

Module 3 Limitations

Pull Data Status Updates: The Pull Data operation is executed through a subprocess triggered by the Flask application. While a lock file prevents multiple simultaneous runs, the webpage does not currently implement asynchronous background status polling. Therefore, status messages do not automatically refresh when the Pull Data job completes and users must manually refresh or click Update Analysis to view updated results.

Button State Behavior: The Pull Data and Update Analysis buttons are designed to disable when scraping is in progress using a file-based locking mechanism. However, in many cases the scraper terminates quickly because GradCafe returns repeated survey pages. Since the job completes rapidly, users may not observe the buttons remaining disabled for an extended period.

Module 2 Scraper Reliability: Module 2 scraping logic was enhanced to meet assignment requirements, including migration from urllib3 to urllib, improved parsing and field extraction, pagination protection logic, and deduplication safeguards. Despite these improvements, GradCafe occasionally returns repeated survey pages when accessed programmatically, which can cause scraper runs to terminate early and limit dataset expansion.

Impact on Module 3: Module 3 functionality is not affected by these scraper limitations because the database was successfully populated using instructor-provided data. All SQL queries and Flask analytics operate correctly using verified datasets, and Pull Data functionality remains operational when valid new data is available.

Scraping Data Accuracy: GradCafe data is user-submitted, anonymous, and not formally validated. Therefore, entries may contain incomplete, inconsistent, or inaccurate information.