*DATA AQUISITION*

*D 205*

*Student ID: 012247127*

*Natallia Zimnitskaya*

***Section A.***

The research question covered in this performance assessment is: “What is the number of male and female customers using the device protection service?”

***1.***

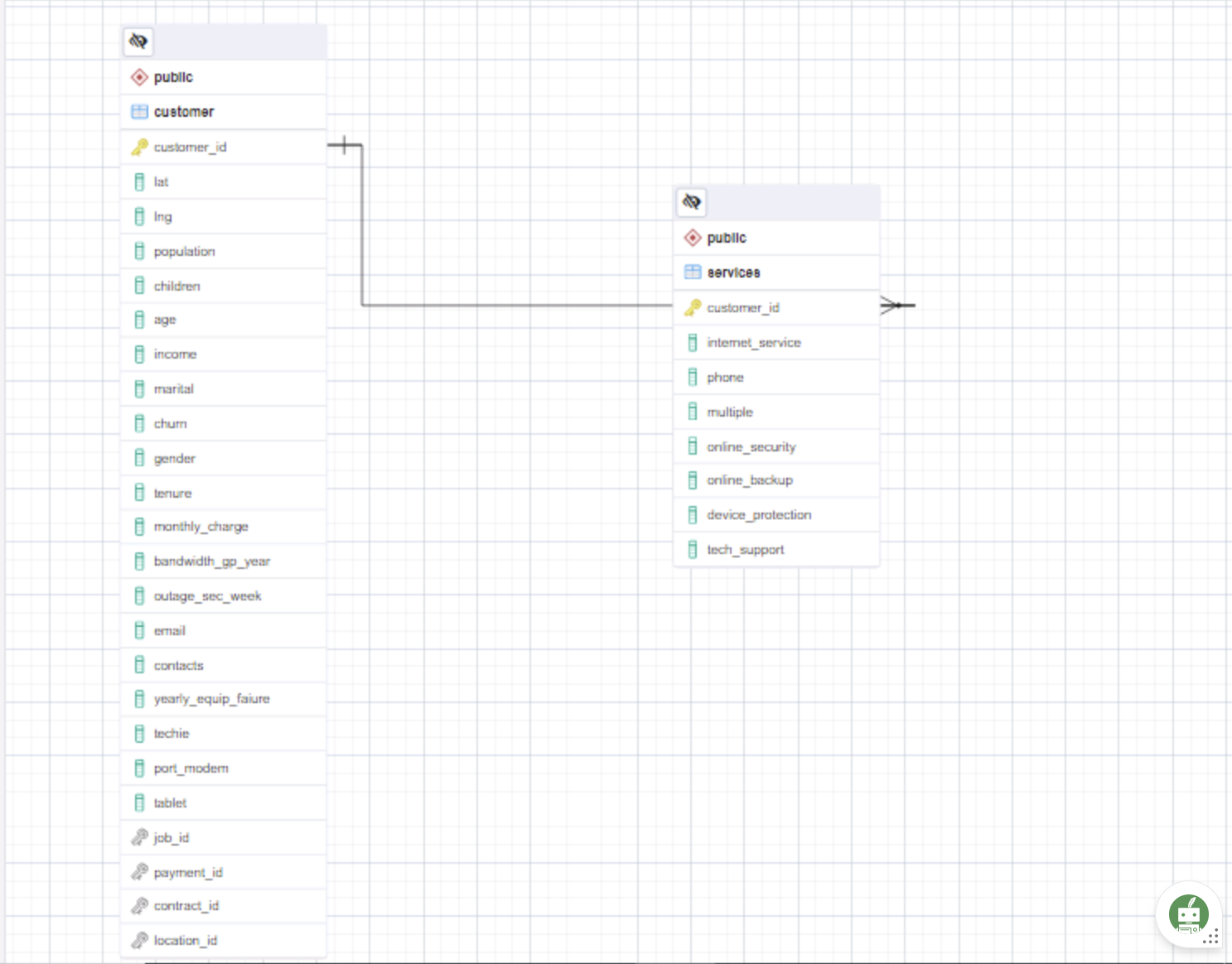
The main goal of this analysis is to determine if there is a relationship between customer gender and having a device\_protection service. Data from the original “customer” table and the Services.csv add-on file will be used to answer the question.

***2.***

The data from the customer table (customer\_id and gender columns) will be utilized. The customer\_id is the primary key to uniquely identify each row in the table, where the gender field provides information about whether the customer using the device protection service is male or female. The Services.csv file contains the data about the types of services provided and the number of customers for each type.

***Section B.***

An Entity Relationship Diagram is created using the Generate ERD option on pgAdmin4.



***1.***

The customer table represents customer data, while the services table represents service-related data. The relationship between customer and services is based on the common customer\_id. The customer\_id column in thecustomer table is the primary key. It contains unique values representing individual customers. In the services table, the customer\_id column is a foreign key that links to the customer table. This means that each row in the services table corresponds to a customer from the customer table. Each service is related to only one customer; however, each can be associated with multiple services (records in the services table). Thus, these two tables are a 1:M (one-to-many) relationship.

***2.***

SQL statement to create a new table “services.”

CREATE TABLE services(

customer\_id VARCHAR(20) PRIMARY KEY,

internet\_service VARCHAR(20),

phone VARCHAR(3),

multiple VARCHAR(3),

online\_security VARCHAR(3),

online\_backup VARCHAR(3),

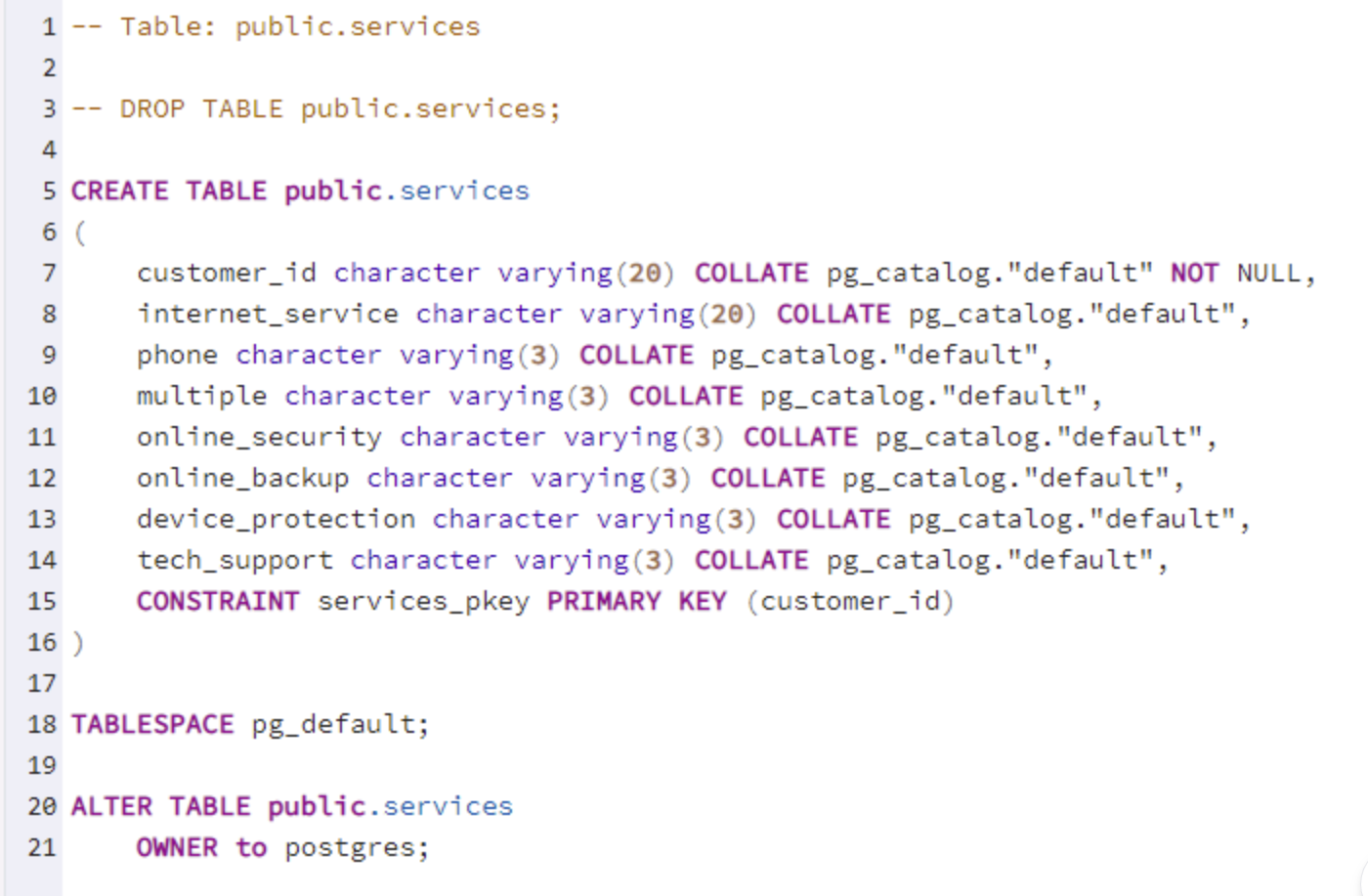
device\_protection VARCHAR(3),

tech\_support VARCHAR(3)

);

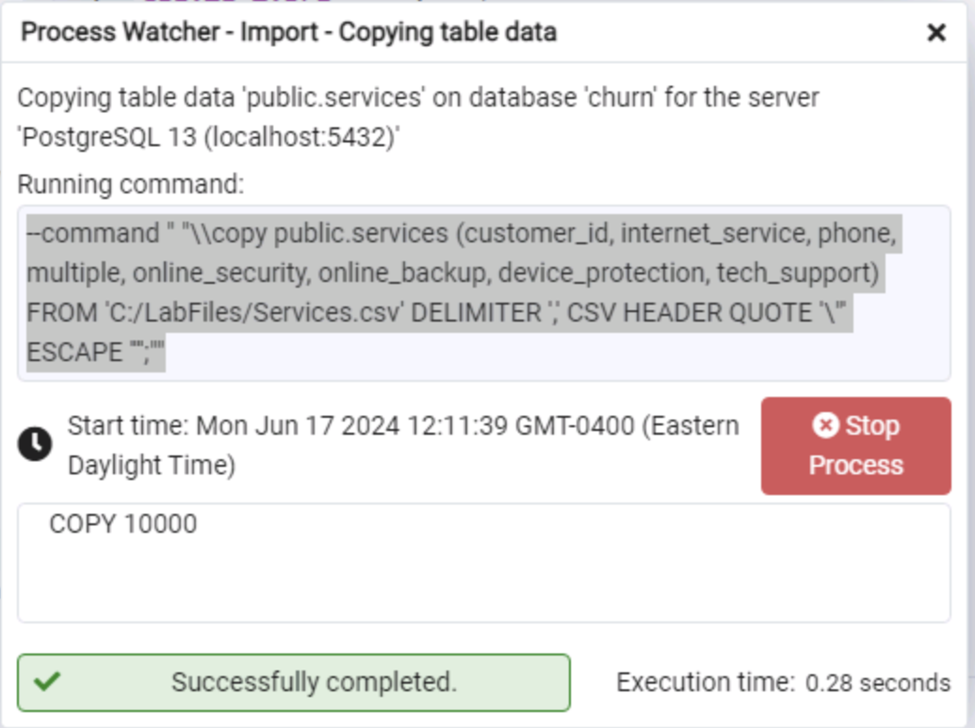
It contains eight columns according to the number of columns in the Services.csv file. The customer\_id is the primary key, and the seven other columns are the type of services provided.

The script below was generated from pgAdmin4.



***3.***

Right-click the customer table and choose Import/Export from the menu. Next, toggle to import, browse the services file, choose CSV format, and choose comma (,) as the delimiter. Next, click the columns tab, deselect the id column, and click OK. Lastly, wait for the import process to finish. The dialog box displaying the import status is below (PostgreSQL Tutorial, 2024c).



--command " "\\copy public.services (customer\_id, internet\_service, phone, multiple, online\_security, online\_backup, device\_protection, tech\_support) FROM 'C:/LabFiles/Services.csv' DELIMITER ',' CSV HEADER QUOTE '\"' ESCAPE '''';""

***Section C***

To answer the research question, “What is the number of male and female customers using the device protection service?” the SQL code was created.

SELECT device\_protection,

COUNT (CASE WHEN gender = 'Male' THEN 1 END) AS male\_customer,

COUNT(CASE WHEN gender = 'Female' THEN 1 END) AS female\_customer,

FROM customer AS c

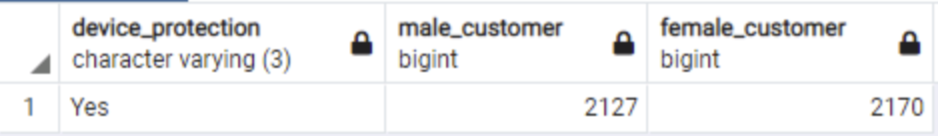
JOIN services AS s

ON c.customer\_id = s.customer\_id

WHERE device\_protection = ‘Yes’

GROUP BY device\_protection;

The final output of the data shows minimal variation between male (2127) and female (2170) customers using device protection services.



***1.***

The CSV file is submitted separately.

***Section D.***

This research question is intended for marketing campaigns for current and potential clients. In this sense, generalized data about trends rather than specific and in-depth information about individual customers is needed for new customer marketing campaigns. However, we don't want our marketing efforts to annoy current clients.

Obtaining and updating the add-on file is necessary for every new billing cycle. This ought to occur every month. It should be updated with every billing cycle to include new clients, clients who have changed their services, and clients who have wholly canceled them. By refreshing the database at this interval, the information will remain current and helpful in addressing the following queries: "How many male and female customers use device protection services?"

***Section E.***

The Panopto video is submitted separately.

***References***

*D205 - Data acquisition. (n.d.). Datacamp. Retrieved June 17, 2024, from* [*https://app.datacamp.com/learn/custom-tracks/custom-d205-data-acquisition*](https://app.datacamp.com/learn/custom-tracks/custom-d205-data-acquisition)

*SQL Sunday D205 April 2023.pptx. (n.d.-b).* [*https://westerngovernorsuniversity-my.sharepoint.com/:p:/g/personal/william\_sewell\_wgu\_edu/EZvWh7S\_6K9HnjS1yDoRXhQBBaL6oZZqrUM6TgsDBbbimQ?e=pdti52*](https://westerngovernorsuniversity-my.sharepoint.com/:p:/g/personal/william_sewell_wgu_edu/EZvWh7S_6K9HnjS1yDoRXhQBBaL6oZZqrUM6TgsDBbbimQ?e=pdti52)

PostgreSQL Tutorial. (2024, February 1). Import CSV file into PosgreSQL table. <https://www.postgresqltutorial.com/postgresql-tutorial/import-csv-file-into-posgresql-table/>

Page, D. (n.d.-b). ERD Tool — pgAdmin 4 8.8 documentation. <https://www.pgadmin.org/docs/pgadmin4/8.8/erd_tool.html>