DA 205- Data Acquisition Performance Assessment

Isiasha Gordon

Research Question

Are customers less than forty years old more likely to have multiple telecom services and how satisfied are they with the number of options they have?

Identifying Data

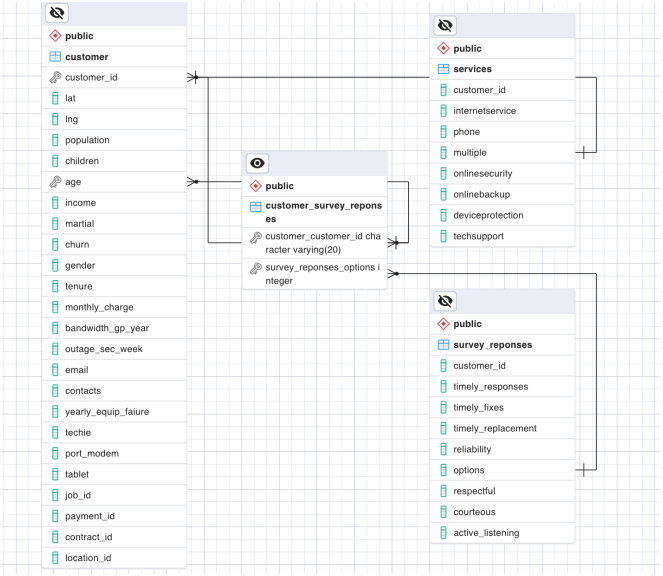
This analysis utilized the multiple service data from the services table and the number of options from the survey responses table by joining the customer id on the customer table. A filter was applied to customer IDs to filter customers aged 40 years old or less.

Tables used to complete this analysis:

- Customer (original database)
- Services (add-on table)
- Survey_responses (add-on table)

Logical Data Model

TABLE	RELEVANT KEYS
Customer	Customer_id
Customer	Age
Survey_response	Customer_id
Suvery_response	Options
Services	Customer_id
Services	Multiple



Code for The Physical Data Model

services table :

```
CREATE TABLE services(
        customer_id varchar(20),
        InternetService varchar(20),
        phone varchar(3),
        Multiple varchar(3),
        OnlineSecurity varchar(3),
        OnlineBackup varchar(3),
        DeviceProtection varchar(3),
        TechSupport varchar(3));
```

survey_responses:

```
CREATE TABLE survey_reponses (
        customer_id varchar(20),
        Timely_Responses int,
        Timely_Fixes int,
        Timely_Replacement int,
        Reliability int,
        options int,
        Respectful int,
        Courteous int,
        Active_Listening int);
```

Loading CSV Data

services table :

--command"\\copy public.services (customer_id, internetservice, phone, multiple, onlinesecurity, onlinebackup, deviceprotection, techsupport) FROM '/Users/igmark/Desktop/WGU Data Files/Services.csv' DELIMITER ',' CSV HEADER QUOTE '\"' ESCAPE '";""

survey_responses:

"--command\\copy public.survey_reponses (customer_id, timely_responses, timely_fixes, timely_replacement, reliability, options, respectful, courteous, active_listening) FROM '/Users/igmark/Desktop/WGU Data Files/Survey_Responses.csv' DELIMITER ',' CSV HEADER QUOTE '\"' ESCAPE '";""

SQL Query

The SQL code below was used to join survey_responses and services on the customer table while filtering for ages less than 40 years old with multiple responses.

```
SELECT c.customer_id, c.age, se.multiple, su.options
FROM customer AS c
LEFT JOIN services AS se
on c.customer_id = se.customer_id
LEFT JOIN survey_reponses AS su
on c.customer_id = su.customer_id
WHERE c.age < 40
AND se.multiple = 'Yes';
```

The query to count the number of customers with multiple services and ages less than 40 is below:

```
SELECT *
FROM
(SELECT COUNT(c.customer_id)
        FROM customer AS c
        LEFT JOIN services AS se
        on c.customer_id = se.customer_id
        LEFT JOIN survey_reponses AS su
        on c.customer_id = su.customer_id
        WHERE c.age < 40
        AND se.multiple = 'Yes')
        AS count1,
(SELECT COUNT(c.customer_id)
        FROM customer AS c
        LEFT JOIN services AS se
        on c.customer_id = se.customer_id
        LEFT JOIN survey_reponses AS su
        on c.customer_id = su.customer_id
        WHERE c.age < 40
        AND se.multiple = 'No')

        AS count2 ;
```

Conclusion

In conclusion, there was no significant relationship between age and receiving multiple services. This conclusion was drawn utilizing the COUNT function on the joined table. There were 1,439 customers less than forty years old with multiple services and 1,680 customers less than forty years old without multiple services.

References

(2022). *Import CSV File Into PostgreSQL Table*. PostgreSQL Tutorial. https://www.postgresqltutorial.com/postgresql-tutorial/import-csv-file-into-posgresql-table/

The pgAdmin Development Team (2023, January 16). *ERD Tool*. PgAdmin.org. https://www.pgadmin.org/docs/pgadmin4/6.19/erd_tool.html

Trocinski, J. (2020, May 19). *Using ON Versus WHERE Clauses to Combine and Filter Data in PostgreSQL Joins*. Pluralsight. https://www.pluralsight.com/guides/using-on-versus-where-clauses-to-combine-and-filter-data-in-postgresql-joins