A1: Dashboard, Data Sources, Data Cleaning

The two datasets I will be using for this assignment are

* The WGU medical dataset is provided by WGU for several of the classes in the MSDA program.
* <https://www.kaggle.com/datasets/ehhhhh/randomly-generated-medical-dataset-for-wgu-d210>

The data set I pulled from Kaggle is a small-scale replica of the WGU medical data set with random entries and values added. I decided to use this “randomized” data set, as so far my luck with the WGU Medical Services tasks has not been fruitful, and I have been slowly suspecting that the WGU dataset is random as well. So my plan is to investigate some results between the assigned data set and a randomized dataset in order to examine how similar they are. The end result is to see if they have similar numbers. If they have similar results and numbers, I will begin to believe the WGU assigned data set is also random, which is why I am not able to find statistically significant insights. However, if the results are vastly different, I can assume that the randomized results are indeed random, but I can also start to infer that the original results are actually legitimate and sourced. If so, it will add validity to my previous assignment results and inquiries regarding whether what I came up with was indeed statistically significant or not. I did not have to do any cleaning for this assignment as the extra data set used was created to be paired with the WGU-provided one.

A2: Installation Guidelines for the Dashboard

Step 1: Install and download Tableau. Get Tableau Desktop. 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 d210.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 drop-down menu.

Navigate to the location on your computer where d210.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 d210.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 Navigation

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, "Intro," is static and serves merely to display the dashboard title, course number, and the researcher's name. To move tabs, simply select one of the four boxes in the middle of the top of the page.

The second tab, labeled "Anxiety Overview," 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, with a specific focus on anxiety. 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 "Back Pain Overview," follows a similar interactive structure. It features visualizations of back pain 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 non-interactive, provides the Key Performance Indicators (KPIs) pertinent to this research project. The KPIs are intended to identify significant differences between the datasets. Should the data exhibit even a slight similarity, it may indicate that the WGU medical data is also random. The dashboard includes six visualizations in total. On the left, it displays the percentages and total amounts of anxiety, arthritis, and obesity from the fabricated dataset. On the right, the same metrics are displayed for the WGU official dataset. Anxiety and obesity data are presented using text tables, while arthritis data is visualized with a pie chart.

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.

B:Panopto Recording

Panopto recordings will be submitted with the rest of the project.

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=abdbc194-7b20-4d27-b658-b1a8017223fe

C1:1. Explain how the purpose and function of your dashboard align with the needs outlined in the data dictionary associated with your chosen data set.

The WGU medical data dictionary indicates a particular interest in predicting the readmission of patients who have been previously hospitalized. However, it does not examine the validity of the data in order to create statistical insights. For this data analysis, I adopted a broader approach. Instead of focusing on readmission rates, I explored demographic and medical condition trends among patients within the WGU Hospital System in comparison to the randomized data set. To do this, I generated three dashboards to investigate the quality and integrity of both data sets and see if they were similar. The first dashboard investigates the trends and data of the patient’s responses to anxiety. Not only does it examine the trends, but it also has conditions based on gender and weight to add more variety to the responses and to examine if the results are coincidences or relevant. The second dashboard does the same as anxiety, but instead of examining the responses to anxiety, I examined their responses to having back pain or not, with filters being available based on gender and their weight status. There is no answer that will immediately jump out regarding whether the data is random or not, but it does give some insight into whether the datasets are similar or not.

1. Explain how the variables in the additional data set enhance the insights that can be drawn from the data set you chose from the provided options.

Thankfully, this data set was created with this task assignment in mind, so all the following variables are the only ones in the data set; however, they are all from the medical services data set provided by WGU, so they were able to be used.

Age

Allergic Rhinitis

Anxiety

Area

Arthritis

Back Pain

Diabetes

Gender

High blood pressure

Hyperlipidemia

Overweight

Reflux esophagitis

State

Stroke

Lattitude

Longitude

Patient

These variables can enhance my insight into the dataset provided because I can cross-compare insights and trends amongst the two. By doing this, I can find out if my models and prediction methods used for other assignments were not efficient and accurate because I established them incorrectly, or if it was just a data quality issue. Since it has been made apparent, the additional data set I used is based on the original one, but with randomized responses. If they have similar trends, I will know that the original dataset itself is skewed and unreliable, which would answer my question on whether my models were good at predicting because the data was not truly accurate or not.

1. Explain two different data representations from your dashboard and how executive leaders can use them to support decision-making.

In the anxiety overview tab, you can see that regardless of gender, the patient is more likely than not to report not having anxiety. This can let executives know not to invest a plethora of money into anxiety-reducing programs or procedures at hospitals to help those with anxiety. I say this because since most people don’t appear to be suffering from it, any increase in costs to reduce it would almost be a waste, as someone coming in will likely not need anxiety assistance. Or, on the contrary, leaders can witness that there are indeed a small number of patients who do suffer from anxiety and proceed to have some form of procedure or plan to help those afflicted.

In the Backpain Overview tab, you can see that patients who were nonbinary and were not overweight had more responses to having backpain compared to those who were binary and overweight. This is an interesting discovery, as you would think you would have more back pain with more weight, but the data says otherwise. With this information, leaders can use this data to create insights about their clients in order to help create new treatments and plans to help remedy what ails the patients.

1. Explain two interactive controls in your dashboard and how each enables the user to modify the presentation of the data.

The two dashboards, anxiety and backpain overview, each have two interactive controls that enable the user to modify the presentation of the data. These two controls are filters for gender and overweight. If you click yes or no on the overweight filters, it will filter all the data the dashboard pulls to only show data that fits the condition you chose. For example, if you are looking at the anxiety overview page and hit "no, the models and visualizations will only be able to pull and read data where the patient is not overweight. The same thing will be done if you hit yes, but instead of the patient not being overweight, the patient does have to be overweight.

Additionally, the controls can do the same thing with gender. There are 3 color options corresponding to the 3 different responses to gender: "male, female, and nonbinary,” and by choosing the corresponding color to your choice, the data will be filtered to only include the specified gender.

1. Describe how you built your dashboard to be accessible for individuals with colorblindness.

In my data analysis, I adjusted my use of color based on insights from a VennGage discussion on color palettes and considerations for colorblindness. I avoided using combinations that could be problematic for individuals with red or green colorblindness, which affects the majority (99%) of colorblind individuals. Instead, I selected colors from the VennGage color wheels: blue for males, pink for females, and yellow for nonbinary genders. These colors are distinguishable for users with red or green colorblindness, as they would perceive them as blue, brown, and yellow, respectively, ensuring the color scheme remains effective for them. Users affected by other types of colorblindness, although less common, can rely on clear labeling provided with the available filters for assistance.

1. Explain how two data representations in your presentation support the story you want to tell.

Two data representations in my presentation that support the story I want to tell are the anxiety and obesity rate KPIs. To demonstrate these, I got the percentages of total responses regarding obesity and anxiety between the two datasets as a percentage for each, respectively. Keeping in mind that we want to see that either the responses were similar in order to prove the WGU data set is faulty or they are drastically different, alluding to the legitimacy of the data provided,.

Regarding anxiety rates, the original WGU-provided data had 68% of total responses as a no for suffering from anxiety, and 32 said yes. On the other hand, for the randomized dataset, it was roughly 50/50 on responses recorded.

Regarding obesity rates, the original WGU provided data, with 29% saying they were not overweight and 71% saying they were. For the randomized dataset, it was again almost a neat 50/50 distribution between the responses.

These data representations support the story that the WGU dataset is not as random as I thought. Ironically, the randomized dataset did 50/50 when given data and had to assign a yes or no as a response, while the data set provided by WGU is a lot more contrasting and understandable when you apply some real-world logistics. Surveying 10,000 people and 5000 of them being both overweight and having anxiety is almost too unfortunate to be plausible. Contrarily, the WGU dataset, with 70% of people surveyed not suffering from obesity or anxiety, seems more realistic and accurately recorded. This doesn't directly state that the original data set isn't random, but at least it points towards the idea that this data isn't as random as I originally assumed.

1. Explain how you used audience analysis to adapt the message in your presentation.

In my presentation to the executive board, which includes the Senior Vice President of Hospital Operations and the Vice President of Research, I will focus on broad findings as highlighted in the WGU data dictionary. Specifically, I cannot draw definitive conclusions about the higher diabetes rates among patients at the WGU Hospital System due to limited expertise and supporting data.

The Vice President of Research is interested in multidisciplinary studies that investigate the reasons behind this phenomenon and potential solutions, aligning with their research-focused role.

The Senior Vice President of Hospital Operations, who oversees hospital management, resource allocation, strategic planning, and patient outcomes, is more concerned with the general demographics of patients in the WGU Hospital System. While the diabetes rates are relevant, their primary focus is on how patient demographics relate to patient care and overall hospital operations.

By tailoring the presentation to these interests, I aim to provide relevant and actionable insights for both executives.

1. Describe how you designed your presentation for universal access by all audiences.

I achieved this by not using incredible jargon so the words and concepts are precise adn straight to the point. I also used simple interactions to add interactive elements that are easy to use.

1. Explain two elements of effective storytelling that you implemented in your presentation and how each element was intended to engage the audience.  
    I utilized visual elements like charts, graphs, and infographics to represent data clearly and concisely. Visual aids help simplify complex information and make it easier for the audience to grasp the key points. Secondly, I utilized an introduction. I began with a strong opening that sets the stage for my presentation. This could be an interesting fact, a relevant anecdote, or a compelling question.