TGM2 — TGM2 TASK 1: DATA ACQUISITION

DATA ACQUISITION — D205

PRFA — TGM2

TASK OVERVIEW   SUBMISSIONS   EVALUATION REPORT

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 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.

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, 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:IDENTIFYING DATA

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not identify which data from both the original data set and the created table from the CSV file are needed to answer the question.	The submission identifies which data from both the original data set and the created table from the CSV file are needed, but the identified data do not provide all the tables, columns, and data types needed to answer the question.	The submission identifies which data from both the original data set and the created table from the CSV file are needed including all the tables, columns, and data types needed to answer the question.

B:ENTITY RELATIONSHIP DIAGRAM

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
An ERD for the database is not provided.	The ERD for the database does not identify m:n relationships and relational constraints.	The ERD for the database that was created identifies m:n relationships and relational constraints.

B1:CODE FOR THE ERD

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
A SQL code is not provided.	The SQL code does not create a table based on the ERD, or it does not correctly specify columns and relevant keys, or it does not include the actual code in text format.	The SQL code creates a table based on the ERD that specifies the columns and relevant keys and includes the actual code in text format.

B2:LOADING CSV DATA

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
SQL code that loads the data from one of the add-on CSV files into the table created in part B1 is not provided.	The SQL code does not correctly load the data from one of the add-on CSV files into the table created in part B1 or it does not include the actual code in text format.	The SQL code correctly loads the data from one of the add-on CSV files into the table created in part B1 and includes the actual code in text format.

C:SQL QUERY

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
A SQL statement or statements for a query or queries that answer the question from part A is not provided.	The SQL statement or statements for a query or queries do not accurately answer the question from part A, or the SQL statement is not in text format.	The SQL statement or statements for a query or queries accurately answer the question from part A, and the SQL statement is in text format.

C1:CSV FILES

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The data file or files are not provided.		The data file or files contain accurate results from the query or

The data file or files do not contain accurate results from the query or queries.

queries.

D:ADD-ON FILE

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not identify the specific time period for how often the add-on file should be acquired and refreshed in the database.	The submission does not accurately and logically 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.	The submission accurately and logically identifies 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.

D1:EXPLANATION OF TIME PERIOD

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not include an explanation of why the time period identified in part D is relevant to the business needs.	The submission does not accurately and logically explain why the time period identified in part D is relevant to the business needs.	The submission accurately and logically explains why the time period identified in part D is relevant to the business needs.

E:PANOPTO VIDEO OF CODE

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
A Panopto video recording of the code used is not provided, or the link provided for the video is not functional.	The Panopto video recording is provided, but a full demonstration of the code used, the code being executed, or the results of the code used in the task is not provided, or the video does not capture both the presenter and the vocalized demonstration.	The Panopto video recording includes a full demonstration of the code used, the code being executed, and the results of the code used in the task. For the duration of the presentation, the video captures both the presenter and the vocalized demonstration.

E1:PANOPTO VIDEO OF PROGRAMS

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT

A Panopto video recording of the programs used is not provided.

The Panopto video recording is provided, but a complete description of the programs used to complete the task is not provided, or the video does not capture both the presenter and the vocalized presentation describing the programs used to complete the task.

The Panopto video recording includes a complete description of the programs used to complete the task. For the duration of the presentation, the video captures both the presenter and the vocalized presentation describing the programs used to complete the task.

F:WEB SOURCES

NOT EVIDENT

A record of the web sources used to acquire data or segments of third-party code to support the application is not provided.

APPROACHING COMPETENCE

The record of the web sources used to acquire data or segments of third-party code to support the application is incomplete or inaccurate. Or the web sources cited are not reliable.

COMPETENT

The record of the web sources used to acquire data or segments of third-party code to support the application is both complete and accurate, and the web sources cited are reliable. Or the candidate stated that they did not use any web sources to acquire data or segments of third-party code.

G:SOURCES

NOT EVIDENT

The submission does not include both in-text citations and a reference list for sources that are quoted, paraphrased, or summarized.

APPROACHING COMPETENCE

The submission includes in-text citations for sources that are quoted, paraphrased, or summarized and a reference list; however, the citations or reference list is incomplete or inaccurate.

COMPETENT

The submission includes in-text citations for sources that are properly quoted, paraphrased, or summarized and a reference list that accurately identifies the author, date, title, and source location as available. Or the candidate stated that they did not use any sources to acquire data or segments of third-party code.

H:PROFESSIONAL COMMUNICATION

NOT EVIDENT

Content is unstructured, is disjointed, or contains pervasive errors in mechanics, usage, or

APPROACHING COMPETENCE

Content is poorly organized, is difficult to follow, or contains

COMPETENT

Content reflects attention to detail, is organized, and focuses on the main ideas as prescribed in

grammar. Vocabulary or tone is unprofessional or distracts from the topic.

errors in mechanics, usage, or grammar that cause confusion. Terminology is misused or ineffective.

the task or chosen by the candidate. Terminology is pertinent, is used correctly, and effectively conveys the intended meaning. Mechanics, usage, and grammar promote accurate interpretation and understanding.

WEB LINKS

Churn Data & Medical Data/Dictionary Data Sets

Labs On Demand

Panopto Access
Sign in using the "WGU" option. If prompted, log in with your WGU student portal credentials, which should forward you to Panopto's website. If you have any problems accessing Panopto, please contact Assessment Services at assessmentservices@wgu.edu. It may take up to two business days to receive your WGU Panopto recording permissions once you have begun the course.

Panopto How-To Videos

Panopto FAQs

Skillable Labs Knowledge Base Article
Please consult this WGU Knowledge Base article for general FAQs regarding your Skillable lab environment.