Lora Milam

Western Governors University

D211 Advanced Data Acquisition

13 October 2023

D211 Performance Assessment

**1 Introduction**

Lora Milam Masters Data Analytics (2/10/2023) Program Mentor:d211@wgu.edu

* 1. **Link to Dashboard**

The dashboard for this Performance Assessment can be found at: <https://public.tableau.com/app/profile/mel.milam/viz/D211_PerformanceAssessment_Dashboard_LoraMilam/D210>

**1.2 Purpose and Function**

The primary needs outlined in the performance assessment instructions were to provide an easily navigable dashboard with broad and understandable insights to support executive decision-making. The datasets chosen for this analysis are the medical data set provided by WGU and a data set from Kaggle, Diabetes Health Indicators Dataset (Teboul). In this analysis, the charts compare high complication risk rates against patient demographics and pre-existing conditions. The purpose of this dashboard is to gain a deeper insight into the connection between patient demographics and pre-existing conditions and their impact on the heightened risk of patient complications.

**2 Business Intelligence Tool Justification**

The choice of Tableau as a business intelligence tool is justified due to its ease of use, strong data visualization capabilities, data integration support, scalability, advanced analytics features, supportive community, security and compliance standards, regular updates, and cost-effectiveness. These attributes make Tableau a valuable asset for businesses seeking to harness the power of data for informed decision-making.

**3 Data Cleaning**

The process of cleaning and preparing the medical data took place within both pgAdmin 4 and Tableau.

The steps to clean and prepare the medical dataset within pgAdmin 4 include:

* Review all tables within medical\_data database
* Detecting any absent data points in the columns that will be utilized in Tableau

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* Detecting any inaccurate data points in the columns that will be utilized in Tableau.
* Established a linkage between pgAdmin 4 and Tableau for the purpose of data transfer.
  + Logged into Tableau and navigated to data source. Select the data source tab at the bottom of the page. Scroll down the page to “To a Server” and choose “PostgreSql” in the expanded list of options.
    - Server: localhost
    - Port: 5432
    - Database: medical\_data
    - Authentication: Username and Password
    - Username: postgres
    - Password: Passw0rd!

The steps to prepare the medical dataset within Tableau include:

* Drag and drop tables “patient”, “location”, “complication”, and “servicesaddon”. The associated underlying SQL queries can be found

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* Underlying SQL for patient table

A screenshot of a computer

Description automatically generated

* Underlying SQL for complication table

A screenshot of a computer

Description automatically generated

* Underlying SQL for location table

A screenshot of a computer

Description automatically generated

* Underlying SQL for servicesaddon table

A screenshot of a computer

Description automatically generated

* Patient-complication relationship

A screenshot of a computer

Description automatically generated

* Patient-location relationship

A screenshot of a computer

Description automatically generated

* Patient-location relationship

A screenshot of a computer

Description automatically generated

The secondary dataset, the Diabetes Health Indicators Dataset, was preemptively cleaned and readied.

The steps to clean and prepare the medical dataset include:

* Download “diabetes health indicators.csv”
* Click “Data” then “New Data Source”
* Scroll to “To a File” and select “Microsoft Excel”
* Select “diabetes health indicators.csv” from Download folder

**4 Dashboard Creation**

The dashboard was constructed from a combination of two stories, each comprising two and four visualizations, respectively. The initial story consists of a color-coded map that illustrates the state-wise percentage of high-risk complications and a corresponding bar chart pinpointing the states with the highest percentage of high-risk complications. Additionally within the story, there are five filters available: state, patient's diabetes status, gender, patient's overweight status, and age. The second story comprises four color-coded charts that depict the male-to-female ratio across various demographics. The chart in the top left corner is a pie chart displaying the distribution of males and females within the secondary dataset. The chart in the top right corner is a bar chart that compares the distribution of positive heart disease attack cases by age. The chart in the bottom left corner is a bar chart comparing the distribution of positive diabetes cases by age. Finally, the chart in the bottom right corner is a bar chart that compares the distribution of positive high blood pressure cases by age. Lastly, the dashboard was crafted to ensure universal accessibility for all viewers. This was achieved by employing color palettes that accommodate those with color blindness, utilizing straightforward yet impactful visuals, and providing an online dashboard to ensure accessibility for all.

**5 Results**

Executive leaders could utilize the map within story one to determine where within the United States there are higher high-risk complications rates. This factor would aid executive leaders in determining where to allocate more resources due to higher probability of patients’ complication. Executive leaders could also utilize story two to determine a consensus of patient demographic that is more prone to health issues which could lead to higher high-risk complication rates. By knowing the most at-risk demographics, executive leaders could then provide more resources to areas with high concentrations of this population.

**6 Limitations**

One limitation of this data analysis is the disparity in the number of entries between the two datasets. The WGU medical dataset comprises 10,000 entries, whereas the secondary dataset consists of 253,681 entries. This difference in scale could potentially result in a loss of detail and resolution, possibly leading to the omission of valuable insights when working with the larger dataset to the small dataset.

Another limitation stems from the fact that the two datasets share only a small subset of features in common. This limited overlap makes it challenging to apply insights gained from the secondary dataset to the primary one when dealing with non-transferable features.

Additionally, the secondary dataset, sourced from Kaggle, may carry potential errors due to inaccurate or unverified responses during data collection and preparation. Similarly, the medical dataset from WGU is unverified and lacks supporting documentation, which makes it difficult to validate the use case and contextual information surrounding the data.

**7 Supporting Documentation**

**7.1 Video**

This can be found within the attached file ‘Panopto Recording’.

**7.2 Sources**

Teboul, Alex. (2021, December). Diabetes Health Indicators Dataset, Version 1.

Retrieved December 19, 2022 from

https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset.

Western Governors University. (n.d.). D211 Advanced Data Acquisition. Salt Lake City.