

Performance Assessment: Data Analysis

Joshua Funderburk

Western Governors University

D211: Advanced Data Acquisition

December 23rd, 2024

Part 1: Data Dashboards

A1: Datasets and Dashboard File

The dashboard uses the following datasets.

1. The *Medical* data set provided by WGU
2. Kaiser Family Foundation's *Hospital Adjusted Expenses per Inpatient Day* data set, located online [here](#).

The .twbx file is included in this submission. Alternatively, the dashboard can be accessed on Tableau public [here](#).

A2: Dashboard Installation

1. Download the d211.zip file to the virtual machine
2. Open Windows File Explorer
3. Navigate to the downloaded file at *C:\Users\LabUser\Downloads*
4. Single click the "d211" folder
5. At the top of File Explorer, click "Compressed Folder Tools"
6. Click Extract All
7. In the text box under "Files will be extract to this folder:" update the file path to be this:
C:\Users\Public\Downloads
8. Click Extract
9. Open pgAdmin4 using the shortcut on the Windows taskbar
10. Click the carrot next to "Servers" on the left side of pgAdmin to expand the menu
11. Right click "medical_data" under the "Databases" drop down
12. Click "Query Tool" from the pop up window
13. Click the "Open File" button from the Query Tool window
14. Navigate the file path to *C:\Users\Public\Downloads\d211*
15. Click on the *d211_queries.sql* file
16. Click "Select"
17. Execute the queries
18. From the Desktop, double click Tableau 2021.4 to open
19. In the top left corner, click File
20. Click Open
21. Navigate the file path to *C:\Users\Public\Downloads\d211*
22. Click on the *D211 - Advanced Data Acquisition .twbx* file
23. Click Open
24. If prompted to sign in to connect to the Postgres server, the username is "postgres" and the password is "Passw0rd!"

A3: Dashboard Navigation

This dashboard is a single-tab dashboard intended to give users insight into Hospital admission data in one consolidated view. It can be broken into six sections and navigated as follows:

1. Key Performance Indicators (KPIs)

Located in the top left

- Card visualizations showing:
 - Average Initial Days - displays typical length of stay
 - Readmission Rate - shows percentage of returning patients
 - Two comparison cards showing average daily hospital admission cost per state for the patient and the hospital admission expense for the hospital (WGU vs. KFF data sets)

2. Admission Types

Located in left middle

- Bar chart
- Compares three categories of patient admissions

3. Primary Services

Located in bottom left

- Tree map
- Tracks frequency of primary medical services during hospital admission

4. Medical Conditions

Located in top right

- Horizontal bar chart
- Displays percentage of hospitalized patients with specific medical conditions

5. Readmission Rates by US State

Located in bottom right

- Interactive US Map visualization
- Hover functionality shows:
 - Specific readmission percentages
 - Count of readmissions
- Color intensity indicates relative rates
- Legend below shows intensity scale

6. Filtering Options

- State selection dropdown
- Readmission status filter (Yes/No)

A4: SQL Code

```
-- Create WGU medical dataset view
CREATE VIEW wgu_medical_data AS
SELECT
    patient.patient_id,
    location.city AS patient_city,
    location.state AS patient_state,
    location.county AS patient_county,
```

```

lpad(location.zip::text, 5, '0') AS patient_zip_code,
patient.lat AS patient_latitude,
patient.lng AS patient_longitude,
patient.population census_population_near_patient_location,
job.job_title,
patient.children AS no_of_children,
patient.age,
patient.income,
patient.marital,
patient.gender,
CASE
    WHEN patient.readmis = 'Yes' THEN TRUE
    WHEN patient.readmis = 'No' THEN FALSE
END AS is_readmitted,
patient.vitd_levels,
patient.doc_visits,
patient.full_meals,
patient.vitd_supp,
CASE
    WHEN patient.soft_drink = 'Yes' THEN TRUE
    WHEN patient.soft_drink = 'No' THEN FALSE
END AS soft_drink,
admission.initial_admission,
CASE
    WHEN patient.highblood = 'Yes' THEN TRUE
    WHEN patient.highblood = 'No' THEN FALSE
END AS has_high_blood_pressure,
CASE
    WHEN patient.stroke = 'Yes' THEN TRUE
    WHEN patient.stroke = 'No' THEN FALSE
END AS has_stroke_history,
complication.complication_risk,
CASE
    WHEN servicesaddon.overweight = 'Yes' THEN TRUE
    WHEN servicesaddon.overweight = 'No' THEN FALSE
END AS is_overweight,
CASE
    WHEN servicesaddon.arthritis = 'Yes' THEN TRUE
    WHEN servicesaddon.arthritis = 'No' THEN FALSE
END AS has_arthritis,
CASE
    WHEN servicesaddon.diabetes = 'Yes' THEN TRUE
    WHEN servicesaddon.diabetes = 'No' THEN FALSE
END AS has_diabetes,
CASE
    WHEN servicesaddon.hyperlipidemia = 'Yes' THEN TRUE
    WHEN servicesaddon.hyperlipidemia = 'No' THEN FALSE
END AS has_hyperlipidemia,
CASE
    WHEN servicesaddon.backpain = 'Yes' THEN TRUE
    WHEN servicesaddon.backpain = 'No' THEN FALSE

```

```

END AS has_back_pain,
CASE
    WHEN servicesaddon.anxiety = 'Yes' THEN TRUE
    WHEN servicesaddon.anxiety = 'No' THEN FALSE
END AS has_anxiety,
CASE
    WHEN servicesaddon.allergic_rhinitis = 'Yes' THEN TRUE
    WHEN servicesaddon.allergic_rhinitis = 'No' THEN FALSE
END AS has_allergic_rhinitis,
CASE
    WHEN servicesaddon.reflux_esophagitis = 'Yes' THEN TRUE
    WHEN servicesaddon.reflux_esophagitis = 'No' THEN FALSE
END AS has_reflux_esophagitis,
CASE
    WHEN servicesaddon.asthma = 'Yes' THEN TRUE
    WHEN servicesaddon.asthma = 'No' THEN FALSE
END AS has_asthma,
servicesaddon.services,
patient.initial_days,
patient.totalcharge AS total_charge,
patient.additional_charges,
survey_responses_addon.item1 AS survey_timely_admission,
survey_responses_addon.item2 AS survey_timely_treatment,
survey_responses_addon.item3 AS survey_timely_visit,
survey_responses_addon.item4 AS survey_reliability,
survey_responses_addon.item5 AS survey_options,
survey_responses_addon.item6 AS survey_hours_of_treatment,
survey_responses_addon.item7 AS survey_courteous_staff,
survey_responses_addon.item8 AS survey_doctor_active_listening
FROM
    patient
    JOIN admission ON admission.admis_id = patient.admis_id
    JOIN complication ON complication.complication_id = patient.compl_id
    JOIN job ON job.job_id = patient.job_id
    JOIN "location" ON location.location_id = patient.location_id
    JOIN servicesaddon ON servicesaddon.patient_id = patient.patient_id
    JOIN survey_responses_addon ON survey_responses_addon.patient_id =
patient.patient_id;

-- Create table for kff data
CREATE TABLE kff_inpatient_expenses (
    state_location VARCHAR(100) NOT NULL,
    expenses_per_day NUMERIC(10, 2) NOT NULL,
    state_abbrev CHAR(2) NOT NULL,

    CONSTRAINT kff_inpatient_expenses_pk PRIMARY KEY (state_abbrev)
)

TABLESPACE pg_default;

```

```
ALTER TABLE public.kff_inpatient_expenses
  OWNER to postgres;

-- Import kff data
COPY kff_inpatient_expenses
FROM 'C:\Users\Public\Downloads\d211\d211_kff.csv'
DELIMITER ','
CSV HEADER;

-- Referential Integrity
ALTER TABLE servicesaddon ADD FOREIGN KEY (patient_id) REFERENCES
patient(patient_id);
ALTER TABLE survey_responses_addon ADD FOREIGN KEY (patient_id) REFERENCES
patient(patient_id);
```

Part 2: Demonstration

See Panopto presentation.

Part 3: Report

C1: Dashboard Alignment

Explain how the purpose and function of your dashboard aligns with the needs of the stakeholders for your chosen dataset.

The medical dataset focuses on hospital readmissions, a critical concern for the associated hospital chain. The analysis aims to quantify the extent of readmission issues and identify actionable insights to address them.

The dashboard is designed to tackle the challenge of CMS penalties by monitoring readmission rates and their contributing factors. By highlighting key correlations and providing detailed analysis via filters, it allows the hospital chain to adopt a data-driven approach to reducing readmissions. The map visualization is particularly useful for a nationwide chain, helping identify high-risk facilities.

The primary purpose of the dashboard is to showcase readmission rates alongside key performance indicators (KPIs) and metrics correlated with readmissions. Notable KPIs include % Readmitted and Average Initial Days, which are prominently highlighted to track readmission rates and average initial hospital stays. Statistical analysis revealed a strong correlation between initial length of stay and readmission, making this metric essential to monitor. Similarly, the Initial Admission Type visualization is critical due to the strong statistically significant relationship between emergency admissions and hospital readmissions. The visualization itself allows users to analyze trends across admission types.

Other visualizations explore primary services during admission, patient medical conditions, and readmissions by U.S. state. These enable users to dissect the data and identify patterns. Filters for Readmission Status and State enhance the dashboard's utility, allowing users to compare characteristics of readmitted versus non-readmitted patients and pinpoint trends by US State.

C2: Business Intelligence Tool

Justify the selection of the business intelligence tool you used.

Tableau is a powerful business intelligence solution for analyzing hospital readmission data across this hospital chain. As an industry leader in data visualization and reporting, it offers exceptional flexibility in its integrations, seamlessly working with flat CSV files, databases, and numerous other data sources. Its out-of-the-box visualizations, including maps and various charts, are perfect for tracking readmission patterns. Tableau's user-friendly interface and strong compatibility with both virtual environments and PostgreSQL databases make it an ideal choice for this performance assessment.

C3: Data Cleaning

Explain the steps used to clean and prepare the data for the analysis.

The following are the steps taken to clean and prepare the data:

- A. Create a view within the medical_data database that joins all of the WGU medical data together. Additionally, update data types and alias columns as needed to follow database best practice standards. The following are steps taken to clean and prepare the data in the view:
 1. Add "patient_" as a prefix to the city, state, and county columns using an alias.
 2. Add leading zeroes to the zip_code column, cast it as text, and add "patient_" as prefix using an alias.
 3. Alias lat to be patient_latitude.
 4. Alias lng to be patient_longitude.
 5. Alias population column as census_population_near_patient_location.
 6. Alias children column as no_of_children.
 7. Use a case statement to make the readmis column boolean (True / False) instead of a text based Yes/No and alias it as is_readmitted.
 8. Use a case statement to make the soft_drink column boolean (True / False) instead of a text based Yes/No.
 9. Use a case statement to make the hignblood column boolean (True / False) instead of a text based Yes/No and alias it as has_high_blood_pressure.
 10. Use a case statement to make the stroke column boolean (True / False) instead of a text based Yes/No and alias it as has_stroke_history.
 11. Use a case statement to make the overweight column boolean (True / False) instead of a text based Yes/No and alias it as is_overweight.
 12. Use a case statement to make the arthritis column boolean (True / False) instead of a text based Yes/No and alias it as has_arthritis.

13. Use a case statement to make the diabetes column boolean (True / False) instead of a text based Yes/No and alias it as has_diabetes.
 14. Use a case statement to make the hyperlipidemia column boolean (True / False) instead of a text based Yes/No and alias it as has_hyperlipidemia.
 15. Use a case statement to make the backpain column boolean (True / False) instead of a text based Yes/No and alias it as has_back_pain.
 16. Use a case statement to make the anxiety column boolean (True / False) instead of a text based Yes/No and alias it as has_anxiety.
 17. Use a case statement to make the allergic_rhinitis column boolean (True / False) instead of a text based Yes/No and alias it as has_allergic_rhinitis.
 18. Use a case statement to make the reflux_esophagitis column boolean (True / False) instead of a text based Yes/No and alias it as has_reflux_esophagitis.
 19. Use a case statement to make the asthma column boolean (True / False) instead of a text based Yes/No and alias it as has_asthma.
 20. Alias totalcharge to be total_charge.
 21. Alias item1 as survey_timely_admission.
 22. Alias item2 as survey_timely_treatment.
 23. Alias item3 as survey_timely_visit.
 24. Alias item4 as survey_reliability.
 25. Alias item5 as survey_options.
 26. Alias item6 as survey_hours_of_treatment.
 27. Alias item7 as survey_courteous_staff.
 28. Alias item8 as survey_doctor_active_listening.
- B. Create a table for the kff_inpatient_expenses data
- C. Import data in to the kff_inpatient_expenses table from a csv

C4: Dashboard Creation

Summarize the steps used to create the dashboard.

The following are the steps taken to prepare the worksheets and dashboard for analysis:

1. Create Calculated Fields

- Create medical condition calculated fields:
 - % Allergic Rhinitis = SUM(IF [Has Allergic Rhinitis] = TRUE THEN 1 ELSE 0 END) / COUNT([Has Allergic Rhinitis])
 - % Anxiety = SUM(IF [Has Anxiety] = TRUE THEN 1 ELSE 0 END) / COUNT([Has Anxiety])
 - % Arthritis = SUM(IF [Has Arthritis] = TRUE THEN 1 ELSE 0 END) / COUNT([Has Arthritis])
 - % Asthma = SUM(IF [Has Asthma] = TRUE THEN 1 ELSE 0 END) / COUNT([Has Asthma])
 - % Back Pain = SUM(IF [Has Back Pain] = TRUE THEN 1 ELSE 0 END) / COUNT([Has Back Pain])

- % Diabetes = SUM(IF [Has Diabetes] = TRUE THEN 1 ELSE 0 END) / COUNT([Has Diabetes])
- % High Blood Pressure = SUM(IF [Has High Blood Pressure] = TRUE THEN 1 ELSE 0 END) / COUNT([Has High Blood Pressure])
- % Hyperlipidemia = SUM(IF [Has Hyperlipidemia] = TRUE THEN 1 ELSE 0 END) / COUNT([Has Hyperlipidemia])
- % Overweight = SUM(IF [Is Overweight] = TRUE THEN 1 ELSE 0 END) / COUNT([Is Overweight])
- % Readmitted = SUM(IF [Is Readmitted] = TRUE THEN 1 ELSE 0 END) / COUNT([Is Readmitted])
- % Reflux Esophagitis = SUM(IF [Has Reflux Esophagitis] = TRUE THEN 1 ELSE 0 END) / COUNT([Has Reflux Esophagitis])
- % Stroke = SUM(IF [Has Stroke History] = TRUE THEN 1 ELSE 0 END) / COUNT([Has Stroke History])
- For each medical condition calculated field, go to Default Properties and set the Number Format as a Percentage
- Create other calculated fields:
 - Avg Daily Hospital Cost Group by State = { INCLUDE [Patient State] : AVG([Total Charge]) }
 - Go to Default Properties and set the Number Format as Currency (Standard)
 - Count Readmitted = SUM(IF [Is Readmitted] = TRUE THEN 1 ELSE 0 END)

2. Create Worksheets

- **% Readmitted by US State**
 1. Rename worksheet to "% Readmitted by US State"
 2. Drag *Patient State* to Detail
 3. Drag % *Readmitted* to Color
 4. Drag % *Readmitted* to Label
 5. Drag *Count Readmitted* to Tooltip
 6. Drag *Patient State* to Filters
 - a. Select "Use all" > "OK"
 - b. Right click *Patient State* > select "Show filter"
 - c. Right click *Patient State* > select "Apply to Worksheets" > select "Apply to All Using Related Data Sources"
 7. Drag *Is Readmitted* to Filters
 - a. Select "Use all" > "OK"
 - b. Right click *Is Readmitted* > select "Show filter"
 - c. Right click *Patient State* > select "Apply to Worksheets" > select "Apply to All Using This Data Source"
- **% Readmitted**
 1. Rename worksheet to "% Readmitted"
 2. Drag % *Readmitted* to text
 3. Click Text

- a. Highlight <AGG(% Readmitted)>
 - b. Change alignment to "Centered"
 - c. Change font size to 36
 - d. Click "OK"
4. Double click the title
 - a. Change to "Centered"
 - b. Click "OK"
- **% of Patients with Specified Medical Condition**
 1. Rename worksheet to "% of Patients with Specified Medical Condition"
 2. Drag % *Allergic Rhinitis* to Rows
 3. Drag % *Anxiety* to Rows
 4. Click "Show Me" > select "horizontal bars"
 5. Drag % *Arthritis* to Measure Values
 6. Drag % *Asthma* to Measure Values
 7. Drag % *Back Pain* to Measure Values
 8. Drag % *Diabetes* to Measure Values
 9. Drag % *High Blood Pressure* to Measure Values
 10. Drag % *Hyperlipidemia* to Measure Values
 11. Drag % *Overweight* to Measure Values
 12. Drag % *Reflux Esophagitis* to Measure Values
 13. Drag % *Stroke* to Measure Values
 14. Right Click X-Axis > Erase "Value" from the Title section
 15. Click X-Axis > Click "Sort Measure Names descending by Measure Values"
 16. Click Label > Click "Show mark labels"
 17. Click *Measure Values* in the Columns pane > Edit Filter > Change "Numbers" to "Percentage"
- **Initial Admission Type**
 1. Rename worksheet to "Initial Admission Type"
 2. Drag *Initial Admission* to Columns
 3. Drag *wgu_medical_data (Count)* to Rows
 4. Right Click Y-Axis > Erase "Count of wgu_medical_data" from the Title section
 5. Click Initial Admission >
 - a. Click "Sort"
 - b. Set "Sort By" equal to "Field"
 - c. Set "Sort Order" to "Descending"
 6. Right click on "Initial Admission" on the chart directly below the "Initial Admission Type" title > select "Hide Field Labels for Columns"
 7. Click Label > Click "Show mark labels"
- **Primary Services**
 1. Rename worksheet to "Primary Services"
 2. Drag *Services* to Label
 3. Drag *wgu_medical_data (Count)* to Color
 4. Drag *wgu_medical_data (Count)* to Size
 5. Drag *wgu_medical_data (Count)* to Label

6. Click Label > Click Font dropdown > Change font size to 20
- **Average Initial Days**
 1. Rename worksheet to “Average Initial Days”
 2. Drag *Initial Days* to Text
 3. Click on SUM(Initial Days) > highlight Measure > change to Average
 4. Double click the title
 - a. Change to “Centered”
 - b. Click “OK”
 5. Click Text
 - a. Highlight <AVG(Initial Days)>
 - b. Change alignment to “Centered”
 - c. Change font size to 36
 - d. Click “OK”
- **Avg State Daily Hospital Inpatient Cost (WGU)**
 1. Rename worksheet to “Avg State Daily Hospital Inpatient Cost (WGU)”
 2. Drag *Avg Daily Hospital Cost Group by State* to Text
 3. Click on SUM(Avg Daily Hospital Cost Group by State) > highlight Measure > change to Average
 4. Double click the title
 - a. Change to “Centered”
 - b. Click “OK”
 5. Click Text
 - a. Highlight <AVG(Avg Daily Hospital Cost Group by State)>
 - b. Change alignment to “Centered”
 - c. Change font size to 36
 - d. Click “OK”
- **Avg State Daily Hospital Expense (KFF)**
 1. Rename worksheet to “Avg State Daily Hospital Expense (KFF)”
 2. Drag *Expenses Per Day* to Text
 3. Click on SUM(Expenses Per Day) > highlight Measure > change to Average
 4. Double click the title
 - a. Change to “Centered”
 - b. Click “OK”
 5. Click Text
 - a. Highlight <AVG(Expenses Per Day)>
 - b. Change alignment to “Centered”
 - c. Change font size to 36
 - d. Click “OK”

3. Create Hospital Admissions Dashboard

1. Rename to “Hospital Admissions”
2. Change Size to Automatic
3. Drag “% Readmitted by US State” map onto the dashboard

4. Drag “% of Patients with Specified Medical Condition” horizontal bar chart above the map
5. Drag “Primary Services” tree map to the left of the geographic map
6. Drag “Initial Admission Type” bar chart to be above the “Primary Services” tree map
7. Drag “% Readmitted” card to the left of the “% of Patients with Specified Medical Condition” horizontal bar chart
 - a. Change card from “Standard” view to “Entire View”
8. Drag “Avg State Daily Hospital Expense (KFF)” card to be underneath the “% Readmitted” card
 - a. Change card from “Standard” view to “Entire View”
9. Drag “Average Initial Days” card to be to the left of the “% Readmitted” card
 - a. Change card from “Standard” view to “Entire View”
10. Drag “Avg State Daily Hospital Inpatient Cost (WGU)” card to be to the left of the “Avg State Daily Hospital Expense (KFF)” card
 - a. Change card from “Standard” view to “Entire View”
11. Place the “Average Initial Days” card and the “% Readmitted” card in to a Horizontal container
 - a. Add a border to the container
 - b. Add an Outer Padding of 4 to all sides
12. Place the “Avg State Daily Hospital Inpatient Cost (WGU)” card and the “Avg State Daily Hospital Expense (KFF)” card in to a Horizontal container
 - a. Add a border to the container
 - b. Add an Outer Padding of 4 to all sides
13. Place “% Readmitted by US State” map and the “% of Patients with Specified Medical Condition” horizontal bar chart in to a vertical container
14. Drag the “% Readmitted” legend from the right pane to be beneath the geographic map in the vertical container
15. Remove “Count of wgu_medical_data” legend from the dashboard
16. Click the “Is Readmitted” filter >
 - a. Click the down arrow for more options
 - b. Hover over “Apply to Worksheets”
 - c. Click “Selected Worksheets...”
 - d. Uncheck “Avg State Daily Hospital Expense (KFF)”
 - e. Click “OK”
17. Resize visualizations as needed

C5: Data Analysis Results

Discuss the results of your data analysis and how it supported the purpose and function of your dashboard.

The most surprising takeaway in this analysis is that there are no true trends between patients that were readmitted and patients that were not readmitted when we compare the specified medical conditions individually. For example, 70.94% of patients in this dataset are overweight.

Of patients that were readmitted, 71.41% are overweight; and for patients that were not readmitted, 70.67% are overweight. Essentially, the same proportion is true for both readmitted and not-readmitted patients. This finding holds true across all 11 medical conditions available in the dataset. This finding well supports the purpose and function of the dashboard as it is giving users the ability to monitor various factors that may be indicative of a patient being readmitted to the hospital. In this instance, there were no trends, but in the future that could change and monitoring the dashboard would allow the user to see that.

Another result of this analysis was finding that the average daily cost for a patient on their initial stay was much higher for patients that were readmitted within 30 days (\$7,722.15) compared to those that were not (\$3,886.85). There are no apparent answers given the other visualizations on the dashboard, but this is the kind of disparity that this dashboard is meant to surface. This is an actionable insight where the action is to dive deeper into the cost disparity and try to determine what things are occurring during those initial visits for patients that were eventually readmitted that are not for others.

C6: Analysis Limitations

Discuss the limitation(s) of your data analysis.

One major limitation with this analysis is the lack of data to support a time series visualization. Typically, one of the first steps an analyst would take in a situation where stakeholders want to track percent readmitted would be to create a line chart that shows how the percentage of patients readmitted has changed over time. This sort of visualization allows users to determine if readmission rate is getting better or worse and also allows them to gauge the success of the actions they have taken to combat readmission.

The external dataset is also limiting for this analysis as it does not break out hospital expenses by patients that were readmitted and those that were not readmitted. This is understandably so, but it would be great to see if the same disparity found with cost between readmitted and not-readmitted also exists with expenses. It could be extremely telling in additional things a hospital stakeholder would want to investigate to reduce readmission.

D: Sources

Hospital adjusted expenses per inpatient day. KFF. (2024, February 1).
<https://www.kff.org/health-costs/state-indicator/expenses-per-inpatient-day/>