WGU

D205

Data Acquisition

Amber Peacock

10/20/2023

1. **Research Question.** For this performance assessment I have decided to explore the following as my research question: What is the top internet service chosen by customers between 25 and 30 years of age? So, for this project, I will be exploring what internet service is most frequently chosen by customers of that age group.

**A1. Identifying Data.** In this scenario, I will need to get information from the services table to find what internet services (column) the customers are choosing. I will also need to use the age column from the customer table. In conjunction, I can use both tables and columns specified to gather the information I need to answer my research question alongside using specific filters in my SELECT statement to obtain information for the desired age group.

1. **Entity Relationship Diagram**. In the picture below, you will see the entity relationship diagram I created in PgAdmin4. I created this by right clicking on the churn database and clicking “generate ERD”. (Intro To PostgreSQL Databases With PgAdmin For Beginners - Full Course, 2019)

A screenshot of a computer

Description automatically generated

**B1. Code for the ERD.** The code I wrote for the add on table “services” can be found below. I created this table by adding in the column names as they are labeled in the add on csv file with the data type. I made sure to include customer\_id as the primary key to show that those values are unique and in the churn database. This will be useful later when importing the csv file and joining the newly created table with the customer table.

**CREATE TABLE** services(

customer\_id **VARCHAR (50) PRIMARY KEY,**

internetservice **VARCHAR (50),**

phone **VARCHAR (20),**

multiple **VARCHAR (10),**

onlinesecurity **VARCHAR (10),**

onlinebackup **VARCHAR (10),**

deviceprotection **VARCHAR (10),**

techsupport **VARCHAR (10)**

);

**B2. Loading the CSV data.**  Below is the SQL code, in text format, that loads the data from one of the add-on CSV files into the table created in part B1. I was able to obtain this code sequence by creating the table seen above in B1 and clicking “more details” on the pop up when the table was successfully created. (Import CSV File Into PostgreSQL Table, 2011)

\\copy public.services (customer\_id, internetservice, phone, multiple, onlinesecurity, onlinebackup, deviceprotection, techsupport) FROM 'C:/LabFiles/Services.csv' DELIMITER ',' CSV HEADER QUOTE '\"' ESCAPE '''';""

1. **SQL query.** To answer the research question that I created in part A, I needed to write a SQL query that would join the tables (customer and services), specified a certain age group, and would show me a table with the count of customers for each internet service type. Below is the query I used to get the answer to my question. As a result of running the query, I found out that the most used internet service for customers between 25 and 30 years of age was fiber optic. Below, a snapshot can be seen of the query results. (SQL for Joining Data, 2017)

SELECT internetservice, COUNT(\*) AS service\_count

FROM services

INNER JOIN customer ON services.customer\_id=customer.customer\_id

WHERE customer.age >=25 AND customer.age <=30

GROUP BY services.internetservice

ORDER BY service\_count DESC;

A screenshot of a computer

Description automatically generated

**C1. CSV files.** Attached to my submission will be an additional CSV file labeled “Part C”. This file captures the results from the query in part C. I got this by saving the results to file from PgAdmin4 to the Labs on Demand system then sending it to myself and saving to my personal computer.

**D. Add on file.** The add-on file should be acquired and refreshed in the database for the data to remain relevant to the business and the question from part A with the addition of each new customer and the query should be ran each quarter of the year. This way as new customers come in, information can be updated to keep up with which services are more popular among that age group.

**D1. Explanation of time period.** The add-on file should be updated as customers are added in so that information is constantly up to date and information is current. The query should be ran quarterly so that the answer to the question can evolve along with the types of services and needs of the customers.

**E & E1. Panopto video of code and programs.** Attached in my submission is my Panopto video showing my code and all programs used to complete this project. <https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=bf398947-a98c-4b1b-9fa9-b0aa01209737>

**F&G. Web Sources.**

# Works Cited

*Import CSV File Into PostgreSQL Table*. (2011). Retrieved from PostgreSQL Tutorial: https://www.postgresqltutorial.com/postgresql-tutorial/import-csv-file-into-posgresql-table/

*Intro To PostgreSQL Databases With PgAdmin For Beginners - Full Course*. (2019, May 27). Retrieved from YouTube: https://www.youtube.com/watch?v=Dd2ej-QKrWY

*SQL for Joining Data*. (2017). Retrieved from DataCamp: https://app.datacamp.com/learn/courses/sql-for-joining-data