DATA SCIENCE PROJECT

MILESTONE 1

The Proposal

Sherry Kosmicki

March 19, 2023

DSC680 Applied Data Science
Professor Catherine Williams

TOPIC

Academic institutions in the United States are required to submit statistics to the department of education to identify student participation, degree completers and whether they received financial aid. Based on these findings, analysis can be done to determine the likelihood of graduating based on your choice of school and financial aid need.

BUSINESS PROBLEM

❖ Is there a relationship between higher graduation rates at a 4-year college and students that receive financial aid verses those that do not receive financial aid?

❖ Are there academic institutions that have higher graduation rates based on there classification of public vs. private non-profit vs. private profit?

❖ In relation to either the school classification or financial aid has there been an increase or decrease in the graduation rates?

DATASETS

- National Center for Education Statistics

- This dataset includes the number of full-time, first-time degree/certificate-seeking undergraduate students in the adjusted cohort, number of completers, and graduation rate for Title IV institutions, level of institution and type of aid received, United States

- This data has been published for related years 2015-2021. I plan to gather all the years available to populate the dataset.

METHODS

Data Project Architecture

Question – Topic Identified

Data – Wrangling and manipulating the data in the analysis format

EDA – Explore visualizations for data understanding

Modeling – Data division, Linear Regression, Logistic Regression, Decision Tree classifier

Interpretation – model accuracy, rework as needed

Communication – results visualizations and storytelling

- Determining a students graduation potential from a specific institution based on their financial choices could be misleading. This could just be one of many factors to determine if a student graduates.

- Title IV institutions could use this information as a marketing tool to coerce populations to attend their school based on graduation rates.

- Government funding could be altered based on the outcomes of these findings.

ETHICAL CONSIDERATIONS

CHALLENGES/ISSUES

- The data is currently in a left to right nested pivot table format. Steps need to be taken to wrangle the data back to a tabular format in order to begin the EDA process.

- There are multiple ways the data could be modeled. Finding the best method could be challenging.

- Converting the data to identify students as receiving financial aid verses not receiving aid could allow for logistic regression modeling.

- Understanding the impact multiple years of survey data has on the model.

- I won't be considering any rise in tuition rates at Title IV institutions during each dataset time period.

REFERENCES

- Source: U.S. Department of Education, National Center for Education Statistics

https://nces.ed.gov/ipeds/Search?query=&query2=&resultType=all&page=1&sortBy=date_desc&overlayTableId=28425

- Calculating financial need

https://studentaid.gov/complete-aid-process/how-calculated

- Rising tuition, financial aid can maintain graduation rates: experts. George Washington University article, Ianne Salvosa, May 26, 2022.

https://www.gwhatchet.com/2022/05/16/rising-tuition-financial-aid-can-maintain-graduation-rates-experts/

- Grants vs. Loans – Study finds that government grants have a positive impact on the graduation rates of low-income students, but unsubsidized loans counter that impact. Scott Jaschik, April 4, 2014

https://www.insidehighered.com/news/2014/04/04/study-shows-positive-imapcts-government-aid-graduation-rates