D211 – Advanced Data Acquisition

WGU M.S. Data Analytics

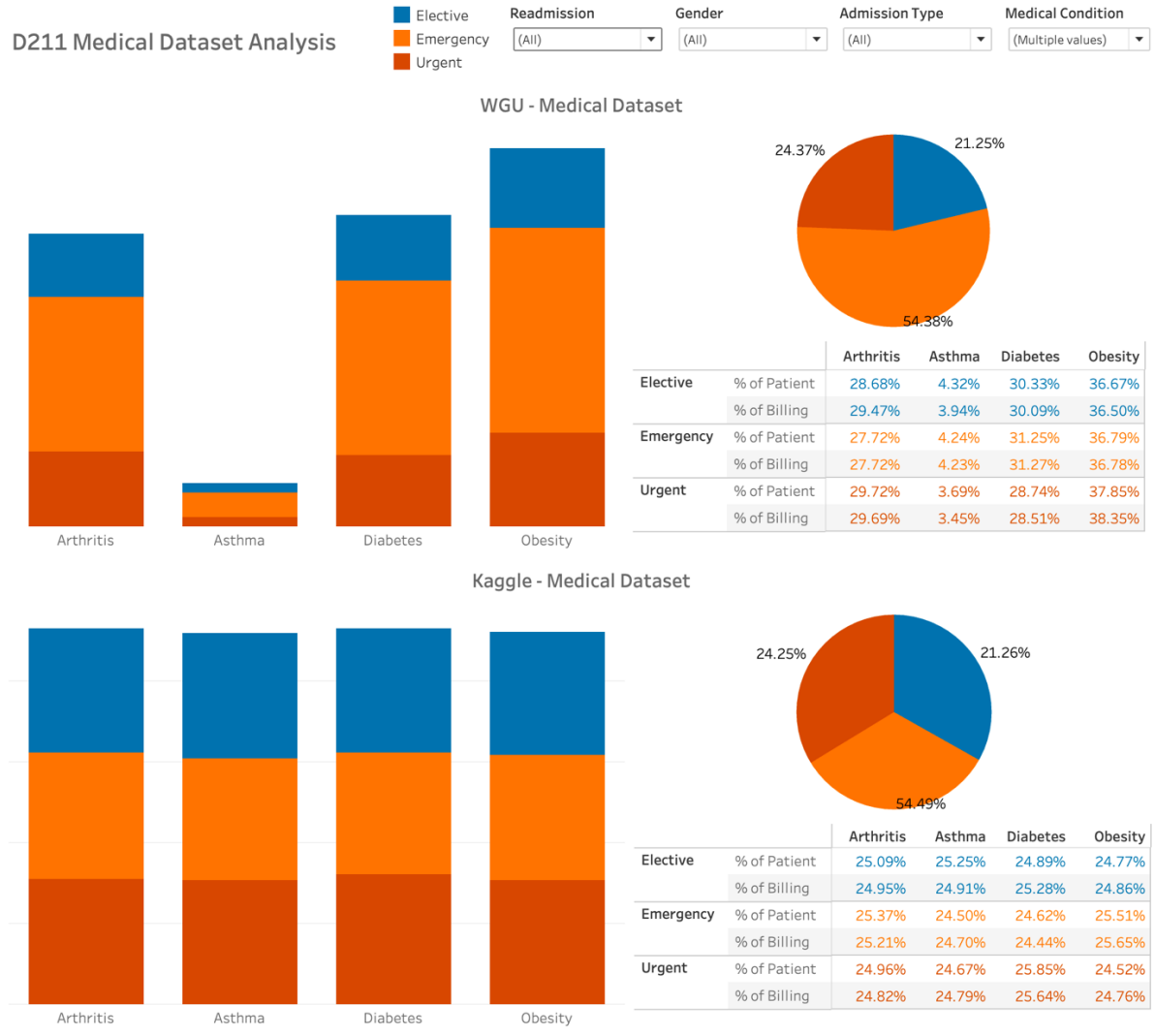
Lyssa Kline

February 20, 2025

A, Part 1: Data Dashboards

The Tableau dashboard for this assignment analyzes medical data. This report compares two datasets side-by-side using three different data representations and four interactive controls for each dataset.

The Tableau dashboard has been uploaded as part of the requirements of this program. See the screenshot of the dashboard below.



		Arthritis	Asthma	Diabetes	Obesity
Elective	% of Patient	28.68%	4.32%	30.33%	36.67%
	% of Billing	29.47%	3.94%	30.09%	36.50%
Emergency	% of Patient	27.72%	4.24%	31.25%	36.79%
	% of Billing	27.72%	4.23%	31.27%	36.78%
Urgent	% of Patient	29.72%	3.69%	28.74%	37.85%
	% of Billing	29.69%	3.45%	28.51%	38.35%

		Arthritis	Asthma	Diabetes	Obesity
Elective	% of Patient	25.09%	25.25%	24.89%	24.77%
	% of Billing	24.95%	24.91%	25.28%	24.86%
Emergency	% of Patient	25.37%	24.50%	24.62%	25.51%
	% of Billing	25.21%	24.70%	24.44%	25.65%
Urgent	% of Patient	24.96%	24.67%	25.85%	24.52%
	% of Billing	24.82%	24.79%	25.64%	24.76%

A1, Data Sources

Two datasets were used for this dashboard analysis. See below.

Dataset #1

- WGU Provided Medical Clean Data
- CSV File: WGU Medical Clean Data

- The dataset consisted of key patient demographics and services performed while patients were admitted to a hospital.

Dataset #2

- Kaggle Open-Source Public Medical Data
- CSV File: Kaggle Medical History Data
- Open-Source Link: Kaggle Dataset Link
- Data 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, Installation Instructions

The dashboard and these instructions are provided in a zip file. This file will allow the evaluators to easily transfer the dashboard from labs on demand to personal machines. The dashboard will be provided in a .twbx format, which can be opened by Tableau Reader and shared. Providing these instructions in a zip folder allows the readers to download the zip file onto labs on demand and have all the details in one spot.

To satisfy the requirements of this course, the mentioned documents will be submitted individually as well as together in a zip folder.

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

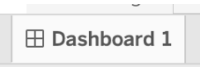
- If transferring to Labs on demand, save the file from the uploaded submission contents into a shared location or email it yourself. Open this location in labs on demand and download it to the machine.
- Navigate to Labs on demand using the WGU portal and open it up on your machine.
- If downloading from Labs on Demand, open a web browser inside Labs on Demand and log onto the WGU portal. Once inside, take the uploaded D211.zip file from the course submission and download it to the machine.
- Once the file is downloaded and saved, navigate to the folder, right-click, and select “Extract All” to extract the files to ‘C:\Users\Public\Downloads’. 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_dataset.csv file, which was saved within the zip file 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.

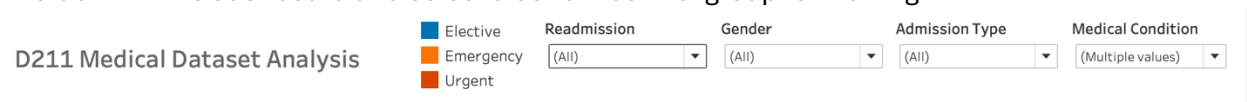
A3, Navigation Instructions

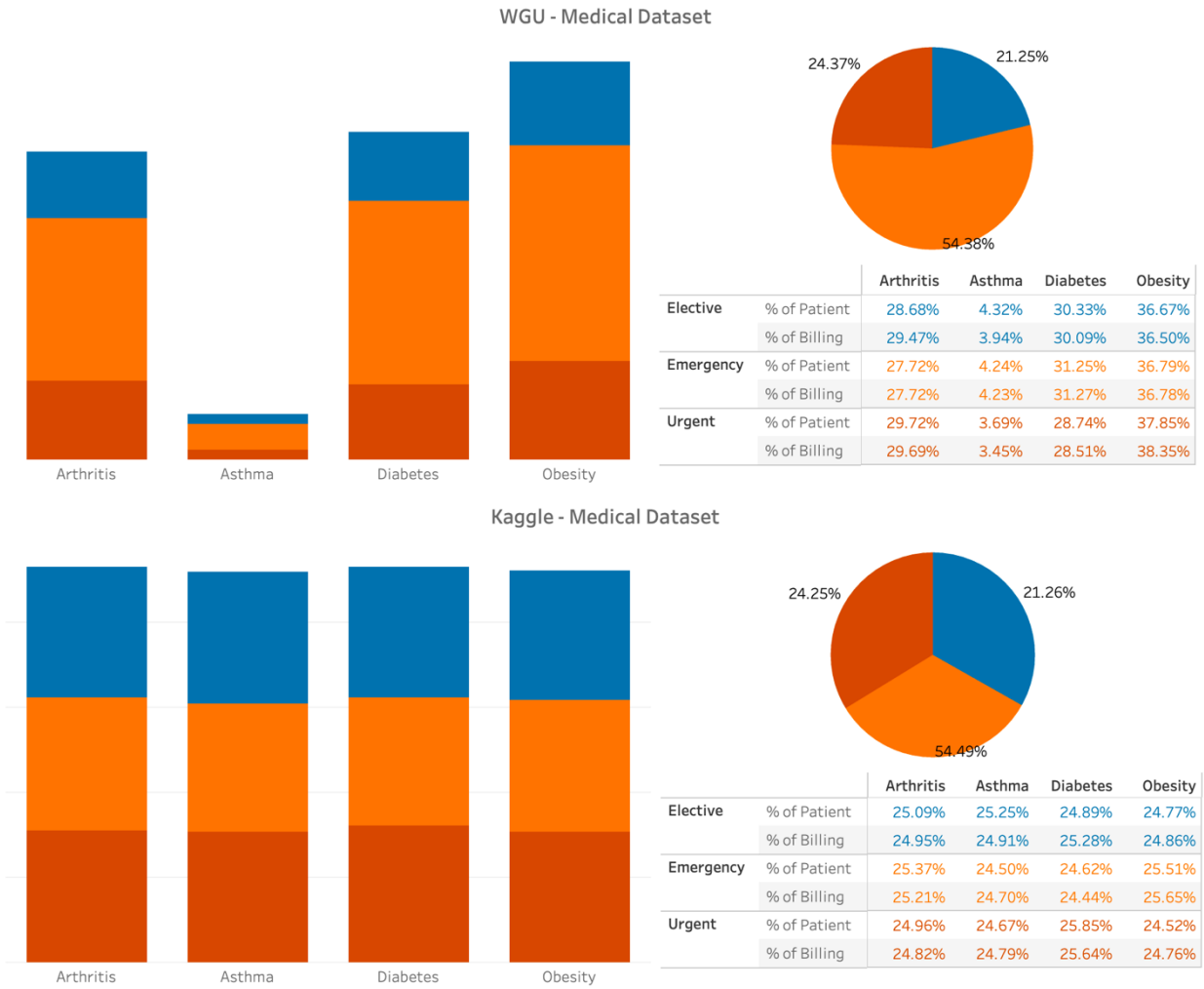
Navigating the dashboard on the Tableau desktop is simple; the instructions provided above will allow you to open the dashboard and begin navigating its interactive controls. Due to the two datasets similarities, the data inside the dashboard has been joined together. By doing so, users can view the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





		Arthritis	Asthma	Diabetes	Obesity
Elective	% of Patient	28.68%	4.32%	30.33%	36.67%
	% of Billing	29.47%	3.94%	30.09%	36.50%
Emergency	% of Patient	27.72%	4.24%	31.25%	36.79%
	% of Billing	27.72%	4.23%	31.27%	36.78%
Urgent	% of Patient	29.72%	3.69%	28.74%	37.85%
	% of Billing	29.69%	3.45%	28.51%	38.35%

		Arthritis	Asthma	Diabetes	Obesity
Elective	% of Patient	25.09%	25.25%	24.89%	24.77%
	% of Billing	24.95%	24.91%	25.28%	24.86%
Emergency	% of Patient	25.37%	24.50%	24.62%	25.51%
	% of Billing	25.21%	24.70%	24.44%	25.65%
Urgent	% of Patient	24.96%	24.67%	25.85%	24.52%
	% of Billing	24.82%	24.79%	25.64%	24.76%

The dashboard is a top-down split comparison with the Web (Kaggle) data source on the bottom and the WGU (provided) data source on the top. The first set of charts shows you a medical condition patient count dispersion; the height of each bar on the bar chart indicates the patient count in each medical condition category. 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 pie charts. These pie charts show the count of patients within each admission type group. This visual allows you to see what the patient's makeup is based on admission type. The color legend for the male and female gender groups can be found in the header section at the top.

The bottom portion of the dashboard shows specific numbers within a chart, indicating the % of patients and the % of billing charges for each of the admission types (elective, emergency, and observation), as well as the medical condition breakout. 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 or charge dispersion is in each category.

A4, SQL Code

Two datasets were used for this analysis, one of which was already loaded to PG Admin and another of which needs to be added. The Medical data set provided by WGU was already provided within the query tool program. The Kaggle dataset needed to be added. To do so, a create table statement will need to be written before uploading the CSV file.

Query used to create a table to house the Kaggle data:

```
-- Table: public.Healthcare_data
-- DROP TABLE public."Healthcare_data";
CREATE TABLE public."Healthcare_data"
(
  "Name" text COLLATE pg_catalog."default" NOT NULL,
  "Age" text COLLATE pg_catalog."default" NOT NULL,
  "Gender" text COLLATE pg_catalog."default" NOT NULL,
  "Blood_Type" text COLLATE pg_catalog."default" NOT NULL,
  "Medical_Condition" text COLLATE pg_catalog."default" NOT NULL,
  "Date_of_Admission" text COLLATE pg_catalog."default" NOT NULL,
  "Doctor" text COLLATE pg_catalog."default" NOT NULL,
  "Hospital" text COLLATE pg_catalog."default" NOT NULL,
  "Insurance_Provider" text COLLATE pg_catalog."default" NOT NULL,
  "Billing_Amount" text COLLATE pg_catalog."default" NOT NULL,
  "Room_Number" text COLLATE pg_catalog."default" NOT NULL,
  "Admission_Type" text COLLATE pg_catalog."default" NOT NULL,
  "Discharge_Date" text COLLATE pg_catalog."default" NOT NULL,
  "Medication" text COLLATE pg_catalog."default" NOT NULL,
  "Test_Results" text COLLATE pg_catalog."default" NOT NULL
)
TABLESPACE pg_default;

ALTER TABLE public."Healthcare_data"
  OWNER to Postgres;
```

After creating the table, you can right-click on the table name and say 'Import/Export. This will allow you to select the CSV file from your saved documents and load it to the table structure. Once loaded, you can begin to query it.

This dashboard was created using two queries: one for the WGU-provided medical dataset and one for the Kaggle dataset. Kaggle data was a straight pull of the necessary fields with a sum of billing amount and a count of patient names. This data was filtered on the medical conditions that overlap with the WGU dataset. The WGU data was also a straight pull with some manipulation required. To get the medical conditions into a single field layout, a case when statement was used. Additionally, a sum of the total charge and a count of patients was performed the same way as the Kaggle set. Within PGAdmin, the medical data was split by tables, requiring two left joins in the query to get the results.

Query used for Kaggle dataset:

```
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')
GROUP BY 1,2,3
```

Query used for WGU dataset:

```
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'
  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."admins_id" = pat."admis_id"

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

GROUP BY 1,2,3
```

B, Part 2: Demonstration

The video recording for this assignment includes a vocalized demonstration of the dashboard, 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 “D211-Kline”.

Panopto video link: https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=4f83defe-266c-462b-af67-b28a016f8a25

C, Part 3: Report

C1, Purpose and Function

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

The dashboard categorizes admissions (elective, emergency, urgent) 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 datasets (WGU vs Kaggle Medical Data), 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 percentage 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.

Since fewer than 1 in 5 hospitals use technology to address readmissions, this dashboard serves as a data-driven foundation to justify investments in AI-driven predictive analytics and patient monitoring. By analyzing trends and comparing multiple datasets, hospitals can create evidence-based strategies to align with CMS regulations and reduce financial penalties.

This tableau 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, Business Intelligence Tool Justification

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 in 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 datasets, in this case, WGU and Kaggle medical datasets 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.

Tableau's visualization and analytical capabilities transform complex healthcare data into clear, actionable insights enabling stakeholders to identify trends, benchmark performance, and drill down into granular details, Tableau supports evidence-based decision-making that can significantly reduce financial penalties and enhance patient care. This makes it a highly effective tool for addressing the diverse challenges faced by hospital administrators, policymakers, and financial officers.

C3, Data Preparation Steps

To get the data in a usable format for this assignment, a few cleaning steps are required. First is loading the public dataset to pgAdmin. Step-by-step instructions for this are provided as part of part 1. Once the public data has been uploaded, it can be manipulated and queried.

There were different steps taken for the public dataset vs the WGU-provided dataset. For the public dataset, a straight 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 dataset.

The WGU data provided required a bit 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 dataset. 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, Dashboard Creation Steps

Creating this dashboard in Tableau involves a structured process from data preparation to visualization. Below is a step-by-step summary of how this dashboard was built.

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 dataset 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 stacked bar chart showing admission type by medical condition, pie charts showing the overall admission type distribution, and tables showing the admission type breakdown by condition. Stacked bar charts are created by dragging the medical condition field to the columns shelf and the patients to the rows shelf. It is then formatted appropriately and color-coded. Pie charts require you to drop the admission type and patient fields in as the measures and then select a pie chart. 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 dataset.

Step three is to add filters and interactivity. For this dashboard, specifically, filters were added for readmission, gender, admission type, and medical condition. I did this 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, pie 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 project, it was not required to publish to Tableau public, so a .twbx file along with downloading instructions was provided as part of the assignment materials.

Dashboard preparation can be tricky; however, it is very important to spend the time to ensure the dashboard flows consistently and looks presentable. Following these structured steps ensures the dashboard is intuitive, interactive, and useful for hospital administrators. It provides actionable insights into patient readmission trends and financial impacts, helping stakeholders make informed decisions to reduce penalties and improve patient care.

C5, Results, Purpose, Function

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 drive up 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 dataset comparison allows you to benchmark the findings. This indicates that the WGU dataset shows higher emergency and urgent admissions for obesity and diabetes, while the Kaggle dataset 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

While the dashboard provides valuable insights, several limitations may impact the accuracy and applicability of the findings. One limitation is that the dataset 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 datasets may lead to inconsistencies in insights. There may be key variables missing, and without these variables, it's challenging to directly correlate admission type with actual readmission rates. Lastly, the dashboard identifies trends but does not use machine learning models to predict future readmission likelihoods.

Despite these 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 granular patient demographics to further refine its usefulness.

D, Web Sources

Healthcare Data: Dummy Data with a Multi-Category Classification Problem on Kaggle is the dataset used for this project. The Kaggle site listed key information about the demographics and contents of the dataset. This site was utilized to gain an understanding of the dataset's variables.

E, Acknowledge the Sources

I acknowledge that no segments of third-party sources were directly stated or copied from the web into this report.