D205: Data Acquisition

Performance Assessment

Western Governor’s University

Mariel Vicente

March 27, 2024

**Performance Assessment**

**A. Research Question**

My research question for this performance assessment is, “How many patients with a known asthma diagnosis also suffer from high blood pressure?” This research question is an example of a business question a medical provider might pose to determine if there is any correlation between two distinct diagnoses. This information could be helpful to the provider when developing an intervention.

**A1. Question Justification**

This research question utilizes patient data from the medical database and data retrieved from an add-on CSV file. The medical database includes which patients reported high blood pressure. The add-on CSV file ‘mservices’ includes diagnosis data for patients with asthma. I will create a table in the medical database and import the data from ‘mservices’ file. Once the table is imported, I will join the tables to determine the count of asthma patients with high blood pressure.

**A2. Identifying Data**

This research question will use data in the ‘patient’ table from the ‘medical\_data’ dataset and the ‘mservices’ add-on CSV file. From the ‘patient’ table, I will be using the ‘patient\_id’ and ‘hignblood’ columns, which both have text data types. This will provide me with the patients that have high blood pressure. From the ‘mservices’ file, I will be using the ‘patient\_id’ and ‘Asthma’ columns, which also have text data types. This will provide me with the patients that have asthma. The ‘patient\_id’ field is the primary key for both tables and will be used for the join.

**B. Entity Relationship Diagram**

A screenshot of a computer

Description automatically generated

**B1. Relationship Discussion**

As shown above, the existing table ‘patient’ and the created table ‘services’ have a 1:1 relationship in the ERD. Both tables share the same primary key, patient\_id. The other columns being used (‘Asthma’ & ‘hignblood’) are uniquely paired to the patient\_id. The results populated in the fields are dependent on the patient, who is identified by the patient\_id.

**B2. Statement for the ERD**

-- Table: public.services

-- DROP TABLE public.services;

CREATE TABLE public.services

(

    patient\_id text COLLATE pg\_catalog."default" NOT NULL,

    "Services" text COLLATE pg\_catalog."default",

    "Overweight" text COLLATE pg\_catalog."default",

    "Arthritis" text COLLATE pg\_catalog."default",

    "Diabetes" text COLLATE pg\_catalog."default",

    "Hyperlidemia" text COLLATE pg\_catalog."default",

    "BackPain" text COLLATE pg\_catalog."default",

    "Anxiety" text COLLATE pg\_catalog."default",

    "Allergic\_rhinitis" text COLLATE pg\_catalog."default",

    "Reflux\_esophagitis" text COLLATE pg\_catalog."default",

    "Asthma" text COLLATE pg\_catalog."default",

PRIMARY KEY (patient\_id),

    CONSTRAINT services\_fkey FOREIGN KEY (patient\_id)

REFERENCES public.patient (patient\_id)

)

TABLESPACE pg\_default;

ALTER TABLE public.services

    OWNER to postgres;

The code above was created by pgadmin. I created a new table in the medical\_data database named ‘services’ using the Create > Table function in pgadmin, which contains all the fields in the ‘mservices’ CSV file.

**B3. Loading CSV Data**

--command " "\\copy public.services (patient\_id, \"Services\", \"Overweight\", \"Arthritis\", \"Diabetes\", \"Hyperlidemia\", \"BackPain\", \"Anxiety\", \"Allergic\_rhinitis\", \"Reflux\_esophagitis\", \"Asthma\") FROM 'C:/LabFiles/Medical/MSERVI~1.CSV' DELIMITER ',' CSV HEADER QUOTE '\"' ESCAPE '''';""

The code above was created by pgadmin. I imported the ‘mservices’ CSV data into the new ‘services’ table I created using the Import/Export function in pgadmin, which then loaded the data into the desired database.

**C. SQL Query**

SELECT COUNT(services."Asthma") AS asth\_HBP

FROM services

LEFT JOIN patient

USING (patient\_id)

WHERE services."Asthma" = 'Yes'

      AND patient.hignblood = 'Yes';

**C1. CSV Files**

A screenshot of a spreadsheet

Description automatically generated

The CSV file containing the results of my SQL query has been submitted as part of the performance assessment. I have also included a screenshot of the results above.

**D. Add-on File Time Period**

The add-on CSV file ‘mservices’ should be acquired and refreshed on a monthly basis for the data to remain relevant to the business activities. As a current employee of a healthcare company, I’ve learned that healthcare companies are interested in seeing trends over time. I believe that updating the count of asthma patients on a monthly basis will allow the healthcare company to assess whether there is a seasonal element to these diagnoses. Monthly is frequent enough to assess changes throughout the year without becoming burdensome to the database schedule.

**D1. Explanation of Time Period**

As a current employee of a healthcare company, I’ve learned that healthcare companies are interested in seeing trends over time. I believe that updating the count of asthma patients on a monthly basis will allow the healthcare company to assess whether there is a seasonal element to these diagnoses. Monthly is frequent enough to assess changes throughout the year without becoming burdensome to the database schedule.

**E. Panopto Video**

The Panopto video recording has been submitted as part of the performance assessment.

**F. Web Sources**

This performance assessment was made possible using information gained from the DataCamp SQL modules and [www.lucidchart.com](http://www.lucidchart.com) website to create an ERD diagram.