Ellie McKinnon (elm8dt), and Grayson Thornton (glt4ae)

**Project Title:** Demystifying the Ivory Tower:a research project on college ROI with data product implementation

**Project Data Set**: Kaggle US Dept of Education: College Scorecard

**Project Description**

Our project aims to help college applicants identify four-year undergraduate universities whose graduates generate the highest return on investment. We will make personalized college recommendations to users based on various attributes of the universities. Our final product will consist of a data tool where individuals can enter personal information and receive a list of schools that will generate the highest return. For instance, users will provide information such as family income level, region preference, and family legacy, and the data tool will return schools ranked by ROI.

Our analysis will help users make more economically informed decisions about where to attend college. In doing so, we aim to help students tackle an extremely important personal decision, while simultaneously shedding light on the pressing societal issue surrounding college tuition and student debt. By providing students with the ROIs of different institutions, we hope to empower our users to lessen the economic pressure on both students and their parents and maximize the financial outcomes of their college decision.

**Plan of Analysis**

For this project, we plan to first conduct a descriptive analysis regarding the current state of college education and general return on investment for students of various ethnic, gender, and geographic backgrounds. With this information, we will attempt to answer the question: “Is going to college a bad investment for students? If so, how economically inefficient is the college education?”

Next, we will to compile all of the variables that impact ROI for education and come up with a comprehensive model. This model will help a student choose the college with the best return on his investment in the future given his economic background and desired college attributes. This model is predictive because it can predict a college’s return on investment based on a student’s profile. It is also prescriptive, because insights generated by our data product can help navigate students to a brighter future economically.

**Data and Variables Overview**

For this project, we are going to use the CollegeScorecard data provided by the U.S. Department of Education. Overall, the dataset records all attributes of college institutions including, but not limited to, its demography, geographic location, and future earnings for students 6 and 10 years after graduation. There are 124,699 entries in the dataset, with 1,731 variables. The dataset records data collected from 1996 to 2013.

We have selected 203 variables out of the original dataset to explore so that our analysis is based only on relevant variables. Data dictionary for those 203 variables are attached in the appendix. The final variables are in the table of this report, which is the result of our condensation of those 203 variables.

Our target variable is return on investment, a calculated variable that is the result of dividing the median 6-year income of college students by the cost of college education during their years of attendance. This project is going to utilize primarily linear regression and decision tree classifiers for the primary data product. However, classification and clustering are used in the process of reducing variables to a manageable size.

The scope of our project and dataset used is very ambitious. We are using a raw dataset that is completely un-cleaned. Therefore, we foresee multiple data quality issues including missing values and inaccurate variable representation in dataset. This could be solved by a vigorous cleaning process and creation of new variables that can more accurately reflect the needs of our project.

|  |  |
| --- | --- |
| **Variable** | **Data Type** |
| Return on investment (Target) | Real |
| Gender | Nominal |
| Family Income | Ordinal |
| Standardized Test Score | Real |
| Desired Area of Study | Nominal |
| Admission Rate | Ordinal (High, Medium, Low) |
| Completion Rate | Ordinal (High, Medium, Low) |
| Debt Burden After Graduation | Ordinal (High, Medium, Low) |
| Region of College | Nominal |

**Appendix: Data Dictionary**

|  |  |
| --- | --- |
| NAME OF DATA ELEMENT | Data type |
| Return On Investment (Target Variable) | float |
| Institution name | string |
| City | string |
| State postcode | string |
| Predominant degree awarded  0 Not classified  1 Predominantly certificate-degree granting  2 Predominantly associate's-degree granting  3 Predominantly bachelor's-degree granting  4 Entirely graduate-degree granting | integer |
| Region (IPEDS) | integer |
| Locale of institution | integer |
| Degree of urbanization of institution | integer |
| Carnegie Classification -- undergraduate profile | integer |
| Carnegie Classification -- size and setting | integer |
| Flag for Historically Black College and University | integer |
|  |  |
| Flag for predominantly black institution | integer |
|  |  |
| Flag for Alaska Native Native Hawaiian serving institution | integer |
|  |  |
| Flag for tribal college and university | integer |
|  |  |
| Flag for Asian American Native American Pacific Islander-serving institution | integer |
|  |  |
| Flag for Hispanic-serving institution | integer |
|  |  |
| Flag for Native American non-tribal institution | integer |
|  |  |
| Flag for men-only college | integer |
|  |  |
| Flag for women-only college | integer |
|  |  |
| Religous affiliation of the institution | integer |
|  |  |
| Admission rate | float |
| Midpoint of SAT scores at the institution (critical reading) | float |
| Midpoint of SAT scores at the institution (math) | float |
| Midpoint of SAT scores at the institution (writing) | float |
| Midpoint of the ACT cumulative score | float |
| Midpoint of the ACT English score | float |
| Midpoint of the ACT math score | float |
| Midpoint of the ACT writing score | float |
| Average SAT equivalent score of students admitted | float |
| Bachelor's degree in Agriculture, Agriculture Operations, And Related Sciences. | integer |
| Bachelor's degree in Natural Resources And Conservation. | integer |
| Bachelor's degree in Architecture And Related Services. | integer |
| Bachelor's degree in Area, Ethnic, Cultural, Gender, And Group Studies. | integer |
| Bachelor's degree in Communication, Journalism, And Related Programs. | integer |
| Bachelor's degree in Communications Technologies/Technicians And Support Services. | integer |
| Bachelor's degree in Computer And Information Sciences And Support Services. | integer |
| Bachelor's degree in Personal And Culinary Services. | integer |
| Bachelor's degree in Education. | integer |
| Bachelor's degree in Engineering. | integer |
| Bachelor's degree in Engineering Technologies And Engineering-Related Fields. | integer |
| Bachelor's degree in Foreign Languages, Literatures, And Linguistics. | integer |
| Bachelor's degree in Family And Consumer Sciences/Human Sciences. | integer |
| Bachelor's degree in Legal Professions And Studies. | integer |
| Bachelor's degree in English Language And Literature/Letters. | integer |
| Bachelor's degree in Liberal Arts And Sciences, General Studies And Humanities. | integer |
| Bachelor's degree in Library Science. | integer |
| Bachelor's degree in Biological And Biomedical Sciences. | integer |
| Bachelor's degree in Mathematics And Statistics. | integer |
| Bachelor's degree in Military Technologies And Applied Sciences. | integer |
| Bachelor's degree in Multi/Interdisciplinary Studies. | integer |
| Bachelor's degree in Parks, Recreation, Leisure, And Fitness Studies. | integer |
| Bachelor's degree in Philosophy And Religious Studies. | integer |
| Bachelor's degree in Theology And Religious Vocations. | integer |
| Bachelor's degree in Physical Sciences. | integer |
| Bachelor's degree in Science Technologies/Technicians. | integer |
| Bachelor's degree in Psychology. | integer |
| Bachelor's degree in Homeland Security, Law Enforcement, Firefighting And Related Protective Services. | integer |
| Bachelor's degree in Public Administration And Social Service Professions. | integer |
| Bachelor's degree in Social Sciences. | integer |
| Bachelor's degree in Construction Trades. | integer |
| Bachelor's degree in Mechanic And Repair Technologies/Technicians. | integer |
| Bachelor's degree in Precision Production. | integer |
| Bachelor's degree in Transportation And Materials Moving. | integer |
| Bachelor's degree in Visual And Performing Arts. | integer |
| Bachelor's degree in Health Professions And Related Programs. | integer |
| Bachelor's degree in Business, Management, Marketing, And Related Support Services. | integer |
| Bachelor's degree in History. | integer |
| Enrollment of undergraduate degree-seeking students | integer |
| Flag for currently operating institution, 0=closed, 1=operating | integer |
| Average net price for $0-$30,000 family income (public institutions) | integer |
| Average net price for $30,001-$48,000 family income (public institutions) | integer |
| Average net price for $48,001-$75,000 family income (public institutions) | integer |
| Average net price for $75,001-$110,000 family income (public institutions) | integer |
| Average net price for $110,000+ family income (public institutions) | integer |
| Average net price for $0-$30,000 family income (private for-profit and nonprofit institutions) | integer |
| Average net price for $30,001-$48,000 family income (private for-profit and nonprofit institutions) | integer |
| Average net price for $48,001-$75,000 family income (private for-profit and nonprofit institutions) | integer |
| Average net price for $75,001-$110,000 family income (private for-profit and nonprofit institutions) | integer |
| Average net price for $110,000+ family income (private for-profit and nonprofit institutions) | integer |
| Average net price for $0-$30,000 family income (program-year institutions) | integer |
| Average net price for $30,001-$48,000 family income (program-year institutions) | integer |
| Average net price for $48,001-$75,000 family income (program-year institutions) | integer |
| Average net price for $75,001-$110,000 family income (program-year institutions) | integer |
| Average net price for $110,000+ family income (program-year institutions) | integer |
| Average net price for $0-$30,000 family income (other academic calendar institutions) | integer |
| Average net price for $30,001-$48,000 family income (other academic calendar institutions) | integer |
| Average net price for $48,001-$75,000 family income (other academic calendar institutions) | integer |
| Average net price for $75,001-$110,000 family income (other academic calendar institutions) | integer |
| Average net price for $110,000+ family income (other academic calendar institutions) | integer |
| Number of Title IV students (public institutions) | integer |
| Number of Title IV students (private for-profit and nonprofit institutions) | integer |
| Number of Title IV students (program-year institutions) | integer |
| Number of Title IV students (other academic calendar institutions) | integer |
| Number of Title IV students, $0-$30,000 family income (public institutions) | integer |
| Number of Title IV students, $30,001-$48,000 family income (public institutions) | integer |
| Number of Title IV students, $48,001-$75,000 family income (public institutions) | integer |
| Number of Title IV students, $75,001-$110,000 family income (public institutions) | integer |
| Number of Title IV students, $110,000+ family income (public institutions) | integer |
| Number of Title IV students, $0-$30,000 family income (private for-profit and nonprofit institutions) | integer |
| Number of Title IV students, $30,001-$48,000 family income (private for-profit and nonprofit institutions) | integer |
| Number of Title IV students, $48,001-$75,000 family income (private for-profit and nonprofit institutions) | integer |
| Number of Title IV students, $75,001-$110,000 family income (private for-profit and nonprofit institutions) | integer |
| Number of Title IV students, $110,000+ family income (private for-profit and nonprofit institutions) | integer |
| Number of Title IV students, $0-$30,000 family income (program-year institutions) | integer |
| Number of Title IV students, $30,001-$48,000 family income (program-year institutions) | integer |
| Number of Title IV students, $48,001-$75,000 family income (program-year institutions) | integer |
| Number of Title IV students, $75,001-$110,000 family income (program-year institutions) | integer |
| Number of Title IV students, $110,000+ family income (program-year institutions) | integer |
| Number of Title IV students, $0-$30,000 family income (other academic calendar institutions) | integer |
| Number of Title IV students, $30,001-$48,000 family income (other academic calendar institutions) | integer |
| Number of Title IV students, $48,001-$75,000 family income (other academic calendar institutions) | integer |
| Number of Title IV students, $75,001-$110,000 family income (other academic calendar institutions) | integer |
| Number of Title IV students, $110,000+ family income (other academic calendar institutions) | integer |
| Average cost of attendance (academic year institutions) | integer |
| Average cost of attendance (program-year institutions) | integer |
| In-state tuition and fees | integer |
| Out-of-state tuition and fees | integer |
| Tuition and fees for program-year institutions | integer |
| Net tuition revenue per full-time equivalent student | integer |
| Proportion of faculty that is full-time | float |
| Percent of low-income (less than $30,000 in nominal family income) students who completed within 8 years at original institution | float |
| Percent of middle-income (between $30,000 and $75,000 in nominal family income) students who completed within 8 years at original institution | float |
| Percent of high-income (above $75,000 in nominal family income) students who transferred to a 4-year institution and completed within 8 years | float |
| Percent of dependent students who completed within 8 years at original institution | float |
| Percent of independent students who completed within 8 years at original institution | float |
| Percent of female students who completed within 8 years at original institution | float |
| Percent of male students who completed within 8 years at original institution | float |
| Percent of students who received a federal loan at the institution and who completed in 8 years at original institution | float |
| Percent of students who never received a federal loan at the institution and who completed in 8 years at original institution | float |
| Percent of first-generation students who completed within 8 years at original institution | float |
| Percent of not-first-generation students who completed within 8 years at original institution | float |
| One-year repayment rate for completers | float |
| One-year repayment rate by family income ($0-30,000) | float |
| One-year repayment rate by family income ($30,000-75,000) | float |
| One-year repayment rate by family income ($75,000+) | float |
| One-year repayment rate for dependent students | float |
| One-year repayment rate for independent students | float |
| One-year repayment rate for females | float |
| One-year repayment rate for males | float |
| One-year repayment rate for first-generation students | float |
| One-year repayment rate for students who are not first-generation | float |
| Three-year repayment rate for completers | float |
| Three-year repayment rate for non-completers | float |
| Three-year repayment rate by family income ($0-30,000) | float |
| Three-year repayment rate by family income ($30,000-75,000) | float |
| Three-year repayment rate by family income ($75,000+) | float |
| Three-year repayment rate for dependent students | float |
| Three-year repayment rate for independent students | float |
| Three-year repayment rate for females | float |
| Three-year repayment rate for males | float |
| Three-year repayment rate for first-generation students | float |
| Three-year repayment rate for students who are not first-generation | float |
| Five-year repayment rate for completers | float |
| Five-year repayment rate for non-completers | float |
| Five-year repayment rate by family income ($0-30,000) | float |
| Five-year repayment rate by family income ($30,000-75,000) | float |
| Five-year repayment rate by family income ($75,000+) | float |
| Five-year repayment rate for dependent students | float |
| Five-year repayment rate for independent students | float |
| Five-year repayment rate for females | float |
| Five-year repayment rate for males | float |
| Five-year repayment rate for first-generation students | float |
| Five-year repayment rate for students who are not first-generation | float |
| Seven-year repayment rate for completers | float |
| Seven-year repayment rate for non-completers | float |
| Seven-year repayment rate by family income ($0-30,000) | float |
| Seven-year repayment rate by family income ($30,000-75,000) | float |
| Seven-year repayment rate by family income ($75,000+) | float |
| Seven-year repayment rate for dependent students | float |
| Seven-year repayment rate for independent students | float |
| Seven-year repayment rate for females | float |
| Seven-year repayment rate for males | float |
| Seven-year repayment rate for first-generation students | float |
| Seven-year repayment rate for students who are not first-generation | float |
| Percentage of aided students whose family income is between $0-$30,000 | float |
| Percentage of students who are financially independent | float |
| Percentage of students who are financially independent and have family incomes between $0-30,000 | float |
| Percentage of students who are financially dependent and have family incomes between $0-30,000 | float |
| Percentage first-generation students | float |
| Aided students with family incomes between $30,001-$48,000 in nominal dollars | float |
| Aided students with family incomes between $48,001-$75,000 in nominal dollars | float |
| Aided students with family incomes between $75,001-$110,000 in nominal dollars | float |
| Aided students with family incomes between $110,001+ in nominal dollars | float |
| Dependent students with family incomes between $30,001-$48,000 in nominal dollars | float |
| Dependent students with family incomes between $48,001-$75,000 in nominal dollars | float |
| Dependent students with family incomes between $75,001-$110,000 in nominal dollars | float |
| Dependent students with family incomes between $110,001+ in nominal dollars | float |
| Independent students with family incomes between $30,001-$48,000 in nominal dollars | float |
| Independent students with family incomes between $48,001-$75,000 in nominal dollars | float |
| Independent students with family incomes between $75,001-$110,000 in nominal dollars | float |
| Independent students with family incomes between $110,001+ in nominal dollars | float |
| Percent of students whose parents' highest educational level is middle school | float |
| Percent of students whose parents' highest educational level is high school | float |
| Percent of students whose parents' highest educational level was is some form of postsecondary education | float |
| The original amount of the loan principal upon entering repayment | float |
| The median debt for students who have completed | float |
| The median debt for students with family income between $0-$30,000 | float |
| The median debt for students with family income between $30,001-$75,000 | float |
| The median debt for students with family income $75,001+ | float |
| The median debt for dependent students | float |
| The median debt for independent students | float |
| The median debt for female students | float |
| The median debt for male students | float |
| The median debt for first-generation students | float |
| The median debt for not-first-generation students | float |
| Cumulative loan debt at the 90th percentile | integer |
| Cumulative loan debt at the 75th percentile | integer |
| Cumulative loan debt at the 25th percentile | integer |
| Cumulative loan debt at the 10th percentile | integer |
| Average family income | integer |
| Median family income | integer |
| Average family income for independent students | integer |
| Median household income | integer |
| Poverty rate, via Census data | float |