Cedarville Graduate Student Loan Statistics

Provided by Jonathan Bitner, PharmD for Steve Winey, Tim Entner, and Christie Bitner

I. Purpose

Background

The Financial Aid team (Steve Winey, Tim Entner, Christie Bitner) is concerned that the 2024-25 FAFSA delay could prevent graduate summer header students from accessing federal aid. Since some students rely on federal loans to pay for living expenses, a delay in aid could lead to a financial crisis.

To prevent this, the Financial Aid team would like to allow certain students to increase their loans for this current semester to cover summer living expenses. Students for consideration are those who have “accepted” federal aid for this current semester and have room to borrow additional funds.

Data collection

Tim and Steve used PUTTY to gather initial data. The first report listed all de-identified graduate students and their active program, and the second report listed de-identified graduate students who completed a FAFSA and “accepted” federal loans. These were saved as two spreadsheets.

The problem

The team asked Jonathan Bitner to help:

1. Remove unnecessary letters at the end of each graduate program abbreviation (for example: change `MSN.NSG` to `MSN`)
2. Merge the two spreadsheets together as Table 1
3. Split up the Unsub and PLUS loans from one column into two columns (for example: Student6008 accepted $33,000 in Unsub loans and \$0 in PLUS loans) as Table 2

In detail, these tables would look like:

- Table 1:
  - Program abbreviation
  - Number of students in that program
  - Number of students that completed the FAFSA
  - Number/percent of students that accepted an Unsubsidized loan
  - Number/percent of students that accepted a PLUS loan
- Table 2:

- Student information (Name, ID, program)
- Total amount of dollars in loans accepted for the year (Unsubsidized versus PLUS)

II. Methodology

Understanding the data

- Discussed with Christie and Tim so I understood their goals and terminology.
- Saved files as `.csv` and loaded into a Python environment (Jupyter Notebook).
- Reviewed files for a general understanding of what was in the files.
  - Summary statistics and data types
  - Discovered and reviewed missing/incomplete data
    - Found 11 loans had no loan fee
    - Replaced missing values with zero
  - Discovered (and removed) undergrad loans

Cleaned the data

- Split `Term` (ex. `2023SU`) into `Year` and `Semester` (ex. (`2023` and `Summer`)
  - Also have the option to save `Year` as `2023-2024`
- Updated variable and value names for clarity
  - Example 1: Changed `ID` to `Student ID` so that anybody picking up this document knows exactly what I mean, also to help differentiate from the `Loan ID` column.
  - Example 2: Changed `DGU` and `DGP` to `Undergrad` and `PLUS`; only one difference between `DGU` and `DGP` is hard to read
- Created categories (ex. `Summer` comes before `Fall`, which comes before `Spring`)
- Removed undesired portion of `Program Abbreviation` (ex. `DNP.DNPPG` to `DNP`)

Merged the data

- If a student did not complete the FAFSA, that row stayed the same
- If a student did complete the FAFSA, they had multiple rows representing each unique loan
- Added new column for `FAFSA Status` that was either `Incomplete` for missing values or `Completed` for `Accepted` or `Declined`
- Added a new column to label all `P1` through `P4` students as `PharmD`
- Reviewed missing values
  - I had Not Missing At Random (NMAR) values, since there was a logical explanation for why they were missing (i.e. If a student did not complete the FAFSA, their column for `Status`, `Loan Type`, etc. was blank)
  - Filled missing categorical values with `None` so those numbers would be included in statistics and charts
    - Except time-related values, such as `Year` and `Semester`, which were left blank and ignored in statistics and charts
  - I had three options for missing quantitative data (i.e. `Term Award`)

    - Leave as is
        - Zero for students that completed the FAFSA and blank otherwise
        - Useful to compare students that did/did not complete the FAFSA
        - This seemed to align best with Christie and Tim's goals
    - Fill all with zero (useful for comparing averages for all students)
    - Fill all as blank (useful for comparing averages for only students that borrowed)

Pivoted the data

- Christie and Tim wanted one row for each student, and a column for PLUS and Unsub loans.
- Decided how to account for duplicate rows for each desired statistic

Visualized the data (see Results)

III. Results

- Created charts (see `.png` files labeled `Page 1`, `Page 2`, `Page 3`
    - Organized from top to bottom, left to right
    - I will refer to as Chart 1.1 for the top left chart on Page 1 and Chart 2.6 for the bottom right chart on Page 2
    - I am able to save charts to any file type (ex `.pdf` or `.png`)
    - I can create charts of any shape
        - Square vs tall or wide, small vs page width vs extra-large, etc.
    - I don't necessarily work with one dataset; I am constantly creating/editing/adding/deleting until I get the customized results I want (see Chart 2.1 and 2.2)

Page 1

Charts 1.1 and 1.2: Number of students in each program

- Relevance to Christie and Tim:
    - This was information they requested, but in a visual format
    - Easy way to see that the number of `PharmD` students is steadily declining
- How I made this/what else I can do:
    - I am able to split up `PharmD` students as desired
    - I am able to further filter results by another variable, such as `Semester` or `FAFSA Status`
    - I can include a table of exact numbers for each chart

Charts 1.3 and 1.4: Percent of students that completed the FAFSA

- Relevance to Christie and Tim:
    - This was information they requested, but in a visual format

    - Easy way to see that more PharmD students completed the FAFSA than the rest of the student body
- How I made this/what else I can do:
    - Pie charts should be used sparingly: No more than four slices, easy to compare slice sizes, must be a proportion, the pie should be a perfect circle and not 3D or distorted, etc.
    - Number of students in each program would not make a good pie chart, instead grouping by program type (i.e. Health, Bible, Misc) would make a good pie chart
    - I also have a table that splits up these results by program

Charts 1.5 and 1.6: Total/Average amounts borrowed

- Relevance to Christie and Tim:
    - This was a bonus, done out of my own curiosity
    - Christie and Tim were surprised to see `P2` and `P3` students borrowed more (on average) than `P4` and `MMSPA` students
    - Could be helpful in comparing student debt vs potential income (i.e. Bible students)
- How I made this/what else I can do:
    - Looking at solely total amounts can present bias in statistics.
    - For example, you would expect that the programs with the most students would have higher totals.
    - I accounted for this by calculating the average borrowed by each student
        - Initially I included all students that completed the FAFSA
        - Tim expected the average for `P4`s to be higher, and requested I recalculate with students that completed the FAFSA but had zero for their `Term Award`
        - Proportions were the same; averages across the board were higher
    - This chart has different colors because there are a number of different ways to create a chart, and this method was easier for the data I wanted to use.

Page 2

Charts 2.1 and 2.2: Number of students with each type of loan

- Relevance to Christie and Tim:
    - This is like a summary of the other charts on the page (with `Both` as a bonus)
    - They expected almost all students with a `PLUS` loan to also have an `Unsubsidized` loan (short by two students)
- How I made this/what else I can do:
    - Note that there is no `Both` in my original dataset; I created four datasets to build these charts
        - The first three were filtered for each `Loan Type`: `Unsubsidized`, `PLUS`, or `None
        - For the fourth, I merged `Unsubsidized` and `PLUS`, removed duplicate names, and filled in all the `Loan Type` values with `Both`

    - I could have built a fifth chart of students that borrowed a `PLUS` loan but not an `Unsubsidized` loan

Charts 2.3 through 2.6: Number of students in each program with a given loan (`PLUS` or `Unsubsidized`)

- Relevance to Christie and Tim:
    - This was information they requested, but in a visual format
    - Proportions are included in the pivot table, and may be more useful
    - Could be useful to see how many students in potentially lower-paying fields are taking out `PLUS` loans
    - We see that zero of the `GCRT` students took out a loan
- How I made this/what else I can do:
    - This is a simple example of building multiple charts at once:
        - Imagine that instead of 2 loan types, there were 2,000
        - Instead of repeating the same task 2,000 times, I can use a `loop`, which is like a repeat button for a set of instructions.
        - Here, I looped through my two datasets from above: `Unsubsidized` and `PLUS`.
        - My instructions were to count the number of students in each program, and repeat for every dataset (in this case two datasets; I could have included the `PharmD` students by adding a few more instructions)

Page 3

Chart 3.1: Relationship between `Term Award` and `Semester`

- This is a boxplot, which shows the distribution of `Term Award`, grouped by `Semester`
    - Breaks up the data into quarters
    - The box shows the middle 50% of data are (i.e. from the 25th percentile to the 75th percentile)
    - Line in the middle of the box indicates the median
    - Lines sticking out of the box (whiskers) indicate the range of the data
    - Diamonds indicate outliers
- Relevance to Christie and Tim:
    - These plots have a positive skew, meaning a handful of students borrowed a lot more than the average student
    - We see that `Summer` has the lowest median, yet two students borrowed more in this semester than `Fall` or `Spring`
    - We can generate questions from this plot: Who were those two students? Why did students borrow slightly more than in `Fall`? What would this chart look like if we excluded students that did not borrow (replace zero with blanks as discussed above)?

Chart 3.2: Relationship between Program and Award, formatted by Semester

- Relevance to Christie and Tim:

    - Bonus chart, done out of my own curiosity
    - We see that `P1` and `P2` students are not taking out `Summer` loans
    - We see that `MMSPA` consistently borrow the same amount throughout the year
- How I made this/what else I can do:
    - With Python, I can pile in multiple variables, and have an extensive variety of formats to distinguish between them
        - For example, with a scatter plot I can use color gradients, change dot size in proportion to the amount borrowed, use an `x` for `PLUS` loans and a `o` for `Unsubsidized` loans; there are so many ways I can customize a chart to whatever is needed.
        - Additionally, I could have added a fourth variable: `Loan Type`. It would be busy, but with another dataset further filtering can be helpful

Appendix

AI. Abbreviations and definitions

Degree names

- DNP: Doctor of Nursing Practice
- GCRT: Graduate Certification (for non-accredited courses)
- MABL: Master of Applied Biblical Leadership
- MAI: Master of Arts in Innovation
- MAT: Master of Athletic Training
- MAWT: Master of Arts in Worship and Theology
- MBA: Master of Business Administration
- MDIV: Master of Divinity
- MMIN: Master of Ministry
- MMSPA: Master of Medicinal Science Physician Assistant
- MSN: Master of Nursing
- P1-4: Graduate Pharmacy Student, Year 1 through 4

University terms

- DGP: Direct Graduate Unsubsidized loan, may also be abbreviated as "Unsub loan" in this document to avoid confusion with DGU.
- DGU: Direct Graduate PLUS loan, may also be abbreviated as "PLUS loan" in this document
- DSS: Direct Subsidized Student loan. Is an undergrad loan, may appear in this document if a student changed from undergraduate to graduate mid-year.
- DUS: Direct Unsubsidized Student loan. Also an undergrad loan.
- FAFSA: Free Application for Federal Student Aid
- SU: Summer term

- FA: Fall term
- SP: Spring term
- Unsubsidized loan
  - Preferred over a PLUS loan
  - Accumulates interest the moment it disburses
  - More favorable interest rates compared to PLUS loan
  - Limits on amount a student can borrow; max depends on program
- PLUS loan
  - Accumulates interest the moment it disburses
  - Higher interest rates than a Direct Unsubsidized Loan
  - No limit to amount a student can borrow (within reason)
- *Note*: A student may take out a PLUS loan, an Unsub loan, or both

Data analysis terms

- NaN: Not a number, a value that is undefined or un-representable (in this case, missing or blank)

AII. File descriptions

There are three `.csv` files, which I named `Stats.csv`, `Loans.csv`, and `Master.csv`.

Loans

Contains 9 columns:

- `ID #`: Student ID number, 7 digits in length. Discrete Quantitative
- `Name`: Student name, formatted "Last, First Middle". Nominal Categorical
- `Term`: The term and year the loan was accepted, formatted "YYYYTT". Ordinal Categorical
- `Loan`: Type of loan, formatted `AAA1`, where `AAA` is the abbreviation for loan type (i.e. `DGU` or `DGP`), and the `1` ranges from `1` to `4` (inclusive), representing how many consecutive loans there were that year. For example:
  - A student takes out one Unsubsidized loan for the year, and it disburses a portion each term. It would show up as `DGU1` for each semester.
  - A student takes out a new Unsubsidized loan each semester. This would show up as `DGU1`, `DGU2`, and `DGU3` for each respective term.
  - A student takes out one Unsubsidized loan for the year, but at some point decides to change or cancel the loan:
    - This would show up as `DGU1` with a non-zero amount in `Term Awd` for each term the student received money.
    - If canceled before receiving the loan, there will be a zero in the `Term Awd` column.
    - In either case, the next loan would start at `DGU2` and increase with each new loan or change.

    - I was told to ignore the digit at the end, but for completion sake have included a description.
- `Term Awd`: The loan amount in US dollars for a given term. Discrete Quantitative
    - I only learned *afterwards* that this variable would better be described as `Amount borrowed`
- `Term Fee`: Fee for taking out a loan, usually a fixed percent of the total loan. Told to ignore this column.
- `Term Disb`: The disbursement, or amount the student received, calculated by `Term Awd` - `Term Fee`. Told to ignore this column.
- `Status`: Whether or not the loan was accepted, rounded to the nearest whole dollar. This specific document should contain all `A`s for "Accepted". Categorical
- `Loan ID`: Number for each unique loan, Discrete Quantitative

Master

Contains 3 columns (first three listed below). They would like me to add three more columns; merge corresponding information from `Loans.csv` to this document. All graduate students in `Loans.csv` have accepted a loan, whereas `Master.csv` contains all graduate students.

- `ID`: Student ID number, 7 digits in length. Discrete Quantitative
- `Name`: Student name, formatted "Last, First Middle". Nominal Categorical
- `GS Program`: Graduate Student Program, see Degree Names above in "Abbreviations" section. Nominal Categorical
- `FAFSA`: Whether or not the student completed the FAFSA. `A` for accepted or `D` for declined. Binary categorical.
- `Unsub Amount`: The loan amount (rounded to the nearest whole dollar) if the student took out an Unsub loan. Discrete Quantitative
- `PLUS Amount`: The loan amount (rounded to the nearest whole dollar) if the student took out a PLUS loan. Discrete Quantitative

Stats

This document contained a sample spreadsheet with blank columns that they wanted me to fill in. Currently has:

- `Program`: The abbreviation for the program formatted as "PROGRAM.XXXX", where ".XXXX" indicates a portion of the string I was told to ignore.
- `# Student`: Total number of students in each program
- `# FAFSA`: Number of students that completed the FAFSA in each program.
- `# Unsub`: Number of students that took out an Unsub loan in each program.
- `# PLUS`: Number of students that took out a PLUS loan in each program.
- `% Unsub`: Percentage of students that took out an Unsub loan in each program, based on total students.
- `% PLUS`: Percentage of students that took out a PLUS loan in each program, based on total students