

Performance Task

Data Analyst, School Programs

TASK

Complete your performance task in response to the scenario below and submit your response to stephanie.franklin@tea.texas.gov.

DIRECTIONS

Review the scenario below, the description of the database, and the associated data files. **Your response should include the following three components:**

- 1) **A SQL query you have written to aggregate and extract necessary data from the database,**
- 2) **A data cleaning/analysis script,**
- 3) **A brief narrative which describes key findings with data visualizations**

For the data cleaning/analysis script, Python is the preferred language but not a requirement. The script can be written in the programming language of your choice. **Your response should demonstrate the following:**

- **Data Aggregation and Extraction:** Demonstrate how you would leverage a database to select, filter, and manipulate the data to prepare it for analysis.
- **Data Analysis:** Demonstrate how you cleaned, prepared, and analyzed the data in response to the scenario.
- **Data Communication:** Communicate key takeaway(s) relevant to the scenario both narratively and visually.

Please be prepared to describe the steps you took and your thought process while preparing your response during the interview. Your response will be evaluated based on your problem-solving approach, the accuracy and relevance of your analysis, and your overall presentation and communication of the result.

SCENARIO

In SY 20-21, TEA required all Texas LEAs (Local Education Agencies such as a school district) to administer either the TX-KEA or mCLASS assessment to Kindergarten students in order to measure Kindergarten readiness. The Tx-KEA and mCLASS diagnostics employ different scales for scoring and evaluate different sets of skills. Initial analysis of Fall 2020 data indicates a 38 percentage point gap in Kindergarten readiness between the two tests; 76% of students who took Tx-KEA were found to be Kindergarten ready compared to just 38% of students who took mCLASS.

Your task is to extract, analyze, and present data that will help the Early Childhood Education team understand to what extent the difference in readiness as measured by the tests is due to differences in the underlying populations of students taking each (as opposed to differences in test design and scoring). As mentioned above, your response should include the following three components:

- 1) A SQL query you have written to aggregate and extract necessary data from the database,
- 2) A data cleaning/analysis script,
- 3) A brief narrative which describes key findings and includes data visualizations

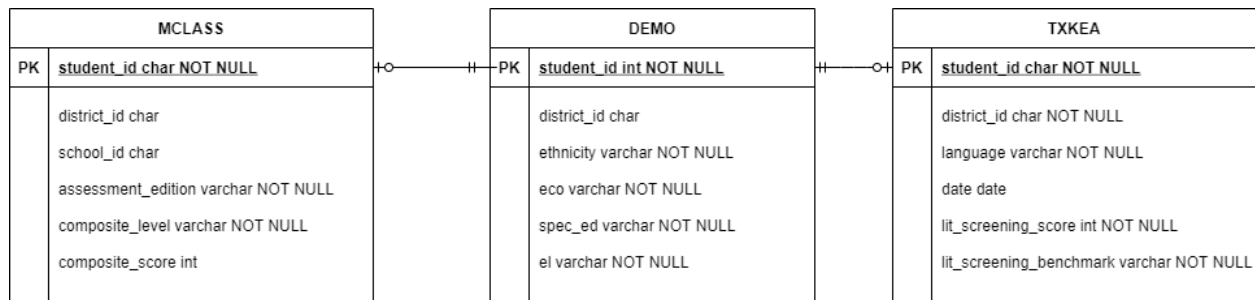
Please limit the analysis to students taking the English version of each diagnostic to avoid complications that arise from differences in the English and Spanish versions of mCLASS.

Component 1:

Below is a description of the database and an entity relationship diagram to help you prepare a SQL query that leverages the database to both extract and aggregate relevant data to prepare for analysis. Your analysis script should take as inputs the attached three data files (currently in csv format) which mimic the tables in the described database.

Database Tables:

- 1) MCLASS – student level performance on mCLASS diagnostic assessment.
 - Field 'assessment_edition' indicates whether the student took the English version of the test (DIBELS) or the Spanish version (IDEL).
 - For field 'composite_level' values of 'At Benchmark' or 'Above Benchmark' indicates Kindergarten readiness.
- 2) TXKEA – student level performance on Tx-KEA diagnostic assessment.
 - Field 'language' indicates whether the student took the English or Spanish version
 - For field 'lit_screening_benchmark' value of 'On-Track' indicates Kindergarten readiness.
- 3) DEMO – student level demographic data.
 - Field 'eco' indicates whether the student is identified as 'economically disadvantaged'
 - Field 'spec_ed' indicates whether the student receives special education services
 - Field 'el' indicates whether the student is identified as an English learner



Component 2:

Demonstrate how you prepared and analyzed the data in response to the scenario. For the data analysis script, Python is the preferred language but not a requirement. The script can be written in the programming language of your choice. Depending on your SQL query (because the database in component 1 is hypothetical), the first part of your data cleaning/analysis script may need to replicate some portion of the initial manipulation your SQL query accomplishes.

Component 3:

Please combine the narrative analysis and visualization in one file. Though Power BI is the preferred visualization software, the data presentation can be done using any software you choose (e.g. Tableau, Looker, Jupyter notebook, etc.).