Data Dashboard and Storytelling

Matthew E. Heino

Representation and Reporting

Data Dashboard and Storytelling

Introduction

The purpose of this paper is to develop appropriate dashboards and provide a cogent story using data from two data sources. This assessment will include visualizations that represent various facets of the data. There is a requirement to bring in data that aims to bolster the conclusions and the questions of the data. The additional dataset will come from an outside source.

All sections for the assessment can be found in this paper. Part III the reflection paper can be found towards the end of this document in the appropriate section. Unless otherwise noted all information for the assessment can be found in the appropriate documents that will be included with the submission.

Background

This assessment will be composed of two datasets. The first dataset will be provided by the university. The dataset that will be used in this assessment will be the medical dataset. This dataset is composed of 50 columns and 10,000 rows of data. The data for this assessment is purported to be cleaned and no further preparation will be needed. It will be used in Tableau without any outside preparation. Any preparation will be handled inside of Tableau, but this will be limited to renaming columns and a few other tasks to help distinguish this set of from the imported dataset.

The next data set will be a data set that will be composed of two CSV files. This data set will be retrieved from Kaggle. Kaggle is a website where there are numerous freely available datasets. The only requirement will be the creation of an account to access some of the features

or datasets of the website. Creating an account is free using the appropriate credentials. A discussion of the dataset and how it was prepared is the subject of another section in this paper.

Part I. Interactive Dashboard

In this section, you will find what has been included in the interactive dashboard. There will be an inclusion of the following types of data representations. File names of the datasets that were used to create the dashboard and the appropriate representations. To aid the user there will be a brief instruction on how to use the dashboard and its functionality.

A1. Copy of the Datasets. In this section, you will find the names of the data files that were used to create the dashboard. The names of the files are:

- **Cleaned_CDC_File_Task_1.csv** – This is the cleaned version of the dataset of the secondary dataset that was sourced from Kaggle. There was some outside data preparation performed on this data. This was to trim down the number of columns and there was a need to combine information from two CSV files. This preparation was to have access to information to better complement the primary data source. The one that was supplied by WGU.
    - This "cleaned" set is composed of the following two CSV files.
        - **demographic.csv**
        - **questionnaire.csv**
        - These were downloaded from the following website:
            - https://www.kaggle.com/datasets/cdc/national-health-and-nutrition-examination-survey

    - These were cleaned using a Python script that joined, created columns that were more user-friendly, and imputed data that was missing from the second data source. It was far easier to use a Python script to handle this task as opposed to using Tableau.
        -
    - The data preparation yielded the cleaned data set that was given in the first bullet point (**Cleaned_CDC_File_Task_1.csv**).
    - The script that was used to create this final is the following:
        - **Heino D210 Task 1.py**
        - This was included with the submission, but it will not be required to run the Tableau submission. This file will be included with the submission. (The reason is that there were no annotations in the rubric about which files should and should not be included with the submission.)
- **medical_clean.csv -** This was the dataset that was provided by the university. This was not altered in any way.

These files can be found in the attachments that accompany the submission of this assessment. When running the dashboard you will only need to include the following files in a suitable directory.

A2. Instructions for Installing Dashboard. To use the dashboard please follow these instructions. You will need to have a copy of the Tableau with a suitable license. The workbook that was created used the Desktop edition. There is a suitable free trial that will allow you to use this version for a period of seven days. Please download the version that suits your needs.

Once you have downloaded and installed the Tableau. You will need to copy the following files to a suitable directory. These files will be included with the submission and are named the following:

- **medical_clean.csv**
- **Cleaned_CDC_File_Task_1.csv**

There are no other CSV files required for this submission. You may start your edition of Tableau. You can connect to the data sources from the main screen of Tableau. You will need to add the two files stated earlier as the sources of the workbook

To load the workbook that has all the worksheets, the dashboard, and the story you will need to load the file called:

- **Heino D210 Workbook .twbx**

This is accomplished by going to the **File** menu → **Open** and choose the file with the name that was given in the previous bullet point. This should open the workbook that contains all the worksheets, the dashboards, and the story.

You will be able to view the all components that make up the dashboard as separate worksheets and the story.

A3. Instructions for Using the Dashboard. To use the dashboard please follow the instructions that are included below. There are essentially four dashboards that are created for this assessment. The dashboards are the following:

- Introduction Dashboard
- Location of Patients Dashboard
- Demographics of Patients Dashboard
- Our Patients vs CDC Patients
- Patient Condition By Number

The dashboards should be found at the bottom of the Tableau. The tabs at the bottom of the page should be ordered with the story first, the dashboards, and then the associated worksheets.

There are no instructions for the use of the **Introduction Dashboard**. To use the **Location of Patients Dashboard**. You will find a series of filters located on the right side of the

page. You are free to choose almost any combination of features to see information about the patients that are available in the dataset. The options to filter are:

- State
- Complication risk
- Re-Admission
- Gender.

This will change the information that is displayed in the various parts of the dashboard. If you click on a city additional information will be displayed as a tool-tip. Information will be given to the patient(s) that live in that particular city.

The next dashboard is the **Demographics of the Patients**. This gives a little more information about the patients that were found in the medical dataset. There isn't any inclusion of the outside dataset. This is found in the last dashboard of the story (**Our Demographics versus CDC Demographics**). The **Demographics of the Patients** dashboard offers some additional filters to filter items like the following:

- State
- Complication Risk
- Re-admission
- Gender
- Number of Children
- Marital Status

These can be chosen in almost any combination and the data will be updated accordingly.

The last dashboard (**Our Patients vs CDC Patients**) is a comparison of the average age of the patients compared to the CDC data that was imported as the second data source. You will need to choose the values for the features separately. This is because this dashboard uses two "unrelated" sources. (There wasn't an easy way to create a suitable association without altering the second source's data significantly.)

These dashboards allow the user to look at information about the patient by clicking on the point or segment of a pie chart to see relevant information about the user. The information will be displayed in the form of a tooltip. This will be illustrated in the accompanying video.

The last dashboard is a summary of information on our patients as compared to the data that was procured from the CDC dataset. There is limited interactivity with this dashboard. You will only be able to scroll up and down to view the data more easily.

Part II. Interactive Dashboard

In this section, there will be a link to the Panopto recording. This will include a presentation of the dashboard. There will be a discussion of the background of the data and how the two datasets relate to each other. There will be a discussion of the results of the analysis that was made using Tableau. A presentation of the representations that support the results that were discovered using Tableau. Finally, there will be a discussion of the insight and a course of action that may be pursued because of the results of the analysis.

The link to the simulated presentation can be found below. The link is:

-

Part III. The Reflection Paper

In this section, there will be a discussion of the purpose of the dashboard. There will be a discussion about the variables that were chosen to create the dashboard. A discussion of the representations that were created in the dashboard and how they can be used. The dashboard will include interactive controls to be used. The dashboard was created with people who have color blindness in mind this will be discussed in the appropriate section. The story will be discussed and how the visual representations support the story. Other pertinent items will be discussed in this part of the assessment.

C1. Purpose and Function. The purpose of the dashboard is to display selected demographics about the patients that were in the dataset. This dashboard is composed of a series of dashboards that adequately display pertinent data about the demographics of the patients that

are recorded in the dataset. The reason for displaying data in multiple dashboards was to better visualize the demographics of the patients. The design rationale will be further discussed in a subsequent section of the assessment.

The dashboard allows interested individuals to see who and where the patients are located. This will help understand what is the makeup of the patients that enter the system. If we understand who are the patient's we can better serve their needs. We can begin to develop services that better address their needs.

The information that was used to develop all came from the datasets that were supplied. The format of the data is currently a CSV file. This form of data communication does not allow for an interested individual to "understand" who the patients are. There is no relationship given between the information. It is easier to see who the patients are when they are depicted in the manner of using maps and other components to better display information about the patient.

Seeing the patients displayed on a map can give the user of the dashboards a better feel of where the patients are located and what demographics are associated with the patient. This association will be covered and seen in the accompanying Panopto presentation.

C2. Explanation of the variables. The variables that were chosen for the creation of the dashboard were all the variables that gave more information about the patient as a person or individual. The dashboards included only information that could be used to directly identify the patient as an individual. The list below is a brief summation of the types of data that were used to create the dashboards.

- Age
- Gender
- Income
- Complication Risk
- City
- State
- Latitude and longitude
- Condition status (Yes/No, high blood

pressure, diabetes, stroke, etc.)

These were mixed with the secondary dataset (CDC). Although some of the data was not an exact match. There was enough of a relationship to help build statistics that were pertinent to the primary dataset.

All these variables were used in an attempt to get a feel for who the patients are, where they are, and the types of conditions the patients may suffer from. This will make it easier to see if there is a specific area that needs to be addressed.

C3. Explanation of the Representation. The representations that are included in the dashboards allow the user to see the distribution of the patients and some information about them. There are a few basic representations that are included with the dashboards. The first ones are the maps. This representation was to show the patients by state. This was done to give the user a visualization of where the patients are located throughout the country and the state. This was the best way to illustrate where the patients are located using another visualization you will not be able to see the relationship between location and the patient.

The other representations are there to give some statistical information either in aggregate or broken down based on how the filters are set. For example, on the Location of Patients and Demographics of the Patients dashboards the user is presented with some important demographics about the readmission of the patients. These are some important statistics to keep in mind when reviewing patient demographics. As an institution, there is a need to make sure the patients are not readmitted. Using this approach it can be helpful to see these statistics as a reminder.

To help better understand the individual patients some tooltips give information about the patient. It allows for more granularity when there is a point of interest in the use of the dashboard.

C4. Explanation of the Interactive Controls. The controls that are included on the dashboards allow the user to filter based on various types of criteria. The common ones that will be found in the dashboards are:

- **State -** This filter allows the user to look at states on their own or as a national map.
- **Complication Risk** – Allows the user to filter based on the complication risk.
- **Re-Admission** – To allow the user to filter based on whether the patient is a readmission.
- **Gender** – To allow the user to filter based on the gender of the patient

Some of these filters have a little more granularity than others. This is to control the context of the filter. Some options do not make sense and this leads to errors or a cluttered visualization if some of the options were included.

C5. Description of Dashboard Creation. The dashboard was created to be freely accessible by individuals who are colorblind by using a color scheme that is visible to those with colorblindness. The color palette was chosen based on the color palette that is readily available within the Tableau platform. Creating the dashboard through the use of colors that should not cause problems. The backgrounds that were used were basic white. This should not cause a problem for any users and it gives the dashboards a nice neat and uniform appearance.

C6. Representations and the Story. The story that wanted to be illustrated with the representations was one that gives a personal feel to the patients. The representations or maps were used to show that the patients were not mere data in a dataset. The best way to do this is to

show them on a map. Using the map representations allows people to see the types of patients that may be near to them. Using location may help the user develop an affinity with the patient as the patient may be someone who is local to them and may even be a neighbor.

Developing the representations using maps and the use of tooltips to display other information makes it easy to see "who" these patients are. They are no longer just numbers or entries in a dataset, but they are people who could be individuals from the same state or even the same city as the user.

These are two of the representations that will add the most to the story because they allow for the use of seeing patients as a group (state or country) or as an individual (using the tooltip). This will aid in making the story of the patients more poignant to those who use the dashboards.

C7. Audience Analysis and the Presentation. The audience for this presentation was executives who need to have as much information at their fingertips as possible. The presentation was to provide a way to “humanize” the patients to the executives who make the policies. These policies influence the patients and giving the executives a way to see information as they pertain to the patients as a group and as an individual was the goal.

Using the maps to display information about the patients helps bring home the point that these patients are not merely data but people who may live near them. Any decisions that these executives make may affect someone in their community.

Executives do need to concern themselves with numbers, but behind the numbers in the medical dataset, there is a person. This presentation wants to instill within the leadership that these people are more than mere numbers and data.

C8. Design for All Users. In the design for the dashboards, there were a few things to keep in mind. The first is there needs to be a dissemination of information in a manner that is

accessible to all individuals. This means that the information that is displayed must not use terms or jargon that may be specific to a department or business role. The terms that were used in this presentation are relatively generic they are not specific to a specific job role or title.

In regards to accessing the dashboards, they are freely accessible using Tableau's software and downloading the appropriate software component like desktop or running through their cloud application.

The color scheme for the dashboard presentations took into account the needs of those who can't see color efficiently. This need was met using the color palette suggested by the Tableau platform.

Using this approach allowed the dashboard to be freely available to the largest audience within the organization. This approach does have one negative caveat. It is not available to those who do not reside within the organization. This could be alleviated by publicly publishing this to the Tableau public community, but this was not done for this submission.

C9. Elements of Effective Storytelling. Two elements of effective storytelling are knowing your audience and providing the relevant data to the audience. Before you can begin to craft an effective data story you need to understand the audience who will consume the presentation. For this presentation, the audience was executive decision-makers in the organization.

With this in mind, you can compose a narrative that pitches the idea of the story that is most appropriate to this segment of the audience. You will need to ask relevant questions about not only the data but about the audience. When comes to questions about the audience you need to answer questions such as (Dunford, 2022):

- What information is relevant to them?

- How does the audience best digest the information given to them?

The audience for this presentation needs to have the information presented quickly and with a sizable amount of information. Keeping in mind that if the audience accesses the dashboards outside of the presentation they should be provided with the information that they are looking for. This information should be presented in an easy-to-understand manner. The information must be role-neutral. This means that the jargon used should be everyday words and not role-specific in any manner.

The audience that was the target of this presentation was the executives. These executives need to have information that is easy to gather from the visual representation. The presentation of the information must be quick precise and relevant to the topic at hand. This was accomplished using maps, associated charts, and tooltips with the appropriate information. The executive can click on an area of interest and view the available information about the patients as a group or look at the patients as an individual.

The way dashboards are set up in a way that makes it easy to expand as more information becomes available. You can add dashboard to accommodate other research questions that may be desired by the executives. This division of the data into separate dashboards allows the executive to ask questions. These questions may be thought of as the executives peruse the given dashboard. These additional dashboards can be added to give a more complete rendering of the story without losing the overall context of the initial story.

The elements that were given in the bullet points were areas that allowed the design of the dashboards and the subsequent story. Keeping these items in mind it allowed for a story that presented the greatest amount of information to the audience. Using the appropriate graphs allowed information to be displayed in a visually relevant manner to the executives. The

executives could garner a lot of information just by glancing at the initial maps, but the executives could drill down further using the appropriate filters to see additional data that may be of interest to them.

References

D. In-text citations

This section will contain any in-text citations that were used to create the Tableau dashboard as well as the accompanying written assessment. This is in addition to resources that may have been provided by the university and those included with the appropriate DataCamp resources.

The following is the citation for the image that was used in the Story. It is the link with information about who created the image as well as the website. I have included as much information as possible to adequately cite this work.

Cottonbro Studio. (2020, March 13). *Group of People Wearing Face Masks*. Pexels. Retrieved December 14, 2023, from https://www.pexels.com/photo/group-of-people-wearing-face-mask-3957992/

Dunford, C. (2022, April 25). *Brent Dykes "Effective Data Storytelling" - Old Street Solutions*. Old Street Solutions. Retrieved December 16, 2023, from https://www.oldstreetsolutions.com/brent-dykes-effective-data-storytelling