

**Group Name:** Daily Screamer

**Group Member:** Oscar Wan, Jason Wang, Wendy Zong, Angela Tao, Sylvia Huang

**Project Name:** Course Evaluation

**Date:** Dec. 13, 2022

**Project Idea Description:**

Goal: build a platform for NYU students to see the quality as well as difficulty of each course.

Course selection and scheduling have always been the top priority for students before semesters start. However, it has been difficult for students to assess the quality and difficulty of a course. Students generally seek such information through ratemyprofessor.com (RMP) and course evaluations (CE) published on Albert, however, there are several issues with these resources:

First, RMP mainly provides ratings for professors but not for classes themselves. Second, it is highly biased. Most of the ratings and comments that professors receive on the website come from students with strong opinions, thus the ratings are highly skewed. Third, the sample size of the ratings are usually very small, which makes the ratings inaccurate.

Albert's course evaluations are available for students and can offer a more objective evaluation of a course, but it is very inconvenient to view them. Furthermore, although students can see the ratings on different attributes of a given course, they cannot see courses that match their expectation for quality and/or difficulty. In other words, Albert CE only takes a course as an input and outputs many attributes but not vice versa. Furthermore, because there are so many attributes, it is difficult to compare courses.

Our product addresses these problems. Through machine learning methods, we create and present to the users new concise metrics that summarize the quality and difficulty of all courses

using CE data. In addition, it is a two-way interaction, meaning that users can define a quality and/or difficulty and then we output matching classes, or they can define a class and we then output their quality and difficulty.

### Data

Using selenium, we scraped Alber's course evaluations for Stern and CAS. Our dataset includes information on 1703 Stern courses and 9667 CAS courses. There are 7 CE ratings for each Stern course, and 17 for CAS. For both schools, we have course code, course name, instructor, response rates for all courses. Furthermore, we scraped all NYU professor ratings and student comments on ratemyprofessor.com. All data is stored in the SQL database.

### Analysis:

Our initial explorations included looking at the distribution of ratings across CE and RMP, comparing sample sizes, and examining the correlation matrix. A very important finding is that on average, professor quality rating on RMP is about 15% lower than that on CE, and with a much larger variation. The average sample size for each rating on CE is almost 3 times of that on RMP (39.37 vs. 13.67).

Most of the features are highly correlated with each other, but "Challenging" correlates much less with all other features. This hints the fact that there will be two prominent principal components. In fact there is. From principal component analysis, we found that the first 2 PCs explained about 82% of the variance in the data, which means 2 questions are all it needs on the course evaluations. From analyzing the loadings plots, we characterized the first 2 PCs as Quality and Difficulty, respectively. This PCA analysis serves as the backbone for our project.

## Frontend

We think streamlit is more convenient than HTML based on our needs of conducting a user friendly interface.

We created 3 functional pages. At the left side of the front page, we have the menu button which allows users to navigate through four possible pages that can satisfies different needs while searching for classes.

- CAS & Stern course summary: This page presents the distribution of the overall rating for both stern and cas classes. We show the graphs of distribution of the overall rating for both stern and cas classes. This page is for the basic understanding of the users, but not considering any actual functionality.
- CAS course search: This page enables users to search courses in NYU CAS. The frontend features including streamlit's checkboxes, sliders, and text boxes. These features are to help users to select and deselect some features of the course, and also allows them to freely navigate the characteristics based on their personal preferences of courses. For cas courses, users can search department courses by difficulty, quality, or both. The two metrics are presented in a scale of 5 based on percentile. For example, difficulty of 1 represents courses in the 0-20th percentile in terms of difficulty, which are the relatively easiest courses within the department. Whereas quality of 5 represents the courses with the highest quality. Users can choose whether to display course evaluation ratings, if not, only the course information and name of professor will be displayed. Users can also choose to look for courses by course code or professor's name to look at the course evaluations and their corresponding difficulty and quality in comparison to other courses.
- Stern course search: Stern course evaluations have different questions than CAS courses,

and there is no difficulty rating for Stern courses. Therefore, there will only be quality rating for course searches, with all else being the same as the CAS page.