DA 205- Data Acquisition Performance Assessment

Isiasha Gordon

Western Governs University

**Table of Contents**

[Section A. Research Question 3](#_Toc125195928)

[Section A1. Identifying Data 3](#_Toc125195929)

[Section B: Logical Data Model 3](#_Toc125195930)

[Section B1: Code for The Physical Data Model 5](#_Toc125195931)

[Section B2: Loading CSV Data 7](#_Toc125195932)

[Section C: SQL Query 8](#_Toc125195933)

[Section C1: CSV File 10](#_Toc125195934)

[Section D: Add-On File 10](#_Toc125195935)

[Section E.  SQL Script 11](#_Toc125195936)

[Section F: Panopto Video 12](#_Toc125195937)

[Section G: Web Sources 12](#_Toc125195938)

[Section H: Sources 12](#_Toc125195939)

[References: 13](#_Toc125195940)

[Section I.  Professional Communication 13](#_Toc125195941)

# Section A. Research Question

This paper, PostgresSQL 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 |

A picture containing chart

Description automatically generated

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\_seriver\_satisfaction.csv.

# Section D: Add-On File

Current data does not support the relationship between age and having multiple services. To increase likely hood 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, 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 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**](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 PostgresSQL. PostgresSQL tutorial provided steps for uploading CSV files to Postgress (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.