**University of Denver – Office of Institutional Research and Analysis**

**Associate Director of Institutional Research Technical Assessment**

Please download the following three CSVs from https://github.com/bennybenibene/DU-Technical-Assessment: fall\_enrollment, grades and program\_data. Please use these files to produce the requested deliverables and answer the provided questions.

To the extent possible, please automate your solution in python. Imagine that you would like to create a program which could generate the result set for the same data inputs but with different values. You can assume that there would be no data issues in the new files beyond what you identify in the provided files.

We are interested in your responses to the questions, but our primary interest is in your process. Below you will find some notes on the files and data as well as some definitions.

**fall\_enrollment.csv:**

DU uses a census-based system for tracking registration where a snapshot is captured at the third week of the 10-week quarter as well as at the end of the term. If no census record is available for a student, it indicates that they were not enrolled. Census records also allow us to compute persistence rates. There should be no more than one enrollment record per student for each term census point combination.

This dataset includes synthesized records for undergraduate students for a single DU term (Fall 2021). The file includes id, basic demographic information, census and program of study information (College, Major, Degree). Note, majors beginning with ‘UN’ represent undeclared majors within specific colleges.

Race/Ethnicity is established based on three fields: race, ethnicity and visa type. All visa types with the exclusion of ‘PR’, ‘RF’ and ‘AS’ are considered ‘international’ when computing an overall race/ethnicity variable. Ethnicity description concerns Hispanic or Latino identities, and anyone indicating such an identity is considered ‘Hispanic or Latino’ for race/ethnicity purposes. Visa type has the highest precedence, followed by ethnicity for the purposes of computing race/ethnicity, otherwise we use the value from the race variable. Please compute a race/ethnicity variable.

**program\_data.csv:**

This file provides a mapping of College, Degree, Major combinations to the Program Code. Program is a combination of degree and academic unit information. Multiple majors may map to individual programs.

**grades.csv:**

The grades file provides letter grades for each student for each class that they completed within the given term. Students will most often have multiple grades per term, reflecting grades for distinct classes. Assume that person-level duplication in this file is intentional and reflects grades for distinct classes. A letter grade to GPA value mapping is provided below. All intervals are half open except the A interval. For example, A: [3.7, 4.0] and A-: [3.3, 3.7)

{A: 4.0, A-: 3.7, B+: 3.3, B: 3.0, B-: 2.7, C+: 2.3, C: 2.0, C-: 1.7, D+: 1.3, D: 1.0, F: 0.0}

**Deliverables:**

1. Please provide a csv file that that provides student-level data. The file should aggregate the provided data in a way that allows you to answer all or most of the questions below. Try to structure this dataset in a way that reduces redundancy while capturing the information critical to generating the necessary result set.
2. Please respond to or address the questions/tasks listed below. You don’t have to format these as a formal presentation. A document with your responses is sufficient.
3. Please provide your code for automating your data cleaning and analysis.
4. Please submit a zipped file of the above materials.

**Questions/Tasks:**

* What is the persistence rate from week three to the end of the term?
* Is there a statistically significant difference in persistence rate between males and females?
* Describe the makeup of the class in terms of race/ethnicity and gender.
* How is age distributed in this class?
* Describe or show the distribution of grades across all students.
* Present a cross tab of average GPA by program.
* Present a visual representation of course grade distributions broken out by broad degree level (BA, BS, BM, BFA).
* What is the proportion of week three undeclared students with a declared major at the end of term?