Data Specialist Performance Task

General Overview

The simulated data you have received pertain to a summer program for 8th and 9th-graders. A subset of 8th-graders at Mercury Middle School participated in a 60-day program in the summer of 2001, and then participated in a similar program tailored to 9th-graders the following summer. For the 9th-grade program, though, the students had matriculated to two different high schools: Venus and Jupiter.

Task Overview

A district official has recently learned about a framework for monitoring student progress and wants to implement it using summer program data. The framework involves looking at a student's attendance and grades to determine whether a student is "On-Track," "Vulnerable," or "Off-Track." You are tasked with completing this project. In the most general sense, you want to communicate clearly to the district the distribution of 8th-grade on-track status and what that on-track status looked like for the same students during their 9th-grade year.

Data Overview

You have access to three tables: 1) General Student Data, 2) Attendance Data, and 3) Academic Data.

1) General Student Data: This table provides a student's district ID number, their current school of enrollment, and their learning style as determined by a district-wide test. The following is a lookup provided by the district for the learning styles:

Mark	Style
V	Visual
A	Auditory
K	Kinesthetic
L	Linguistic

2) Attendance Data: This table provides a student's district ID number, date value, day of week, an attendance binary (1=present;0=absent), school at the time of enrollment, and school year. 2001 and 2002 data are available.

3) Academic Data: This table provides a student's district ID number, course number for each course he or she took, the letter grade they received in the course, school at the time of enrollment, and school year.

The Framework

If a student has an attendance rate of <90% and a GPA <3.0, then they are considered "Off-Track." If a student has either an attendance rate of <90% and a GPA <3.0, but not both, then they are considered "Vulnerable." If neither, "On-Track."

Attendance rates are calculated by dividing the number of days present by total days enrolled. Thus, if student A's att_binary values are 1,1,1,0, then his/her attendance rate would be 75%.

GPAs are calculated using a generic gpa point system (displayed below). A student's GPA is calculated by averaging a student's gpa points across courses. Thus, if student A's grades are A,B,C, his/her gpa points would be 4.0,3.0,2.0, from which we can calculate a GPA of 3.0.

Grade	GPA Points
A+	4
A	4
A-	3.7
B+	3.3
B	3
B-	2.7
C+	2.3
C	2
C-	1.7
D+	1.3
D	1
D-	0.7
F	0

The Task

The task has five pieces that will help us understand your ability to analyze data, communicate findings, and share analysis techniques.

1. Determine the on-track distribution for each high school's students. Specifically, how many students fall into each category?

2. Provide a list of student IDs for students whose on-track status has changed and what that change is.

3. Many of our partners understand analysis better when it is presented visually. Create a one page visualization for each high school showing the on-track status of its 9th graders grouped by their 8th grade on-track status. Label your visualizations such that it can be understood without you there to talk someone through it.

4. As a data analyst, you like to provide districts with insight beyond what they request, especially if that insight could be used to help students. Create one additional visualization that could be used to familiarize educators with something interesting you discovered in the data. Label your visualization such that it can be understood without you there to talk someone through it.

5. Describe your process for accomplishing the above. If you encounter any issues in the course of accomplishing the above, include what those issues were and how you managed them.

Submit items 1, 3, and 4 in a Google Slides deck that the district official could use to introduce to share on-track results with other district educators. Submit item 2 in a Google Sheets workbook that the district official could share with educators in order to help them target students for interventions. Submit item 5 in a Google doc that the RISE data team can use to understand how you approached this performance task. When you are ready to submit your work, put your three files into a folder and share the folder to data@ctrise.org.

Expectations

Our work relies upon the use of scripting languages to automate processes with data sources that get updated nightly. As such, you are required to complete this task using a scripting language or SQL. Please include whatever contextual information you feel is sufficient to recreate the process as we intend to test the submitted script with the data in its original state.

Final Notes

The data were created to simulate the kind of tasks the RISE data team encounters regularly and some of the student data will be messy or even missing. The purpose of this task is to allow us to understand your approach to analyzing and presenting data: There is no single "solution" to the tasks. We look forward to seeing what you discover.

If you have any questions throughout the process, please send them to data@ctrise.org and we'll be happy to help.