**Western Governors University (WGU)**

**D210: Representation and Reporting**

**Natallia Zimnitskaya | ID: 012247127**

**Master of Science, Data Analytics**

**Part 1: Interactive Data Dashboard**

**A1.** The dashboard with my data analysis can be found at <https://public.tableau.com/views/RepresentationandReporting_17297992839080/KPI?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link>

**A2.** The internal dataset “medical\_clean.csv” was provided by WGU, and I used this dataset for the D208 and D209 performance assessments. Please see the attached “medical\_clean.cvs” file.

The external dataset “National Health and Nutrition Examination Survey 2013-2014” provided by the Centers for Disease Control and Prevention was found on Kaggle <https://www.kaggle.com/datasets/cdc/national-health-and-nutrition-examination-survey>.

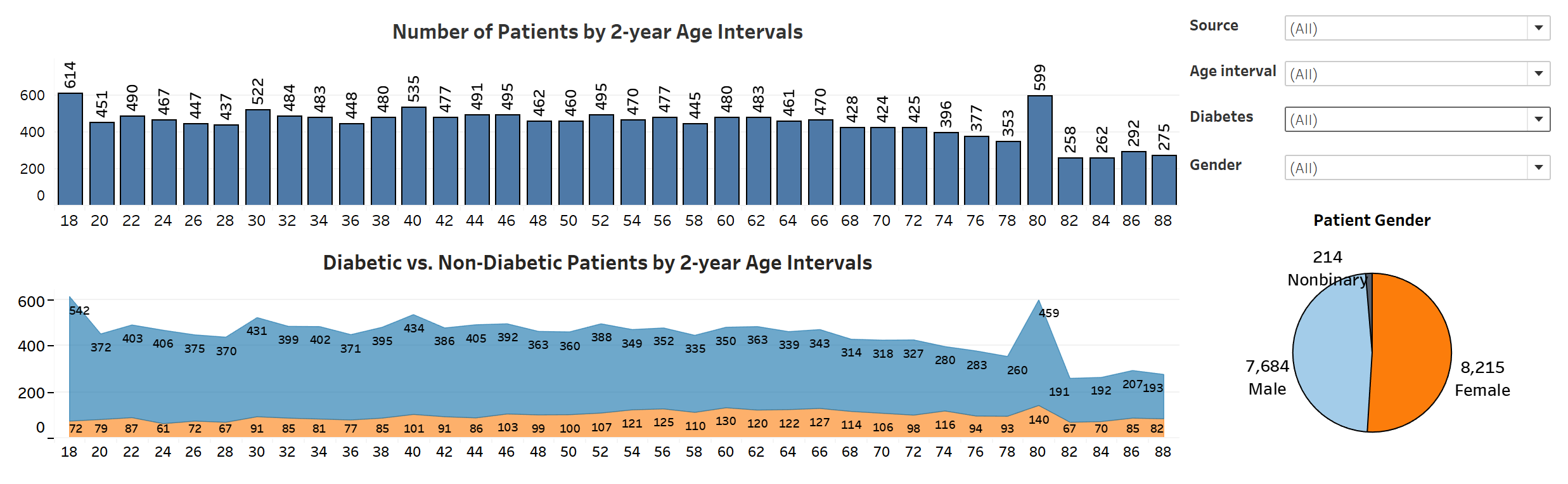
From the “medical\_clean.csv” file, I kept columns related to the assessment: Age, Gender, Diabetes, Hyperlipidemia, HighBlood, Overweight, and Stroke.

The external “National Health and Nutrition Examination Survey 2013-2014” file contains five datasets: Demographic dataset, Examination dataset, Dietary data - total nutrient intake dataset, Laboratory dataset, Questionnaire dataset, and Medication dataset. I got Age and Gender from the “demographic.csv” dataset and medical conditions from the “questionnaire.csv”. I used the Demographic Variable List and Questionnaire Variable List provided by the CDC (*NHANES 2013-2014 Demographics Variable List*, n.d.; *NHANES 2013-2014 Questionnaire Variable List*, n.d.).

**A3.** There is no need for any installation. The dashboard can be easily accessed by following this link: <https://public.tableau.com/views/RepresentationandReporting_17297992839080/KPI?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link>

**A4. Dashboard navigation.**

Dashboard “Exploring Patient Demographics and Health Conditions”



Bar Chart “Number of Patients by 2-Year Age Intervals”:

* Use the "Age Interval" filter to focus on specific age ranges.
* Apply the "Source" filter to compare data from WGU and CDC.
* Adjust the "Gender" filter to see how age distribution varies between males, females, and non-binary.
* Use the "Diabetes" filter to analyze the number of diabetic vs. non-diabetic patients in each age interval.

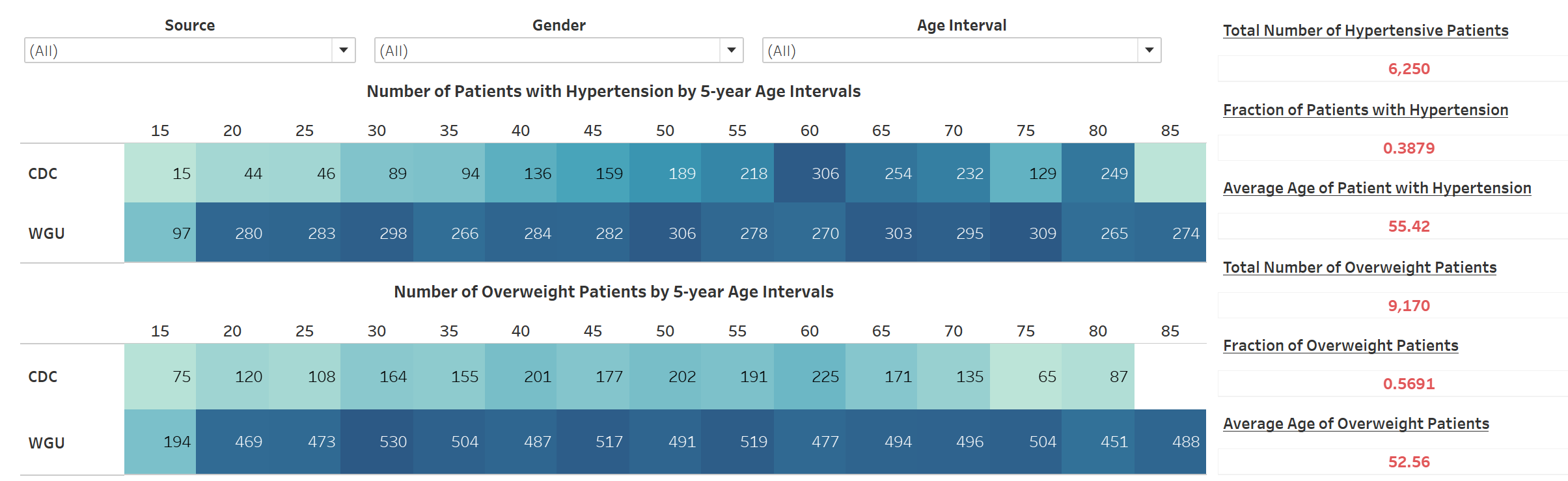
Pie Chart “Patient Gender”:

* Utilize the "Source" filter to differentiate the gender distribution between WGU and CDC datasets.
* Combine with the "Age Interval" filter for gender-specific insights in particular age groups.

Area Chart “Diabetic vs. Non-Diabetic Patients by 2-Year Age Intervals”:

* + - * Manipulate the "Diabetes" filter to compare trends across diabetic and non-diabetic patients.
      * Explore how these trends differ with age using the "Age Interval" filter.
      * Adjust the "Gender" and "Source" filters to uncover additional patterns.

Dashboard “Focused Health Condition Analysis”



Highlighted Tables “Number of Hypertensive Patients by 5-year Age Intervals” and “Number of Overweight Patients by 5-year Age Intervals”:

* "Source" compares data between WGU and CDC or sees combined results.
* Leverage the "Age Intervals" and "Gender" filters to gain a deeper understanding.

KPIs:

* Examine the total number of hypertensive and overweight patients.
* Examine the proportion of the population dealing with hypertension or overweight issues.
* Adjust filters to view metrics for specific sources, age groups, or genders.
* Use these KPIs to quickly grasp key health statistics and compare them across patient segments.

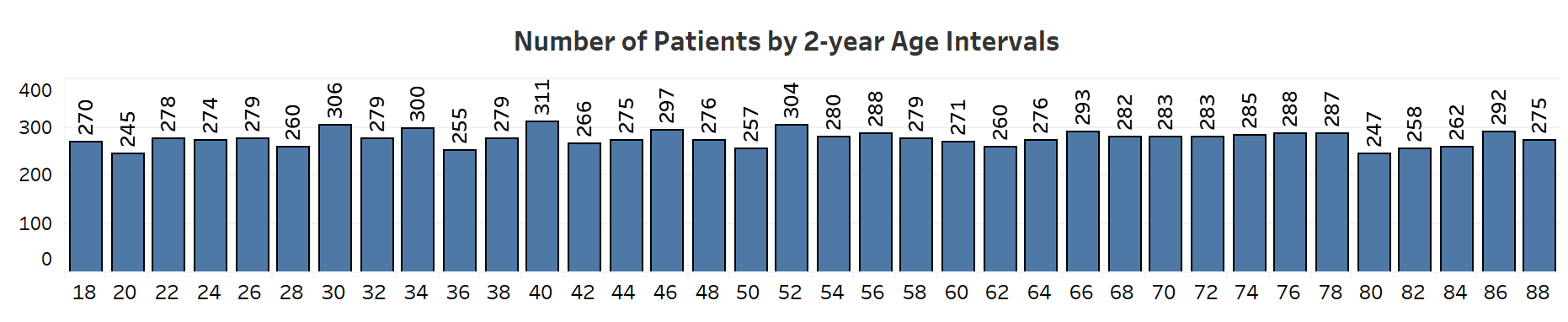
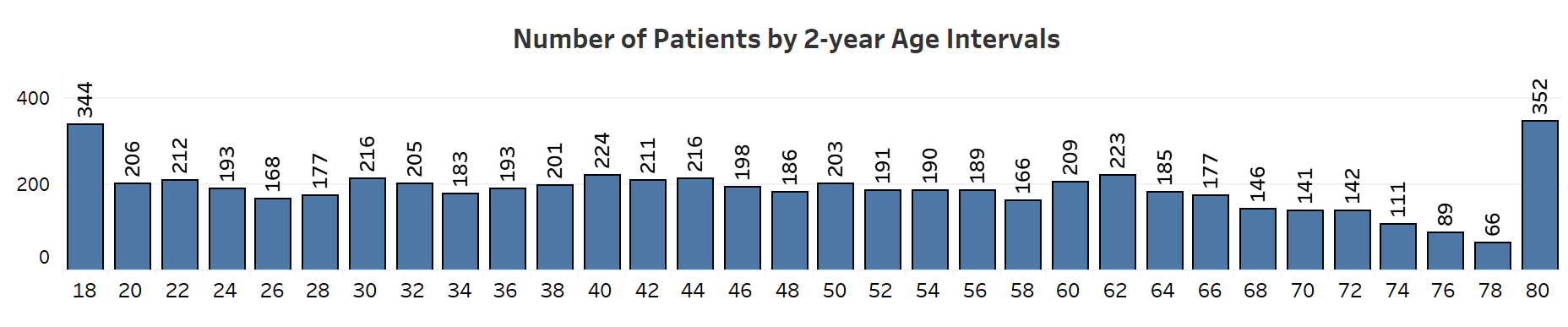
**Part 2: Storytelling with Data**

**B.** Link to Panopto presentation <https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=efb5ff30-7cdf-47af-b897-b21401779926>

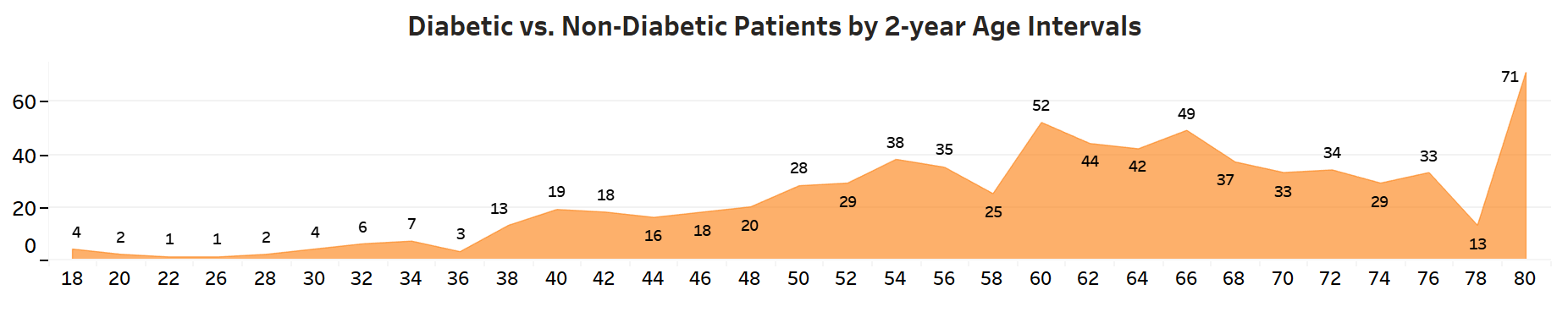
**Part 3: Reflection Paper**

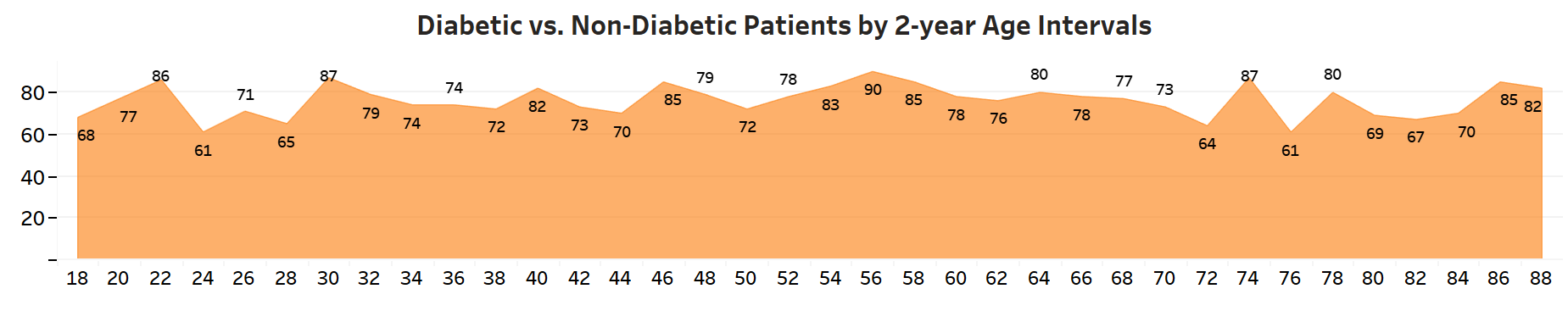
**C1.** My dashboard perfectly aligns with the needs outlined in the data dictionary by focusing on crucial health metrics like age, gender, diabetes, hypertension, and overweight conditions. For instance, the Dashboard “Exploring Patient Demographics and Health Conditions” includes a bar chart showing the "Number of Patients by 2-Year Age Intervals," a pie chart for the "Gender of Patients," and a line chart comparing "Diabetic vs. Non-Diabetic Patients by 2-Year Age Intervals." These visualizations make it easy to identify patterns and trends.

Let's see a couple of examples. The images below show how the number of patients differs between the CDC and WGU datasets. The upper bar chart demonstrates the number of patients in the CDC dataset, while the lower one shows the number of patients in the WGU dataset. In the CDC dataset, we see that the lowest numbers of patients are in the 76-77 and 78-79 age intervals, and the highest numbers are within the 18-19 and 80-81 age intervals. In contrast, WGU patients are distributed evenly across the age groups. Additionally, the WGU dataset has a higher total number of patients.



The next two images demonstrate how the number of diabetic patients differs in each age group in the WGU and CDC datasets. The upper image shows the CDC dataset with the number of patients with diabetes starting to rise with age, around 38 years old, reaching its first peak at 60 years. Compared to this, the WGU diabetic patients demonstrate no correlation between age and diagnosis.





By presenting this information clearly and concisely, this dashboard gives stakeholders the insights they

need to make informed healthcare decisions and improve patient outcomes. This data-driven approach ensures alignment with the data dictionary's focus on understanding and addressing key health metrics.

**C2.** Combining the CDC data with the WGU data gives us a clearer and more detailed picture of patient health. Since both datasets have the same variables, adding them together creates a larger and more diverse sample. This helps us spot unique health patterns that might be missed in just one dataset. For example, while the WGU data shows no age correlation for diabetes, the CDC data shows that diabetes cases rise with age. This difference highlights the need for different healthcare approaches: age-specific interventions for one dataset and generalized strategies for the other. Cross-referencing the data from both sources also helps us validate the results, ensuring our findings are reliable. The additional variables let us look deeper into how age and gender affect health outcomes, leading to better healthcare decisions. Overall, integrating these datasets gives us a fuller understanding of patient health, which helps us make more informed decisions and improve healthcare programs.

**C3.** My “Focused Health Condition Analysis” dashboard features two key data representations: highlight tables showing the “Number of Hypertensive Patients by 5-Year Age Intervals” and the “Number of Overweight Patients by 5-Year Age Intervals.” Executive leaders can use the hypertension table to identify age groups with the highest prevalence of hypertension, allowing them to allocate resources and develop targeted programs for those specific age brackets. The overweight table reveals the distribution of overweight conditions across different age intervals, helping leaders focus on weight management and preventive health measures.

The dashboard also includes filters such as Source, Age Intervals, and Gender. This enhances the analysis by allowing leaders to drill down into specific population subsets, identify unique patterns, and tailor interventions accordingly. Additionally, the six KPIs—Total Number of Hypertensive Patients, Fraction of Hypertensive Patients, Average Age of Hypertensive Patients, Total Number of Overweight Patients, Fraction of Overweight Patients, and Average Age of Overweight Patients—provide quick, actionable metrics that highlight the overall burden of these conditions.

These data representations, filters, and KPIs enable leaders to make informed, strategic decisions to improve patient care and health outcomes.

**C4.** In my dashboards, the Source and Diabetes filters are two essential interactive controls that help users customize their data view. The Source filter allows users to compare data from the WGU and CDC datasets or view a combined dataset, enabling them to spot discrepancies or similarities between the two sources. This is crucial for validating findings and understanding the differences in health patterns across datasets. The Diabetes filter, on the other hand, lets users isolate data for diabetic and non-diabetic patients, helping to identify specific trends and conditions associated with diabetes. These filters enhance the interactivity and personalization of the dashboard, allowing users to tailor the data presentation to their needs and gain deeper insights.

**C5.** To ensure my dashboard is accessible to individuals with colorblindness, I chose color palettes that are distinguishable for all color vision deficiencies, such as colorblind-safe palettes. 

Additionally, I included text labels to clarify the data further. This multi-faceted approach ensures that the dashboard is inclusive and interpretable by everyone, regardless of their ability to perceive color.

**C6.** The story I aim to convey is the significant differences between the CDC and WGU datasets. Two key data representations in my presentation illustrate this effectively.

Firstly, the comparison of diabetes prevalence highlights a rise in diabetes cases starting at age 38 and peaking at 60 in the CDC dataset, whereas the WGU dataset shows no age correlation. This discrepancy emphasizes the need for age-specific interventions in the CDC context and a more generalized approach for WGU.

Secondly, the analysis of hypertensive and overweight conditions reveals that these conditions increase with age in the CDC data, especially peaking in the 60-64 age group. Conversely, the WGU dataset shows no age-related pattern, reinforcing the message that healthcare strategies in WGU should be broadly inclusive rather than age-targeted.

These data representations underscore the differences between the two datasets, supporting the narrative that healthcare approaches must be tailored to each dataset's specific demographic and health profiles.

**C7.** I used audience analysis to tailor my presentation to resonate with executive leaders and healthcare stakeholders. By understanding their need for actionable insights and strategic decision-making, I focused on presenting data clearly and concisely. Visualizations like highlight tables, line charts, and KPIs were chosen to provide quick, impactful insights. This approach ensures the message is engaging and informative, helping the audience understand the importance of targeted interventions and resource allocation in improving patient care.

**C8.** I prioritized clarity and inclusivity to design my presentation for universal access by all audiences. I used clear, readable fonts and high-contrast color schemes to ensure visibility for everyone, including those with visual impairments. Recognizing the needs of colorblind individuals, I incorporated colorblind-safe palettes and included patterns and textures to differentiate data points effectively. Additionally, Tableau Public makes this presentation accessible to everyone online, making it easy to share and view. Considering these factors, I ensured the presentation was easily understandable and accessible to a diverse audience, making the data insights relevant and impactful for all viewers.

**C9.** Drawing from the elements of effective storytelling in my presentation, I utilized narrative and visuals to engage the audience effectively. In the narrative, I emphasized two main conclusions: the disparities between the patient populations of the CDC and WGU datasets and the differing patterns of health conditions within these populations. By highlighting these key points, I guided the audience through the data, helping them grasp the significance of the findings and their implications for healthcare strategies.

Additionally, I employed visual storytelling by incorporating clear and concise visualizations such as highlight tables, bar charts, and line charts. These visuals made complex data easily digestible, capturing the audience's attention and facilitating better understanding. Providing multiple data visualizations for the primary points ensured the narrative was well-supported and engaging. These elements created a compelling data story that helped stakeholders make informed decisions.

**D. Sources:**

*Best practices for telling great stories*. (n.d.). Tableau. https://help.tableau.com/current/pro/desktop/en-us/story\_best\_practices.htm

*Https://app.datacamp.com/*. (n.d.). Retrieved October 1, 2024, from <https://app.datacamp.com/learn/custom-tracks/custom-d210-representation-and-reporting>

*NHANES 2013-2014 Demographics Variable list. (n.d.). https://wwwn.cdc.gov/Nchs/Nhanes/Search/variablelist.aspx?Component=Demographics&CycleBeginYear=2013*

*National Health and Nutrition Examination Survey*. (2017, January 26). Kaggle. https://www.kaggle.com/datasets/cdc/national-health-and-nutrition-examination-survey

*NHANES 2013-2014 Questionnaire Variable list*. (n.d.). https://wwwn.cdc.gov/Nchs/Nhanes/Search/variablelist.aspx?Component=Questionnaire&CycleBeginYear=2013

Seto, F. (2024, April 15). *How to build Powerful Tableau KPI Dashboard (Step by Step)*. Coupler.io Blog. https://blog.coupler.io/tableau-kpi-dashboard/

Shaffer, J. (2016, April). 5 Tips on Designing Colorblind-Friendly Visualizations. *https://www.tableau.com/*. https://www.tableau.com/blog/examining-data-viz-rules-dont-use-red-green-together