TGM3 — TGM3 TASK 1: DATA ACQUISITION

DATA ACQUISITION — D205

PRFA — TGM3

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 dataset. This process involves the exploration of the source and target datasets 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. The existing tables and CSV data files are located on this Labs on Demand website, as well.

REQUIREMENTS

Your submission must represent your original work and understanding of the course material. Most performance assessment submissions are automatically scanned through the WGU similarity checker. Students are strongly encouraged to wait for the similarity report to generate after uploading their work and then review it to ensure Academic Authenticity guidelines are met before submitting the file for evaluation. See Understanding Similarity Reports for more information.

Grammarly Note:

Professional Communication will be automatically assessed through Grammarly for Education in most performance assessments before a student submits work for evaluation. Students are strongly encouraged to review the Grammarly for Education feedback prior to submitting work for evaluation, as the overall submission will not pass without this aspect passing. See Use Grammarly for Education Effectively for more information.

Microsoft Files Note:
Write your paper in Microsoft Word (.doc or .docx) unless another Microsoft product, or pdf, is specified in the task directions. Tasks may not be submitted as cloud links, such as links to Google Docs, Google Slides, OneDrive, etc. All supporting documentation, such as screenshots and proof of experience, should be collected in a pdf file and submitted separately from the main file. For more information, please see Computer System and Technology Requirements.

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.

Perform the following tasks to combine an existing dataset with 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. Justify how your question provided in part A will be answered.
   2. Identify which data from **both** the original dataset and the add-on CSV file are needed to answer the question provided in part A, 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 1:1, 1:M, or M:N relationships and relational constraints.
   1. Describe the relationship between the existing table and the table created in part B and discuss any issues with the relationship in the ERD.
   2. Write a SQL statement, in text format, that creates a table for the add-on CSV based on the ERD and specifies the columns and relevant keys.
   3. Write a SQL statement, in text format, that loads the data from **one** of the add-on CSV files into the table created in part B.

   Note: Do not include SQL statements as a screenshot.

C. Write **one or more** SQL statements in text format for a query or queries that answer the question provided in part A.

   Note: Do not include SQL statements as a screenshot.

   1. Provide data 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 activities and the question from part A.
   1. Explain why the time period identified in part D is relevant to the specific business activities.

E. Provide a Panopto video recording that can be used in a preproduction code review that includes **each** of the following:
   - You as the presenter on the screen

Note: The audiovisual recording should feature you visibly presenting the material (i.e., not in voiceover or embedded video) and should simultaneously capture both you and the LOD.

- An introduction of yourself by stating your name and student ID number
- A demonstration of each of the following:
  - **All** code used
  - The code being executed
  - The results of **all** code used in the task
- A vocalized description of **all** programs used to complete the task

Note: For instructions on how to access and use Panopto, use the "Panopto How-To Videos" web link. 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 "Data Analytics TGMx | D205 (Student Creators)." 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 **two** web sources for any code used to support the application. Be sure the web sources are reliable.

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

Note: No student input is required for part G. This is for evaluator review and comments.

File Restrictions

File name may contain only letters, numbers, spaces, and these symbols: ! - _ . * ' ( )
File size limit: 200 MB
File types allowed: doc, docx, rtf, xls, xlsx, ppt, pptx, odt, pdf, csv, txt, qt, mov, mpg, avi, mp3, wav, mp4, wma, flv, asf, mpeg, wmv, m4v, svg, tif, tiff, jpeg, jpg, gif, png, zip, rar, tar, 7z

RUBRIC

A:QUESTION

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
A question is not provided.	The question cannot be answered using SQL to acquire data from *both* the original database and the add-on CSV file data or does not require data from *both* the original database and the add-on CSV data, or the question is illogical.	The question is logical and requires using SQL to acquire data from the original database and the add-on CSV file data.

A1:QUESTION JUSTIFICATION