**Performance Assessment**

SLM1 — TASK 1: DATA ANALYSIS

Bader Ale

Department of Information Technology, Western Governors University

D211 Advanced Data Acquisition

September 2024

Table of Contents

[Part I: Purpose and Function 3](#_Toc176937312)

[Part II: Business Intelligence Tool 3](#_Toc176937313)

[Part III: Data Cleaning Steps 3](#_Toc176937314)

[Part IV: Dashboard Creation Steps 4](#_Toc176937315)

[Part IV(a): Opening a Packaged Workbook File 14](#_Toc176937316)

[Part V: SQL Scripts 14](#_Toc176937317)

[Part VI: Data Streams 16](#_Toc176937318)

[Part VII: Data Points 16](#_Toc176937319)

[Part VIII: Database Creation 17](#_Toc176937320)

[Part IX:Referential Integrity 17](#_Toc176937321)

[Part X: Results 17](#_Toc176937322)

[Part XI: Limitations 18](#_Toc176937323)

# Part I: Purpose and Function

The purpose of the dashboard is to explore the readmission rates for the hospitals and try to uncover the root cause for said readmissions. For this assessment, both internal datasets and a publicly available dataset from the US. Census Bureau was used. Moreover, the datasets were joined, and a tableau dashboard was created to facilitate insights into the readmission rates.   
Our stakeholders encompass key figures within the organization, including the Senior Vice President of Hospital Operations (SVP), Vice President of Research (VP), and a Panel of Regional Vice Presidents (Regional VPs). By thoroughly analyzing the data and identifying the root cause, we can make informed decisions to reduce readmissions and potentially alleviate the financial burdens associated with them.

# Part II: Business Intelligence Tool

The business intelligence tool used was Tableau. Tableau serves as a valuable tool for presenting data insights to stakeholders due to its user-friendly interface, interactive dashboards, and dynamic visualizations, which make complex information easily understandable and accessible. Stakeholders can explore data from various angles in real-time, customize visualizations to their preferences, and collaborate effectively with others. Making the raw data in a format that is easily digestible for stakeholders is key to transforming data into actionable results (Leung, 2021).

# Part III: Data Cleaning Steps

In order to prepare the dataset for the analysis, some data cleaning steps needed to be performed. Firstly, detecting and treating null values is a key step before transferring creating the Tableau dashboard. This ensures that when the visualizations are made a complete picture can be presented to the stakeholders and no NULL values shown. Furthermore, treating duplicated values is also a step paramount to data analysis. If there exists duplicated values in the dataset then the information would be incorrect again depicting erroneous information that could be used by stakeholders for decision making. Several modules within the SciKit learn python package, such as the *sklearn.preprocessing* allow for such transformation of the dataset before finalization (scikit-learn developers, 2024).

# Part IV: Dashboard Creation Steps

In order to create the dashboard, several steps needed to be taken. Firstly, the appropriate database in PostgreSQL was created along with all the tables necessary. The tables created on the local server are shown in the picture below from pgAdmin:

A screenshot of a computer

Description automatically generatedAs shown in the figure above, the appropriate tables were created, namely *admission*, *complication*, *job*, *location*, *patient*, *servicesaddon* and *survey\_responses\_addon*. Next, the links were created between the tables as appropriate to produce the correct data connections. An entity relationship diagram (ERD) was then generated to verify these connections. A screenshot is A screenshot of a computer

Description automatically generatedshown below.

An external dataset was also used for the Tableau dashboard – census population dataset per state from the US Census Department was downloaded and saved for later use. This contained the latest population statistics for every state.

The first step in creating the dashboard is to import the necessary data sources. From the initial welcome screen, select “Microsoft Excel” as the first data source in order to import the population by state file from the US Census Department.

A screenshot of a computer

Description automatically generated

Next, to add the PostgresSQL tables, add another data source but this time select the “PostgresSQL” option under “To a Server”. If it is not shown, click the “More” button to expand the options and choose there.

A screenshot of a computer

Description automatically generated

In order to connect to the Postgres server on the LabsOnDemand environment, certain information must be entered to that Tableau can communicate with the local server. Using the information below, fill in the information in the Tableau dialog box in order to connect to the server. Note, for this dashboard, the Databaseis *medical\_data* and not *churn.*

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

After successfully connecting the Postgres local server and uploading the Microsoft Excel populationbyState file, the connections should be similar to what is shown below.

A screenshot of a computer

Description automatically generated

After connecting the data sources, the patient and location tables should be dragged to the center of the Tableau screen in order to create the custom join. The *patient* table was joined with the *location* table on the *location\_id* column using the Custom SQL Query command from within Tableau – the SQL query can be found in the text file called “Custom SQL Code”. Right clicking either patient or location shows a dropdown box where “Convert to Custom SQL” is shown. Copy and paste the code from the text file into the box that shows in order to create a join between the patient table and the location. The join will be created using the *location\_id* column that is found in both patient and location.

A screenshot of a computer

Description automatically generated

After the code is executed, one can proceed to the creation of the various sheets to finally make the dashboard. The first sheet will contain the number of readmissions versus the population of each state. In the upper section, place the readmissions and population vales by dragging from the panel on the left to the appropriate locations, as shown below. Convert the measures of these to “COUNT” and “SUM” respectively as shown in the image below. Place State in the Columns section.



Moreover, change the Marks section to display the variable changes by color for both as shown below.

A screenshot of a computer

Description automatically generated

The second sheet, named Readmission by State is created using the map feature of Tableau. On the left side, search for the latitude and longitude variable names and place them as shown below. This will create a map of all the states where there is data.

A green rectangle with white text

Description automatically generated

A map of the united states

Description automatically generated

On the left side “Marks” panel, the variable readmission should be placed three times along with State – the first instance shows colors based on the variable quantity (dark blue to light blue. The second instance of readmission is a calculation of the quantity of readmission for that state based on a percentage of the total number. This can be achieved by clicking the arrow next to the variable name and selecting Quick Table Calculation and selecting Percentage of Total.

A screenshot of a computer

Description automatically generated

The third worksheet to create before the dashboard is the Total Charges by State – only the Total Charges and State variable will be used. Add State to the Columns section and Total Charges to the Rows. Total Charges should be changed to an average calculation by right clicking and selecting “Average” under ‘Measure’ as shown below.

A screenshot of a graph

Description automatically generated

The final step is creating the actual dashboard using the previously created worksheets. On the bottom of the Tableau window there are three icons to create new features – choose the icon with the four quadrants, as shown below, to open a new dashboard.

A screenshot of a computer

Description automatically generated

The dashboard screen will look like the image below. The created worksheets will appear on the left panel. These can be dragged and dropped anywhere on the white space in order to create the dashboard with the existing worksheets. On the left side pane, simply click and drag the worksheet onto the white space and Tableau will place the worksheet. By dragging the mouse over the workspace, a gray section will appear indicating the potential layout of the selected worksheet. For this dashboard, the sheet with the map was selected first followed by Total Charges by State on the right and finally the Population versus Readmission by State on the lower portion of the dashboard.

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

Finally, export the dashboard and all the sheets as a “Packaged Workbook” – click on File on the upper menu and scroll down to “Export Packaged Workbook”.

A screenshot of a computer

Description automatically generated

This query retrieved all the information from the patient table, and also provide me with the zip code, city, state, and county for each patient by matching each patient's location identifier with the location identifier in the location table. The patient table had alias ‘p’ and the location table had alias ‘l’. An inner join was created with the location table on *location\_id.* This created a new table that was used in the creation of the dashboard.

In order to create the dashboard, both the census dataset and the newly created joined table was used. This allowed for population analysis as well as with the initial medical dataset. The dashboard created showed population vs readmissions per state as shown below.

In order to open the dashboard, double click the Tableau Packaged Workbook (.twbx) as it contains the workbook and all dependencies. Upon opening, it will ask for a user name and password for the database. The following information is needed:

*Server: localhost  
 Port: 5432*

*Database: medical\_data*

*Username: postgres*

*Password: Passw0rd! (zero)*

After this, the dashboard can be found by click the button as shown below:



Navigating the dashboard is simple – by click on the state in the map on the left, all the other pertinent information about that state will be updated on the other sections. For example, clicking on Florida on the map will update the bar graph on the right that shows the average total charges as well as updating the section on the bottom showing the population and readmissions for Florida.

A screenshot of a computer

Description automatically generated

## Part IV(a): Opening a Packaged Workbook File

In order to open the packaged work Tableau file, the following are the steps to be followed: Firstly, ensure that your system has Tableau Desktop installed. You may download it from Tableau's official website if you don't already have it. Follow the instructions to completely install Tableau. Once installed, open Tableau and either select "Open" from the start-up screen or via File > Open menu option. Now find D211 Advanced Data Acquisition.twbx (Tableau Packaged Workbook) in local drive of your computer. Click on "Open" to open out workbook and Tableau will automatically load workbook with required sources including populationByState.xls which is Excel file provided. If it asks to verify connecting to data, make sure that the connection is proper under Data → Data Connections. You are able to view a dashboard via an appropriate tab at the bottom portion of the interface and then you will see a functional dashboard ready to use in Tableau.

# Part V: SQL Scripts

The SQL scripts played an important role in bringing the data for the dashboard. Specifically, a custom SQL query was used to join the patient and location tables in the database by matching the location\_id field in both. This allowed to pull patient information and link it to their geographic location (like zip code, city, and state). Once this join was done, a complete table that could be used in the dashboard to display where patients were from and how those locations tied into their hospital readmission data. This process made sure that only the relevant information was pulled into Tableau, which helped the dashboard present the right insights without any unnecessary or mismatched data. The SQL script is shown below.

*SELECT*

*p.\*,*

*l.zip,*

*l.city,*

*l.state,*

*l.county*

*FROM*

*patient p*

*INNER JOIN*

*location l*

*ON*

*p.location\_id = l.location\_id*

This SQL query retrieves a combined set of data from two tables: patient and location. It performs the following operations:

SELECT p.\*, l.zip, l.city, l.state, l.county: This line is specifying the columns to be selected. p.\* means all columns from the patient table, and l.zip, l.city, l.state, l.county are specific columns selected from the location table.

FROM patient p: This line indicates that the data is being selected from the patient table, and it is giving this table an alias p that is used to refer to it in other parts of the query.

INNER JOIN location l: This tells the SQL engine to perform an inner join operation with the location table, which is given an alias l. An inner join only includes rows from both tables that have matching values in the columns being joined on.

ON p.location\_id = l.location\_id: This line specifies the condition for the join. It tells the database to match rows from the patient table with rows from the location table where the location\_id column in the patient table is equal to the location\_id column in the location table.

# Part VI: Data Streams

For this project, two main data streams were used. The first was the hospital’s internal data, which included patient records and their locations, and the second was an external dataset from the U.S. Census Bureau, providing population data by state. Before combining the two datasets, data cleaning had to be performed—removing empty values, fixing duplicates, and making sure everything was consistent. The internal data was processed using Python, while the external data came as an Excel file. Once both were cleaned, Tableau was used to link the patient data with state population figures, so correct analysis on how hospital readmission rates compared to the population in each state. This gave a clear picture of which states were experiencing higher readmission rates and where focus and resources allocation was needed.

# Part VII: Data Points

The alignment of data points was crucial in making sure the dashboard presented the right insights. Combining the hospital’s readmission rates with population data from the U.S. Census Bureau was key. To do this, a custom SQL query in Tableau was used to match the location\_id from the patient data with the corresponding state data from the population dataset. This ensured that every readmission data point was linked to the correct state, allowing visualization on how population sizes compared with the readmission rates. This alignment helped identify any states that stood out, especially where readmissions were disproportionately high relative to the population size, which pointed to areas that needed more attention or resources.

# Part VIII: Database Creation

The database that supported the project was built in PostgreSQL. Several tables were created to hold all the necessary data, such as patient records and location details. Using pgAdmin, tables like admission, complication, patient, location, and survey\_responses\_addon were created. Each table was linked using foreign key relationships to ensure all the data was properly connected. For instance, the location\_id in the patient table was linked to the location table, so a join could be easily created later in Tableau. This setup allowed management of the data efficiently and made it easy to pull the necessary information into Tableau for analysis.

# Part IX:Referential Integrity

Maintaining referential integrity in the database was key to ensuring the data relationships stayed consistent. In order to do this, a foreign key constraints key was created, especially between the patient and location tables. This meant that every patient record in the patient table was tied to a valid entry in the location table. So, for example, no patient could be listed without a valid geographic location. This also ensured that any changes in one table wouldn’t break the connections between the tables. These steps helped guarantee that the data in the dashboard was accurate and that the relationships between the data points were consistent.

# Part X: Results

The provided Tableau dashboard serves as an analytical tool for visualizing the relationship between hospital readmission rates and the population by state, presenting health care executives with actionable data. The map highlights the geographic spread of readmissions, while the bar chart allows for an at-a-glance comparison of these rates in relation to state populations, facilitating the identification of any disproportionate figures. Through its interactive capabilities, the dashboard permits a granular examination of the data, enabling the leadership to focus efforts on improving patient outcomes and pinpointing where to allocate resources effectively.

The dashboard is designed to meet the needs of hospital stakeholders by offering a clear view of readmission rates, total charges, and population data across various states. By visualizing key information such as the number of readmissions and the financial impact in each state, it helps hospital administrators and financial officers identify areas with higher readmissions and costs. This insight allows them to prioritize their efforts and resources to reduce readmissions and avoid penalties from CMS.

Moreover, the dashboard helps stakeholders make informed decisions by showing how each state compares in terms of readmission rates relative to its population. This makes it easier to tailor strategies for different regions and allocate resources where they are most needed, leading to better patient outcomes and ensuring compliance with CMS guidelines.

Such a resource is pivotal for detecting regional issues swiftly and monitoring the impact of health interventions over time. Hospital administrators can leverage this information to compare current performance to established benchmarks and other pertinent health care indicators. This empowers them to make well-informed choices about enhancing care quality and managing hospital resources adeptly, fostering an initiative-taking stance in health care administration. The data visualization provided by the dashboard lays the groundwork for informed and responsive decision-making in the complex landscape of health care services.

# Part XI: Limitations

One limitation of this analysis and dashboard is its reliance on the accuracy and completeness of the data. If the data from hospitals or states is incomplete, outdated, or inaccurate, the insights generated may be misleading, potentially resulting in poor decision-making. Additionally, the dashboard focuses mainly on quantitative metrics like readmission rates and costs, without taking into account qualitative factors such as patient demographics, social determinants of health, or varying healthcare practices across states, which can all influence readmissions.

Another limitation is the lack of insight into the effectiveness of readmission reduction strategies already in place. The dashboard doesn’t account for any ongoing interventions or the reasons behind high readmission rates, which might be influenced by factors like the complexity of patient cases, hospital resources, or regional healthcare infrastructure. Without these contextual elements, the data alone might not provide a full picture, limiting the effectiveness of strategic decisions based solely on the dashboard.

Works Cited

Leung, K. (2021, September). *How Tableau Helps Your Organization Achieve Greater Data Insights*. Retrieved March 2024, from Datacamp: Rader Analytics: https://www.datacamp.com/blog/how-tableau-helps-your-organization-achieve-greater-data-insights

scikit-learn developers. (2024). *scikit Learn*. Retrieved 2024, from https://scikit-learn.org/stable/modules/preprocessing.html: https://scikit-learn.org/stable/modules/preprocessing.html