COMPETENCIES

**4034.3.1** : **Dimension, Quality, Relations, and Limitations**

The graduate examines the data available for analysis to determine their dimension, quality, relations, and limitations.

**4034.3.2** : **Physical Data Models**

The graduate implements physical data models.

**4034.3.3** : **Table Operations**

The graduate performs table operations and queries within the context of data acquisition for analysis.

INTRODUCTION

A common undertaking for a data analyst is to connect an external data collection to an existing data set. This process involves the exploration of the source and target data sets to merge the new data in a meaningful and logical way.

You will replicate this common process in this task. You will take external data from a comma-separated values (CSV) file and design conceptual, logical, and physical models that describe the data. You will load the new data into an existing database alongside the existing data and run queries across it.

You will work on this performance assessment on the Labs on Demand website, which can be found in the Web Links section. The data sets, associated data dictionary, and CSV file are located at the [Labs on Demand](https://lrps.wgu.edu/provision/271283956) site.

REQUIREMENTS

*Your submission must be your original work. No more than a combined total of 30% of the submission and no more than a 10% match to any one individual source can be directly quoted or closely paraphrased from sources, even if cited correctly. The similarity report that is provided when you submit your task can be used as a guide.  
  
You must use the rubric to direct the creation of your submission because it provides detailed criteria that will be used to evaluate your work. Each requirement below may be evaluated by more than one rubric aspect. The rubric aspect titles may contain hyperlinks to relevant portions of the course.  
  
Tasks may****not****be submitted as cloud links, such as links to Google Docs, Google Slides, OneDrive, etc., unless specified in the task requirements. All other submissions must be file types that are uploaded and submitted as attachments (e.g., .docx, .pdf, .ppt).*

Perform the following tasks to combine an existing data set with some external data:

A.  Provide a question that can be answered using structured query language (SQL) to acquire data from *both* the original database and the add-on CSV file data. The question should require data from *both* data sources.

1.  Identify which data from the original data set and the add-on CSV file are needed to answer the question including all tables, columns, and data types.

B.  Create an entity relationship diagram (ERD) for the add-on CSV file and any other tables and columns used to answer the question from part A by evaluating the data contained in the file and identifying the m:n relationships and relational constraints.

1.  Write SQL code, in text format, that creates a table based on the ERD and specifies the columns and relevant keys.

2.  Write SQL code, in text format, that loads the data from one of the add-on CSV files into the table created in part B1.

*Note: Do not include SQL code as a screenshot.*

C.  Write a SQL statement or statements in text format for a query or queries that answer the question from part A.

*Note: Do not include SQL statements as a screenshot.*

1.  Provide a data file or files that capture the results from the query or queries.

D.  Identify the specific time period for how often the add-on file should be acquired and refreshed in the database for the data to remain relevant to the business and the question from part A.

1.  Explain why the time period identified in part D is relevant to the business needs.

E.  Provide a Panopto video recording that includes the presenter and a vocalized demonstration showing all code used, the code being executed, and the results of all code used in the task.

1.  Include a vocalized demonstration within the Panopto video recording provided in part E that describes the programs used to complete the task.

*Note: For instructions on how to access and use Panopto, use the "Panopto How-To Videos" web link provided below. To access Panopto's website, navigate to the web link titled "Panopto Access" and then choose to log in using the “WGU” option. If prompted, log in using your WGU student portal credentials, and then it will forward you to Panopto’s website.*

*To submit your recording, upload it to the Panopto drop box titled “Master of Science, Data Analytics TGM2 | D205 (Student Creators) [assignments]." Once the recording has been uploaded and processed in Panopto's system, retrieve the URL of the recording from Panopto and copy and paste it into the Links option. Upload the remaining task requirements using the Attachments option.*

F.  Acknowledge web sources used to acquire data or segments of third-party code to support the application. Be sure the web sources are reliable.

*Note: Submit web sources for part F separate from the sources in part G, or state none were used.*

G.  Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.

H.  Demonstrate professional communication in the content and presentation of your submission.