

BUDT 758X Project Proposal:

Exploring data analytics in the Student debt crisis in the US.

Group 29 (Tachy Terps 

Introduction:

The US Student loan debts have reached a whopping \$1.6 Trillion and continue growing exponentially as more and more students apply for Federal aids and grants. The student loans have crossed the total amounts of auto loans and credit card loans! The student debt crisis has caught the attention of a lot of top bureaucrats including President Trump, Elizabeth Warren, and Bernie Sanders. This resulted in Elizabeth Warren introducing a bill titled the "Student Loan Debt Relief Act of 2019." The bill stated a plan that would cancel student loan debt for more than 95% of borrowers, and would entirely cancel student loan debt for more than 75% of Americans with student loan debt.

The population of interest is undergraduate and graduate students of US educational institutions who avail education loans. As the sampling is random and has enough observations the results of this study could be generalized for the population of interest.

Hasan Minaj (a popular standup comedian) recently spoke about the crisis on his show The Patriot Act. You should definitely watch it to grasp the depth of the problem: [Click Here!](#)

Questions of Interest:

- What is the student debt amount per year in the United States?
- What is the average student debt per educational institution/city/state?
- Is there a relationship between the SAT score of an undergraduate and the loan repayment rate.
- Can we identify an evident relation between the loan repayment rate and the race(Caucasians, Hispanics, African-Americans, and Asians) of a person?
- Does the institution a student graduate from effect the loan repayment rate?
- What are the changes observed on the loan repayment rate after earning for 10 years?
- Does a pay gap exist amongst genders graduating from the same institution?
- Do students at private for-profit institutions have a lower rate of federal loan repayment compared to students at non-profit institutions?

Data Processing and Analysis

Dataset Description:

The college Scorecard site provides the data behind the College Scorecard, as well as other data on federal financial aid and earnings information. These data provide insights into the performance of schools eligible to receive federal financial aid. There are 7804 cases available on the website, but as the data has missing and invalid values the data has to be cleansed and offer a look at the outcomes of students at those schools. The files include data from 1996 through 2017 for all degree-granting institutions of higher education but we will be using data from 2007 to 2017.

The important columns for the analysis are: INSTNM, CITY, STABBR, ZIP, LATITUDE, LONGITUDE, MENONLY, WOMENONLY, ADM_RATE, SAT_AVG, UGDS_WHITE, UGDS_BLACK, UGDS_HISP, UGDS_ASIAN, TUITIONFEE_IN, TUITIONFEE_OUT, TUITIONFEE_PROG, TUITFTE, INEXPFTE

Source: <https://collegescorecard.ed.gov/data/>

Data Processing Tasks

- Merging datasets
- Data Cleaning
- Managing and changing missing data with a dummy and average values respectively
- Sorting, filtering data records
- Data visualization
- Data interpretation
- Regression analysis

Data Analysis:

What kinds of visualizations and modeling do you need to do to answer your questions of interest? Maps? Charts? Will your visuals be static or interactive?

We will be creating:

1. A static heat map of the United States to show the average student debt per state
2. A time series to show the rise in student debts over the years
3. Finding correlations between factors
4. Histograms for comparisons
5. A linear model to understand regression

Expected Findings

We are expecting to understand why the student debt crisis has become a hot topic right now, especially with the upcoming elections. We are expected to draw information from the correlation between different factors like repayment rates, loan amounts based on state or city or institution.

We expect to infer if any racial disparities, gender bias, or any other factors exist when deciding a loan repayment loan.

Project Timeline

Task	Task Lead	Due Date
Data Cleaning, Merging and Extraction	Prachi	10/31
Handling NULL values and missing values	Chinmay	11/05
Data Formatting	Madhura	11/12
Data Visualization	Prachi	12/01
Data Interpretation	Madhura	12/03