

DA 205- Data Acquisition Performance Assessment

Western Governors University

Table of Contents

| | |
|---|-----------|
| <i>Section A. Research Question</i> | 3 |
| Section A1. Identifying Data | 3 |
| <i>Section B: Logical Data Model</i> | 4 |
| Section B1: Code for The Physical Data Model | 5 |
| Section B2: Loading CSV Data | 6 |
| <i>Section C: SQL Query</i> | 6 |
| Section C1: CSV File | 7 |
| <i>Section D: Add-On File</i> | 7 |
| <i>Section E. SQL Script</i> | 8 |
| <i>Section F: Panopto Video</i> | 8 |
| <i>Section G: Web Sources</i> | 9 |
| <i>Section H: Sources</i> | 9 |
| References: | 9 |
| <i>Section I. Professional Communication</i> | 10 |

Section A. Research Question

This paper, PostgreSQL will be used to answer the 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?

Section A1. Identifying Data

The main goal of this analysis is to determine if there is a relationship between age and having multiple services and the significance of having multiple services available.

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.

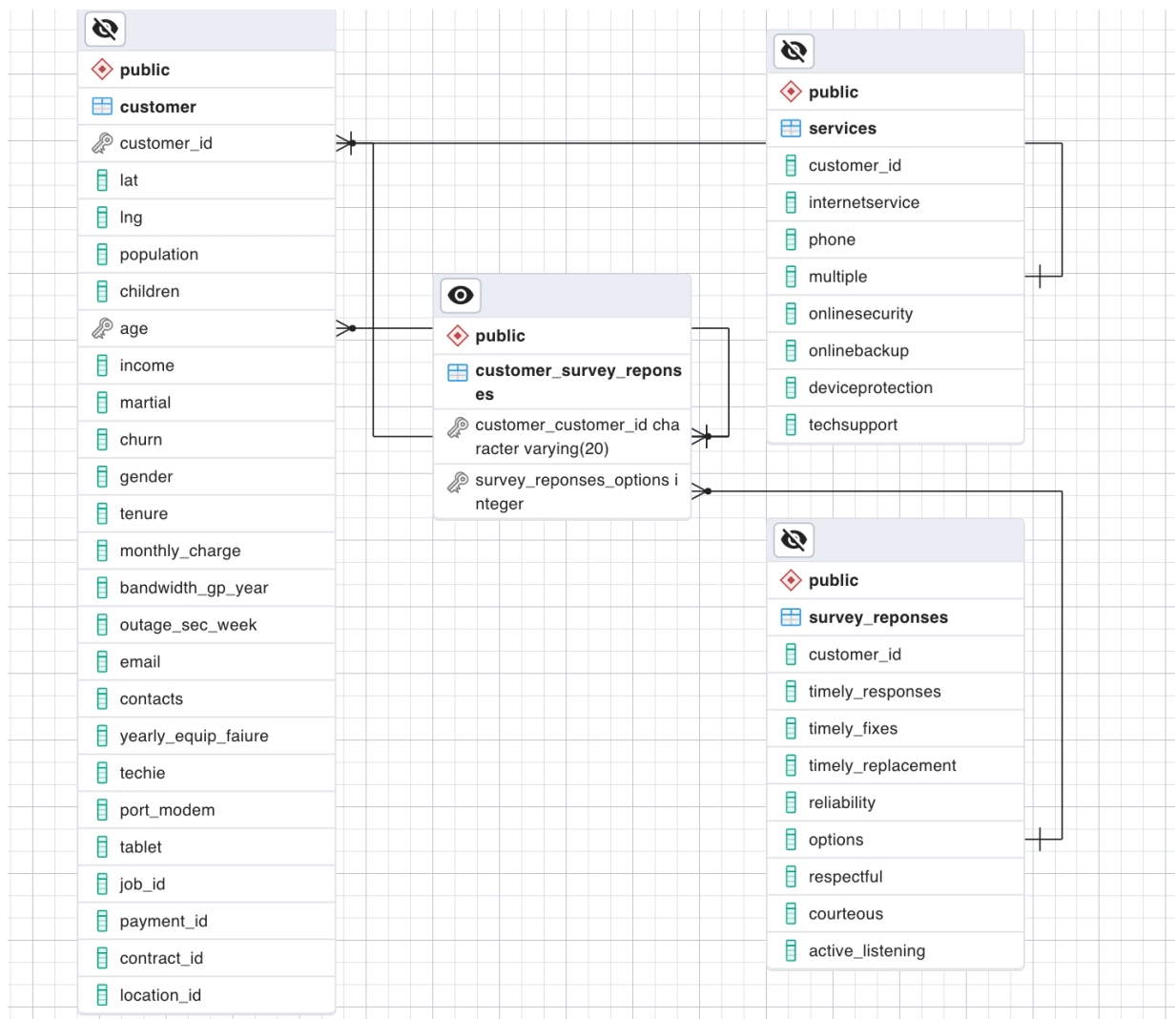
Tables used to complete this analysis:

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

Section B: Logical Data Model

The logical data model below illustrates the relational constraints of the add-on CSV files (services and survey_responses). This model illustrates the following relationships:

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



The above ERD was created using the pdAdmin guide (2023).

Section B1: Code for The Physical Data Model

The following SQL code was used to create tables for the add-on tables services and survey_responses:

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);
```

Section B2: Loading CSV Data

The following commands were executed to create the SQL tables above.

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
'\"';"
```

Section C: 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 ;
```

Section C1: CSV File

The CSV files for the table joins and count that capture the results from queries will be attached. The file is named customer_server_satisfaction.csv.

Section D: Add-On File

Current data does not support the relationship between age and having multiple services. To increase likelihood of a positive correlation increased and expanded data is needed. More

data on customer satisfaction and the need for multiple services could provide insight into relevant business needs and goals. This data should be collected, analyzed, and refreshed in the database quarterly to obtain the most up-to-date business feedback for growth and development.

Section E. SQL Script

The following scripts/commands were used to perform the process of loading the add-on data to the original database:

services table:

```
--command"\copy public.services (customer_id, internet-service, phone, multiple, online-security, online-backup, device-protection, tech-support) FROM '/Users/igmark/Desktop/WGU Data Files/Services.csv' DELIMITER ',' CSV HEADER QUOTE '\" ESCAPE '\";
```

survey responses:

```
--command"\copy public.survey_responses (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 '\";
```

Section F: Panopto Video

Click the following for the Panopto video describing the data and data collection process:

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=88f12fff-72ea-4570-bd48-af920161249e>

Section G: Web Sources

No web sources were used to acquire data or segments of third-party code to support the application.

Section H: Sources

Three CSV files were used to create the SQL queries listed above. The CSV add-on files include services and survey_responses. The original database CSV file was the customer file. These files were imported into PostgreSQL. PostgreSQL tutorial provided steps for uploading CSV files to Postgres (2022). The LEFT JOIN feature was utilized to join the add-on files to the original CSV table and a filter for customers aged forty years old or less was implemented using the WHERE and ON clauses (Triocinski, 2020).

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>

Section I. Professional Communication

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.