

**Advanced Data Acquisition**

Jason Willis  
College of Information Technology,  
Western Governors University  
Dr. Daniel Smith  
May 30<sup>th</sup>, 2022

## Table of Contents for Each Rubric

***Part I: Data Dashboard***

<i>Both Data Sets, Installation/Navigation, SQL Code</i> .....	3
--	---

***Part II: Presentation and Setting up Dashboard***

<i>Presentation, Technical Env, Dashboard Functionality, Scripts, Data Streams/Points/Creation, Referential Integrity</i> .....	7
---	---

***Part III: Data Cleaning, Analysis and Results***

<i>Data Alignment, Cleaning, Analysis Results and Limits</i> .....	8
--	---

***Part IV: Sources and Demonstration***

<i>Video</i> .....	17
<i>In-text Citations and References</i> .....	17
<i>Sources</i> .....	18

### **Hospital Readmission Problem**

For our chain of hospitals to lower readmission concerns, we need to identify patients who have increased risk of rehospitalization within a month of their release. According to Schuller (2020), non-obese adults were 21% less likely to be readmitted than obese adults. A readmission study by Gert, et. al. (2002) showed a correlation between longer initial hospital stays and readmission. Within the provided dataset, I'm leveraging these studies to help create my hypothetical question and shape my approach in finding potential patient groups with a statistically significant chance for readmission outcomes.

After viewing the provided `medical_clean.csv` data set and accompanying data dictionary, there seems to be some patient groupings which are aligned with the research mentioned above. For instance, the following patient data fields: Initial patient admin days, Total Charges, and Initial Says (inpatient) both caught my attention and were underscored by the research mentioned above. While my initial feelings towards these variables might make them feel related, are they?

### **A – Data Dashboard**

The Tableau data dashboard supports data driven decision-making by connecting chosen data sources, building worksheets, and dashboards. The data has been focused to view patient readmissions with additional data points to try to find significant correlations and/or improvement opportunities.

### **A1 – Both Data Sets**

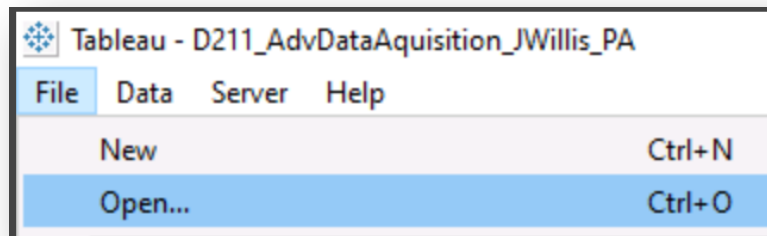
The following data sets were used in the Tableau dashboard:

- `hosp_ratings.csv` – provided by Kaggle (Bouchet, J, et. al. 2022)

- admission.xlsx, complication.xlsx, job.xlsx, location.xlsx, patient.xlsx, serviceaddon.xlsx, and survey\_response\_addon.xlsx – provided by WGU
- *Note: these data sets are provided in the package folder.*

## A2 – Installation Instructions

To open the provided dashboard, please open Tableau. Choose File → Open (Figure 1) and navigate to the saved file “D211\_AdvDataAquisition\_JWillis\_PA.twbx” select and choose open. Additionally, you could navigate directly to the saved Tableau packaged workbook executable and double-click the file to open in Tableau. You may need to input the password, found in Figure 3 below.



*Figure 1 - Open Saved \*.twbx File*

Once opened, the dashboard should load and look similar to Figure 2 below.



Figure 2 - Tableau Dashboard

Within the lab, a connection needs to be created between Tableau and PostgreSQL for seven of the eight tables. A connection popup should appear to input PostgreSQL credentials, which are found on the right side of the virtual machine window. (Figure 2). If not, the following SQL query code selects the data from the following tables (Figure 3):

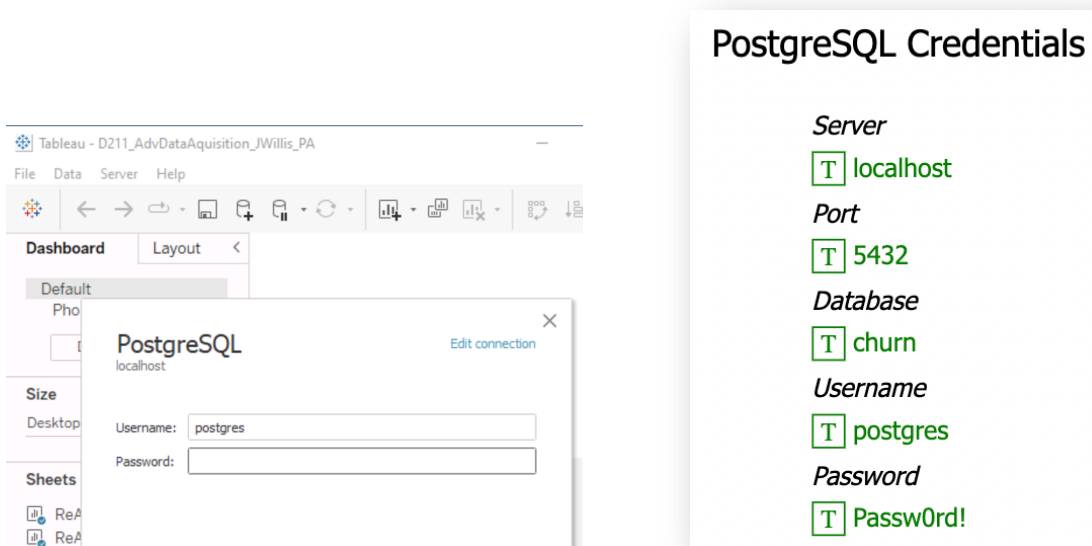


Figure 3 - Tableau Data Connection

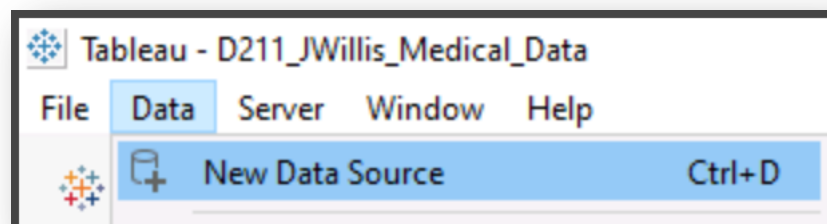


Figure 4 - Tableau: Add New Data Source

### A3 – Navigation Instructions

Once connected, you should see the workbook, complete with the following tabs:

- Data Source
- Summary Dashboard
- ReAdmins per Service Add-On
- ReAdmins per National Mortality Avg by State

- ReAdmins per Gender and Avg Age
- ReAdmins per Complication Risk, Initial Admission per Initial Days

On the bottom of Figure 2, notice the dashboard and sheet tabs. Each sheet provides a focused look at an individual visualizations or view. The ‘Summary Dashboard’ tab is a Tableau dashboard, which provides a collection of all sheet views.

#### A4 – SQL Code

Once Tableau is connected to PostgreSQL, the following SQL query code selects the data from the following tables in order to view within the business tool:

- Admission:
  - `SELECT "admission"."admins_id" AS "admins_id",  
CAST("admission"."initial_admission" AS TEXT) AS "initial_admission" FROM  
"public"."admission" "admission"`
- Patient
  - `SELECT "patient"."additional_charges" AS "additional_charges",  
patient"."admis_id" AS "admis_id", "patient"."age" AS "age", patient"."children"  
AS "children", "patient"."compl_id" AS "compl_id", "patient"."doc_visits" AS  
“doc_visits”, "patient"."full_meals" AS "full_meals", CAST("patient"."gender"  
AS TEXT) AS "gender", CAST("patient"."highblood" AS TEXT) AS  
"highblood", "patient"."income" AS "income", "patient"."initial_days" AS  
"initial_days", "patient"."job_id" AS "job_id", "patient"."lat" AS "lat",  
"patient"."lng" AS "lng", "patient"."location_id" AS "location_id",  
CAST("patient"."marital" AS TEXT) AS "marital", CAST("patient"."patient_id"  
AS TEXT) AS "patient_id", "patient"."population" AS "population",`

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

- **Complication**

- SELECT "complication"."complication\_id" AS "complication\_id",  
CAST("complication"."complication\_risk" AS TEXT) AS "complication\_risk"  
FROM "public"."complication" "complication"

- **Job**

- SELECT "job"."job\_id" AS "job\_id (job)", CAST("job"."job\_title" AS TEXT)  
AS "job\_title" FROM "public"."job" "job"

- **Location**

- SELECT CAST("location"."city" AS TEXT) AS "city",  
CAST("location"."county" AS TEXT) AS "county", "location"."location\_id" AS  
"location\_id (location)", CAST("location"."state" AS TEXT) AS "state",  
"location"."zip" AS "zip"  
FROM "public"."location" "location"

- **Services Add-on**

- SELECT CAST("servicesaddon"."allergic\_rhinitis" AS TEXT) AS  
"allergic\_rhinitis", CAST("servicesaddon"."anxiety" AS TEXT) AS "anxiety",  
CAST("servicesaddon"."arthritis" AS TEXT) AS "arthritis",  
CAST("servicesaddon"."asthma" AS TEXT) AS "asthma",  
CAST("servicesaddon"."backpain" AS TEXT) AS "backpain",



```

CAST("servicesaddon"."diabetes" AS TEXT) AS "diabetes",
CAST("servicesaddon"."hyperlipidemia" AS TEXT) AS "hyperlipidemia",
CAST("servicesaddon"."overweight" AS TEXT) AS "overweight",
CAST("servicesaddon"."patient_id" AS TEXT) AS "patient_id (servicesaddon)",
CAST("servicesaddon"."reflux_esophagitis" AS TEXT) AS "reflux_esophagitis",
CAST("servicesaddon"."services" AS TEXT) AS "services" FROM
"public"."servicesaddon" "servicesaddon"

```

- **Survey Responses Add-on**

- SELECT "survey\_responses\_addon"."item1" AS "item1",
 "survey\_responses\_addon"."item2" AS "item2",
 "survey\_responses\_addon"."item3" AS "item3",
 "survey\_responses\_addon"."item4" AS "item4",
 "survey\_responses\_addon"."item5" AS "item5",
 "survey\_responses\_addon"."item6" AS "item6",
 "survey\_responses\_addon"."item7" AS "item7",
 "survey\_responses\_addon"."item8" AS "item8",
 CAST("survey\_responses\_addon"."patient\_id" AS TEXT) AS "patient\_id
 (survey\_responses\_addon)" FROM "public"."survey\_responses\_addon"
 "survey\_responses\_addon"

## **B – Panopto Presentation**

The link to my Panopto multimedia presentation is located here:

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=789bae39-4871-4f3e-a163-aea60017c570>



**B1 – Technical Environment**

The technical environment is made up of six tables in PostgreSQL and one comma separated text file. Connecting to PostgreSQL, described in A2 – Installation Instructions, should be all that is necessary since the hosp\_ratings.csv is packaged within the Tableau workbook executable.

Tableau was used to create worksheets and a summary dashboard. The worksheets combine to form the dashboard and provide a convenient view to see all visualizations and draw conclusions.

**B2 – Demonstrate Dashboard Functionality**

The presentation demonstrates the project dashboard's functionality.

**B3 – SQL Scripts**

Section A4 – SQL Code has the scripts used to create the dashboards. The Panopto presentation will go over what the scripts are doing. An example, they all tend to select the data fields needed to view, some cast fields into text data types, and explicitly lists the table the data is requested from.

**B4 – Data Streams**

Data streams were connected via the data source tab. Seven tables were connected from PostgreSQL and hosp\_ratings.csv was connected directly. The data field joins are listed in B5 – Data Points.

**B5 – Data Points**

Data was joined on the following fields and tables:

- Admissions table was joined on the Admin Id field to the patient table
- The Patient table joins:
  - The complication table where compl id = Complication Id
  - The job table where job id = job id
  - The location table where location id = location id
  - The service add-on table where Patient Id = Patient Id
  - The survey response add-on where Patient Id = Patient Id
- The locations table joined the hosp\_ratings table where Zip = ZIP Code

## B6 – Data Base Creation

Please see section A4 to understand how the databases were created.

## B7 – Referential Integrity

Within the provided assignment folder is an ERD image file. The ERD, which is covered in the presentation, shows that each primary and foreign key are mapped. For example, in the SQL code to create the admission table, a primary key is created (Figure 5). Viewing snippets from the ERD, you can see there is a one-to-many relationship from public.admission.admins\_id to public.patient.admins\_id.

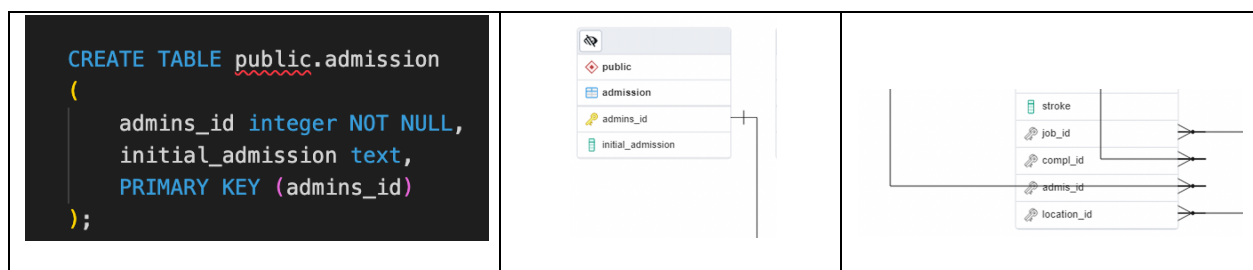


Figure 5 - Admissions Code, Table PK View and Patient Table FK View

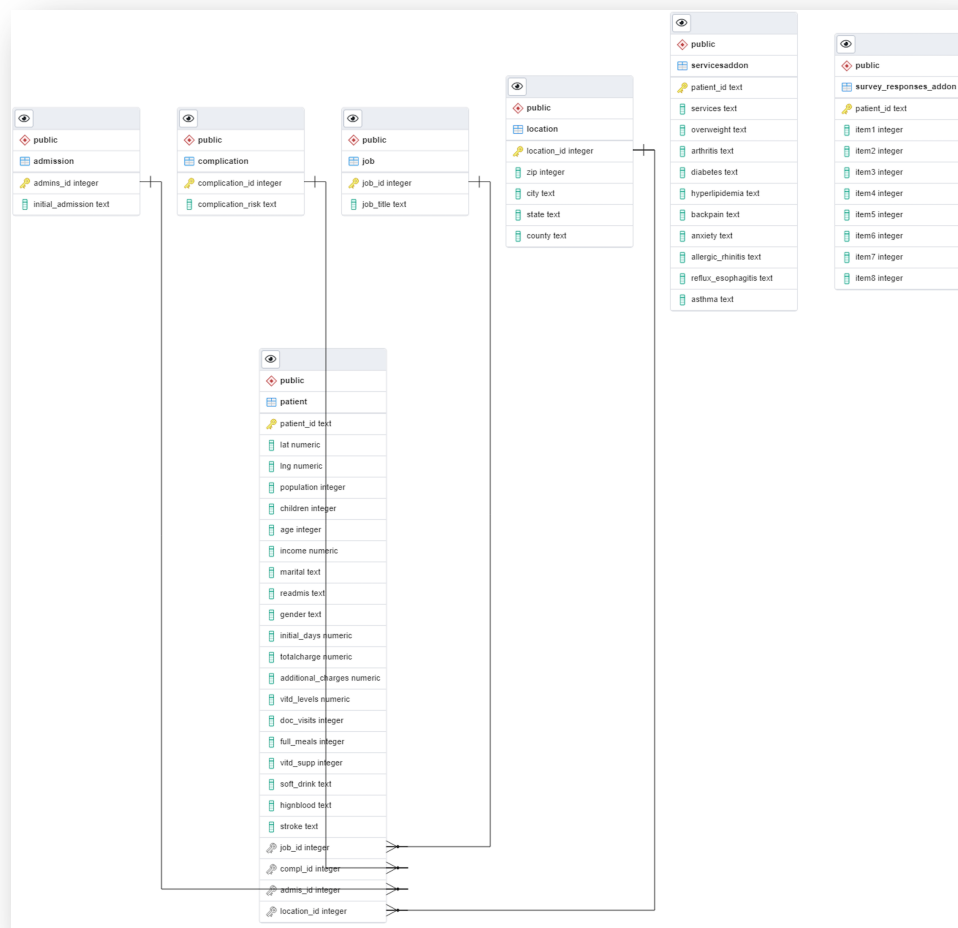


Figure 6 - ERD Full View

## C – Written Report

Please see section C4 for data exploration, section A4 for advanced SQL operations, and section C5 for data analysis.

### C1 – Dashboard Alignment

The dashboard meets certain decision-making needs for broad observations; yet, allows the user to drill into data by geolocation for a base view of the data which encompasses specific

state/regions. As such, using the interactive controls, seen in Figure 2, within the summary dashboard can help isolate by singular or multiple States.

### **C2 – Business Intelligence Tool**

Tableau is a sound business intelligence tool. Selecting and importing datasets was quite simple, even from multiple sources. I've been able to perform early level data exploration, simple data wrangling, SQL joins were automagically performed in most cases by Tableau as well! The tool performs many low-level tasks and gets out of your way while you can focus on displaying visualizations for your audience.

### **C3 – Data Cleaning**

Some data cleaning was needed to remove null values from displaying. This was rather simple within the worksheets.

### **C4 – Dashboard Creation**

Summary dashboard creation is relatively intuitive. Once you have created the worksheets, simply choose which ones you would like to include. Then, it was a matter of tweaking the layout for symmetry.

### **C5 – Data Analysis Results**

The submission should be accurate, logical and complete both in personal viewing and with the help of my presentation providing a walkthrough.

### **C6 – Analysis Limitations**

Since zip codes were used to join the two data sets, limitations in the data derived from the availability in information from linked zip codes. Additionally, we compare to national ratings, but they don't really provide insight as to "why" a rating is given.

**D – Web Sources**

Please see the sources and references section hereafter.

**E – Sources**

References are located on the last page, in-text citations are provided throughout the report.

Additionally, the following sources were used not as a literary reference but one to understand how to use Tableau: [LinkedIn Learning](#) and [DataCamp](#).

### References

- Bouchet, J., GSD, Sujval, R. (2022). Hospital Ratings: The official Dataset Used on Medicare.gov for Hospital Quality Comparison. Kaggle.  
<https://www.kaggle.com/datasets/center-for-medicare-and-medicaid/hospital-ratings>
- Gert P Westert, Ronald J Lagoe, Ilmo Keskimäki, Alastair Leyland, Mark Murphy,  
An international study of hospital readmissions and related utilization in Europe and the  
USA, *Health Policy*, Volume 61, Issue 3, 2002, Pages 269-278, ISSN 0168-8510,  
[https://doi.org/10.1016/S0168-8510\(01\)00236-6](https://doi.org/10.1016/S0168-8510(01)00236-6).  
<https://www.sciencedirect.com/science/article/pii/S0168851001002366>
- pgAdmin Documentation. (2022) pgAdmin.org Website. Retrieved From  
[https://www.pgadmin.org/docs/pgadmin4/latest/erd\\_tool.html](https://www.pgadmin.org/docs/pgadmin4/latest/erd_tool.html)
- Schuller K. A. (2020). Is obesity a risk factor for readmission after acute myocardial  
infarction? *Journal of healthcare quality research*, 35(1), 4–11.  
<https://doi.org/10.1016/j.jhqr.2019.09.002>
- Tableau Desktop and Web Authoring Help. (2022) Tableau Website. Retrieved From  
<https://help.tableau.com/current/pro/desktop/en-us/default.htm>