A1:Data Sets

I have attached my data analysis dashboard to this submission. The twod datasets used for this assignment are Randomly generated medical Dataset for WGU D210 and the original medical dataset provided by wgu.

A2:Dashboard Installation

Step 1: Install and download Tableau Get Tableau Deskto. Select the version that is compatible with your operating system (Mac or Windows). After entering your information, click "Download" on the form.

Step 2:For Windows: Double-click the downloaded.exe file to launch the installer (it should be in your Downloads folder). Mac users can drag the Tableau icon into their Applications folder after finding the downloaded.dmg file (which is often located in your Downloads folder) and double-clicking it to launch the installer.

Step 3: agree to the licensing terms. Use the default directory or select the installation directory. After selecting "Install," watch for the installation to finish.

Start Tableau:

On Windows: Locate Tableau Desktop in your Start menu or by clicking the shortcut that was generated for it on your desktop.

On a Mac: After navigating to the Applications folder, double-click Tableau.

Step 3: Open the Tableau File d211.twb

1. **Open Tableau Desktop:**

Launch Tableau Desktop from your desktop shortcut or Applications folder.

1. **Open the .twb File:**

Click on "File" in the top left corner of the Tableau interface.

Select "Open..." from the dropdown menu.

Navigate to the location on your computer where d211.twb is stored.

Select the file and click "Open".

Step 4: Explore the Dashboard

1. **View the Dashboard:**

Once the file is open, you should see the Tableau dashboard and all the visualizations contained in d211.twb.

1. **Interact with the Dashboard:**

Click on different elements to filter data, hover over charts to see details, and use any interactive features built into the dashboard

A3:Dashboard Usage

The structure of the dashboard can be conceptualized as a series of four tabs, analogous to those found in a web browser. The initial tab, "Introduction," is static and serves merely to display the dashboard title, course number, and the researcher's name. To move tabs, simply select 1 of the 4 boxes in the middle of the top of the page.

The second tab, labeled "Back Pain," is interactive and presents visualizations of anxiety-related responses derived from both WGU’s dataset and a fabricated dataset. This comparison aims to juxtapose WGU’s authentic data with a synthetic dataset encompassing the same variables. The visualizations are segmented and filtered based on gender and obesity responses. Users can select male, female, or nonbinary options to update all charts to reflect the chosen gender. Additionally, an overweight filter allows users to select "yes" or "no," thereby visualizing responses that meet the specified condition.

The third tab, designated "Arthritis ," follows a similar interactive structure. It features visualizations of Arthritis responses from both WGU’s dataset and a fabricated dataset. The objective is to compare WGU's authentic data with the synthetic dataset, with a particular focus on back pain. These charts are also segmented and filtered by gender and obesity responses. Users can select male, female, or nonbinary options to update all charts accordingly. The overweight filter similarly allows users to select "yes" or "no" to visualize responses fitting the specified condition.

The fourth tab which is named Diabetes follows the same formula as the previous two but has the focus on the variable diabetes. Finally the last tab Anxiety and Obesity, gives percentages of their responses by data set but without filters.

This structured approach facilitates a comprehensive analysis and comparison of the datasets, providing insights into the nature of the data and the reliability of the findings.

A4: SQL CODE

Attached to the submission are two text files with the sql code needed to create the tables Randomtable and WGUtable

CREATE TABLE finalunion AS SELECT\* FROM WGUtable UNION SELECT \*FROM Randomtable;

B:Panopto

C1:Dashboard Alignment

The WGU medical data dictionary focuses on predicting the readmission of previously hospitalized patients but does not assess the validity of the data to generate statistical insights. For this analysis, I took a broader approach, moving beyond readmission rates to examine demographic and medical condition trends among patients within the WGU Hospital System compared to a randomized data set. To investigate the data quality and integrity of both datasets, I generated four dashboards.

The first dashboard explores trends and data related to patients' responses to back pain, including conditions based on gender and weight. This variety helps determine if the results are coincidental or relevant. The second dashboard analyzes responses to arthritis, also filtered by gender and weight status. While no immediate answers regarding data randomness are apparent, the insights gathered help compare the datasets.

The third dashboard examines responses to diabetes, following the same trend analysis as the previous dashboards. The fourth and final dashboard looks at responses to anxiety and obesity without any filters. These dashboards collectively aim to assess the data quality of the WGU dataset by comparing it to the randomized data set, highlighting similarities and differences.

C2: justify Business Intelligence Tool

For this analysis, I used Tableau Desktop as the business intelligence tool. Tableau enables the straightforward creation of intricate and interactive data visualizations. This functionality aids in effectively communicating the results of data analysis and simplifies the process of exploring complex data to uncover new relationships or noteworthy patterns.

C3: Data Preparation

For this assignment, I did not have to do any meticulous data preparation. All I had to do was rename the columns in the original raw datasets to the same names and dropped all the columns in the Medical\_clean.csv file except Gender, Stroke,Arthritis,Backpain,Diabetes,overweight,Anxiety, Allergic Rhinitis, High blood pressure, Refulx esophagitis, stroke, state, and hyperlipidemia because those were the only variables that were similar in the randomized data set. I was able to do this without using sql coding by doing the data cleaning in tableau prep. Dropping the columns not being mutually used eased the union and sql process and allowed for easy transitioning into the next stage of of the project.

C4: Dashboard Creation

Gender filter

1. Drag Gender Into Rows
2. Drag Gender into Color under marks

Stroke Filter

1.Drag Stroke Into Rows

2. Drag Stroke into color under Marks

Arthritis WGU

1. Drag Arthritis into rows
2. Drag Gender into columns
3. Drag Extract (Count) next to Gender in Columns
4. Drag Arthrits Into Filters and unselect the null box.
5. Drag Arthritis On Color
6. Drag Arthritis Onto Label, rightclick, hover over measure, then select count

Arthritis Random

Drag Arthritis Random into rows

Drag Gender into columns

Drag Extract (Count) next to Gender in Columns

Drag Arthrits Into Filters and unselect the null box.

Drag Arthritis Random On Color

1. Drag Arthritis Random Onto Label, rightclick, hover over measure, then select count

Back Pain WGU

Drag Back Pain WGU into rows

Drag Gender into columns

Drag Extract (Count) next to Gender in Columns

Drag Back Pain WGU Into Filters and unselect the null box.

Drag Back Pain WGU On Color

Drag Back Pain WGU Onto Label, rightclick, hover over measure, then select count

Back Pain Random

Drag Back Pain Random into rows

Drag Gender into columns

Drag Extract (Count) next to Gender in Columns

Drag Back Pain Random Into Filters and unselect the null box.

Drag Back Pain Random On Color

Drag Back Pain Random Onto Label, rightclick, hover over measure, then select count

Diabetes WGU

Drag Diabetes WGU into rows

Drag Gender into columns

Drag Extract (Count) next to Gender in Columns

Drag Diabetes WGU Into Filters and unselect the null box.

Drag Diabetes WGU On Color

Drag Diabetes WGU Onto Label, rightclick, hover over measure, then select count

Diabetes Random

Drag Diabetes Random into rows

Drag Gender into columns

Drag Extract (Count) next to Gender in Columns

Drag Diabetes Random Into Filters and unselect the null box.

Drag Diabetes Random On Color

Drag Diabetes Random Onto Label, rightclick, hover over measure, then select count

WGU Overweight

Drag Overweight into Rows

Drag Overweight into Filter and unselect Null

Drag Overweight to Text

Drag Overweight onto Label, rightclick, hover over quick calculations and select percent of total

Random Overweight

Drag Random Overweight into Rows

Drag Random Overweight into Filter and unselect Null

Drag Random Overweight to Text

Drag Random Overweight onto Label, rightclick, hover over quick calculations and select percent of total

Anxiety

Drag Anxiety into Rows

Drag Anxiety into Filter and unselect Null

Drag Anxiety to Text

Drag Anxiety onto Label, rightclick, hover over quick calculations and select percent of total

Anxiety Random

Drag Anxiety Random into Rows

Drag Anxiety Random into Filter and unselect Null

Drag Anxiety Random to Text

Drag Anxiety Random onto Label, rightclick, hover over quick calculations and select percent of total

Dashboard 1

Drag Back Pain Random and Back Pain WGU to the top left and right of screen

Drag the stroke and Gender filters to the bottom left and right

Click each filter, move the cursos to the top right of the box and select the use as filter button

Dashboard 2

Drag Arthritis Random and WGU Arthritis Responses to the top left and right of screen

Drag the stroke and Gender filters to the bottom left and right

Click each filter, move the cursos to the top right of the box and select the use as filter button

Dashboard 3

Drag Diabetes WGU and Diabetes Random to the top left and right of screen

Drag the stroke and Gender filters to the bottom left and right

Click each filter, move the cursor to the top right of the box and select the use as filter button

Dashboard 4

Drag Random Overweight to the top left

Drag WGU Anxiety Rate to the top right

Drag WGU Overweight to the bottom left

Drag Random Anxiety Rate to the bottom right

C5:Results of data Analysis

After looking at the different distributions of answers in both the random and provided data sets, I noticed that the randomized data set almost always shows a 50% split in responses. There’s never a major skew or any variable with a noticeable difference in answers. This is somewhat helpful because the visualizations from the randomized data set act as a visual average or middle ground when compared to the other data sets. Overall, I think the fact that the randomized data set gives perfect splits in responses actually makes me believe more in the accuracy of the original data set. At least the original data set has some reasonable differences between answers. These differences lead me to think that the WGU data set likely contains real survey answers rather than being randomly generated.

C6:Limitations of Data Analysis

Despite the insights gained from comparing the distributions of the random and provided data sets, there are notable limitations to my analysis. Firstly, the reliance on the 50% split in the randomized data set as a baseline may oversimplify the complexity inherent in genuine data, which often contains subtle patterns and anomalies that are not captured in a purely random distribution. Additionally, while the noticeable differences in the WGU data set suggest authenticity, this assumption could be flawed without further validation techniques to confirm the data’s accuracy and integrity. Moreover, the analysis does not account for potential biases or errors in data collection methods, which could affect the validity of the conclusions drawn. These limitations highlight the need for more comprehensive methods and cross-validation with other data sources to ensure the robustness of the findings.

Sources

<https://www.kaggle.com/datasets/ehhhhh/randomly-generated-medical-dataset-for-wgu-d210>

The randomized data set I use for my analysis and dashboard