**D211 Task 1: Part 3**

**Data Analysis Report**

Amanda Terzin

D211 Advanced Data Acquisition

David Gagner

May 22, 2022

**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.**

The scenario for the medical data set states that I have been "asked to build a data dashboard to enable executive leaders to explore the data, identify trends, and compare key metrics" for a hospital chain to evaluate patient data for readmission insights. The dashboard I created includes the 10,000 patient files from the hospital chain and 2018 readmission data from the Centers for Medicare and Medicaid Services (CMS). CMS maintains statistics for all hospitals in the United States and the data set has readmission rates for each state broken down into age groups and gender (CMS, n.d.). The dashboard allows the user to explore the readmission rate by state, age, gender, length of stay, medical conditions, and total charges in the primary set and compare rates by age, gender, and state with the CMS data set.

At the top of the dashboard, there is a summary of our patient data by the number of patients, the number of those patients readmitted, the readmission overall rate, and the rate by gender and age group. All these statistics update with filter selection throughout the dashboard. If a user selects diabetes, the numbers show only patients with yes in the diabetes field. CMS data by overall rate, gender, and age are to the right of the hospital data for comparison and will filter by state. Four visualizations represent the readmission data for the initial length of stay (LOS), total charge, medical conditions, and state. Using this information, users can look for trends and patterns of readmitted patients and see how the hospital chain is measuring up against other hospitals.

# 2. Justify the selection of the business intelligence tool you used.

I used Tableau to make my dashboard. It can connect to the SQL database and the data can be refreshed as the database is updated. There are multiple ways to share dashboards and workbooks with others. Dashboards can enable non-technical individuals to explore data independently. (Loth, 2019)

# 3. Explain the steps used to clean and prepare the data for the analysis.

The data did not require much cleaning. There were no nulls, duplicates, or inappropriate entries. One column name had a misspelling. The external data was loaded into a new table in the medical\_data database.

1. There was one misspelled column name, "hignblood" which was changed to "highblood".

ALTER TABLE patient

RENAME COLUMN hignblood TO highblood;

1. A table was created for the external data set.

CREATE TABLE IF NOT EXISTS public.cms

(

state text COLLATE pg\_catalog."default" NOT NULL,

age integer,

less\_65 integer,

a65\_to\_74 integer,

a75\_to\_84 integer,

a85\_more integer,

fem integer,

male integer,

CONSTRAINT cms\_pkey PRIMARY KEY (state)

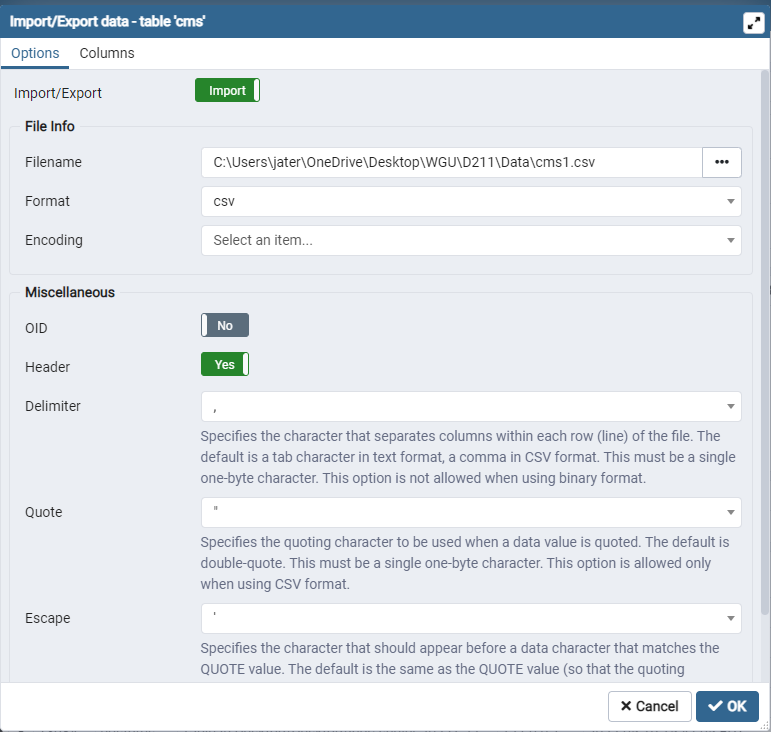
)

TABLESPACE pg\_default;

ALTER TABLE IF EXISTS public.cms

OWNER to postgres;

1. External data was loaded using pgAdmin.

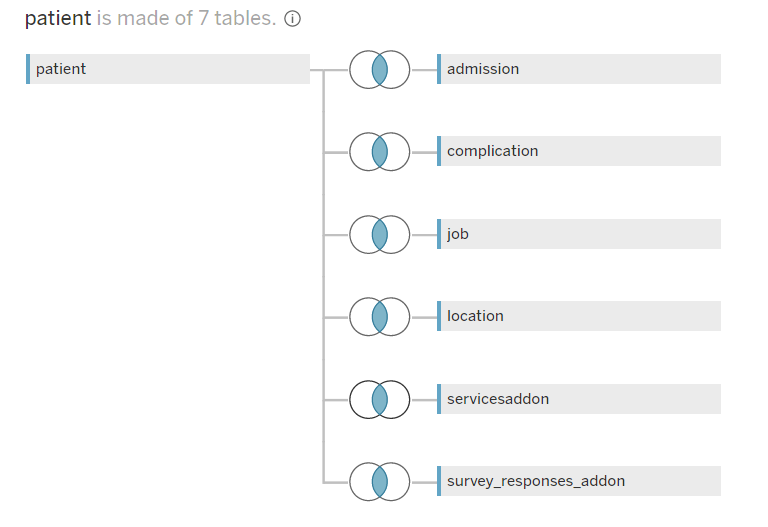


**4.  Summarize the steps used to create the dashboards.**

I created the dashboard in Tableau using the following steps:

1. Connect to the data source through the PostgreSQL server
2. Create custom SQL to join medical\_data tables on the primary and foreign keys

(https://help.tableau.com/current/pro/desktop/en-us/customsql.htm)



SELECT "patient"."additional\_charges" AS "additional\_charges",

"admission"."admins\_id" AS "admins\_id",

"patient"."admis\_id" AS "admis\_id",

"patient"."age" AS "age",

CAST("servicesaddon"."allergic\_rhinitis" AS TEXT) AS "allergic\_rhinitis",

CAST("servicesaddon"."anxiety" AS TEXT) AS "anxiety",

CAST("servicesaddon"."arthritis" AS TEXT) AS "arthritis",

CAST("servicesaddon"."asthma" AS TEXT) AS "asthma",

CAST("servicesaddon"."backpain" AS TEXT) AS "backpain",

"patient"."children" AS "children",

CAST("location"."city" AS TEXT) AS "city",

"patient"."compl\_id" AS "compl\_id",

"complication"."complication\_id" AS "complication\_id",

CAST("complication"."complication\_risk" AS TEXT) AS "complication\_risk",

CAST("location"."county" AS TEXT) AS "county",

CAST("servicesaddon"."diabetes" AS TEXT) AS "diabetes",

"patient"."doc\_visits" AS "doc\_visits",

"patient"."full\_meals" AS "full\_meals",

CAST("patient"."gender" AS TEXT) AS "gender",

CAST("patient"."highblood" AS TEXT) AS "highblood",

CAST("servicesaddon"."hyperlipidemia" AS TEXT) AS "hyperlipidemia",

"patient"."income" AS "income",

CAST("admission"."initial\_admission" AS TEXT) AS "initial\_admission",

"patient"."initial\_days" AS "initial\_days",

"survey\_responses\_addon"."item1" AS "item1",

"survey\_responses\_addon"."item2" AS "item2",

"survey\_responses\_addon"."item3" AS "item3",

"survey\_responses\_addon"."item4" AS "item4",

"survey\_responses\_addon"."item5" AS "item5",

"survey\_responses\_addon"."item6" AS "item6",

"survey\_responses\_addon"."item7" AS "item7",

"survey\_responses\_addon"."item8" AS "item8",

"job"."job\_id" AS "job\_id (job)",

"patient"."job\_id" AS "job\_id",

CAST("job"."job\_title" AS TEXT) AS "job\_title",

"patient"."lat" AS "lat",

"patient"."lng" AS "lng",

"location"."location\_id" AS "location\_id (location)",

"patient"."location\_id" AS "location\_id",

CAST("patient"."marital" AS TEXT) AS "marital",

CAST("servicesaddon"."overweight" AS TEXT) AS "overweight",

CAST("servicesaddon"."patient\_id" AS TEXT) AS "patient\_id (servicesaddon)",

CAST("survey\_responses\_addon"."patient\_id" AS TEXT) AS "patient\_id

(survey\_responses\_addon)",

CAST("patient"."patient\_id" AS TEXT) AS "patient\_id",

"patient"."population" AS "population",

CAST("patient"."readmis" AS TEXT) AS "readmis",

CAST("servicesaddon"."reflux\_esophagitis" AS TEXT) AS "reflux\_esophagitis",

CAST("servicesaddon"."services" AS TEXT) AS "services",

CAST("patient"."soft\_drink" AS TEXT) AS "soft\_drink",

CAST("location"."state" AS TEXT) AS "state",

CAST("patient"."stroke" AS TEXT) AS "stroke",

"patient"."totalcharge" AS "totalcharge",

"patient"."vitd\_levels" AS "vitd\_levels",

"patient"."vitd\_supp" AS "vitd\_supp",

"location"."zip" AS "zip"

FROM "public"."patient" "patient"

INNER JOIN "public"."admission" "admission" ON ("patient"."admis\_id" =

"admission"."admins\_id")

INNER JOIN "public"."complication" "complication" ON ("patient"."compl\_id" =

"complication"."complication\_id")

INNER JOIN "public"."job" "job" ON ("patient"."job\_id" = "job"."job\_id")

INNER JOIN "public"."location" "location" ON ("patient"."location\_id" =

"location"."location\_id")

INNER JOIN "public"."servicesaddon" "servicesaddon" ON ("patient"."patient\_id"

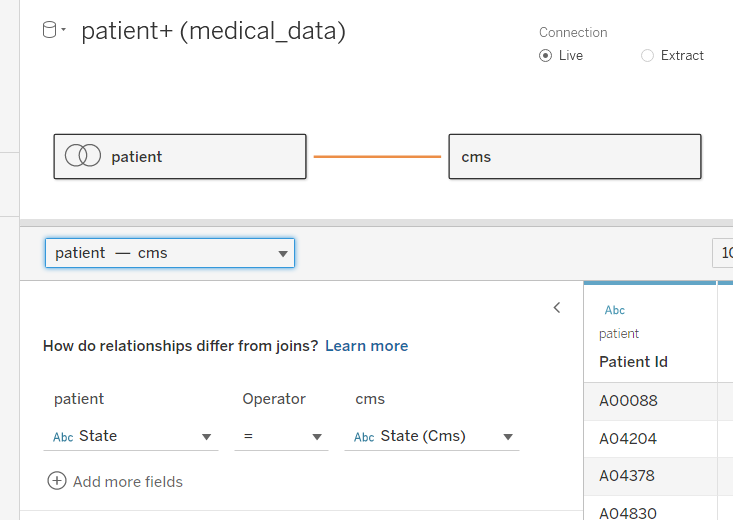
= CAST("servicesaddon"."patient\_id" AS TEXT))

INNER JOIN "public"."survey\_responses\_addon" "survey\_responses\_addon" ON

("patient"."patient\_id" = CAST("survey\_responses\_addon"."patient\_id" AS

TEXT))

1. Create a relationship between medical\_data state and cms state



1. Rename columns Initial LOS and Readmission for clarity
2. Create calculated fields for the number of patients, number of readmitted patients, readmission rate, medical conditions\_yes, CMS admission rate, and CMS rate by age groups
3. Create groups for medical\_data age to facilitate direct comparison to the CMS age groups
4. Create bins for initial LOS to make a bar graph
5. Create sheets for use in the dashboard
6. Place sheets on the dashboard and format
7. Set the "use as a filter" settings for sheets
8. Place and format drop-down filters
9. Place instruction boxes
10. Run optimizer and address warnings

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

The analysis of the data shows that the hospital system's readmission rate is more than twice the CMS rates for state, age, and gender. Performance is poor compared to other hospitals. These statistics support the need to implement protocols and procedures to lower readmission rates. Lowering rates will avoid CMS penalties and improve patient outcomes.

Additionally, the analysis shows that none of the underlying medical conditions recorded for patients had any predictive value. This supports the need to look closer at how the diagnostic criteria are used in the inclusion of the associated diagnoses of the patients and what quality controls are in place to assure the accuracy of the past medical history they maintain. It is a known problem that one of the downsides to the Electronic Medical Record (EMR), is that the problem lists are frequently not appropriately updated with each encounter. (Smith, 2017)

There were two data points for the patients that did hold significate predictive value for readmission, "Initial\_days" and "Totalcharge". Unfortunately, this is not very helpful in practice as it does not identify which patients are at risk of readmission on day one of the initial hospitalization to be able to implement protocols to reduce readmission. It does however demonstrate the additional cost of higher LOS and readmitted patients and further supports decision-making by showing the opportunity to reduce costs with reduced LOS and readmission.

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

The CMS data does not include any medical condition data about the patients to compare with the medical conditions in the primary data set. There is no way to determine if there is a disparity in severity between the patients in the two sets that would explain the difference in readmission rates. Additionally, information on confounding variables, such as the reasons for initial and subsequent hospitalizations, is not included in either dataset. (Bruce, 2020)

**References**

Bruce, Peter, et al. (2020). Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python, O'Reilly Media, Incorporated, 2020.

Centers for Medicare and Medicaid Services (CMS) (n.d.).[*Mapping Medicare Disparities by Population - Centers for Medicare & Medicaid Services Data (cms.gov)*](https://data.cms.gov/tools/mapping-medicare-disparities-by-population)

Loth, A. (2019). *Visual analytics with Tableau.* John Wiley & Sons, Incorporated.

Smith, Donna. (2017). *HIMagine That: The problem list is a problem.* [HIMagine That: The problem list is a problem - 3M Inside Angle](https://insideangle.3m.com/his/blog-post/himagine-problem-list-problem/)