Performance Assessment 1

D210 – Representation and Reporting

Jessica Hosey

MSDA, College of Information Technology

Western Governors University

August 11th, 2024

Part 1: Interactive Data Dashboard

A. Provide a copy of your interactive Tableau dashboard.

See the attachments in my submission.

A. 1. Provide both data sets that serve as the data source for the dashboard.

See the attachments in my submission. There are three: one for the WGU dataset, one for the Behavior Risk Surveillance System Survey dataset, and a combined dataset with both datasets.

A. 2. Provide step-by-step instructions to guide users through the dashboard installation.

Tableau is an excellent tool for the analytics world as it is easily viewed online by anyone. Since this is a task for a course and I am not dealing with data that must be protected due to company policies, the dashboard can be seen on Tableau Public using the link I have provided in my submission.

To install it, you need to click the link provided in my submission, and that is it. There are no additional steps to take as it is on Tableau Public.

A. 3. Provide instructions to help users navigate the dashboard.

After clicking the link to the dashboard on Tableau Public (the link in my submission), navigating the interactive dashboard is simple. Once open, the dashboard's main page is an introduction to the dashboard. On the top, there are four grey boxes with numbers. These numbers are the different "tabs," like on a web browser, that you can use to navigate the dashboard's contents.

The first tab is the introduction; it has the title of my analysis and information about both datasets used to complete this presentation. This has two images of the organizations from which the data came and the presenter's name.

The second tab is baseline statistics about the datasets to give an idea about our patients. BRFSS is the dataset from the CDC, and WGU is the WGU medical dataset. This tab contains information on gender, age, and number of children. This tab lets you interact with the visualizations by clicking on any parts. For example, if you are interested in only female data in either dataset, then click the portion of the pie graph for females. This transforms the page into information only about female survey responses from that dataset. To return to the original design, click the white space near the pie graph until it returns to the original look.

The third tab is information about each of the KPIs: number of patients with diabetes, high blood pressure, and who are overweight (BMI > 25). Each of these heatmaps shows the number of each KPI by age and by data source (WGU or BRFSS). This tab lets you interact with the visualizations by clicking on any parts. For example, if you were interested in data about WGU patients who are 50-54 years old and considered overweight, click the number "491" found under the age group 50-54 of WGU in the third KPI section. This gives you the rate of overweight individuals in the WGU dataset. You could further select the gender in the upper right corner also to see the rates amongst males or females. In addition, this would change the Diabetes and High Blood Pressure KPIs diagrams to show those results amongst that

demographic of patients. Again, to return to the original design, click the white space near the heatmap until it returns to the original look.

The fourth tab is a concluding tab that summarizes crucial messages and recommended steps that the group should take after this analysis.

Part 2: Storytelling with Data

B. Panopto Video

See the files in my submission.

Part 3: Reflection Paper

C.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.

When searching for a dataset related to the WGU dataset, I knew that Kaggle had several publicly available datasets for people to use through practice and to show off their skills. The WGU dataset is a hospital's data from one day of operation and contains patient details from who walked into the hospital that day. The CDC completed in 2015, a Behavior Risk Assessment via telephone, and the dataset was released for public use later. The CDC dataset contains much information, from how many children an individual must exercise daily. It had over 45,000 rows, over four times the size of the WGU dataset. A few columns in the CDC dataset were like the WGU dataset (the list of variables is in the table below). My dashboard is focused on the rate of diabetes of individuals in both datasets and what factors (like overweight, age, and high blood pressure) might be contributing to a patient's diabetes diagnosis.

List of variables that align with the WGU and the CDC Behavior Risk Survey Dataset:		
WGU – Medical Dataset:	What I named the variable:	CDC – BRFSS Dataset:
Age	Age	_AGE5YR
Income	Income	INCOME2
State	StateName	State
Gender	Gender	SEX
Marital	MaritalStatus	MARITAL
Children	Children	CHILDREN
HighBlood	HighBP	BPHIGH4
Stroke	Stroke	CVDSTRK3
Arthritis	Arthritis	HAVARTH3
Diabetes	Diabetes	DIABETE3
Overweight	Overweight	_RFBMI5

C.2. 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.

The CDC dataset is much larger than the WGU dataset, so it is adding more data to the mix to help produce more insight into why patients are diagnosed people with diabetes and how a company might want to gain public health attention in preventing this disease in the future.

When choosing variables that align with the WGU dataset, I was more focused on what variables matched and what variables would answer my analysis question. So here are all the variables and their context below:

- _AGE5YR – This was the Age column; it provided a better distribution of ages than the WGU dataset provides.
- INCOME2 – This is the Income column; since the BRFSS dataset is much larger than the WGU, this provided some insights into whether how much you make can influence your chances of being diagnosed with diabetes. However, Income is not a biological alignment, so I did not include this analysis in the final dashboard.
- State – This is the state column; in the BRFSS dataset, responses from all US territories were added, whereas the WGU only has one patient from Puerto Rico. It was interesting to see which states had a higher rate of diabetes than others.
- SEX – This is the Gender column; it only provided Male or Female responses; no Nonbinary responses were in the BRFSS dataset as it was from 2015 (an era that shunned anyone different from the typical boy or girl labels).
- MARITAL – This was the marital status column; it was used to see the general distribution of responses in the BRFSS data compared to the WGU data.
- CHILDREN – Number of Children the caller or patient has/had. There were several responses from individuals with 20+ children. I did not remove these during data cleaning but excluded them from the final presentation as they messed with the view of the visualization.
- BPHIGH4 – This was the column for high blood pressure; it aligned perfectly with the WGU dataset.
- CVDSTRK3 – This is the stroke column. The BRFSS dataset had a few questions (columns) related to stroke, so I picked the one that indicated whether the individual had a stroke to best align with the WGU dataset.
- HAVARTH3 – This is the arthritis column; this was added as it aligned with the WGU dataset.
- DIABETE3 – This is the diabetes column; the BRFSS dataset had a few diabetes-related columns, which was the closest in the context of the diabetes column in the WGU dataset.
- _RFBMI5 – This is the overweight column; the BRFSS dataset did not have an overweight column, but it had a BMI calculation column completed during the BRFSS analysis using height and weight. This was easier to transform into an overweight column identical to the context of the WGU overweight column for our analysis.

C.3. Explain two data representations from your dashboard and how executive leaders can use them to support decision-making.

One of the visualizations allows the data to be broken down into 5-year age groups interactively. This allows the user to select a core group that they might want to focus on for insights into the rate of

diabetes, high blood pressure, or overweightness. Executive leaders could use this to see what specific age groups have the highest rate of diabetes and choose to target them in their next marketing campaign.

Another visualization allows the user to see the number of children a patient has. This could inform leaders about the number of patients with high blood pressure caused by the frustrations their children might cause before that doctor's appointment. All jokes aside, this visualization allows you to take this larger dataset and select key metrics to pull a smaller sample size to see the results.

C.4. Explain two interactive controls in your dashboard and how each enables the user to modify the data's presentation.

As stated previously in part A3, someone can interact with this dashboard in several different ways. In tab 2, the user can select an age group from either the WGU or BRFSS dataset to see the gender demographics of that specific group. You would also see the distribution of the number of children those individuals have in that age group.

Another interactive control is found in tab 3, where a user could select either males or females on the top right, and the dashboard transforms and calculates metrics for that specific demographic. You could take another step further and choose a particular condition, such as KPI, diabetes, high blood pressure, or being overweight, and select an age group. This will transform and calculate again for the user and provide them with the number of patients and the rate of all the conditions on that page.

C.5. Describe how you built your dashboard to be accessible for color-blind individuals.

There are three types of colorblindness that individuals can have, and each one affects the individual's ability to see specific colors ("How to design a dashboard accessible to color-blind users"). During the creation of the dashboard, the focus was to stay away from greens, reds, and yellows. This allows anyone, regardless of having a color-blind gene, to interact and visualize the dashboard contents. Changing the color on Tableau is incredibly easy, and they even have a palette explicitly designed for this crucial part of the storytelling journey.

C.6. Explain how two data representations in your presentation support your story.

In tab 3 of my presentation, the heatmaps and the rate of each condition are the two data representations that best support the story of my analysis question, "What factors contribute to a patient's diabetes diagnosis?" If we select the age group in the BRFSS data for diabetes, it allows us to see what individuals with diabetes also deal with. In this case, individuals in the BRFSS dataset typically have an approximately 50% chance of also having high blood pressure and about a 27% chance of being overweight (BMI > 25). If I select the 80+ age group with diabetes in the AGU dataset, more people with diabetes are overweight, and only 40% have high blood pressure. This gives us a good picture of which of the three conditions may have led to a diabetic diagnosis for the patient.

C.7. Explain how you used audience analysis to adapt the message in your presentation.

This task indicated that the presentation was geared towards the audience of executive leaders at a company. These individuals are well-versed in the topic and the terms used in healthcare. Most hospital executives have a background in healthcare before choosing to elevate their standing to a leadership role. They are not always good at seeing data and numbers, but they understand diabetes-associated terms. Therefore, to tailor my presentation, I used concise, direct statements covering the main points of my analysis (last tab of the dashboard). In addition, I made sure to include information that is needed for this company's next steps. Within the visualizations, I used titles that reference the information best, like 'Diabetes,' rather than using data analytics table terms, like 'HAVARTH3,' an executive would have no idea what this column title references.

C.8. Describe how you designed your presentation for universal access by all audiences.

What is excellent about Tableau Public is that anyone with an internet connection can view the full dashboard presentation. In addition, they could download the dashboard from Tableau Public and play with the sheets, dashboards, and data to create their own story from my creation. Another ease of access comes from Panopto videos, which allows me to present topics that could be paused, sped up, or have closed captions for hard-of-hearing individuals.

C.9. Explain two elements of compelling storytelling that you implemented in your presentation and how each element was intended to engage the audience.

The biggest thing about data analytics is trying to get someone to care about what you are presenting. Why should they listen to me talk about diabetes if it doesn't affect them? The data shows older individuals have higher rates of diabetes than younger populations due to the body's biological mechanisms starting to degrade. You need to hook the audience and then tell the story.

In my presentation, I created a title with my analysis question; this gives the user/listener a general idea or goal (Microsoft 2024) of what this presentation is all about and what they will probably find.

Different visualizations are used in the presentation to pull users' eyes into the data story. These visualizations are incredibly interactive, so users can guide themselves or others with a straightforward click to draw evidence from the dashboard.

D. Sources

Alex the Analyst. (2021, October 5). How to Create Visualizations in Tableau | Tableau Tutorials for Beginners [Video]. YouTube. https://youtu.be/ebjDwX8sUMk?si=B5sjnQ9snKddYPsC

Centers for Disease Control and Prevention. (2016). Behavioral Risk Factor Surveillance System: 2015 codebook. https://www.cdc.gov/brfss/annual_data/2015/pdf/codebook15_llcp.pdf

Databox. (2022, July 14). Dashboard presentation: How to create and present dashboards effectively. Databox. Retrieved August 11, 2024, from https://databox.com/dashboard-presentation

Kaggle. (n.d.). Behavioral Risk Factor Surveillance System. Retrieved August 11, 2024, from https://www.kaggle.com/datasets/cdc/behavioral-risk-factor-surveillance-system?select=2015.csv

LinkedIn Community. (n.d.). How to design a dashboard accessible to color-blind users [Post]. LinkedIn. https://www.linkedin.com/advice/3/how-can-you-design-dashboard-accessible-users-color-undsf#:~:text=Creating%20a%20color%2Dblind%2Dfriendly,icons%20to%20support%20color%20cues

Microsoft. (2024). Data storytelling with Power BI. Microsoft. Retrieved August 11, 2024, from https://powerbi.microsoft.com/en-us/data-storytelling/

University of Pittsburgh. (n.d.). *Audience analysis*. Department of Communication, University of Pittsburgh. Retrieved August 11, 2024, from https://www.comm.pitt.edu/oral-comm-lab/audience-analysis