ReadyCT
Performance Task - - Data and Research Specialist

Congratulations! You have advanced to the next stage in the hiring process for the data and research specialist position with ReadyCT.

We are excited to learn more about your qualifications as they pertain to this position. This next step in the process provides an opportunity to better gauge whether this role aligns with your professional experience and interests.

While the time required will vary based on your skill level, you can expect to spend between 1 and 3 hours on this task.

General Overview

Below is a hypothetical scenario, including a fictional program, students, schools, districts, external stakeholders, and datasets, which you will use to respond to the prompts provided. Please respond to the prompts to the best of your ability using the information provided.

At ReadyCT, we primarily use Python for our data analysis tasks and generally expect candidates to use Python as well. However, if you prefer to use a different programming language or method, that's perfectly fine. Just be sure to clearly explain what you did and how you did it. Additionally, we appreciate comments in your code that explain your process, as it helps us better understand your approach.

Provided Datasets

You have been given view only access to a Google Sheets spreadsheet containing three different datasets – please download the data for your own use on this task. Following is the explanation of the data provided:

1. **Students Dataset**

   The **Students** dataset contains information about 998 hypothetical students across two districts: Gotham City Public Schools (including James Gordon Center for Learning and Lucius Fox High School) and Neverland Public Schools (including Lost Boys Academy and Captain Hook High School). This dataset contains the following data:

   - **id number:** A unique identifier for each student.
   - **first name:** The student's first name.
   - **last name:** The student's last name.
   - **grade:** The student's current grade level.
   - **school:** The name of the school the student is enrolled in.

- **district:** The district that the student's school belongs to.
- **dist_code:** A code representing the district.
- **sys_id:** A system identifier for the student, used to track cross district data in our system.
- **pathway:** The career pathway or program that the student is enrolled in.

2. **Events Dataset**

   The **Events** dataset contains data about various **Work-Based Learning (WBL) events** that students attended, such as resume reviews, worksite tours, skills workshops, guest speaker sessions, and mock interviews. This dataset is essential for tracking student participation in these career-related activities.

   - **event_id:** A unique identifier for each event.
   - **school:** The name of the school associated with the event.
   - **event_date:** The date on which the event took place.
   - **location:** The physical location where the event occurred.
   - **school_hours:** An indicator showing whether the event took place during or outside school hours.
   - **event_type:** The type of event (e.g., resume review, worksite tour, skills workshop, guest speaker, mock interview).

3. **Student_Events Dataset**

   The **Student_Events** dataset records the attendance of students at various WBL events. It acts as a bridge between the **Students** and **Events** datasets, linking students to the specific events they attended.

   - **student_id:** A unique identifier for the student who attended the event (matches the **sys_id** in the Students dataset).
   - **event_id:** A unique identifier for the event that the student attended (matches the **event_id** in the Events dataset).

Task Instructions

Task 1: Event Attendance List for Principals

The principals of Captain Hook High School and James Gordon Center for Learning have requested a list of events along with the students who attended them. The list should include the student's details as well as information about the events they attended.

Please use Python join the provided datasets and generate the required list. The output should include the following columns:

- First Name
- Last Name
- Student ID
- Grade
- Pathway
- Location
- School Hours
- Event Type

Please provide your Python code along with the resulting dataset.

Task 2: Flagging WBL Participants

In our data system, any student who has participated in a Work-Based Learning (WBL) event is considered a **WBL participant**. Using the provided datasets, flag all students who have participated in any WBL event.

Please provide an updated **Students** dataset (or a new dataset) with an additional column, **WBL_Participant**, that indicates whether a student participated in a WBL event. Please include any Python code used.

Task 3: Counting WBL Events per Student

Understanding the level of engagement of students in Work-Based Learning (WBL) activities is crucial for evaluating program impact. To this end, we need to count the number of WBL events each student has participated in.

Using the provided datasets, count the number of WBL events that each student has participated in. Please provide an updated **Students** dataset (or a new dataset) with an additional column, **WBL_Event_Count**, indicating the number of WBL events each student has participated in. Please include any Python code used.

Task 4: Identifying Pathway Students Not Engaged in WBL Events

It is expected that all students enrolled in a pathway participate in at least one Work-Based Learning (WBL) event during a given school year. The school administrators need a list of students from each school who are in a pathway but have not yet participated in any WBL event.

Using the provided datasets, identify and list all pathway students from each school who have not participated in any WBL event. Please provide a new dataset listing these students. Ensure the dataset is structured in a way that can be easily distributed to school staff for each school. Include any Python code used.

Submission Instructions:

Be sure to:

1. **Clearly label your submission documents/files with your name.**
2. Provide the datasets in separate tabs in a single Excel workbook (if possible) If you can't do this, send them in clearly labeled separate files.
3. **Clearly label** each dataset and code with the corresponding task number (e.g., Task 1, Task 2, etc.).
4. Include **comments in your code** where necessary to ensure we can easily understand how the code relates to each task.

Thank you for your time and effort in completing this performance task. We look forward to reviewing your work!