**A:QUESTION**

The main question I asked in this project is the following: “How many male customers in each state are without internet subscriptions??”

**A1:QUESTION JUSTIFICATION**

I decided to ask this specific question since it can be relevant in identifying how many males per state are without internet service. It can be subsequently used to structure an advertising campaign encouraging males in those locations without internet subscription to sign up for an internet service and also assist the company in identifying where to establish fiber optic infrastructure.

## A2:IDENTIFYING DATA

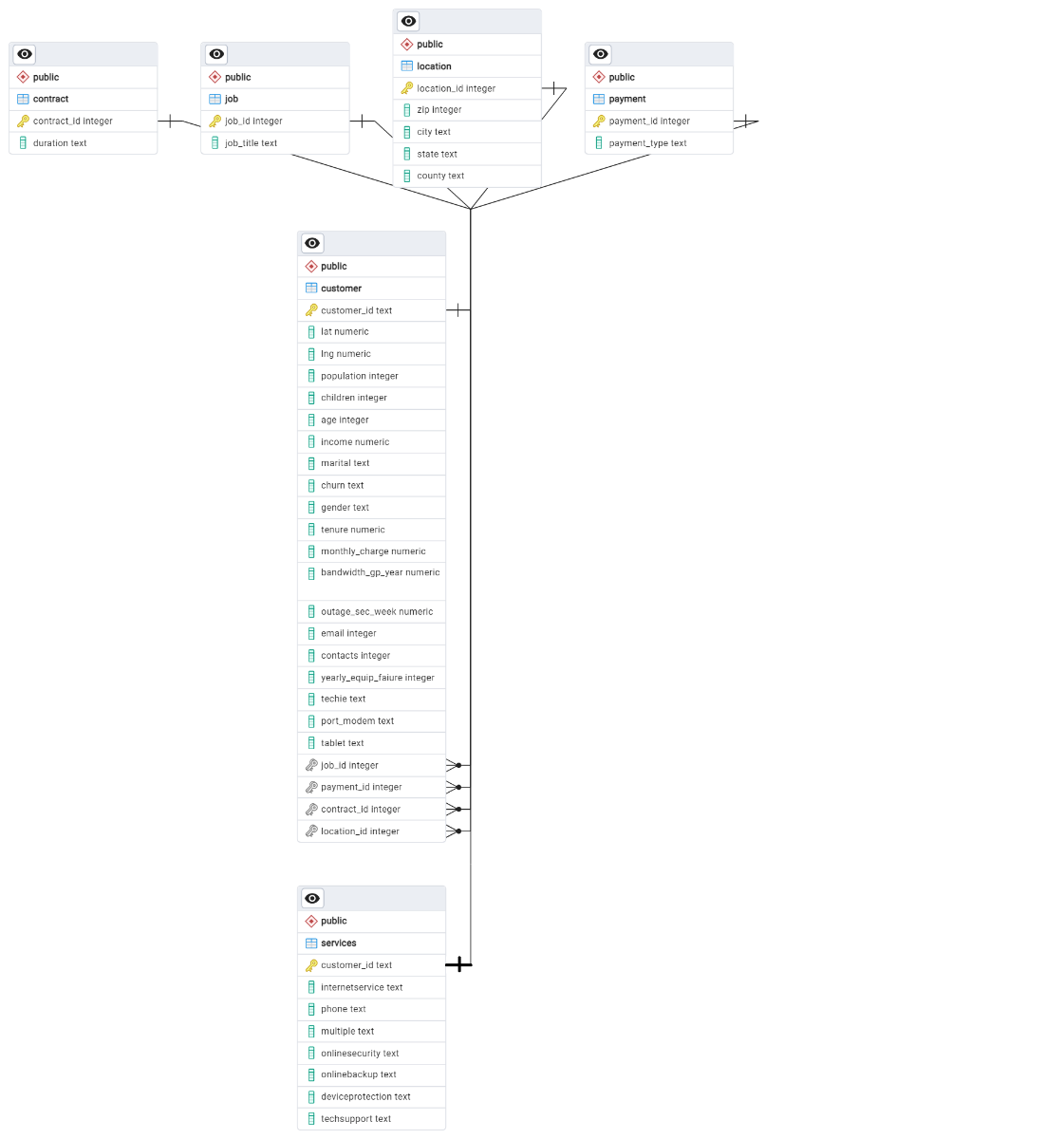
I required the usage of the customers and location tables that were provided in the original database as well as the additional services table. The customer\_id column from the customer table and the customer\_id column from the services table were required to perform the necessary joins for proper data retrieval.

Similarly, the location\_id columns in both the customer and location tables were also required for the same reason. The customer\_id from the customer table was also required in the usage of a count function that was ultimately aliased as the “males\_without\_internet\_service” column in the final table. The “gender” column in the customer table and the “internetservice” column in the services table were required in the usage of a “WHERE” statement to properly filter for the necessary data.

The data types for all of the aforementioned columns are as follows: customer.customer\_id (text), services.customer\_id (text), customer.location\_id (integer), location.location\_id (integer), customer.gender (text), services.internetservice (text), males\_without\_internetservice (bigint).

The final required columns were the “state (text)” and “males\_without\_internet\_service (bigint)” columns that were generated by the query required to answer the research question. These results were saved in a .csv file named “number\_of\_males\_without\_internet\_service\_per\_state.csv”.

***B:ENTITY RELATIONSHIP DIAGRAM***



**B1:RELATIONSHIP DISCUSSION**

The relationship between the services table and the customer table is 1:1 since each customer\_id in the services table only appears once in the customer table and vice versa. The customer\_id column in the services table is the primary key of that table while the customer\_id column in the customer table is the foreign key of that specific table.

I also used the location table, which has a 1:M relationship with the customer table. This is because each location id (location\_id column entries) in the location table appears once while location\_id column entries in the customer table can show the same location multiple times due to the same states/locations having multiple customers.

**B2:STATEMENT FOR THE ERD**

CREATE TABLE public.services

(

customer\_id text NOT NULL,

InternetService text NOT NULL,

Phone text NOT NULL,

Multiple text NOT NULL,

OnlineSecurity text NOT NULL,

OnlineBackup text NOT NULL,

DeviceProtection text NOT NULL,

TechSupport text NOT NULL,

PRIMARY KEY (customer\_id),

CONSTRAINT customer\_id\_fkey FOREIGN KEY (customer\_id)

REFERENCES public.customer (customer\_id) MATCH SIMPLE

ON UPDATE NO ACTION

ON DELETE NO ACTION

NOT VALID

);

I also added the following SQL query to properly keep and ensure the services table ownership to the public schema. Subsequently, I created a simple query to check for successful importation of the data from the services.csv file into the services table.

ALTER TABLE public.services

OWNER to postgres;

SELECT \*

FROM services;

**B3:LOADING CSV DATA**

COPY Services

FROM ‘C:\LabFiles\Services.csv’

WITH

(FORMAT CSV,

HEADER,

DELIMITER ‘,’);

**C:SQL QUERY**

SELECT l.state,

COUNT (c.customer\_id) AS males\_without\_internet\_service

FROM customer c

INNER JOIN services s

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

INNER JOIN location l

ON c.location\_id = l.location\_id

WHERE gender = 'Male' AND internetservice = 'None'

GROUP BY l.state

ORDER BY l.state, males\_without\_internet\_service DESC;

**C1:CSV FILES**

The final table that shows the number of male subscribers without internet subscriptions per state will be contained within the attached “number\_of\_males\_without\_internet\_service\_per\_state.csv” file in my submission.

**D & D1: ADD-ON FILE TIME PERIOD AND EXPLANATION OF TIME PERIOD**

I propose that a refresh add-on file be acquired and loaded in a monthly interval since most significant changes in customer service internet subscription counts tend to occur in that general time frame and services billing statements tend to occur in a monthly fashion.

Additionally, to keep the integrity of the primary key and foreign key relationships between the customer and services table, respectively, new customer IDs should be added to the customer table before adding that same customer ID to the services table if they become a subscriber to either DSL or Fiber Optic internet services.

**E:PANOPTO VIDEO**

My Panopto submission is viewable at the following link:

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=1f918fea-58b3-403e-afd6-b1a200461685>

**F:WEB SOURCES**

*Backup dialog¶*. Backup Dialog - pgAdmin 4 8.8 documentation. (n.d.). https://www.pgadmin.org/docs/pgadmin4/8.8/backup\_dialog.html

*Erd tool¶*. ERD Tool - pgAdmin 4 8.8 documentation. (n.d.). https://www.pgadmin.org/docs/pgadmin4/8.8/erd\_tool.html