Dameia Brewster
WGU D211
Advanced Data Acquisition
Task 1: Data Analysis
January 7, 2023

Part 1: Data Dashboards

1. Datasets (csv files are attached to submission):
    - medical_clean.csv
    - healthcare_dataset.csv - https://www.kaggle.com/code/hainescity/healthcare-dataset-eda/notebook

2. How to install dashboard:

The tableau package workbook is provided with the report and can be installed following the instructions below:

- Open the d211.twbx file/workbook once download is complete.
- When prompted for the linkage to PostgresSQL, input the following information:
    - Server: localhost
    - Port: 5432
    - Database: medical_data
    - Authentication: username and password
    - Username: postgres
    - Password: Passw0rd!
- Sign in and installation should complete.

3. How to navigate Dashboard 1:

Patient Admission by Gender

The data presented on the left and right reveals the percentages of patients by gender that were admitted in both the Kaggle data and WGU data set. We can notice that both data sets have about the same percentage of females and males admitted.

The filters give the option of diving deeper into the analysis to select a specific category that explains the reason for admission.

Patient Admission by Age

The data presented on the left and right identifies each patient by ‘name’ or ‘patient id’ categorized by age. The filters give the option to evaluate an individual’s age easily if given name or patient id by typing into the search bar.

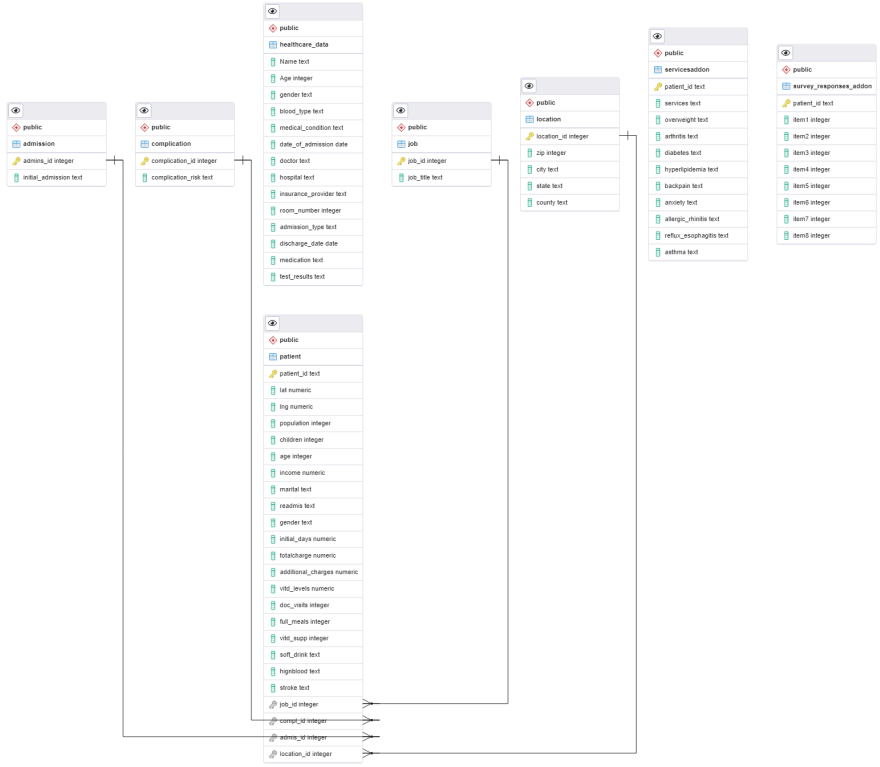
We can observe that both data set’s oldest patients were around 80 years old.

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

-See codes at the bottom of the report.

Part 2: Demonstration

- B3-B7: An ERD was used to help identify the primary and foreign keys:



Link to Panopto:
https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=dc01595a-9db3-4cee-92f5-b1060130a961

Part 3: Report

C1. Explain how the purpose and function of your dashboard aligns with the needs outlined in the data dictionary associated with your chosen data set.

There are many issues with individuals not receiving the proper health care or following routine check up procedures which start at an early age. This could be caused by individuals being over confident and underprepared. The primary focus of this dashboard is age and gender admissions in hospitals.

The main objective of this dashboard is to observe the ages and genders of patients that are admitted into the hospitals to identify which individuals could be at risk. These results can help hospitals correctly treat patients based on their age and gender related health conditions.

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

Tableau is the business intelligence tool selected for this report. This business intelligence tool is very user friendly and integrates with any data source. Its comprehensive features enables anyone to analyze data and create clear visualizations based on that data. Tableau users can visually interact with data to find insights faster, and make critical decisions that better serve customers, improve products, or even save lives (*Excel vs Tableau,* n.d). In addition, its data relationship function allows data sources to retain their level of detail and helps provide a more flexible approach to joining and refining the data.

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

Preparing and cleaning the data are one of the first steps of analysis, but in this instance, the medical data was provided through thePostgreSQL database. An ERD was used to help identify the primary and foreign keys. Also, the external dataset, 'healthcare_dataset' was loaded as an additional table created in the medical_data database.

C4. Summarize the steps used to create the dashboards.

The only tables used from this analysis were patient, admission, and the health care data. The patient and admission were joined in Tableau using the custom SQL code to select the columns that are needed for visualizations. The relationship between both datasets was established by the age and gender variables. The tables used from the analysis created the percentage of Patient Admission by Gender and Patient Admission by Age visualizations. The 'admission type' was also used as a filter to help identify the reasoning behind intial admission.

Interactive controls were also created to allow modifications to help further explore both datasets.

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

The pie charts labeled 'Patient Admission by Gender' shows an identical percentage of patients between genders. This could be due to the total amount of data collected being similar between data sets.

The bar graphs labeled 'Patient Admission by Age', we can observe that between both data set's, the oldest patients were around 80 years old. For the data set on the left's youngest patient was years old and for the new data set on the right, the youngest admitted was. In relation to the health and body changes, there is a huge difference, which could explain the fluctuation in the bar graphs.

For stakeholders, knowing the reasoning behind the numerous age groups and genders admitted, should be taken into account along with identifying if there are any indicators that might show which ages and genders are more admitted and if it could be an age or gender-specific health condition that may have caused admittance.

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

The limitation of the analysis would be the need for more data from the external data set related to the medical data set. The external health care data set provides a closer look into more hospital information on the patient, but it doesn't go into detail about the patient's demographics like the medical data set. This caused the external data to only contain a limited amount of related variables.

Another limitation would be the tableau version on the virtual machine not being updated which may have caused an issue on why the 'type of admission' from both data sets could not be joined due to it being the third join. This was attempted, and once the join was added, the sheets and dashboards went blank. Once this issue was further looked into, it was learned that the version of tableau has some issues when joining multiple tables.

D. Sources

Excel vs Tableau: Transforming bi and analytics for Excel users (no date) *Tableau*. Available at: https://www.tableau.com/solutions/customer/tableau-vs-excel (Accessed: 28 January 2024).

Hainescity (2023) *Healthcare dataset - eda*, *Kaggle*. Available at: https://www.kaggle.com/code/hainescity/healthcare-dataset-eda/notebook (Accessed: 28 January 2024).

YouTube. (2023). *YouTube*. Retrieved December 18, 2023, from https://www.youtube.com/watch?v=yN5EY3wUzh8.

Custom SQL connections. Tableau. (n.d.). https://www.tableau.com/drive/custom-sql

PostgresSQL:

```
BEGIN;


CREATE TABLE public.admission
(
    admins_id integer NOT NULL,
    initial_admission text,
    PRIMARY KEY (admins_id)
);

CREATE TABLE public.complication
(
    complication_id integer NOT NULL,
    complication_risk text,
    PRIMARY KEY (complication_id)
);

CREATE TABLE public.healthcare_data
(
    "Name" text,
    "Age" integer,
    gender text,
    blood_type text,
    medical_condition text,
    date_of_admission date,
    doctor text,
    hospital text,
    insurance_provider text,
    room_number integer,
    admission_type text,
    discharge_date date,
    medication text,
    test_results text
);

CREATE TABLE public.job
```

```
(
    job_id integer NOT NULL,
    job_title text,
    PRIMARY KEY (job_id)
);

CREATE TABLE public.location
(
    location_id integer NOT NULL,
    zip integer,
    city text,
    state text,
    county text,
    PRIMARY KEY (location_id)
);

CREATE TABLE public.patient
(
    patient_id text NOT NULL,
    lat numeric,
    lng numeric,
    population integer,
    children integer,
    age integer,
    income numeric,
    marital text,
    readmis text,
    gender text,
    initial_days numeric,
    totalcharge numeric,
    additional_charges numeric,
    vitd_levels numeric,
    doc_visits integer,
    full_meals integer,
    vitd_supp integer,
    soft_drink text,
    hignblood text,
    stroke text,
    job_id integer,
    compl_id integer,
    admis_id integer,
    location_id integer,
    PRIMARY KEY (patient_id)
);
```

```
CREATE TABLE public.servicesaddon
(
    patient_id text NOT NULL,
    services text,
    overweight text,
    arthritis text,
    diabetes text,
    hyperlipidemia text,
    backpain text,
    anxiety text,
    allergic_rhinitis text,
    reflux_esophagitis text,
    asthma text,
    PRIMARY KEY (patient_id)
);

CREATE TABLE public.survey_responses_addon
(
    patient_id text NOT NULL,
    item1 integer,
    item2 integer,
    item3 integer,
    item4 integer,
    item5 integer,
    item6 integer,
    item7 integer,
    item8 integer,
    PRIMARY KEY (patient_id)
);

ALTER TABLE public.patient
    ADD FOREIGN KEY (admis_id)
    REFERENCES public.admission (admins_id)
    NOT VALID;


ALTER TABLE public.patient
    ADD FOREIGN KEY (compl_id)
    REFERENCES public.complication (complication_id)
    NOT VALID;


ALTER TABLE public.patient
```

```
    ADD FOREIGN KEY (job_id)
    REFERENCES public.job (job_id)
    NOT VALID;


ALTER TABLE public.patient
    ADD FOREIGN KEY (location_id)
    REFERENCES public.location (location_id)
    NOT VALID;

END;
```

Tableau SQL:

```
SELECT "patient"."additional_charges" AS "additional_charges",
    "admission"."admins_id" AS "admins_id",
    "patient"."admis_id" AS "admis_id",
    "patient"."age" AS "age",
    "patient"."children" AS "children",
    "patient"."compl_id" AS "compl_id",
    "patient"."doc_visits" AS "doc_visits",
    "patient"."full_meals" AS "full_meals",
    CAST("patient"."gender" AS TEXT) AS "gender",
    CAST("patient"."hignblood" AS TEXT) AS "hignblood",
    "patient"."income" AS "income",
    CAST("admission"."initial_admission" AS TEXT) AS "initial_admission",
    "patient"."initial_days" AS "initial_days",
    "patient"."job_id" AS "job_id",
    "patient"."lat" AS "lat",
    "patient"."lng" AS "lng",
    "patient"."location_id" AS "location_id",
    CAST("patient"."marital" AS TEXT) AS "marital",
    CAST("patient"."patient_id" AS TEXT) AS "patient_id",
    "patient"."population" AS "population",
    CAST("patient"."readmis" AS TEXT) AS "readmis",
    CAST("patient"."soft_drink" AS TEXT) AS "soft_drink",
    CAST("patient"."stroke" AS TEXT) AS "stroke",
    "patient"."totalcharge" AS "totalcharge",
    "patient"."vitd_levels" AS "vitd_levels",
    "patient"."vitd_supp" AS "vitd_supp"
  FROM "public"."patient" "patient"
    INNER JOIN "public"."admission" "admission" ON ("patient"."admis_id" =
"admission"."admins_id")
) "Custom SQL Query"
```

```
  INNER JOIN "public"."healthcare_data" "healthcare_data" ON ((CAST("Custom SQL
Query"."gender" AS TEXT) = CAST("healthcare_data"."gender" AS TEXT)) AND ("Custom SQL
Query"."age" = "healthcare_data"."Age"))
```