SLM2 — SLM2 TASK 1: DATA ANALYSIS

ADVANCED DATA ACQUISITION — D211
PRFA — SLM2

COMPETENCIES

4034.4.1 : Advanced SQL Operations

The graduate applies advanced SQL operations to integrate multiple data sources.

4034.4.2 : Explore Data

The graduate explores data acquisition.

INTRODUCTION

The data analyst's job does not end once the data has been analyzed. A vital skill for data analysts is to represent and report the data to stakeholders. In this task, you will demonstrate your ability to identify actionable insights from data and communicate them using effective storytelling methods.

This task consists of three parts:

In part 1, you will use SQL and a business intelligence tool to create data dashboards to support executive decision-making. Your dashboards should enable leaders to answer a specific research question on an organizational need. It is recommended, but not required, that the business intelligence tool be Tableau.

Your dashboard will integrate data from two sources:

1. Provided dataset: You will select one of the datasets, either churn or medical_data, which are already loaded into PostgreSQL and accessible with pgAdmin4 or Tableau in the Labs on Demand (LOD). The Web Links section at the bottom of the task page has the link for accessing LOD.

   Note: It is recommended, but not required, that you choose the same dataset, either churn or medical_data, that you used for your performance assessments in previous courses. A scenario for **each** dataset is below.

2. Additional dataset: You will identify an external public dataset as an additional dataset for your dashboard. This dataset should contain data and variables that complement the dataset you chose from the provided options and should enhance the insights you can gain from the provided dataset. Recommended sources include U.S. census data, Kaggle, or other public data repositories. You are responsible for ensuring that you have the rights to use the dataset.

In part 2, you will give a simulated presentation to a panel of peers. You will describe the use of SQL in preparing the data from the two datasets for the analysis and then demonstrate the functionality of the SQL scripts and other codes that supported the creation of the dashboards.

In part 3, you will write a reflection paper to outline the organizational need and summarize the findings from the analysis.

SCENARIO

Scenario: Telecommunications Churn Data

In the telecommunications industry, customers can choose from multiple service providers and actively switch from one provider to another. Customer "churn" is defined as the percentage of customers who stopped using a provider's product or service during a certain time frame. In this highly competitive market, some telecommunications industries can experience average annual churn rates as high as 25 percent. Given that it costs 10 times more to acquire a new customer than to retain an existing one, customer retention has now become even more important than customer acquisition. For many providers, retaining highly profitable customers is the number one business goal. To reduce customer churn, telecommunications companies need to predict which customers are at high risk of churn. You are an analyst on a team of analysts in a popular telecommunications company, which serves customers in all regions of the United States. You have been asked to analyze the dataset to explore the data, identify trends, and compare key metrics.

Scenario: Medical Readmission Data

In the medical industry, readmission of patients is such a problem that an external organization, Centers for Medicare and Medicaid Services or CMS, penalizes hospitals for excessive readmissions. When it comes to readmission penalties, studies show that many hospitals are overconfident and underprepared. The percentage of hospitals penalized for readmissions has increased each year since CMS began imposing penalties, and, according to the CMS reporting, as much as 78 percent of hospitals were fined in fiscal year 2015. However, three-quarters of hospitals feel confident in their ability to reduce readmissions, and only 55 percent of them anticipate receiving a penalty this year. Given the historical trend and the addition of COPD and hip and knee replacement to the list of medical conditions measured, the percentage of hospitals penalized will likely be much higher than 55 percent. Additionally, although hospitals are applying various reduction strategies, fewer than 1 in 5 utilize technology that is specific to reducing their readmissions, so they may not be doing all that they can. You are an analyst on a team of analysts for a popular medical hospital chain with patients in almost every state in the United States. You have been asked to investigate the extent to which readmission is a problem for this chain of hospitals.

Note: The original reason for hospitalization is not provided in the data. The purpose of the analysis is to predict readmission based on other conditions and factors of the patient.

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.

Part 1: Data Dashboards

A. Provide a copy of your dashboard that supports executive decision-making.
   1. Provide *both* datasets that serve as the data source for the dashboard.
   2. Provide step-by-step instructions on how to have the dashboard appear fully operational on a default version in Labs on Demand.

      Note: Online platforms such as Tableau Public are not permissible.

   3. Provide step-by-step instructions to help users navigate the dashboard.
   4. Provide *all* SQL code or other code supporting the dashboard in text format.

Part 2: Demonstration

B. Provide a link to a Panopto multimedia presentation in which you present the dashboards to an audience of data analytics peers. You should do *all* of the following in your presentation:
   1. Describe the technical environment used to create the dashboard.
   2. Demonstrate the functionality of the dashboard.
   3. Explain the SQL scripts used to support the creation of the dashboard.
   4. Explain how the data streams were prepared to support the analysis.
   5. Describe how data were aligned with other data points.
   6. Demonstrate how the database was created.
   7. Explain how referential integrity was enforced in the database.

   Note: The audiovisual recording should feature you presenting the material on screen (i.e., not in voice-over or embedded video) and should simultaneously capture both you and your multimedia presentation.

   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, Advanced Data Acquisition SLMx | D211 (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.

Part 3: Report

C. Write a report to outline the exploration of the data, the use of advanced SQL operations, and the analysis of the data. Do the following as part of your report:
1. Explain how the purpose and function of your dashboard aligns with the needs of the stakeholders for your chosen dataset.
2. Justify the selection of the business intelligence tool you used.
3. Explain the steps used to clean and prepare the data for the analysis.
4. Summarize the steps used to create the dashboard.
5. Discuss the results of your data analysis and how it supported the purpose and function of your dashboard.
6. Discuss the limitation(s) of your data analysis.

D. List the web source(s) used to acquire data or segments of third-party code to support the application. Ensure the web sources are reliable.

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

F. 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, 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:DATA DASHBOARDS

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
A copy of the data dashboards is not provided.	The data dashboard is not operational or does not support executive decision-making.	The data dashboard is operational and supports executive decision-making.

A1:DATASETS AND DASHBOARD FILE

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT