# CSC 805 Group Project Phase 3

Gio Jung
Kenji Madden
Sai Saketh Bavisetti
Areeb Abbasi

## System architecture

Our system is using Python to implement such a visualization of our project. We used “Pandas” to load the data set (CSV file). Furthermore, the visualization tools we used are “matplotlib”, “seaborn”, and “plotly” to create the visualization charts. Below is the figures for each tools (not able for “seaborn”)

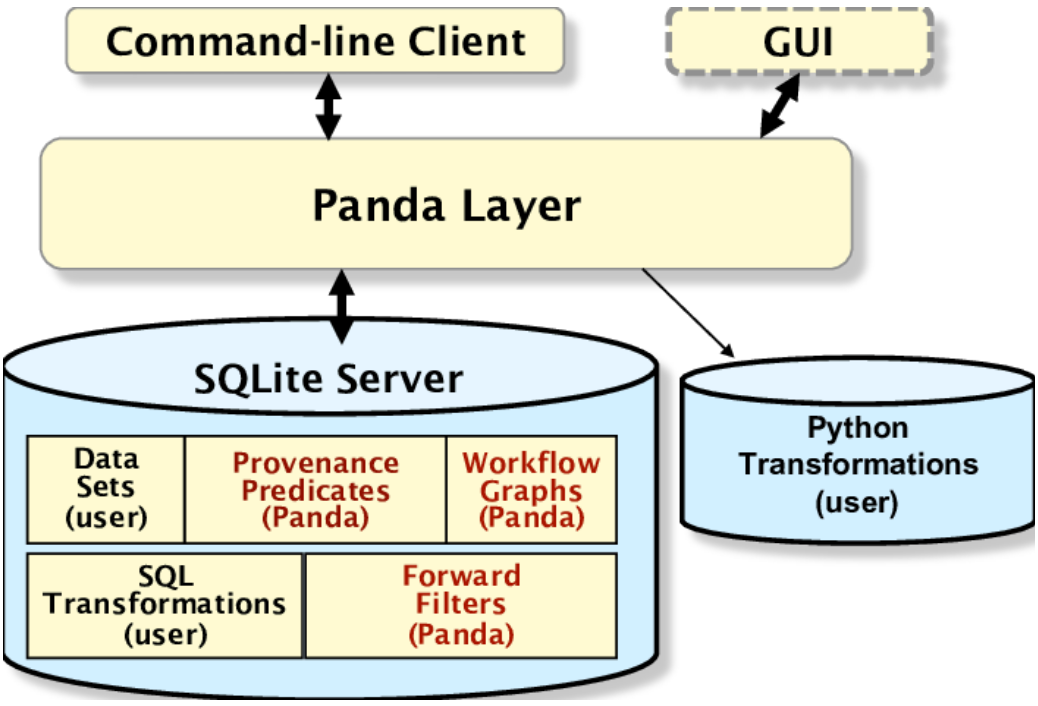



Diagram of Pandas

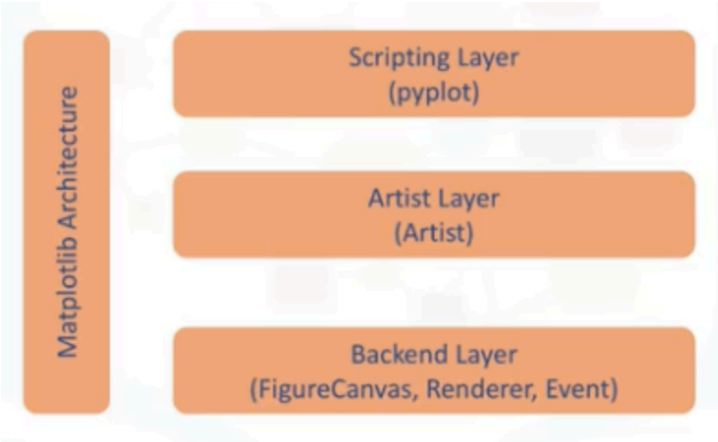



Diagram of Matplotlib

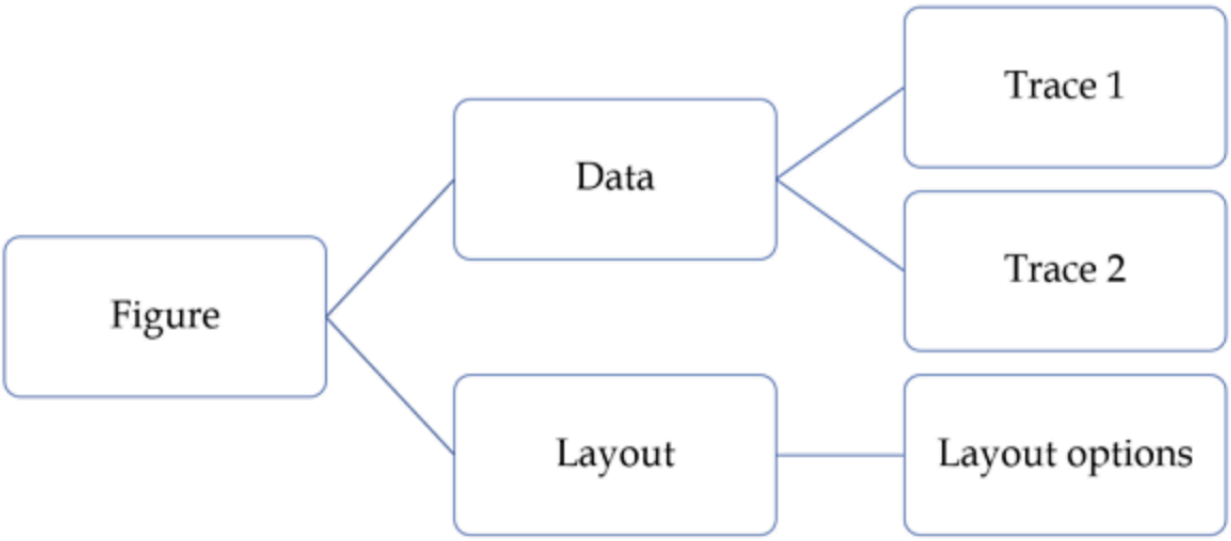



Diagram of Ploty

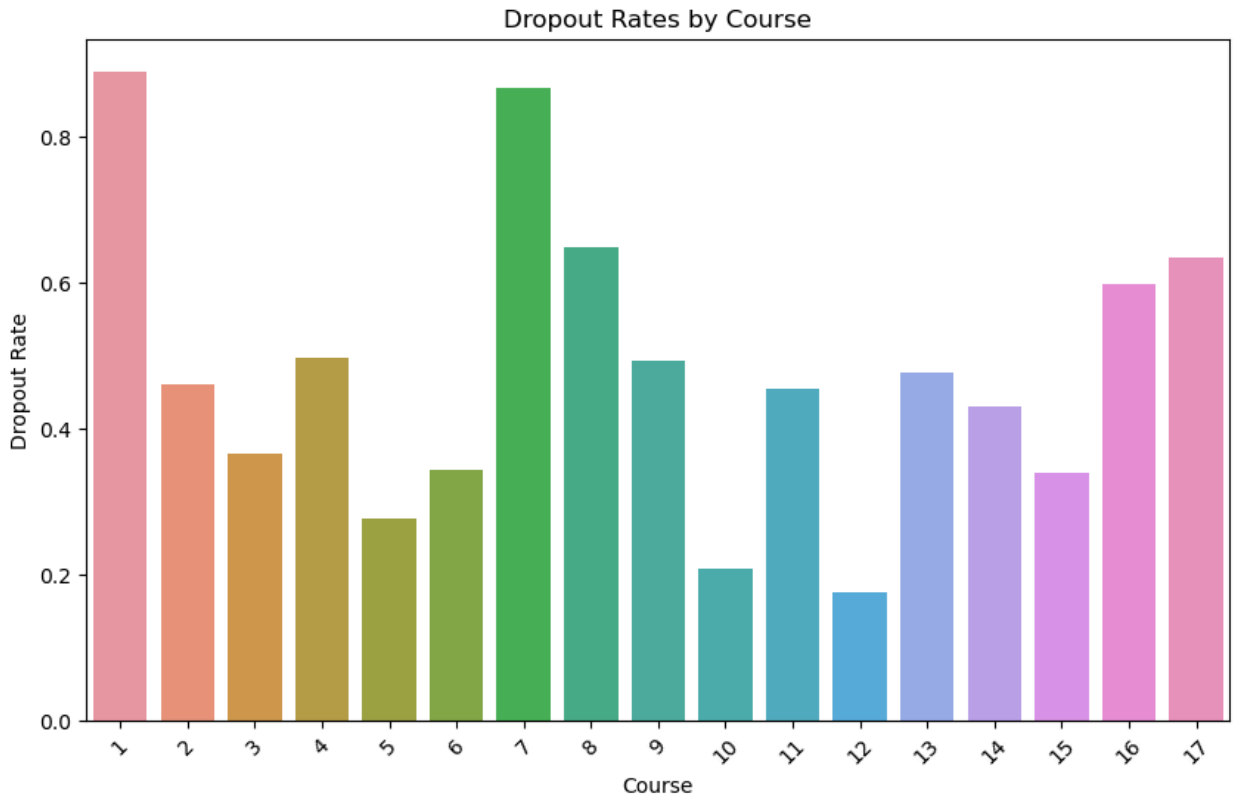
## Data set description

The dataset shows detailed statistics about various undergraduate students such as demographics, socioeconomic status, and academic performance. These statistics can be used to make assessments on academic performance and analyze factors of academic success or failure. The datasets allow us to make multiple assessments based on different data and cross-compare between sets to identify independent and compounding trends. Finally, the dataset can be used to identify institution problems, if there are disproportionate results based on enrollment.

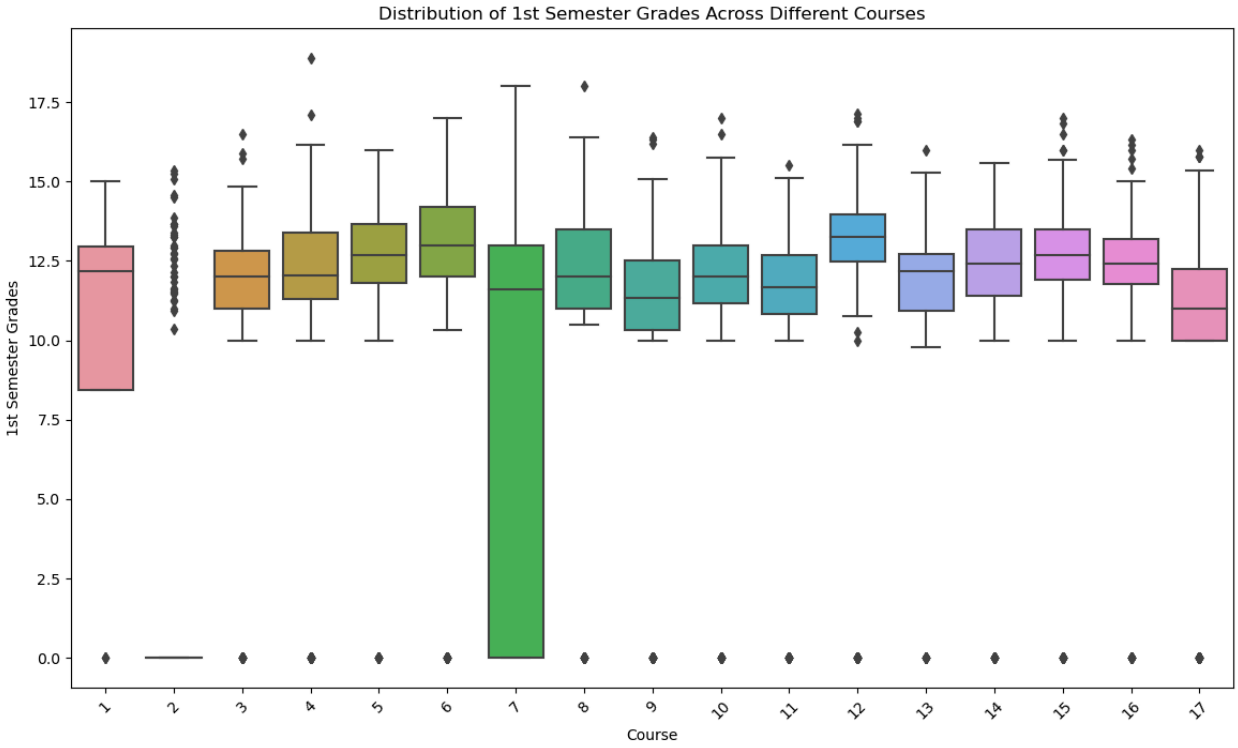
## System description

Our system will utilize the provided dataset to generate visualizations, aiding in the identification of patterns that can be used to prevent student dropout by providing emotional support and financial aid. The visualizations include dropout rates by course, distribution of 1st semester grades across different courses, and scatter plots showing the relationship between various economic indicators and dropout rates. By exploring these patterns, we aim to compare the insights gained from visual analysis with human analysis and assess the accuracy of our predictive model.

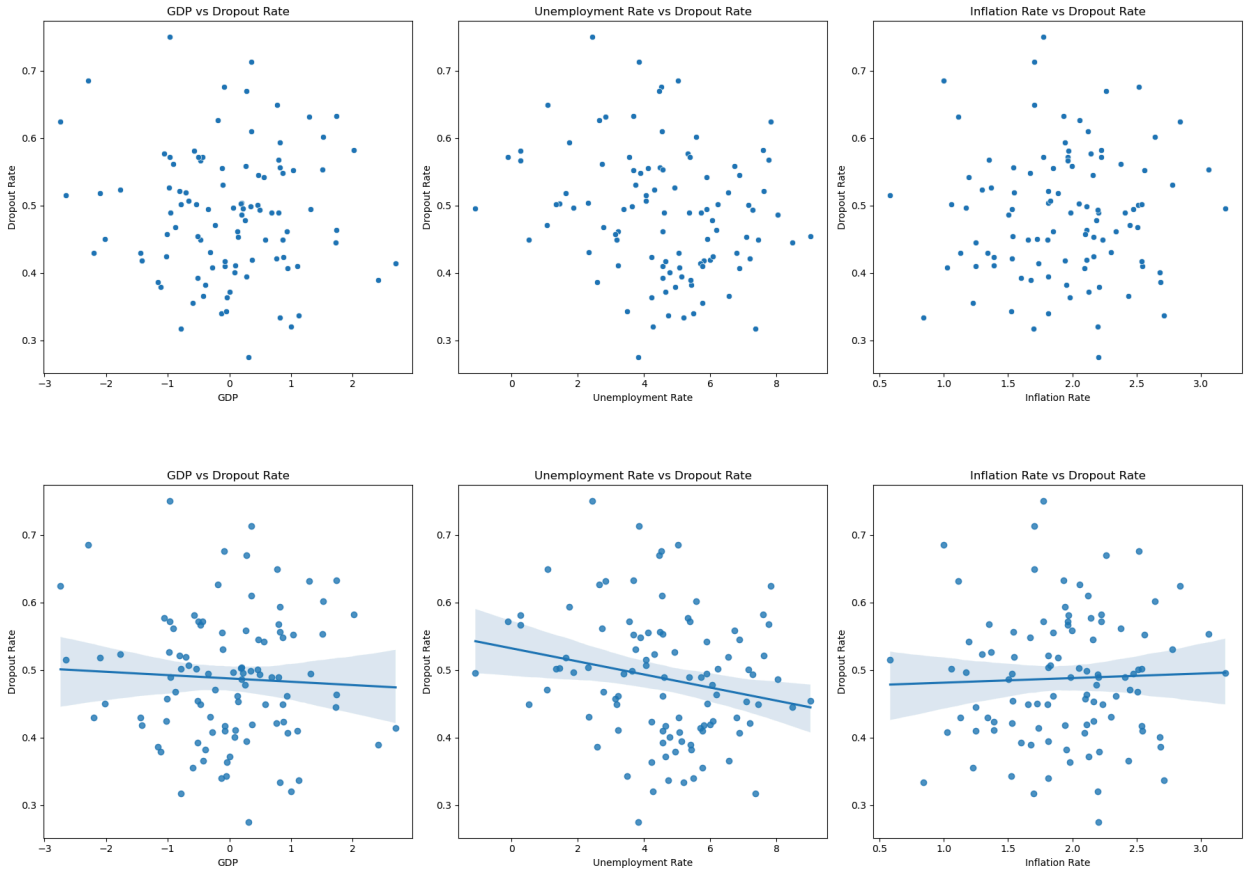
## Screenshots of each view and 2-3 sentences description of each visualization of the screenshot



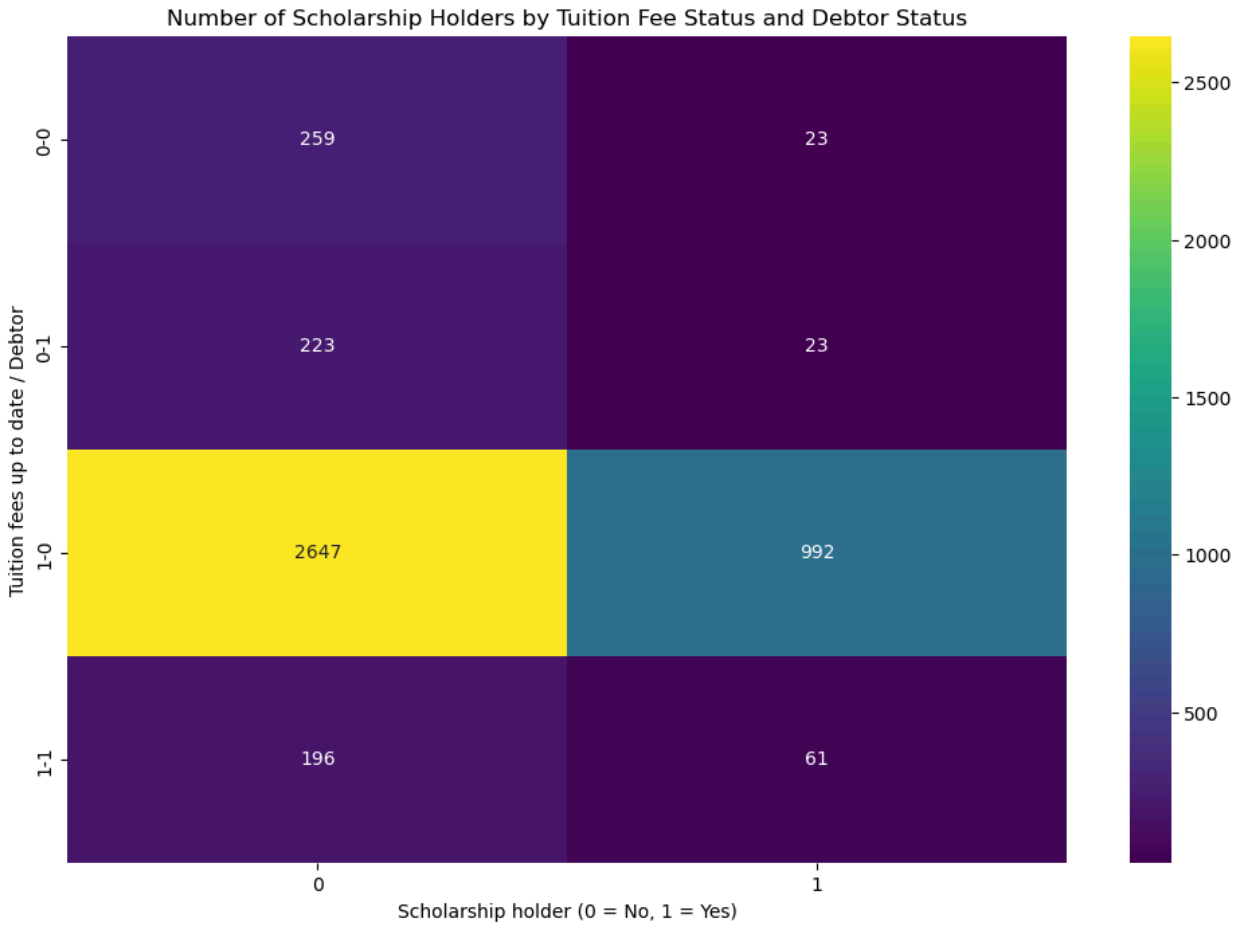

Firstly we wanted to look at the distribution of dropouts by courses to see if there were outliers classes that may have driven people to drop out. The graph shows that there are two significant outlier classes with three above-average classes. This is important because we can identify for certain that two classes have disproportionate influence on dropout statistics.




When viewing these same courses' grade distribution we can see that both 1 and 7 have the largest distribution and course 7 has a significant lower deviation in grade distribution. This is a clear sign the people who take course 7 have a higher chance of failing and dropping out.




These next three graphs show the distribution of GDP, unemployment, and inflation rate vs dropout rate. These are general economic indicators, but only the center graph shows any relation since both GDP and Inflation do not seem to affect the drop out rate at all. However, the unemployment rate shows a slight relationship as it increases the dropout rate decreases. I suspect that higher unemployment motivates people to stick with school.




This visualization represents the cross product of the number of students who either paid the tuition and who either are debtors among the non-scholarship holders and the scholarship holders. Basically, 0 and 1 refer to Yes and no for each attribute (Tuition fees up to date, Debtor, and Scholarship holder). Furthermore, there is a color variable to easily identify each aggregation of the data block. As we can see, students who are not scholarship holders and are not debtors but have paid for their tuition comprise the largest subset of data.

## Link to Demo Video

https://drive.google.com/file/d/1iPp28H3Mby2eH39VoMRJDnK1w_AYM60h/view

## Link to System Source

https://github.com/GioJung97/CSC805-Project.git

## Individual Work Report

- Gio Jung : I managed to split the tasks and contributed to writing down the report.
- Kenji Madden: I contributed to writing down the report.
- Sai Saketh Bavisetti : I contributed to setting up the system and implemented the code. Also, I created a demo video.
- Areeb Abbasi : I contributed double checking the code and modified it.