**University of Denver Technical Assessment Report**

Author:  
Paul Martinez, Ph.D.

8/13/2025

*Acknowledgements*

I would like to thank the University of Denver’s Office of Institutional Research & Analysis for designing a technical assessment that reflects real-world institutional research needs while challenging analytical skill. This work was completed independently, but guided by the principles of automation, reproducibility, and clarity that underpin effective research practice.

*Executive Summary*

This report documents the design, development, and execution of a fully automated Python workflow that takes raw enrollment and academic performance data, cleans and consolidates it into a student-level dataset, and then produces all required tables, visualizations, and outputs in one reproducible run.

The purpose was not only to answer the given assessment questions, but to build a reusable, parameter-independent process capable of producing the same deliverables for future academic terms with minimal user intervention—requiring only directory path updates.

Two separate Python scripts were created:

1. Cleaning Script — Automates dataset ingestion, merging, cleaning, flag creation, and student-level collapsing.
2. Analysis Script — Automates generation of statistical tables and graphics, outputting each to a separate sheet in one Excel workbook.

When run in order, these scripts take the process from raw data to final report, ensuring that the results are accurate, transparent, and reproducible.

*Process Overview*

Phase 1 — Cleaning Script

* Reads three input datasets (Fall Enrollment, Grades, Program Data).
* Checks file presence and merges datasets on cleaned keys.
* Resolves formatting inconsistencies and applies standard naming conventions.
* Generates all required analytical flags: persistence, undeclared→declared, preferred program, gender, race/ethnicity, and age.
* Reduces the dataset to one record per student-term to remove redundancy.
* Adapts to variable term codes without altering core code.
* Output: student\_level\_final\_data.csv — a consolidated student-level dataset ready for analysis.

Phase 2 — Analysis Script

* Reads the cleaned dataset and produces all deliverables specified in the assessment instructions.
* Outputs include:
  + Persistence rate
  + Two-proportion z-test for gender differences
  + Race/ethnicity and gender composition
  + Age distribution statistics and visualization
  + Grade distribution statistics and visualization
  + Average GPA by program (table only)
  + Degree-level GPA distributions (visualization and binned table)
  + Proportion of WK3 undeclared students declared by EOT
* Output: DU\_IRA\_Assessment\_Report.xlsx — a single Excel workbook with each table or visualization on its own sheet.

*Conclusion*

This workflow is designed to be executed in sequence and with minimal edits for each new term’s data—requiring only a path update before running the cleaning and analysis scripts. The structure provides a sustainable and repeatable process for institutional reporting, ensuring both speed and consistency while reducing the risk of manual error. With these tools in place, reports that once required hours of manual work can now be produced in minutes, ready for immediate review.

*AI Disclaimer*

*In the interest of transparency, I want to clearly state that AI was used as a supporting tool in the development of both the code and the written report. All process design, coding structure, automation strategy, and overall analytical approach were my own ideas. At no point was any student-level data uploaded into the AI tool. My practice when using AI is to work exclusively through conversational guidance; if I encounter a challenge that cannot be addressed through discussion alone, I create a synthetic dataset to replicate the structural problem without exposing any sensitive information.*

*Overall, I view AI as a collaborative assistant—helpful in enhancing clarity and efficiency, but not a replacement for domain knowledge or critical thinking. Like any tool, it has limitations and errors, and its value comes from being used judiciously, with the human analyst maintaining full control over design decisions and interpretations.*