

**Performance Assessment: Advanced Data Acquisition**

Gabriela Howell

Master of Science Data Analytics, Western Governors University

D211 – Advanced Data Acquisition

Professor Sewell

July 19, 2024

**Part 1: Data Dashboards****A. Provide a copy of your dashboard that supports executive decision-making.**

A copy of the Dashboard workbook is attached with the submissions called :  
PA\_Tableau\_workbook.twbx

**1. Provide both datasets that serve as the data source for the dashboard.**

The two datasets I used are “medical\_clean” from WGU and “CDC\_Dataset” from the CDC. Both are included in the folder provided with the upload.

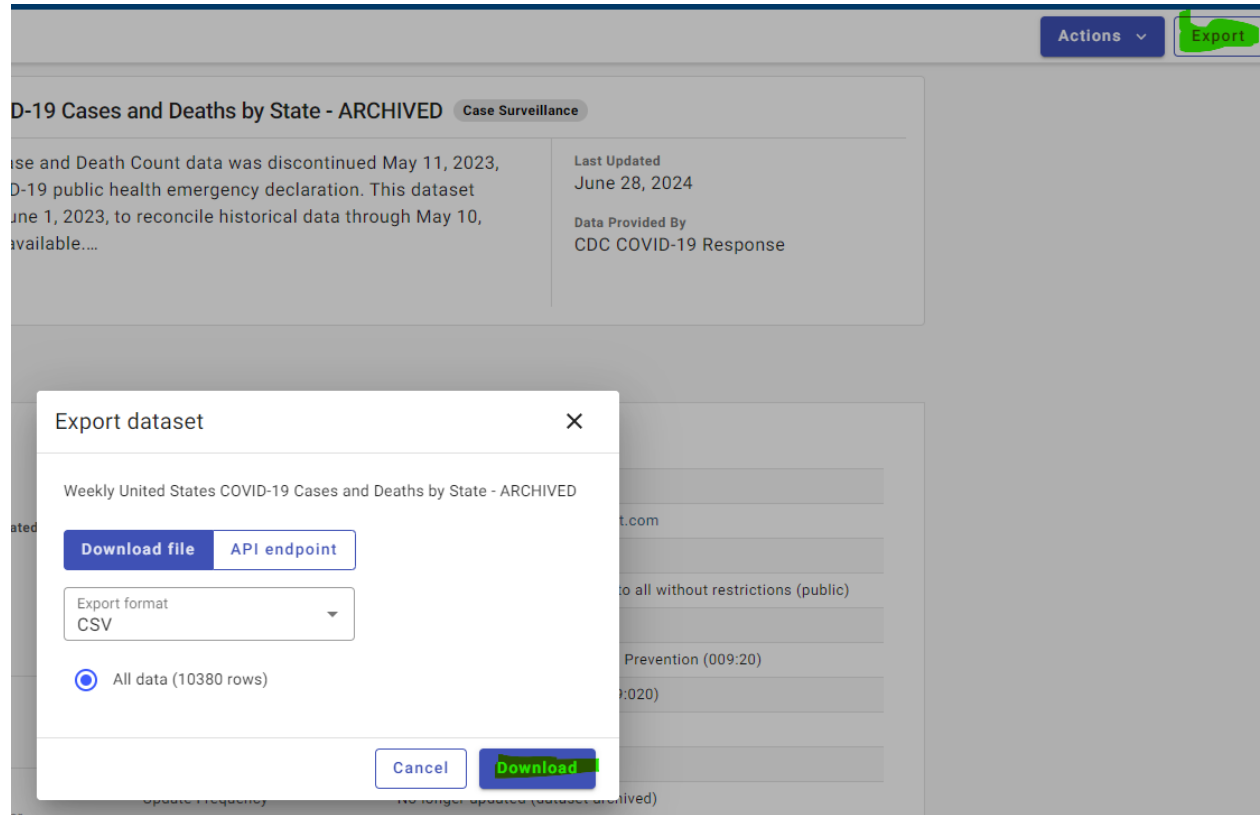
**2. Provide step-by-step instructions on how to have the dashboard appear fully operational on a default version in Labs on Demand.****1. Open Labs on Demand:**

- Access the Labs on Demand platform.

**2. Download External Data:**

- Open this [link](#) to reach the external data.
- Select the “Export” option at the top right.
- When the popup appears, select “Download.”

- Ensure the CSV file is saved in C:\Users\LabUsers\Downloads.



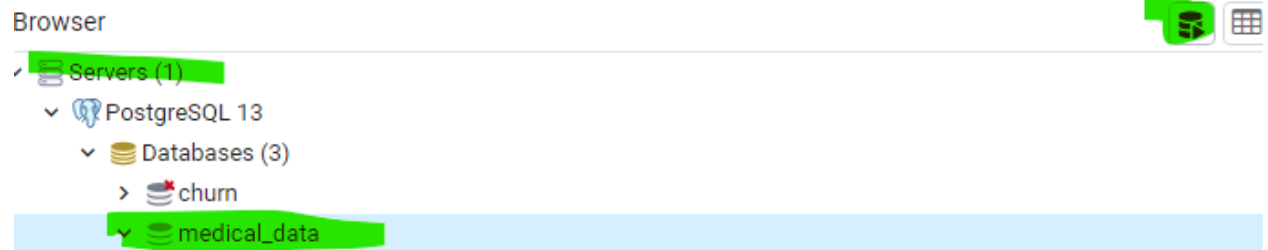
### 3. Prepare the SQL Script:

- Locate the included “postgres.txt” script and have it ready.

### 4. Open pgAdmin4:

- Connect to the Database.
- Click on medical\_data.
- Click on the database symbol with a start arrow (hover to see “Query Tool”).
- Open the postgres.txt script in the Query Tool.

- Run each step of the SQL script.



## 5. Import Data:

- When it comes to importing the data in Step 3 of the script:
  - Run the query.
  - If cdc\_data\_staging does not appear, right-click and refresh.
  - Right-click on cdc\_data\_staging and select “Import/Export.”
  - Select the “Import” toggle at the top.
  - Select the file from C:\Users\LabUser\Downloads\Weekly\_United\_States\_COVID-19\_Cases\_and\_Deaths\_by\_State\_-\_ARCHIVED\_20240728.csv.
  - Set the header toggle to “Yes” and the delimiter to a comma “,”.
  - Click “OK.”

Import/Export data - table 'cdc\_data\_staging'

Options Columns

Import/Export **Import**

**File Info**

Filename C:\Users\LabUser\Downloads\Weekly\_United\_States\_COVID-19\_Ca ...

Format CSV

Encoding Select an item...

**Miscellaneous**

OID No

Header Yes

Delimiter ,

Specifies the character that separates columns within each row (line) of the file. The default is a tab character in text format, a comma in CSV format. This must be a single one-byte character. This option is not allowed when using binary format.

Cancel OK

## 6. Verify and Complete Data Import:

- Verify that the data is in the staging table.
- Proceed with steps 5 to 10 of the SQL script, running each query one at a time.
- Refresh the tables in medical\_data to ensure cdc\_data\_staging is gone, and only cdc\_data remains.

## 7. Open the Tableau Dashboard:

- Open the provided Tableau called "PA\_Tableau\_workbook.twbx" file.
- When prompted, log in with:
  - Username: postgres
  - Password: Passw0rd!
- Wait for the dashboard to load.

- You should see the dashboard titled “ Healthcare and Covid-19 Statewide Doctor Visits and Death Rate.”

### **3. Provide step-by-step instructions to help users navigate the dashboard.**

#### **1. Opening Tableau:**

- Ensure you are on the Dashboard tab, “Healthcare and Covid-19 Statewide Doctor Visits and Death Rate.”

#### **2. Interacting with the Dashboard:**

- COVID-19 Deaths by State and Year: Use the dropdown menu below the “Deaths by Year” section or select the desired year from the “Deaths by Year” bar chart. This will adjust the map to display COVID-19 death data for different states in the “Deaths by State” section.
- Doctor Visits by State: In the ‘Doc Visits by Population’ section, click on one of the population bars in the bar chart. This will filter the chart to show only the selected bin, and the U.S. map will update to reflect the count of doctor visits by population. Hover over state circles to view the average doctor visits and total population sum for each state.

#### **3. Exploring Data Representations:**

- Charts and Graphs: Hover over any chart or graph to see detailed data points.

- Legends: Use the legends next to each visualization to understand the color coding and categories.

#### **4. Navigating Between Views:**

- Tabs: If there are multiple tabs on the top, click on each tab to switch between different views and data representations.

#### **5. Interacting with KPIs:**

- Metrics Display: Review the key performance indicators (KPIs) displayed at the top of the dashboard for quick insights.
- Detailed View: Click on any KPI tab for a more detailed breakdown of the data.

#### **6. Exporting and Sharing Data:**

- Download Options: Use the download button (usually in the upper-right corner) to export data visualizations or reports as needed.
- Share dashboard: Adjacent to the download button, you'll find a three-circle arrow symbol. Click on it to copy the link, use the letter icon for emailing, and share via Twitter or Facebook.

#### **7. Help and Support:**

- Tooltips: Hover over question marks or information icons for additional guidance and explanations.

**4. Provide all SQL code or other code supporting the dashboard in text format.**

--From pgAdmin

--If run into issues please restart and drop tables

DROP TABLE public.cdc\_data\_staging;

DROP TABLE public.cdc\_data;

-- Step 1: Create Final Table

```
CREATE TABLE public.cdc_data ( id SERIAL PRIMARY KEY, date_updated DATE, state VARCHAR(2),
start_date DATE, end_date DATE, tot_cases INT CHECK (tot_cases >= 0), new_cases INT CHECK (new_cases
>= 0), tot_deaths INT CHECK (tot_deaths >= 0), new_deaths INT CHECK (new_deaths >= 0),
new_historic_cases INT CHECK (new_historic_cases >= 0), new_historic_deaths INT CHECK
(new_historic_deaths >= 0) );
```

-- Step 2: Create Staging Table

```
CREATE TABLE public.cdc_data_staging ( date_updated VARCHAR(10), state VARCHAR(3), start_date
VARCHAR(10), end_date VARCHAR(10), tot_cases INT, new_cases INT, tot_deaths INT, new_deaths INT,
new_historic_cases INT, new_historic_deaths INT );
```

-- Step 3: Load Data into Staging Table



```
--command " " "\\copy public.cdc_data_staging (date_updated, state, start_date, end_date, tot_cases,  
new_cases, tot_deaths, new_deaths, new_historic_cases, new_historic_deaths) FROM  
'C:/Users/LabUser/DOWNLO~1/WEEKLY~1.CSV' DELIMITER ',' CSV HEADER QUOTE '\"' ESCAPE '\"','\"'
```

-- Step 4: Verify Data in Staging Table

```
SELECT * FROM public.cdc_data_staging LIMIT 10;
```

-- Step 5: Remove Non-US States

-- Select unique states that are not in the list of U.S. state abbreviations

```
WITH us_states AS ( SELECT 'AL' AS state UNION ALL SELECT 'AK' UNION ALL SELECT 'AZ' UNION ALL  
SELECT 'AR' UNION ALL SELECT 'CA' UNION ALL SELECT 'CO' UNION ALL SELECT 'CT' UNION ALL SELECT  
'DE' UNION ALL SELECT 'FL' UNION ALL SELECT 'GA' UNION ALL SELECT 'HI' UNION ALL SELECT 'ID'  
UNION ALL SELECT 'IL' UNION ALL SELECT 'IN' UNION ALL SELECT 'IA' UNION ALL SELECT 'KS' UNION  
ALL SELECT 'KY' UNION ALL SELECT 'LA' UNION ALL SELECT 'ME' UNION ALL SELECT 'MD' UNION ALL  
SELECT 'MA' UNION ALL SELECT 'MI' UNION ALL SELECT 'MN' UNION ALL SELECT 'MS' UNION ALL  
SELECT 'MO' UNION ALL SELECT 'MT' UNION ALL SELECT 'NE' UNION ALL SELECT 'NV' UNION ALL  
SELECT 'NH' UNION ALL SELECT 'NJ' UNION ALL SELECT 'NM' UNION ALL SELECT 'NY' UNION ALL SELECT  
'NC' UNION ALL SELECT 'ND' UNION ALL SELECT 'OH' UNION ALL SELECT 'OK' UNION ALL SELECT 'OR'  
UNION ALL SELECT 'PA' UNION ALL SELECT 'RI' UNION ALL SELECT 'SC' UNION ALL SELECT 'SD' UNION  
ALL SELECT 'TN' UNION ALL SELECT 'TX' UNION ALL SELECT 'UT' UNION ALL SELECT 'VT' UNION ALL  
SELECT 'VA' UNION ALL SELECT 'WA' UNION ALL SELECT 'WV' UNION ALL SELECT 'WI' UNION ALL SELECT  
'WY' )
```

```
SELECT DISTINCT state FROM public.cdc_data_staging WHERE state NOT IN (SELECT state FROM  
us_states);
```

```
--Delete the Non-US States
```

```
WITH us_states AS ( SELECT 'AL' AS state UNION ALL SELECT 'AK' UNION ALL SELECT 'AZ' UNION ALL  
SELECT 'AR' UNION ALL SELECT 'CA' UNION ALL SELECT 'CO' UNION ALL SELECT 'CT' UNION ALL SELECT  
'DE' UNION ALL SELECT 'FL' UNION ALL SELECT 'GA' UNION ALL SELECT 'HI' UNION ALL SELECT 'ID'  
UNION ALL SELECT 'IL' UNION ALL SELECT 'IN' UNION ALL SELECT 'IA' UNION ALL SELECT 'KS' UNION  
ALL SELECT 'KY' UNION ALL SELECT 'LA' UNION ALL SELECT 'ME' UNION ALL SELECT 'MD' UNION ALL  
SELECT 'MA' UNION ALL SELECT 'MI' UNION ALL SELECT 'MN' UNION ALL SELECT 'MS' UNION ALL  
SELECT 'MO' UNION ALL SELECT 'MT' UNION ALL SELECT 'NE' UNION ALL SELECT 'NV' UNION ALL  
SELECT 'NH' UNION ALL SELECT 'NJ' UNION ALL SELECT 'NM' UNION ALL SELECT 'NY' UNION ALL SELECT  
'NC' UNION ALL SELECT 'ND' UNION ALL SELECT 'OH' UNION ALL SELECT 'OK' UNION ALL SELECT 'OR'  
UNION ALL SELECT 'PA' UNION ALL SELECT 'RI' UNION ALL SELECT 'SC' UNION ALL SELECT 'SD' UNION  
ALL SELECT 'TN' UNION ALL SELECT 'TX' UNION ALL SELECT 'UT' UNION ALL SELECT 'VT' UNION ALL  
SELECT 'VA' UNION ALL SELECT 'WA' UNION ALL SELECT 'WV' UNION ALL SELECT 'WI' UNION ALL SELECT  
'WY' )
```

```
DELETE FROM public.cdc_data_staging WHERE state NOT IN (SELECT state FROM us_states);
```

```
-- Step 6: Check for Invalid Dates
```

```
SELECT * FROM public.cdc_data_staging WHERE TO_DATE(date_updated, 'MM/DD/YYYY') IS NULL OR  
TO_DATE(start_date, 'MM/DD/YYYY') IS NULL OR TO_DATE(end_date, 'MM/DD/YYYY') IS NULL;
```

-- Check for Rows Violating Constraints

```
SELECT * FROM public.cdc_data_staging WHERE tot_cases < 0 OR new_cases < 0 OR tot_deaths < 0 OR  
new_deaths < 0 OR new_historic_cases < 0 OR new_historic_deaths < 0;
```

-- Remove Rows Violating Constraints

```
DELETE FROM public.cdc_data_staging WHERE tot_cases < 0 OR new_cases < 0 OR tot_deaths < 0 OR  
new_deaths < 0 OR new_historic_cases < 0 OR new_historic_deaths < 0;
```

-- Verify that no rows with invalid data remain

```
SELECT * FROM public.cdc_data_staging WHERE tot_cases < 0 OR new_cases < 0 OR tot_deaths < 0 OR  
new_deaths < 0 OR new_historic_cases < 0 OR new_historic_deaths < 0;
```

-- Step 7: Insert Data into Final Table

```
INSERT INTO public.cdc_data (date_updated, state, start_date, end_date, tot_cases, new_cases, tot_deaths,  
new_deaths, new_historic_cases, new_historic_deaths) SELECT TO_DATE(date_updated, 'MM/DD/YYYY'),  
state, TO_DATE(start_date, 'MM/DD/YYYY'), TO_DATE(end_date, 'MM/DD/YYYY'), tot_cases, new_cases,  
tot_deaths, new_deaths, new_historic_cases, new_historic_deaths FROM public.cdc_data_staging;
```

-- Step 8: Set the Owner of the Final Table

```
ALTER TABLE IF EXISTS public.cdc_data OWNER TO postgres;
```

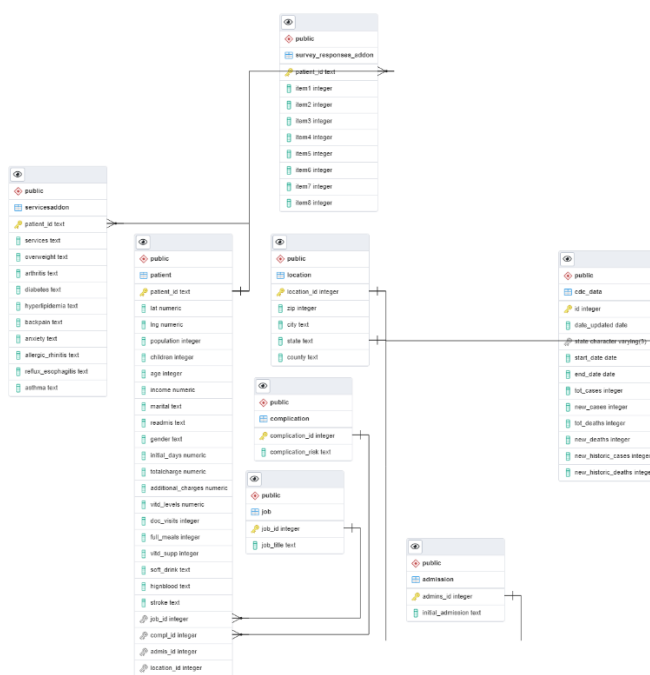
```
-- Step 9: Drop the Staging Table
```

```
DROP TABLE public.cdc_data_staging;
```

```
-- Step 10: Verify Data in Final Table
```

```
SELECT * FROM public.cdc_data LIMIT 10;
```

```
--ERD of the Medical DB
```



```
--From Tableau
```

```
SELECT "cdc_data"."date_updated" AS "date_updated",
```

```
"cdc_data"."end_date" AS "end_date",
```

```
"cdc_data"."id" AS "id",  
  
"cdc_data"."new_cases" AS "new_cases",  
  
"cdc_data"."new_deaths" AS "new_deaths",  
  
"cdc_data"."new_historic_cases" AS "new_historic_cases",  
  
"cdc_data"."new_historic_deaths" AS "new_historic_deaths",  
  
"cdc_data"."start_date" AS "start_date",  
  
"cdc_data"."state" AS "state",  
  
"cdc_data"."tot_cases" AS "tot_cases",  
  
"cdc_data"."tot_deaths" AS "tot_deaths"  
  
FROM "public"."cdc_data" "cdc_data"
```

## Part 2: Demonstration

### B. Provide a link to a Panopto:

The video is online and is located here on this link: [D211 Video GH](#)

## Part 3: Report

**C. Write a report to outline the exploration of the data, the use of advanced SQL operations, and the analysis of the data. Do the following as part of your report:**

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

The purpose of this dashboard is to track and display COVID-19 deaths across the United States by state, as well as doctor visits per population in each state. This aligns with the needs specified in the data dictionary by offering clear and actionable insights into critical public health metrics, which are critical for medical facilities. By focusing on key health indicators, the dashboard supports executives in monitoring and making data-driven decisions during health crises. Its emphasis on relevant metrics (COVID-19 deaths and doctor visits) and avoidance of unnecessary clutter ensures that the information is easily comprehensible for the audience (Knafllic, 2016).

**2. Justify the selection of the business intelligence tool you used.**

I chose Tableau because it's great for visualizing data and making it easy for people to understand. Tableau is user-friendly with its drag-and-drop design, which helps in showing and analyzing data clearly.

To get the data ready for Tableau, I used pgAdmin to import the CDC data and combine it into one source. This way, I could set up connections and join tables in Tableau to give a clear and complete view of the data.

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

The first step of cleaning and preparing the data involved several key tasks:

- Loading Data: Imported raw data into a staging table.

- Removing Non-US States: Deleted rows that did not correspond to valid US states.
- Handling Invalid Dates: Identified and removed rows with invalid date formats.
- Removing Negative Values: Deleted rows with negative values in key columns like tot\_cases and new\_deaths.

These steps ensured that the data was accurate, consistent, and ready for analysis use in Tableau.

#### **4. Summarize the steps used to create the dashboard.**

Steps to Create the Dashboard:

- Data Import: Loaded cleaned data from pgAdmin4 into Tableau.
- Data Integration: Established relationships between different tables within Tableau.
- Visualization Design: Created various visualizations, including bar charts, pie charts, and interactive maps.
- Dashboard Assembly: Combined individual visualizations into a cohesive dashboard layout.
- Interactivity: Added filters and interactive elements to allow users to explore data dynamically.

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

Data Analysis Results:

- COVID-19 Impact by State: Revealed state-specific trends in COVID-19 deaths, highlighting areas needing more resources.

- Doctor Visits Analysis: Showed patterns in healthcare utilization across different population groups.

The analysis supported the dashboard's purpose by providing actionable insights into healthcare trends and resource allocation needs. It enabled stakeholders to make data-driven decisions to improve health outcomes and manage the impact of COVID-19 effectively.

## **6. Discuss the limitation(s) of your data analysis.**

Data analysis has several limitations. First, the timeliness of the data can be an issue, as outdated information might affect the accuracy of real-time decisions. Additionally, data completeness is crucial; incomplete data can result in gaps in the analysis, which can impact the overall insights drawn. Lastly, depending on a single data source can restrict the breadth of the analysis, potentially leading to missed opportunities for incorporating other relevant datasets.

## **D. List the web source(s) used to acquire data or segments of third-party code to support the application. Ensure the web sources are reliable.**

Centers for Disease Control and Prevention. (n.d.-b). *Weekly United States covid-19 cases and deaths by state - archived*. Centers for Disease Control and Prevention.

[https://data.cdc.gov/Case-Surveillance/Weekly-United-States-COVID-19-Cases-and-Deaths-by-/pwn4-m3yp/about\\_data](https://data.cdc.gov/Case-Surveillance/Weekly-United-States-COVID-19-Cases-and-Deaths-by-/pwn4-m3yp/about_data)

Connect to a Custom SQL Query. (n.d.).

Tableau.<https://help.tableau.com/current/pro/desktop/en-us/customsql.htm>



Lacroix, H. (n.d.). Analyzing Data in Tableau. Datacamp. from

<https://app.datacamp.com/learn/courses/analyzing-data-in-tableau>

Page, D. (n.d.). ERD Tool — pgAdmin 4 7.0 documentation.

[https://www.pgadmin.org/docs/pgadmin4/development/erd\\_to\\_ol.html](https://www.pgadmin.org/docs/pgadmin4/development/erd_to_ol.html)

**E. Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.**

Knafllic, C. N. (2015). Storytelling with data: A data visualization guide for business

professionals. Wiley .(Chapter 1, Chapters 3-5, Chapters 7 - 8).

<https://ebookcentral.proquest.com/lib/westerngovernorsebooks/reader.action?docID=4187267&ppg=1>