

D211: Advanced Data Acquisition

Department of Information Technology, Western Governors University

By Andrew Fagundes

Professor William Sewell

June 6th, 2025

Table of Contents

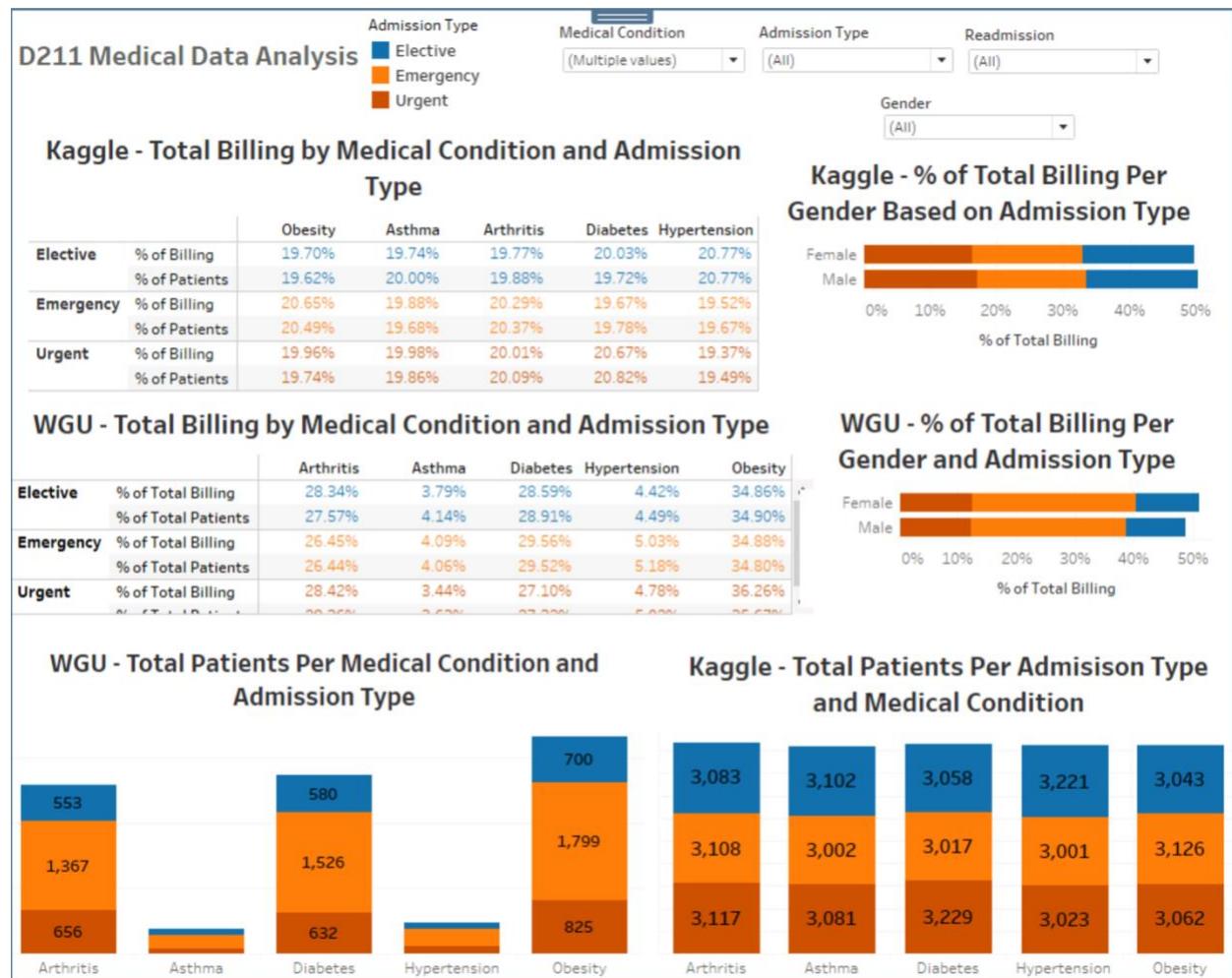
PART I: INTERACTIVE DATA DASHBOARDS	3
A1: DATA SETS FOR DASHBOARD	3
A2: STEP-BY-STEP INSTRUCTIONS FOR OPERATIONAL DASHBOARD IN LABS ON DEMAND (ONLINE PLATFORMS SUCH AS TABLEAU PUBLIC ARE NOT PERMISSIBLE).....	4
A3: USER INSTRUCTIONS FOR DASHBOARD NAVIGATION.....	6
A4. SQL CODE FOR DASHBOARD (PROVIDE SCREENSHOTS; THIS INCLUDES SQL STATEMENTS WITHIN YOUR VISUALIZATIONS)	8
PART II: PRESENTATION	11
PART III: REPORT	11
C1: PURPOSE AND FUNCTION OF DASHBOARDS (THERE IS NO DATA DICTIONARY, INSPECT MEDICAL DATA DATABASE AND ITS TABLES. HOW ARE THEY RELATED AMONG THEMSELVES AND THEN TO THE EXTERNAL DATASET THAT YOU ARE USING?).....	11
C2: JUSTIFICATION OF BUSINESS INTELLIGENCE TOOL USED	12
C3: STEPS USED TO CLEAN AND PREPARE DATA FOR ANALYSIS	13
C4: SUMMARY OF STEPS USED TO CREATE DASHBOARD.....	14
C5: RESULTS OF DATA ANALYSIS AND HOW IT SUPPORTS PURPOSE OF DASHBOARD (WILL HAVE TO REVIEW MY DASHBOARD TO SEE IF THERE IS DIFFERENT INSIGHTS)	15
C6: LIMITATIONS OF DATA ANALYSIS (REVIEW DATA SETS TO SEE IF THERE IS ANYTHING TO ADD)	17
REFERENCES (NO THIRD-PARTY CODE REFERENCES USED)	18

Part I: Interactive Data Dashboards

A1: Data Sets for Dashboard

See attached file “WGU 211 Medical Data Dashboard.twbx” or screenshot

below:



The Tableau dashboard for this assessment includes a Comparison Analysis performed on medical data. This report contains a side-by-side comparison using two different data sets, three different data representations and four interactive controls.

For this task, I used the WGU provided medical_clean data set, and a 3rd party data set obtained via Kaggle: Healthcare Data: Dummy Data with Multi Category Classification Problem

(Patil, 2024). Please see the attached CSV file of my data set titled:

Healthcare_Data_Treated_FiNaL_v3.csv for the 3rd party dataset used in this task. The WGU data set consisted of key patient demographics and services performed while admitted into a hospital. The Kaggle data set has been created to serve as a resource for data science, this data contains key patient demographics and services performed while admitted into a hospital. Data was chosen due to its similarity to the WGU-provided data.

A2: Step-by-Step Instructions for Operational Dashboard in Labs on Demand

The dashboard will be provided in a .twbx format, which can be opened by Tableau Reader and shared.

Please see the following steps to download, access, and open the relevant files for this course on Labs on Demand.

- If transferring to Labs on Demand, save the file from the uploaded submission contents into a shared location or email it yourself. Open this location in Labs on Demand and download it to the machine.
- Navigate to Labs on Demand using the WGU portal and open it up on your machine.
- If downloading from Labs on Demand, open a web browser inside Labs on Demand and log onto the WGU portal. Once inside, take the uploaded D211 files from the course submission and download it to the machine.
- Once the file is downloaded and saved, navigate to the files, and double click to being viewing the content. This will extract the public dataset, queries, the .twbx tableau file, and these instructions.

- Next, double-click on the pgAdmin icon to open it. For the Tableau to properly work on each machine, you must first get the publicly selected dataset loaded to pgAdmin. To do so, inside pgAdmin, navigate to servers, PostgreSQL, Databases, medical data, public, and Tables.
- Once there, click on Tables and select 'Query Tool', open the query file that was downloaded insert the 'Create Table' statement to the query tool, and hit the play 'run' button.
- After running this query, you can right-click on the server drop-down at the top and click 'refresh'; the created table should appear in the database now.
- Navigate back to the created table, right-click on the table name, and click 'Import/Export'. Now you are in the import data tool. Step one is to select 'import' at the top, then navigate to the Healthcare_Data_Treated_FiNaL_v3.csv file, which was saved within the downloads location. Select yes for the header under the miscellaneous section. And then click ok.
- Once you have received a 'success' pop-up, it means your data is loaded and good to go.
- Next, navigate to the .twbx file that was saved, open it, and click Yes on the pop-up that shows up to run the queries.
- There are queries already attached to the Tableau file however, you must connect them to the correct database for it to work. While in the workbook, select the 'data source' tab. Here, you will put in the server login information.
 - Server: localhost
 - Port: 5432

- Database: medical_data
 - Authentication: Username and Password
 - Username: postgres
 - Password: Passw0rd!
- The attached queries should now execute properly and allow you to access the data. You are now free to navigate through the dashboard tabs.

A3: User Instructions for Dashboard Navigation

Navigating the dashboard on the Tableau desktop is simple; the instructions provided above will allow you to open the dashboard and begin navigating its interactive controls. Due to the two data sets' similarities, the data inside the dashboard has been joined together. By doing so, users can view the same metrics between both data sets side by side with the filters working for both outputs.

On the Tableau desktop, there will be multiple tabs at the bottom. The key one to view is the dashboard, which will have a little square icon next to it. Each of the contents within the dashboard is created as its own 'tab' and then pulled together into containers on the dashboard. You can see each of these next to the dashboard tab.

The section at the top that looks like a header is where you will find the title, color legends, and variable filters. The dashboard is an analysis of medical data, therefore, the title 'D211 Medical Dataset Analysis' has been chosen. This indicates that the dashboard is showing a dispersion of medical patient data. Within this section, you get an overview of the two different data sets contained, one of which being the WGU-provided Medical Data, the other being a publicly posted Kaggle source. Next to this information, you can see the color legends selected

for the various variables that are used throughout the dashboard. Lastly, in this section, there are four filters: readmission type, gender, admission type, and medical condition. These filters allow users to interact with the dashboard and select a certain control group for viewing.

The dashboard is a top-down split comparison with the Web (Kaggle) data source on the top and the WGU (provided) data source on the bottom. The first set of charts shows you a medical condition patient count dispersion; the height of each bar on the bar chart indicates the percentage of total billing in each medical condition category per gender. An additional variable was dropped on top of these bar charts to show you how the total charge/patient count is impacted by the admission types. The admission types are 'Elective', 'Emergency', and 'Observation'. The legend for which color corresponds to which variable is in the header at the top.

Next to the bar graphs, you can see two text tables. These text tables illustrate the percentage of total billing and percentage of patients that exist for patients dependent on their medical conditions and admission types. This is color coded based on admission type. The color legend for admission type is located at the top of the dashboard.

The bottom portion of the dashboard shows total number of patients per admission type and medical within vertical stacked bar charts. This section guides users into a high-level view of numbers, allowing users to have a better grasp of how many patients are included within each dataset and what the count is in each category. This is also color coded and can be referred to the color legend at the top of the dashboard. This is used to understand the patients that are most at risk.

A4. SQL Code for Dashboard (provide screenshots; this includes SQL statements within your visualizations)


```

-- FOR KAGGLE DATA SET
SELECT
    "Gender", "Medical_Condition", "Admission_Type", SUM(CAST("Billing_Amount" AS numeric)) AS Total_Billing,
    COUNT("Name") AS Patient_Count
FROM public."Healthcare_data"
WHERE "Medical_Condition" IN ('Asthma', 'Diabetes', 'Obesity', 'Arthritis', 'Hypertension')
GROUP BY 1, 2, 3

```

Lab Custom SQL

```

-- FOR WGU DATA SET
SELECT pat."gender" AS gender,
    CASE WHEN serv."diabetes" = 'Yes' THEN 'Diabetes'
        WHEN serv."arthritis" = 'Yes' THEN 'Arthritis'
        WHEN serv."overweight" = 'Yes' THEN 'Obesity'
        WHEN serv."asthma" = 'Yes' THEN 'Asthma'
        WHEN pat."highblood" = 'Yes' THEN 'Hypertension'
    END AS Medical_Condition,
    CASE WHEN admis."initial_admission" = 'Emergency Admission' THEN 'Emergency'
        WHEN admis."initial_admission" = 'Observation Admission' THEN 'Urgent'
        WHEN admis."initial_admission" = 'Elective Admission' THEN 'Elective'
    END AS Admission_Type,
    pat."readmis" AS Readmission,
    SUM(pat."totalcharge") AS Total_Billing,
    COUNT(pat."patient_id") AS Patient_Count
FROM public."admission" AS admis

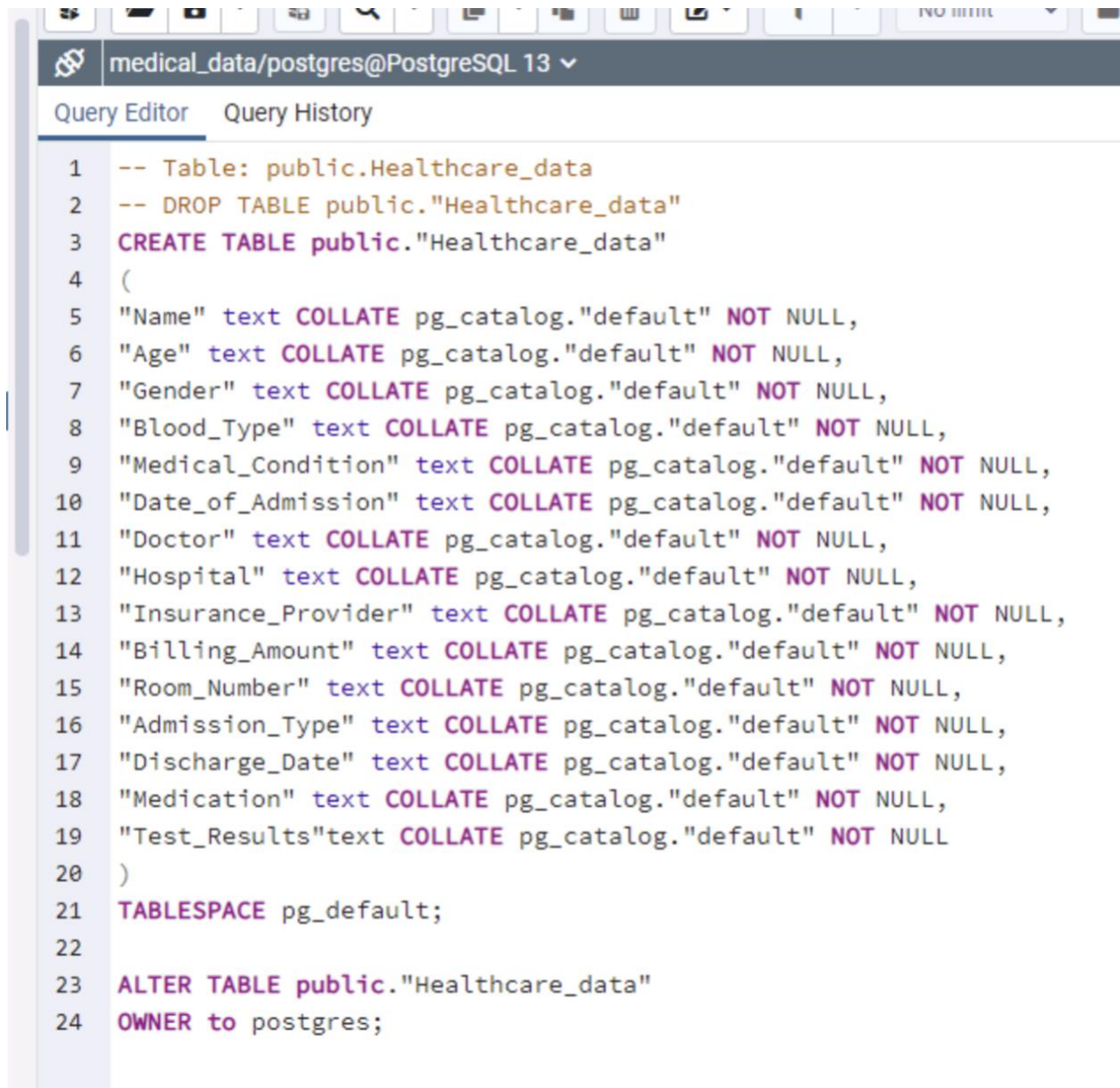
LEFT JOIN public."patient" AS pat
    ON admis."admis_id" = pat."admis_id"

LEFT JOIN public."servicesaddon" AS serv
    ON pat."patient_id" = serv."patient_id"

GROUP BY 1, 2, 3, 4

```

Lab Workspace



The screenshot shows a PostgreSQL Query Editor window with the title bar 'medical_data/postgres@PostgreSQL 13'. The window has two tabs: 'Query Editor' and 'Query History'. The 'Query Editor' tab is active, displaying a SQL script. The script is as follows:

```
1  -- Table: public.Healthcare_data
2  -- DROP TABLE public."Healthcare_data"
3  CREATE TABLE public."Healthcare_data"
4  (
5    "Name" text COLLATE pg_catalog."default" NOT NULL,
6    "Age" text COLLATE pg_catalog."default" NOT NULL,
7    "Gender" text COLLATE pg_catalog."default" NOT NULL,
8    "Blood_Type" text COLLATE pg_catalog."default" NOT NULL,
9    "Medical_Condition" text COLLATE pg_catalog."default" NOT NULL,
10   "Date_of_Admission" text COLLATE pg_catalog."default" NOT NULL,
11   "Doctor" text COLLATE pg_catalog."default" NOT NULL,
12   "Hospital" text COLLATE pg_catalog."default" NOT NULL,
13   "Insurance_Provider" text COLLATE pg_catalog."default" NOT NULL,
14   "Billing_Amount" text COLLATE pg_catalog."default" NOT NULL,
15   "Room_Number" text COLLATE pg_catalog."default" NOT NULL,
16   "Admission_Type" text COLLATE pg_catalog."default" NOT NULL,
17   "Discharge_Date" text COLLATE pg_catalog."default" NOT NULL,
18   "Medication" text COLLATE pg_catalog."default" NOT NULL,
19   "Test_Results" text COLLATE pg_catalog."default" NOT NULL
20 )
21 TABLESPACE pg_default;
22
23 ALTER TABLE public."Healthcare_data"
24 OWNER to postgres;
```

Part II: Presentation

The video recording for this assignment includes a vocal presentation of the dashboard being used, and a discussion of the charts created within it. The video recording for this project can be found inside the Panopto drop box titled “Master of Science, Advanced Data Acquisition SLMx | D211 (Student Creators) [assignments].”. Panopto video link:

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=85aaab83-782e-472c-94e3-b2f4013ea88c>.

Part III: Report

C1: Purpose and Function of Dashboards

The purpose and function of this dashboard align with the needs of stakeholders such as hospital administrators, policymakers, and financial officers by providing data-driven insights into patient readmissions, admission types, and associated billing percentages.

The dashboard categorizes admission types across various medical conditions, which helps stakeholders identify trends in readmissions. Since CMS penalizes hospitals for excessive readmissions, understanding which conditions have higher readmission rates helps mitigate financial penalties. By comparing different data sets, stakeholders can benchmark performance against industry trends.

First, the billing percentage breakdown for different admission types provides insights into the financial burden per condition. This helps financial teams forecast potential penalties and develop cost-saving strategies and allows decision-makers to allocate resources effectively, ensuring that high-risk conditions receive more preventive care interventions. Second, the stacked bar charts clearly show which medical conditions contribute the most to different

admission types. Conditions like diabetes and obesity have a higher likelihood of emergency and urgent admissions, signaling potential areas for targeted intervention. This enables hospitals to implement preventive care programs focused on high-risk conditions to reduce readmission rates. Lastly, the interactive filters (Readmission, Gender, Admission Type, and Medical Condition) allow stakeholders to drill down into specific groups. This helps hospital administrators assess whether certain demographics or conditions have disproportionately high readmission rates. Additionally, this can aid in developing customized patient care plans and deploying technological interventions that can reduce readmissions.

By analyzing trends and comparing multiple data sets, hospitals can create evidence-based strategies to align with CMS regulation and reduce financial penalties. This dashboard empowers stakeholders with actionable insights to tackle readmission challenges, financial risks, and operational inefficiencies. It provides real-time data visualization to pinpoint problem areas and helps the hospital chain optimize patient outcomes while reducing penalties from CMS.

C2: Justification of Business Intelligence Tool Used

Tableau is a powerful Business Intelligence (BI) tool that helps align analysts with stakeholders' needs due to its capabilities in data visualizations, interactivity, and ease of use. Tableau transforms raw data into interactive and visually appealing dashboards that are easy to interpret. Dashboards can be tailored to different audiences, ensuring that stakeholders only see the relevant KPIs for their needs.

Tableau allows you to connect to various data sources seamlessly. For this project, it was required to set up a connection to PG-Admin to Tableau due to the connection capabilities. This

made it the best program for this analysis. Tableau dashboards can be saved as a .twbx file and sent between people, making it easy to connect and make changes to dashboards.

Tableau is suited to address the complex, multi-faceted needs of hospital administrators, policymakers, and financial officers. Tableau aligns with stakeholder requirements by using integrated data-driven insights, customized visualizations for needs, and strategic alignment with regulatory and investment goals. Tableau can combine multiple data sets, in this case, WGU and Kaggle medical data sets were combined, this allows stakeholders to benchmark their hospital's performance against industry standards. Tableau's ability to offer dynamic visualizations allowed this dashboard to offer immediate insights into patient readmissions and billing percentages, ensuring that decision-makers always have current information available. With options to filter by readmission status, gender, admission type, and medical condition, the dashboard lets users drill down into specific segments. This level of detail supports targeted interventions and personalized care plans for stakeholders.

C3: Steps Used to Clean and Prepare Data for Analysis

To get the data in a usable format for this assignment, a few cleaning steps are required. First is loading the public data set to pgAdmin. Step-by-step instructions for this are provided as part of section A. Once the public data has been uploaded, it can be manipulated and queried.

There were different steps taken for the public data set vs. the WGU provided data set. For the public data set, a complete pull of the necessary fields, including a sum of total charge and a count of patients, was done. A filter was applied to the medical condition to pull only the same medical conditions that overlap with the WGU data set.

The WGU data provided required more manipulation to get it in the same format as the public data. This consisted of a few fields being pulled in directly and two case-when statements to get the admission types and medical conditions in a single-column format with the same naming conventions as the public data set. Additionally, two joins were required in this data due to it being split up within the pgAdmin server.

After being properly queried, these are posted in Tableau and executed there, so just the query output gets pulled into Tableau. Now that the data has been loaded, manipulated and pulled into the dashboard, we can continue to build out the dashboard for analysis.

C4: Summary of Steps Used to Create Dashboard

Step one is to prepare the data for import. The data was obtained from two sources: the WGU data was obtained from the WGU course materials, and the Kaggle data set was obtained from the Kaggle website. Once obtained, the data was loaded to PGAdmin, reviewed, queried, and loaded to Tableau.

Step two is to create key visualizations in Tableau. The visualizations consisted of a vertical stacked bar chart showing admission type by medical condition, a horizontal bar chart illustrating percentage of total billing based on gender and admission type, and text tables showing the admission type breakdown by condition. Bar charts are useful to compare values of a single measure or to rank its values from the top or the bottom (Loth 2019). Stacked bar charts are created by dragging the medical condition field to the columns shelf and the patients to the row shelf. It is then formatted appropriately and color-coded. The horizontal bar chart required the percentage of total billing to be located in the columns shelf and the gender variable located in rows. For text tables, the best way to do this is to double click the measures you want and then

adjust things in the columns/rows accordingly. Tableau will create a text chart for you, and you can adjust the color formatting to highlight differences if needed. Color and size formatting was done on each chart. Each of these tables was duplicated for the appropriate data set.

Step three is to add filters and interactivity. For this dashboard, specifically, filters were added for readmission, gender, admission type, and medical condition. This was done by adding the filters to one page. Once inserted into the dashboard, you can allow the filter to select every table on the dashboard. Additionally, the fields are interactive, allowing users to toggle categories.

Step four requires you to assemble the dashboard and format it. When doing so, it is easiest to drop vertical and horizontal containers into the layout and then drag and drop each table where you want it. Tableau requires some hacking and adjusting to get it perfect, this step usually requires the most time. A structured layout requires you to drag and arrange the bar charts and tables. Add label sections clearly, and use titles, annotations, and tooltips for additional details. Next, this dashboard was formatted and adjusted for readability, a consistent color scheme was selected, and bold headers, proper spacing, and alignment were all applied.

The last step in the dashboard preparation was to publish and share. For this assessment, it was not required to publish to Tableau Public, so a .twbx file along with instructions written in this assessment was provided as part of the assignment materials.

C5: Results of Data Analysis and How It Supports Purpose of Dashboard (will have to review my dashboard to see if there is different insights)

The results of the data analysis provide a clear understanding of patient admission trends, readmission risks, and financial implications, aligning with the dashboard's purpose—to help

stakeholders identify areas of concern, mitigate CMS penalties, and develop intervention strategies.

The results include key findings, relevance to stakeholders, and financial impact. A few key findings to highlight are that Emergency and Urgent admissions account for a significant portion of total hospital visits across all conditions, and Obesity and Diabetes have the highest percentage of Emergency and Urgent admissions, indicating that these conditions may contribute to higher readmission rates. Since CMS penalizes hospitals with excessive readmissions, hospitals need to focus on these high-risk conditions to minimize financial penalties. Additionally, the billing percentage closely mirrors the percentage of patient admissions, meaning that high readmission conditions also increase costs. Obesity and Diabetes require more expensive treatments, as seen in the billing distribution across admission types. These insights help hospital administrators optimize financial planning and justify investments in preventive care to reduce CMS penalties.

The data set comparison allows you to benchmark the findings. This indicates that the WGU data set shows higher emergency and urgent admissions for obesity and diabetes, while the Kaggle data set presents a more evenly distributed admission type breakdown. This suggests differences in hospital management strategies, patient demographics, or regional healthcare accessibility. Benchmarking against Kaggle helps identify best practices that may reduce readmission risks.

Lastly, the effectiveness of the features allows for ease of use between parties. Filters for Readmission, Gender, Admission Type, and Medical Condition allow users to explore specific patient groups. Stakeholders can quickly identify which patient categories are most at risk, making the dashboard a valuable tool for customized policymaking and intervention strategies.

C6: Limitations of Data Analysis

While the dashboard provides valuable insights, several limitations may impact the accuracy and applicability of the findings. One limitation is that the data set may not fully represent all hospitals within the chain or across different regions. Another limitation is that differences in data collection methods between the WGU and Kaggle data sets may lead to inconsistent insights. There may be key variables missing such as other medical conditions that could be found in WGU's medical data like back pain and stroke. Without these variables, it is challenging to directly correlate admission types with readmission rates. Lastly, the dashboard identifies trends but does not use machine learning models to predict future readmission likelihoods.

Despite the limitations, the dashboard effectively highlights trends in patient admissions, identifies high-risk conditions, and offers financial insights that are crucial for reducing CMS penalties and improving patient care. Future enhancements could include real-time data integration, predictive modeling, and more specific patient demographics to further refine its usefulness.

References (No third-party code references used)

Loth, A., Vogel, N., Sparkes, S. (2019). Visual Analytics with Tableau. John Wiley & Sons, Incorporated.

Patil, Prasad (2024). "Healthcare Dataset: Dummy data with Multi Category Classification Problem." Kaggle. <https://www.kaggle.com/datasets/prasad22/healthcare-dataset?resource=download>.