DATA DASHBOARDS

DATASETS USED IN THIS DASHBOARD

NCHS - Leading Causes of Death: United States: https://catalog.data.gov/dataset/nchs-leading-causes-of-death-united-states **(File attached as: deaths_clean.csv)**

STEP-BY-STEP INSTRUCTIONS TO INSTALL DASHBOARD FILE ON LABS ON DEMAND

1. Open Labs on Demand
2. Obtain LOD_files.zip. Extract all contents of the .zip file on the Labs on Demand desktop.
3. Launch pgAdmin4. Expand Servers, then expand medical_data. Expand Schemas, then Public and locate the **Tables**.
4. Right-click on the Tables, then click the **Query Tool**.
5. From the extracted LOD_files, drag cleaning.sql into the Query Editor. It should populate with SQL codes. If none appeared, open cleaning.sql in WordPad, then copy & paste SQL code into Query Editor. Then, execute the code.
6. Right-click on the Tables and click **Refresh**. A new table called leading_causes should be added to the Tables.
7. Right-click on leading_causes table. Click **Import/Export…**
8. Make sure the option is on Import by toggling the Import/Export button. On Filename, select the deaths_clean.csv from the extracted LOD_files. Toggle Yes for header.
9. Launch Tableau 2021.4 found in the Labs on Demand desktop.
10. Download or obtain the Tableau Dashboard file. In this case, the name of the file is D211_rraneses.twbx from the extracted LOD_files.
11. In Tableau, click **Open A Workbook**. Navigate to the location of D211_rraneses.twbx. Then, click the workbook file.
12. A pop-up will appear that establishes a connection to the PostgreSQL database. Input Username as **postgres** and Password as **Passw0rd!** (the ! is included). Then, click Sign In.

HOW TO NAVIGATE THE DASHBOARD

1. Average Total Charges by Gender
   - This visualization displays the average total charges incurred by patients, categorized by gender.
   - To interpret this visualization, look at the table with the values in dollars. It is also color-coded to have dark orange as the highest charge and light blue as the lowest charge.
2. Services Utilization
   - This visualization shows the utilization of different services provided to patients.
   - Each service is represented by a segment of a pie chart. The pie chart shows the specific value of utilization of that service.
3. Prevalent Medical Conditions
   - This visualization ranks the prevalent medical conditions in descending order based on their frequency.

- The height of each bar indicates the frequency of that medical condition. The graph is also color-coded in gradient where yellow has the highest frequency and dark green has the lowest frequency.
- The user can identify the most prevalent medical conditions by looking at the highest bar.

4. Leading Cause of Death
   - This visualization displays the top causes of death, ranked in descending order based on their prevalence.
   - Each cause of death is represented by a bar, and the bar's height indicates the cause's frequency. The graph is also color-coded in gradient where light blue has the highest frequency has the highest frequency and dark blue has the lowest frequency.
   - The user can identify the most common causes of death by examining the top-ranking bars.
5. Leading Causes by State
   - This visualization shows each state's leading cause of death, providing insights into regional variations in mortality patterns.
   - The user can observe how the leading causes of death differ across states and identify common trends or disparities.
   - The cause of death is represented by the different color bars, with blue assigned to cancer and orange assigned to heart disease.
   - Hover over the bar to view the total count of the leading cause of that state.
6. Average Age-Adjusted Death Rate per Year
   - This visualization displays the average age-adjusted death rate over time, allowing users to analyze trends in mortality rates.
   - The x-axis represents the years, and the y-axis represents the average age-adjusted death rate.
   - The user can observe how the death rate changes and identify significant fluctuations or patterns.
7. Most Utilized Service vs. the Leading Cause of Death per State
   - This visualization shows how service utilization relates to the leading cause of death.
   - The user can identify regions where certain services are more commonly utilized and where specific causes of death are more prevalent.
   - The dashboard contains two maps with each representing one of the most utilized services. The colors of each state on each map represent the leading cause of death, with blue indicating cancer and orange indicating heart disease.
   - The user can start by selecting one of the two maps to focus on a specific service: blood work or intravenous. Each map will display the corresponding service’s utilization across different states.
   - The user can hover over each state to view the state’s name, the most utilized service, and the leading cause of death.

SQL CODES

CODE TO GENERATE ERD

```
-- This script was generated by a beta version of the ERD tool in pgAdmin 4.
-- Please log an issue at
https://redmine.postgresql.org/projects/pgadmin4/issues/new if you find any
bugs, including reproduction steps.
BEGIN;


CREATE TABLE public.admission
(
    admins_id integer NOT NULL,
    initial_admission text,
    PRIMARY KEY (admins_id)
);

CREATE TABLE public.complication
(
    complication_id integer NOT NULL,
    complication_risk text,
    PRIMARY KEY (complication_id)
);

CREATE TABLE public.job
(
    job_id integer NOT NULL,
    job_title text,
    PRIMARY KEY (job_id)
);

CREATE TABLE public.leading_causes
(
    year integer,
    "113cause" character varying(255),
    cause character varying(255),
    state character varying(255),
    deaths integer,
    age_adjusted numeric,
    state_initials character varying(2),
    PRIMARY KEY (state_initials)
);

CREATE TABLE public.location
(
    location_id integer NOT NULL,
    zip integer,
    city text,
    state text,
```

```
    county text,
    PRIMARY KEY (location_id)
);

CREATE TABLE public.patient
(
    patient_id text NOT NULL,
    lat numeric,
    lng numeric,
    population integer,
    children integer,
    age integer,
    income numeric,
    marital text,
    readmis text,
    gender text,
    initial_days numeric,
    totalcharge numeric,
    additional_charges numeric,
    vitd_levels numeric,
    doc_visits integer,
    full_meals integer,
    vitd_supp integer,
    soft_drink text,
    highblood text,
    stroke text,
    job_id integer,
    compl_id integer,
    admis_id integer,
    location_id integer,
    PRIMARY KEY (patient_id)
);

CREATE TABLE public.servicesaddon
(
    patient_id text NOT NULL,
    services text,
    overweight text,
    arthritis text,
    diabetes text,
    hyperlipidemia text,
    backpain text,
    anxiety text,
    allergic_rhinitis text,
    reflux_esophagitis text,
    asthma text,
    PRIMARY KEY (patient_id)
);

CREATE TABLE public.survey_responses_addon
```

```
(
    patient_id text NOT NULL,
    item1 integer,
    item2 integer,
    item3 integer,
    item4 integer,
    item5 integer,
    item6 integer,
    item7 integer,
    item8 integer,
    PRIMARY KEY (patient_id)
);

ALTER TABLE public.patient
    ADD FOREIGN KEY (admis_id)
    REFERENCES public.admission (admins_id)
    NOT VALID;


ALTER TABLE public.patient
    ADD FOREIGN KEY (compl_id)
    REFERENCES public.complication (complication_id)
    NOT VALID;


ALTER TABLE public.patient
    ADD FOREIGN KEY (job_id)
    REFERENCES public.job (job_id)
    NOT VALID;


ALTER TABLE public.patient
    ADD FOREIGN KEY (location_id)
    REFERENCES public.location (location_id)
    NOT VALID;


ALTER TABLE public.leading_causes
    ADD FOREIGN KEY (state_initials)
    REFERENCES public.location (state)
    NOT VALID;


ALTER TABLE public.servicesaddon
    ADD FOREIGN KEY (patient_id)
    REFERENCES public.patient (patient_id)
    NOT VALID;


ALTER TABLE public.survey_responses_addon
```

```
    ADD FOREIGN KEY (patient_id)
    REFERENCES public.patient (patient_id)
    NOT VALID;

END;
```

CODE USED IN PGADMIN4

```
-- more cleaning
-- change prefer not to answer to non-binary
UPDATE patient
SET gender = CASE
        WHEN gender = 'Prefer not to answer' THEN 'Non-binary'
        ELSE gender
        END;

-- correct spelling to highblood
ALTER TABLE patient
RENAME COLUMN hignblood to highblood;

-- make a table for the external dataset
CREATE TABLE leading_causes (
   Year INTEGER,
   "113cause" VARCHAR(255),
   Cause VARCHAR(255),
   State VARCHAR(255),
   Deaths INTEGER,
   age_adjusted NUMERIC,
   State_Initials VARCHAR(2)
);

-- WGU KPIs
-- Average Total Charges + Additional Charges by Gender
SELECT
   Gender,
   AVG(TotalCharges) AS "Average Total Charges"
FROM (
   SELECT
```

```
    Gender,
    (TotalCharge + Additional_charges) AS TotalCharges
  FROM
    patient
) as TotalCharges
GROUP BY Gender
ORDER BY "Average Total Charges" DESC;

-- Services Utilization
SELECT Services, COUNT(Services) AS "Service Utilization"
FROM servicesaddon
GROUP BY Services
ORDER BY "Service Utilization" DESC;

-- Most Prevalent Medical Condition
SELECT
  'highblood' AS Condition,
  SUM(CASE WHEN highblood = 'Yes' THEN 1 ELSE 0 END) AS Yes_Count
FROM
  patient
UNION ALL
SELECT
  'Stroke' AS Condition,
  SUM(CASE WHEN Stroke = 'Yes' THEN 1 ELSE 0 END) AS Yes_Count
FROM
  patient
UNION ALL
SELECT
  'Overweight' AS Condition,
  SUM(CASE WHEN Overweight = 'Yes' THEN 1 ELSE 0 END) AS Yes_Count
FROM
  servicesaddon
UNION ALL
SELECT
  'Arthritis' AS Condition,
```

```
    SUM(CASE WHEN Arthritis = 'Yes' THEN 1 ELSE 0 END) AS Yes_Count
FROM
    servicesaddon
UNION ALL
SELECT
    'Diabetes' AS Condition,
    SUM(CASE WHEN Diabetes = 'Yes' THEN 1 ELSE 0 END) AS Yes_Count
FROM
    servicesaddon
UNION ALL
SELECT
    'Hyperlipidemia' AS Condition,
    SUM(CASE WHEN Hyperlipidemia = 'Yes' THEN 1 ELSE 0 END) AS Yes_Count
FROM
    servicesaddon
UNION ALL
SELECT
    'BackPain' AS Condition,
    SUM(CASE WHEN BackPain = 'Yes' THEN 1 ELSE 0 END) AS Yes_Count
FROM
    servicesaddon
UNION ALL
SELECT
    'Anxiety' AS Condition,
    SUM(CASE WHEN Anxiety = 'Yes' THEN 1 ELSE 0 END) AS Yes_Count
FROM
    servicesaddon
UNION ALL
SELECT
    'Allergic_rhinitis' AS Condition,
    SUM(CASE WHEN Allergic_rhinitis = 'Yes' THEN 1 ELSE 0 END) AS Yes_Count
FROM
    servicesaddon
UNION ALL
SELECT
```

```
    'Reflux_esophagitis' AS Condition,
    SUM(CASE WHEN Reflux_esophagitis = 'Yes' THEN 1 ELSE 0 END) AS Yes_Count
FROM
    servicesaddon
UNION ALL
SELECT
    'Asthma' AS Condition,
    SUM(CASE WHEN Asthma = 'Yes' THEN 1 ELSE 0 END) AS Yes_Count
FROM
    servicesaddon
ORDER BY
    Yes_Count DESC;

-- CDC KPI
-- Leading Cause of Death
SELECT cause, SUM(Deaths) AS "Total Deaths"
FROM leading_causes
WHERE cause != 'All causes'
GROUP BY cause
ORDER BY "Total Deaths" DESC;

-- Leading Cause per State
SELECT state, cause, total_deaths
FROM (
    SELECT state, cause, SUM(deaths) AS total_deaths,
        ROW_NUMBER() OVER (PARTITION BY state ORDER BY SUM(deaths) DESC) AS rank
    FROM leading_causes
            WHERE cause != 'All causes'
    GROUP BY state, cause
) as ranked_causes
WHERE rank = 1
ORDER BY total_deaths desc;

-- Average Age Adjusted per Year
SELECT year, ROUND(AVG(age_adjusted), 2) AS "Average Age Adjusted Death Rate"
```

```
FROM leading_causes
GROUP BY Year
ORDER BY Year;

-- use both datasets
-- most utilized services vs the leading cause of deaths per state
WITH most_utilized_services AS (
   SELECT l.state, s.services,
        ROW_NUMBER() OVER (PARTITION BY l.state ORDER BY COUNT(*) DESC) AS service_rank
   FROM location l
   JOIN patient p ON l.location_id = p.location_id
   JOIN servicesaddon s ON p.patient_id = s.patient_id
   GROUP BY l.state, s.services
),
leading_causes_per_state AS (
   SELECT state_initials, cause,
        ROW_NUMBER() OVER (PARTITION BY state_initials ORDER BY SUM(deaths) DESC) AS cause_rank
   FROM leading_causes
   WHERE cause != 'All causes'
   GROUP BY state_initials, cause
)
SELECT m.state,
      m.services AS most_utilized_service,
      lc.cause AS leading_cause_of_death
FROM most_utilized_services m
JOIN leading_causes_per_state lc ON m.state = lc.state_initials
WHERE m.service_rank = 1 AND lc.cause_rank = 1;
```

REFLECTION PAPER

DASHBOARD ALIGNMENT

The purpose and function of the dashboard align closely with the needs of the stakeholders interested in understanding mortality patterns and healthcare utilization across different demographics and regions.

The Average Total Charges by Gender aims to understand the financial impact of healthcare services based on gender. This provides a clear comparison of the average total charges incurred by patients. Stakeholders can identify any disparities in healthcare costs between genders. The Services Utilization analyzes the utilization of different healthcare services. This presents a visual representation of service

utilization through a pie chart. Stakeholders can quickly grasp the distribution of services utilized and identify any areas of high or low utilization. Prevalent Medical Conditions can identify the most common medical conditions. This ranks medical conditions by frequency, providing stakeholders with insights into prevalent health issues. The Leading Cause of Death helps to understand the primary causes of mortality. This allows stakeholders to prioritize areas for intervention based on the frequency of specific causes. The Leading Causes by State helps to explore the regional variations in mortality patterns. This enables stakeholders to understand regional disparities and tailor healthcare interventions accordingly. The Average Age-Adjusted Death Rate makes it easier to analyze trends in mortality rates over time. This provides a historical perspective on the death rate, which helps in the identification of trends and anomalies. The Most Utilized Service vs Leading Cause of Death per State investigates the relationship between service utilization and mortality patterns at the state level. This enables stakeholders to visualize how service utilization correlates with leading causes of death across different states. Stakeholders can identify the potential areas for healthcare resource allocation.

Overall, the dashboard serves as a comprehensive tool for stakeholders to gain insights and allow them to make informed decision-making and resource allocations based on data-driven insights.

BUSINESS INTELLIGENCE TOOL

Tableau, the primary tool used for this analysis, is an ideal business intelligence tool for multiple reasons.

Firstly, the tool has a user-friendly interface that makes it easy for users to create interactive visualizations and dashboards without requiring extensive technical expertise. The drag-and-drop functionality and intuitive design make it accessible to users across various skill levels.

Furthermore, as seen in this analysis, Tableau offers seamless connectivity to various data sources such as PostgreSQL. More data sources also include databases, spreadsheets, and cloud services. This allows users to access and analyze data from multiple sources within a single platform.

Lastly, Tableau provides a wide range of visualizations, including but not limited to bar charts, line charts, scatter plots, and maps. Users can create visually compelling and insightful visualizations to represent complex data in a clear and understandable format.

DATA CLEANING

To clean the external dataset from the CDC, I used Python. First, I used exploratory data analysis to check the info(), head(), and is.null() to find the missing values in the dataset. I noticed that the column ‘113 Cause Name’ has values that are serial numbers in parentheses not necessary for the analysis. Using the re module, I removed the parentheses and the values inside the column ‘113 Cause Name’. Additionally, the external dataset has the states’ full names as opposed to the medical_data states that only have the state initials. My solution is to define a dictionary of state_initials that maps full state names to their respective state initials. Then, I added a new column, ‘State_initials,’ to the data frame based on the mapping done. Lastly, I removed rows where the ‘State’ column contained the value ‘United States’. This row was not necessary for the analysis and made the calculations incorrect.

Some minor cleaning for the internal dataset was done in pgAdmin4. First, I updated the gender column in the patient table. Every occurrence of Prefer not to answer was changed to Non-binary. Then, I corrected the column name of hignblood to highblood in the patient table. Lastly, I created a table for

leading causes with the specified columns to store the external dataset. The columns match the fields in the external dataset to ensure compatibility when importing the data.

DASHBOARD CREATION

The 1st dashboard consists of three sheets. Utilize multiple custom SQL queries to calculate the average overall charges incurred by patients (categorized by gender), the utilization numbers of different services provided to patients, and the frequency of prevalent medical conditions. Create a table visualization for average overall charges, a pie chart for the service utilization, and a bar chart for the prevalent medical conditions. After all sheets are made, combine them into the 1st dashboard. Make sure to add the labels and legends.

The 2nd dashboard also consists of three sheets. Utilize multiple custom SQL queries to determine the top causes of death and their prevalence, determine the leading cause of death for each state, and calculate the average age-adjusted death rate over time. Create bar charts for the leading cause of death and leading cause by state and a line chart for the average age-adjusted death rate. After all sheets are made, combine them into the 2nd dashboard. Make sure to add the label and legends.

The 3rd dashboard consists of only one sheet. Employ a custom SQL theory to explore the relationship between service utilization and leading causes of death. Create two maps, each representing one of the most utilized services (blood work and intravenous). Color-code the states on each map based on the leading cause of death. Include tooltips to display the state name, most utilized service, and leading cause of death when hovering over each state. Place this sheet into a 3rd dashboard.

DATA ANALYSIS RESULTS

The analysis of the dashboards provides valuable insights into various aspects of healthcare and mortality trends.

The 1st dashboard is the KPI for the medical_data dataset. Firstly, the average total charges by gender reveal a significant difference in average total charges between genders, with non-binary individuals incurring the highest charges. This insight can help healthcare providers identify potential disparities in healthcare access and support efforts toward equitable healthcare provisions. Additionally, blood work emerges as the most utilized among patients in service utilization. This indicates its importance in diagnostic and preventive healthcare measures. This information is crucial for healthcare facilities to allocate resources effectively. In the prevalent medical conditions overweight emerges as the most prevalent medical condition among patients. This highlights the importance of addressing obesity-related health issues in healthcare strategies.

The 2nd dashboard is the KPI for the external dataset, the leading_causes dataset. The leading cause of death is heart disease. The underscores the significance of cardiovascular health initiatives and interventions. Hospital leadership can use this information to prioritize funding and strategies to manage their heart disease. Moreover, the leading causes by state reveal regional variation in leading causes of death, with heart disease being predominant across most states. This insight can inform state-level public health policies, resource allocation, and intervention strategies tailored to address specific mortality trends within each region. The average age-adjusted death rate per year gradually declines over the years, indicating overall improvements in healthcare initiatives. This trend underscores the effectiveness of healthcare interventions and advancements in medical treatments.

The last dashboard is the most utilized service versus the leading cause of death per state. This dashboard uses both the medical_data and leading_causes datasets. This dashboard highlights the correlation between the most utilized healthcare service and each state's leading cause of death. This insight can guide healthcare resource allocation and intervention strategies. This ensures that healthcare services are aligned with the predominant health needs and mortality trends in each state.

ANALYSIS LIMITATIONS

While the data analysis can provide valuable insights, it is essential to acknowledge some limitations that may affect the generalization of the findings.

The datasets used in the analysis may not represent the entire population or healthcare landscape comprehensively. There could be inherent biases in the selection of patients, healthcare facilities, or regions included in the datasets. For example, if certain demographic groups or geographic areas are underrepresented in the data, the analysis results may not accurately reflect their utilization patterns or mortality trends.

Furthermore, the analysis may not capture temporal changes in healthcare practices, policies, or demographics. Healthcare utilization patterns and mortality trends could evolve due to advancements in medical technology or shifts in population demographics. Without considering temporal changes, the analysis may provide a static snapshot that overlooks important dynamics over time.

Lastly, the analysis may focus on specific aspects of healthcare utilization or mortality trends. This potentially overlooks systemic issues that influence population health outcomes. A comprehensive understanding of healthcare dynamics requires consideration of factors such as healthcare access, quality of care, and healthcare policies.