SLM1 — TASK 1: DATA ANALYSIS

**ADVANCED DATA ACQUISITION — D211**

**PRFA — SLM1**

Jason Smith  
Student ID: 293963  
[jsmi268@wgu.edu](mailto:jsmi268@wgu.edu)  
WGU – MSDA  
Program Mentor: Christiana Okhipo

**Part 1: Interactive Data Dashboard**

**A.  Dashboard link:**

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=b476b746-5410-4cfa-bbbc-ae9a0005890a>

**A1.  Datasets:**

The following two datasets have been uploaded.

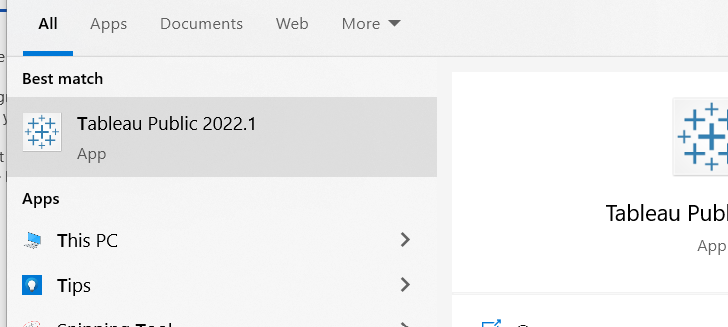
**Original file:** “med\_data\_clean.csv”

**Additional file:** “acs2015countydata.csv” (Muonneutrino, 2019)

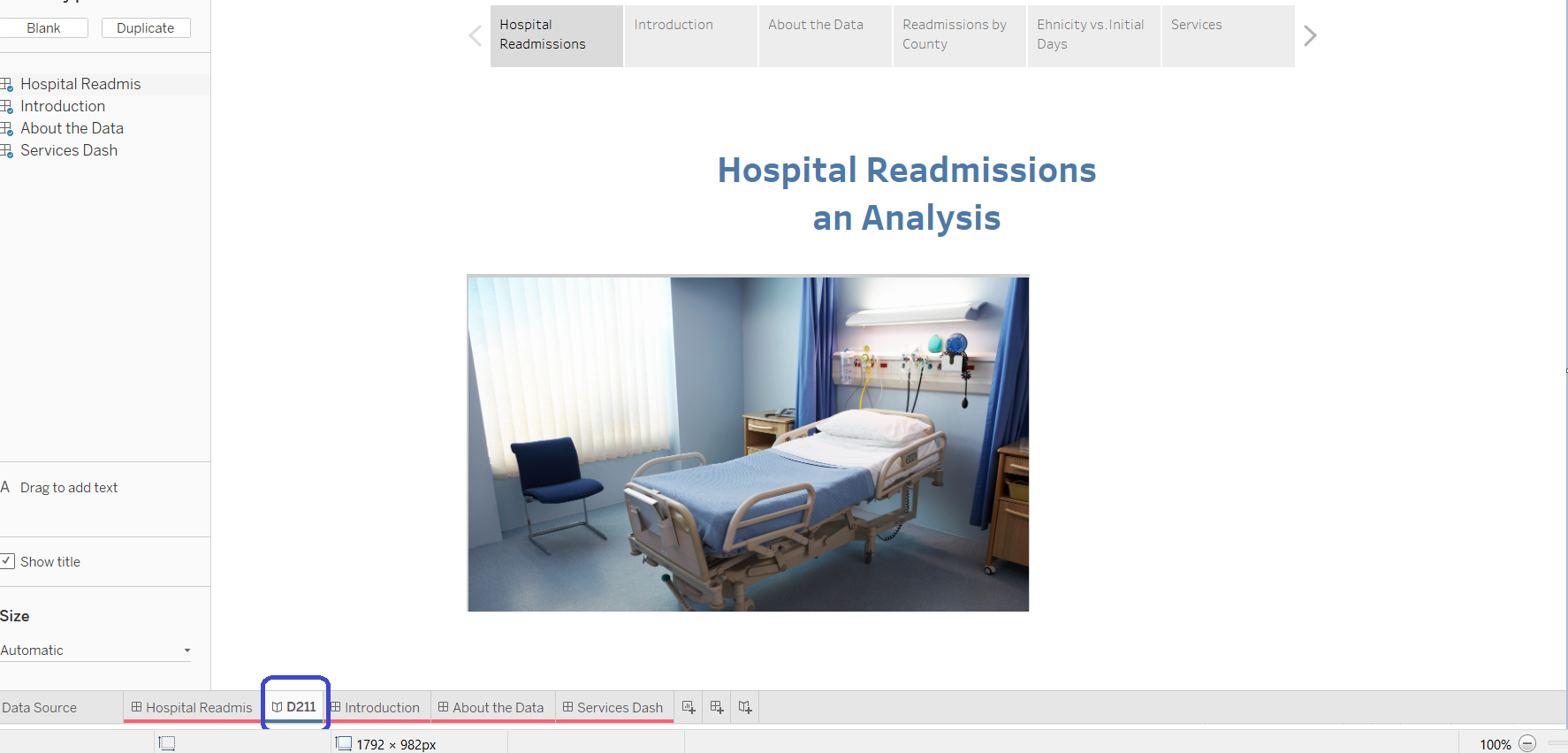
**A2.  Dashboard installation:**

**\*\*** Please note the following steps assume the “med\_data” schema is already loaded into Postgres environment.

1. Download the attached file ‘D211 Advanced Data Acquisition.twb.
2. Click the Start button, then type Tableau.
3. Click on “Tableau Public 2022.1, see below:



1. Once Tableau open, go to File, Open.
2. Navigate to the folder where you downloaded the “twb” file, select the file and click Open.
3. Click on the D211 story board link at the bottom of the page. See the following screenshot, the D211 story board is highlighted in blue.



**A3.  Dashboard navigation:**

Once the dashboard is open in a browser, clicking on the individual tabs located at the top of the page or the adjacent arrows will page you (frontwards or backwards) through the dashboard story analysis.

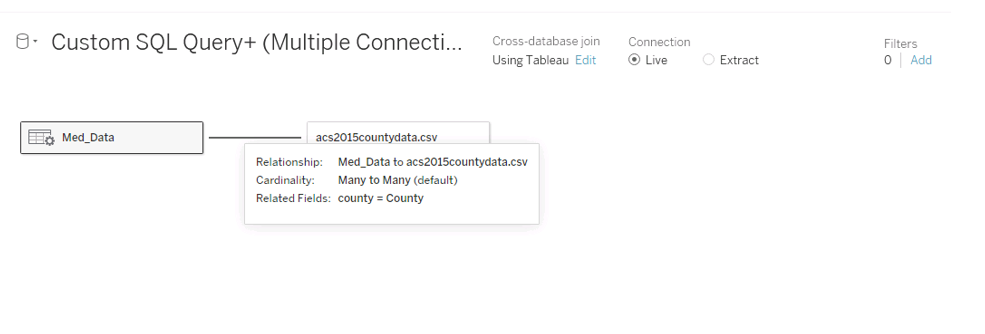
The tabs included in the storyboard include:

* Hospital Readmissions

an Analysis

* Introduction
* About the data
* Ethnicity vs. Initial Days
* Services

**A4. Provide a copy of all SQL code and other code supporting the dashboards.**

1. The attached file, “DDL.sql” contains the database-schema-creation statements.
2. The file “query.sql” contains the query that was used to extract the data that is presented in Tableau. The following tables were accessed in the query: patient, location, complication, and “servicesaddon”. These tables were interconnected with inner joins.
3. The file “acs2015countydata.csv” was connected to the “med\_data” database instance in Tableau by using a cross-database join on county, see below: 

**Part 2: Demonstration**

**B.  Link to a Panopto multimedia presentation:** [**link**](https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=b476b746-5410-4cfa-bbbc-ae9a0005890a)

**Part 3: Reflection Paper**

**C.  Reflection paper:**

**C1.  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 census dataset “acs2015countydata.csv” (Muonneutrino, 2019), was used in combination with the medical dataset for this analysis, see attached. Both datasets are scoped nationally for the year 2015. City and state attributes are common in both datasets and were used in the join condition. The census data attributes that are used in my analysis consist of:

* Asian Alone - Numeric
* Black Alone - Numeric
* Hispanic or Latino Alone - Numeric
* Native Hawaiian Alone - Numeric
* White Alone - Numeric

**C2.  Justify the selection of the business intelligence tool you used.**

* Postgres database was used to store the medical data.
* PGAdmin4 was used to run queries against the database.
* Tableau’s custom SQL feature was used to select a subset of the data. Tableau was also used to link the SQL and the CSV data sources and to display visualizations of data relationships.

**C3.  Explain the steps used to clean and prepare the data for the analysis.**

1. The medical dataset is contained in a database instance called “Medical\_Data”.
   * Connected to the Postgres database using PGAdmin and ran SQL queries to determine if null values exist. For example:  
     select \* from patient where readmis is NULL;

No null values were observed.

* + Ran SQL queries to determine if data appears inaccurate. For example:

Select \* from patient.

* + A SQL query was used to extract the hospital data from 4 tables. Refer to attached file “query.sql”.
  + Logged into Tableau and navigate to data source. Selected the data source tab at the bottom of the page then searched down the page for “To a Server” and choose Postgres in the expanded list of options. Clicked Custom Query and pasted the contents of “query.sql” into the text box. Renamed custom query to MedData then clicked OK.

1. The census data is contained in a file called “acs2015countydata.csv”.
   * Clicked add data source and selected “acs2015countydata.csv”.
   * The two data sources were then joined on county with a many to many cardinality. See below: Graphical user interface, text, application

     Description automatically generated

**C4.  Explain two interactive controls in your dashboard and how each enables the user to modify the presentation of the data.**

The “Readmissions by County” story point consists of a map chart along with the top 10 ranking of readmissions per county. By clicking an item in the county ranking chart, will change the section criteria on the map chart. Further, the map chart allows filtering by county, state, and readmission (yes or no).

The “Services” story point contains the Services vs. Readmissions tree map and a table of displaying similar information. There are several interactive controls that stakeholders would find useful. Clicking on a specific portion of the tree map will focus the display of the table on the same aspect. The opposite is true as will. Clicking on a specific portion of the table will display complementary information in the tree map. This tab also contains slide filters for various demographics, such as for Hispanic, White, Pacific, Native, Child Poverty, Black, and Asians.

**C5.  Describe how you built your dashboard to be accessible for individuals with colorblindness.**

I took several steps to build dashboards that are accessible by individuals with colorblindness. The “County Readmissions”, Services, and Initial days with Demographics tabs all use color coding that is accessible by the color blind. In addition, the County Readmissions chart and services tree map uses size to indicated readmission density.

**C4.  Summarize the steps used to create the dashboards.**

A story board was created to present a useful tool to Hospital executives which provide a window as to which U.S. based patients are most likely to be readmitted. The medical dataset was combined with U.S. Census data for the year 2015. The story board presents the audience with the question being analyzed, an explanation of the data used, and dashboards that demonstrate various effects on hospital readmissions.

**C5.  Discuss the results of your data analysis and how it supports executive decision-making.**

Various trends related to readmissions can be ascertained from the story board.

The “Introduction” story point specifies the target question being analyzed, “Which patients are most likely to be readmitted”? In addition, a reason prompting this study was also provided.

The “About the Data” story point provides an explanation of the data used in the analysis and where it came from.

The “Readmissions by County” story point provides a ranked chart that displays the top ten counties with the highest readmission rates. The map chart uses size and color to display the highest concentrations of readmissions. Additionally, demographics can be filtered on to see these effects.

The “Ethnicity vs. Initial Days” story point displays the relationships that initial days and ethnicity have on readmissions. Patients that have an initial stay of more than 45 days are most likely to be readmitted down the road. White patients are much more likely to visit hospitals as compared to other nationalities. For this reason, Whites are more likely to be readmitted.

The “Services” story point shows the relationship that different service types have on readmissions. Patients that are provided injection-type services (Bloodwork, Intravenous) are much more likely to be readmitted.

There are several trends this data analysis provides that can aid hospital executives to determine which patients are more likely to be readmitted.

**C6.  Discuss the limitation(s) of your data analysis.**

Census data may not always be accurate due to error in model-based estimates and non-truthful responses (United States Census Bureau,10/08/2021). On a similar note, the accuracy of the medical data is unverified as well.

In addition, Tableau formatting is limited to 16 columns. Tableau’s parameters are always static and need to be manually updated when data is modified (Absent Data).

**Sources:**

1. Muonneutrino (2019), US Census Demographic Data (2015): <https://www.kaggle.com/datasets/muonneutrino/us-census-demographic-data>

2. Hospital, http://weknowyourdreams.com/single/hospital/hospital-09

3. Hospital bed, http://weknowyourdreams.com/single/hospital/hospital-09

4. Saunders ND, Nichols SD, Antiporda MA, Johnson K, Walker K, Nilsson R, Graham L, Old M, Klisovic RB, Penza S, Schmidt CR. (October 15, 2011–April 15, 2012). Examination of Unplanned 30-Day Readmissions to a Comprehensive Cancer Hospital. <https://ascopubs.org/doi/10.1200/JOP.2014.001546>

5. Absent Data (), Advantages and Disadvantages of Tableau,   
<https://www.absentdata.com/advantages-and-disadvantages-of-tableau/>

6. United States Census Bureau,(October 8, 2021), Source and Accuracy,   
https://www.census.gov/programs-surveys/sahie/technical-documentation/source-and-accuracy.html