Laycee Glass

SLM2 Task 1

Advanced Data Acquisition

12/10/24

Western Governors University

## SLM Performance Assessment

### Student Information

Student Name: Laycee Glass

ID Number: 001896026

Date: 12/10/24

### Part I: Data Dashboards

### A1. Datasets Used

For this task, I used two datasets. The first was the medical dataset provided by WGU for most courses in this program, and the second was a publicly available dataset on Kaggle titled “Health Reentry Analysis” (2024). Both are provided in CSV form with this submission.

### A2. Dashboard Instructions

In order to access the created dashboard, the following steps can be taken:

1. Open up Labs on Demand. All it to initialize on your device.
2. Download the attached .twbx file onto your device. Once it has been downloaded, email it to yourself.
3. Once back in Labs on Demand, open up and login to your email. Find the .twbx file you emailed yourself and download it. This should send it to your Downloads folder.
4. Open the Tableau application (Tableau 2021.4) and go to File. Select Open…This should prompt you to make a selection from a window where you can now select the previously downloaded .twbx file.
5. Once selected and opened, this should allow for interaction with the dashboard.

### A3. Dashboard Navigation

I recommend going into presentation mode (F7) to display. For the two dashboards, there is a scroll bar to scroll up and down on the page as needed.

The first dashboard has 3 charts about different patient demographics that give the title of the demographic being compared between datasets and the percent of each. Hovering over the bars will allow you to see the specific percent of the total number of records for that dataset.

The second dashboard also has 3 charts. The income level and readmission stacked bar chart has largely the same features as the bar charts on the first dashboard. Specific percentages can be found by hovering on the portion of the bar you want to know about. Average length of stay versus medical diagnoses is a straightforward line chart labelled to easily see the amount of days patient stay during admission. A highlight map also allows for easy identification about where patients are being discharge, filtered by insurance type.

### A4. SQL Code

The SQL code used to create the table “raddata” for the dashboard in PgAdmin4 is as follows:

CREATE TABLE public.raddata (

patient\_id VARCHAR,

admission\_date VARCHAR,

num\_diagnoses VARCHAR,

num\_procedures VARCHAR,

num\_external\_injuries VARCHAR,

length\_of\_stay VARCHAR,

median\_household\_income VARCHAR,

insurance\_type VARCHAR,

age INTEGER,

gender VARCHAR,

readmitted VARCHAR,

diagnoses VARCHAR,

external\_injuries VARCHAR,

home VARCHAR,

facility VARCHAR,

homeless VARCHAR,

discharge\_date VARCHAR

)

TABLESPACE pg\_default;

ALTER TABLE public.raddata

OWNER TO postgres;

The SQL code used to perform the union in Tableau is as follows:

SELECT "t0"."Table Name" AS "Table Name",

  "t0"."additional\_charges" AS "additional\_charges",

  "t0"."admis\_id" AS "admis\_id",

  "t0"."admission\_date" AS "admission\_date",

  "t0"."age" AS "age",

  "t0"."children" AS "children",

  "t0"."compl\_id" AS "compl\_id",

  "t0"."diagnoses" AS "diagnoses",

  "t0"."discharge\_date" AS "discharge\_date",

  "t0"."doc\_visits" AS "doc\_visits",

  "t0"."external\_injuries" AS "external\_injuries",

  "t0"."facility" AS "facility",

  "t0"."full\_meals" AS "full\_meals",

  "t0"."gender" AS "gender",

  "t0"."hignblood" AS "hignblood",

  "t0"."home" AS "home",

  "t0"."homeless" AS "homeless",

  "t0"."income" AS "income",

  "t0"."initial\_days" AS "initial\_days",

  "t0"."insurance\_type" AS "insurance\_type",

  "t0"."job\_id" AS "job\_id",

  "t0"."lat" AS "lat",

  "t0"."length\_of\_stay" AS "length\_of\_stay",

  "t0"."lng" AS "lng",

  "t0"."location\_id" AS "location\_id",

  "t0"."marital" AS "marital",

  "t0"."median\_household\_income" AS "median\_household\_income",

  "t0"."num\_diagnoses" AS "num\_diagnoses",

  "t0"."num\_external\_injuries" AS "num\_external\_injuries",

  "t0"."num\_procedures" AS "num\_procedures",

  "t0"."patient\_id" AS "patient\_id",

  "t0"."population" AS "population",

  "t0"."readmis" AS "readmis",

  "t0"."readmitted" AS "readmitted",

  "t0"."soft\_drink" AS "soft\_drink",

  "t0"."stroke" AS "stroke",

  "t0"."totalcharge" AS "totalcharge",

  "t0"."vitd\_levels" AS "vitd\_levels",

  "t0"."vitd\_supp" AS "vitd\_supp"

FROM (

  SELECT "t1"."Table Name" AS "Table Name", "t1"."additional\_charges" AS "additional\_charges", "t1"."admis\_id" AS "admis\_id", "t1"."admission\_date" AS "admission\_date", "t1"."age" AS "age", "t1"."children" AS "children", "t1"."compl\_id" AS "compl\_id", "t1"."diagnoses" AS "diagnoses", "t1"."discharge\_date" AS "discharge\_date", "t1"."doc\_visits" AS "doc\_visits", "t1"."external\_injuries" AS "external\_injuries", "t1"."facility" AS "facility", "t1"."full\_meals" AS "full\_meals", "t1"."gender" AS "gender", "t1"."hignblood" AS "hignblood", "t1"."home" AS "home", "t1"."homeless" AS "homeless", "t1"."income" AS "income", "t1"."initial\_days" AS "initial\_days", "t1"."insurance\_type" AS "insurance\_type", "t1"."job\_id" AS "job\_id", "t1"."lat" AS "lat", "t1"."length\_of\_stay" AS "length\_of\_stay", "t1"."lng" AS "lng", "t1"."location\_id" AS "location\_id", "t1"."marital" AS "marital", "t1"."median\_household\_income" AS "median\_household\_income", "t1"."num\_diagnoses" AS "num\_diagnoses", "t1"."num\_external\_injuries" AS "num\_external\_injuries", "t1"."num\_procedures" AS "num\_procedures", "t1"."patient\_id" AS "patient\_id", "t1"."population" AS "population", "t1"."readmis" AS "readmis", "t1"."readmitted" AS "readmitted", "t1"."soft\_drink" AS "soft\_drink", "t1"."stroke" AS "stroke", "t1"."totalcharge" AS "totalcharge", "t1"."vitd\_levels" AS "vitd\_levels", "t1"."vitd\_supp" AS "vitd\_supp"

  FROM (

    SELECT ('patient'::text) AS "Table Name",

      "patient"."additional\_charges" AS "additional\_charges",

      "patient"."admis\_id" AS "admis\_id",

      CAST(NULL AS TEXT) AS "admission\_date",

      "patient"."age" AS "age",

      "patient"."children" AS "children",

      "patient"."compl\_id" AS "compl\_id",

      CAST(NULL AS TEXT) AS "diagnoses",

      CAST(NULL AS TEXT) AS "discharge\_date",

      "patient"."doc\_visits" AS "doc\_visits",

      CAST(NULL AS TEXT) AS "external\_injuries",

      CAST(NULL AS TEXT) AS "facility",

      "patient"."full\_meals" AS "full\_meals",

      CAST("patient"."gender" AS TEXT) AS "gender",

      CAST("patient"."hignblood" AS TEXT) AS "hignblood",

      CAST(NULL AS TEXT) AS "home",

      CAST(NULL AS TEXT) AS "homeless",

      "patient"."income" AS "income",

      "patient"."initial\_days" AS "initial\_days",

      CAST(NULL AS TEXT) AS "insurance\_type",

      "patient"."job\_id" AS "job\_id",

      "patient"."lat" AS "lat",

      CAST(NULL AS TEXT) AS "length\_of\_stay",

      "patient"."lng" AS "lng",

      "patient"."location\_id" AS "location\_id",

      CAST("patient"."marital" AS TEXT) AS "marital",

      CAST(NULL AS TEXT) AS "median\_household\_income",

      CAST(NULL AS TEXT) AS "num\_diagnoses",

      CAST(NULL AS TEXT) AS "num\_external\_injuries",

      CAST(NULL AS TEXT) AS "num\_procedures",

      CAST("patient"."patient\_id" AS TEXT) AS "patient\_id",

      "patient"."population" AS "population",

      CAST("patient"."readmis" AS TEXT) AS "readmis",

      CAST(NULL AS TEXT) AS "readmitted",

      CAST("patient"."soft\_drink" AS TEXT) AS "soft\_drink",

      CAST("patient"."stroke" AS TEXT) AS "stroke",

      "patient"."totalcharge" AS "totalcharge",

      "patient"."vitd\_levels" AS "vitd\_levels",

      "patient"."vitd\_supp" AS "vitd\_supp"

    FROM "public"."patient" "patient"

  ) "t1"

   UNION  ALL

  SELECT "t2"."Table Name" AS "Table Name", "t2"."additional\_charges" AS "additional\_charges", "t2"."admis\_id" AS "admis\_id", "t2"."admission\_date" AS "admission\_date", "t2"."age" AS "age", "t2"."children" AS "children", "t2"."compl\_id" AS "compl\_id", "t2"."diagnoses" AS "diagnoses", "t2"."discharge\_date" AS "discharge\_date", "t2"."doc\_visits" AS "doc\_visits", "t2"."external\_injuries" AS "external\_injuries", "t2"."facility" AS "facility", "t2"."full\_meals" AS "full\_meals", "t2"."gender" AS "gender", "t2"."hignblood" AS "hignblood", "t2"."home" AS "home", "t2"."homeless" AS "homeless", "t2"."income" AS "income", "t2"."initial\_days" AS "initial\_days", "t2"."insurance\_type" AS "insurance\_type", "t2"."job\_id" AS "job\_id", "t2"."lat" AS "lat", "t2"."length\_of\_stay" AS "length\_of\_stay", "t2"."lng" AS "lng", "t2"."location\_id" AS "location\_id", "t2"."marital" AS "marital", "t2"."median\_household\_income" AS "median\_household\_income", "t2"."num\_diagnoses" AS "num\_diagnoses", "t2"."num\_external\_injuries" AS "num\_external\_injuries", "t2"."num\_procedures" AS "num\_procedures", "t2"."patient\_id" AS "patient\_id", "t2"."population" AS "population", "t2"."readmis" AS "readmis", "t2"."readmitted" AS "readmitted", "t2"."soft\_drink" AS "soft\_drink", "t2"."stroke" AS "stroke", "t2"."totalcharge" AS "totalcharge", "t2"."vitd\_levels" AS "vitd\_levels", "t2"."vitd\_supp" AS "vitd\_supp"

  FROM (

    SELECT ('raddata'::text) AS "Table Name",

      CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION) AS "additional\_charges",

      CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT) AS "admis\_id",

      "raddata"."admission\_date" AS "admission\_date",

      "raddata"."age" AS "age",

      CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT) AS "children",

      CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT) AS "compl\_id",

      "raddata"."diagnoses" AS "diagnoses",

      "raddata"."discharge\_date" AS "discharge\_date",

      CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT) AS "doc\_visits",

      "raddata"."external\_injuries" AS "external\_injuries",

      "raddata"."facility" AS "facility",

      CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT) AS "full\_meals",

      "raddata"."gender" AS "gender",

      CAST(NULL AS TEXT) AS "hignblood",

      "raddata"."home" AS "home",

      "raddata"."homeless" AS "homeless",

      CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION) AS "income",

      CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION) AS "initial\_days",

      "raddata"."insurance\_type" AS "insurance\_type",

      CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT) AS "job\_id",

      CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION) AS "lat",

      "raddata"."length\_of\_stay" AS "length\_of\_stay",

      CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION) AS "lng",

      CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT) AS "location\_id",

      CAST(NULL AS TEXT) AS "marital",

      "raddata"."median\_household\_income" AS "median\_household\_income",

      "raddata"."num\_diagnoses" AS "num\_diagnoses",

      "raddata"."num\_external\_injuries" AS "num\_external\_injuries",

      "raddata"."num\_procedures" AS "num\_procedures",

      "raddata"."patient\_id" AS "patient\_id",

      CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT) AS "population",

      CAST(NULL AS TEXT) AS "readmis",

      "raddata"."readmitted" AS "readmitted",

      CAST(NULL AS TEXT) AS "soft\_drink",

      CAST(NULL AS TEXT) AS "stroke",

      CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION) AS "totalcharge",

      CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION) AS "vitd\_levels",

      CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT) AS "vitd\_supp"

    FROM "public"."raddata" "raddata"

  ) "t2"

) "t0"

### Part II: Demonstration

### B. Panopto Recording

I have attached a Panopto recording with this submission which demonstrates the functionality of the code/dashboards, and it also addresses parts B1-B7 of this task.

### Part III: Report

### C1. Purpose and Function

The WGU medical dataset scenario was very interested in investigating readmission rates of patients. I chose an external dataset that gave a lot of data in this same vein (demographics, admission date, readmission, diagnoses, length of stay, etc.) Comparing the demographics of patients of both datasets could help to provide understanding of patient populations, which can help tailor services and improve patient engagement. For example, if there is a high percentage of geriatric patients, more chronic disease management may be required, whereas a higher percentage of younger/pediatric patients could benefit from more preventative services. Other factors like discharge location, length of stay, and medical conditions/diagnoses would help hospitals identify trends for better resource management, care coordination, and cost control.

### C2. Justification

Tableau was used for this task for various reasons. The first was its ability to connect to several different data sources, including PostgreSQL, which was also employed for this task. Tableau also has a user-friendly interface that allows for easy creation of complex visualizations and customizable dashboards that can be tailored to the specific needs of the creator.

### C3. Data Preparation

Data preparation for this task was not a lengthy process. Both datasets were already very clean, and all of the cleaning/preparation I did took place in Tableau because of this.

The first thing I did was to verify that all 10000 records of the medical dataset were there as identified in the data dictionary.

In Tableau, I ended up needing to change the “readmitted” column in the “raddata” table from 1/0 values to Yes/No values for visualization purposes. I also had to create calculated fields for my purposes as well. This included creating a source column to differentiate between the WGU dataset and the external dataset, creating age categories (0-18, 18-35, etc), multiplying the median income column in the external dataset by 1000 to match the format of the WGU income column, creating a discharge location column from the three columns in the external dataset (home, facility, homeless), and creating unified fields for income and readmission status.

### C4. Dashboard Creation

Once the tables had been put together via union in Tableau, I was able to create visuals. The layout for these visuals is shown as follows:

Sheet 1: Patient Gender, WGU vs RaD Data

Columns: Source, gender

Rows: CNT(Union), computed using gender, % of total

Color: gender

Chart: Side-by-side bar chart

For this visual, I created a calculated field titled Source. The code for this field is below:

IF [Table Name] = ‘patient’ THEN ‘WGU Data’

ELSE ‘RaD Data’

END

Sheet 2: Patient Age, WGU vs Rad Data

Columns: Source, Age Cats

Rows: CNT(Union), computed using Age Cats, % of total

Color: Age Cats

Chart: Side-by-side bar chart

For this visual, I created a calculated field titled Age Cats. The code for this is below:

IF [age] <= 18 THEN ‘0-18’

ELSEIF [age] <= 35 THEN ‘19-35’

ELSEIF [age] <= 50 THEN ‘36-50’

ELSEIF [age] <= 65 THEN ‘51-65’

ELSEIF [age] <= 80 THEN ‘66-80’

ELSE ‘80+’

END

Sheet 5: Patient Income Level, WGU vs RaD Data

Columns: Source, Unified Income (bin)

Rows: CNT(Union), computed using income, % of total

Color: Unified Income (bin)

Chart: Side-by-side bar chart

For this visual, I created two calculated fields titled Adjusted Median Income and Unified Income. The code for these is below:

[median\_household\_income] \* 1000

IF ISNULL ([Adjusted Median Income)] THEN [income]

ELSE [Adjusted Median Income]

END

The bins for the income were increments of 15,000.

Sheet 3: Discharge Location vs Insurance Type

Columns: Discharge Location

Rows: insurance\_type

Color: CNT(Union)

Label: CNT(Union)

Filters: insurance\_type (NULL values)

Chart: Highlight (square)

For this visual, I created a calculated field titled Discharge Location. The code for this is below:

IF [home] = ‘yes’ THEN ‘Home’

ELSEIF [facility] = ‘yes’ THEN ‘Facility’

ELSEIF [homeless] = ‘yes’ THEN ‘Homeless’

ELSE ‘Unknown’

END

Sheet 4: Average Length of Stay vs Medical Diagnoses

Columns: diagnoses

Rows: AVG(length\_of\_stay)

Color: AVG(length\_of\_stay)

Label: AVG(length\_of\_stay)

Filters: length\_of\_stay (NULL values)

Chart: Line

Sheet 6: Income Level vs Readmission

Columns: Unified Income (bin)

Rows: CNT(Union)

Color: Unified Readmission Status

Label: CNT(Union), computed using Unified Readmission Status, % of total

Chart: Stacked bar chart

For this visual, I used the aforementioned Unified Income calculated field, along with creating a new field titled Unified Readmission Status. The code for this is below:

IF NOT ISNULL([readmis]) THEN [readmis]

ELSEIF STR([readmitted)] = “1” THEN ‘Yes’

ELSEIF STR([readmitted)] = “0” THEN ‘No’

END

Once the visualizations were finished, I then created my two dashboards and formatted them in a way I found most pleasing to the eye, whilst still capturing all the data I was trying to show. The layout for the dashboards is as follows:

Dashboard 1: Patient Demographics Comparison

Sheet 1: Patient Gender, bottom right

Sheet 2: Patient Age, left half

Sheet 5: Patient Income Level, top right

Dashboard 2: Patient Outcomes and Care Trends

Sheet 3: Discharge Location, top left

Sheet 4: Length of Stay, bottom half

Sheet 6: Readmission Status, top right

### C5. Results

There were a few insights to gain from the dashboards created during this task. For patient demographics, the age and gender breakdown were roughly the same for both datasets. There was slight variance in income levels between the two with a majority of WGU patients making less than 75k, though this may be attributed to the differences in data (WGU data had single income, whereas external data had household income).

The most frequent discharge location was home for patients with a private insurance, though 6 were discharged homeless. This was a surprise because private insurance has a high cost and would be a struggle to afford during homelessness.

The longest average length of stay was associated with chronic obstructive pulmonary disease at 16.769 days, and the shortest was associated with diabetes at 14.737 days. All diagnoses were within about 2 days of each other during admission.

The highest percentage of readmission happened in lower outcome levels (less than 60k). There were some higher percentages in the higher levels (150k), but this was due to such low volume in total patients in this level. So a few readmissions seemed to have skewed the percentages.

### C6. Limitations

One of the biggest limitations for this analysis was the number of records. The WGU dataset contained 10000 records, and the external dataset contained just 1000 records. The analysis would be much more generalizable with even more records.

Another limitation I noticed was the exclusion of young patients. Neither dataset had any minors in it, which is not representative of the US population where minors do get hospitalized and readmitted. This also affects generalizability of this analysis.

### D. Code Sources

No outside sources were used for the code in this task.

### E. Text Sources

D211 Webinar with Dr. Sewell

Kaggle Dataset “Health Reentry Analysis”