Performance Assessment for
D205 Data Acquisition

Drew Mendez
MSDA Western Governors University
D205: Data Acquisition
Dr. William Sewell
March 2, 2024

A. Research Question

The research question for this assignment is:

> "Were there more clients that churned who were enrolled in the online backup service than clients that churned and were not enrolled in the online backup service?"

A1. Identifying the Data

Two fields with the text datatype are required to answer this question: the 'churn' field from the customer table, and the 'online_backup' field from the services table.

B. The Entity Relationship Diagram

Below is the ERD for the tables used to answer this research question. In this diagram, customer_id is a primary key for both the customer and the services tables. The customer table has customer_id as a foreign key, which references the primary key of the services table to enforce referential integrity (Grossenbacher et al., n.d.). There is a one-to-one relationship between the services table and the customer table via the customer_id field. The ERD tool is in beta and thus the forks cannot be removed in the diagram below.

B1. SQL Code to Create Add-On Table

Below is the SQL code used to generate the services table, with primary key ‘customer_id’ defined inline.

```
CREATE TABLE services (
        customer_id TEXT PRIMARY KEY,
        internet_service TEXT,
        phone TEXT,
        multiple TEXT,
        online_security TEXT,
        online_backup TEXT,
        device_protection TEXT,
        tech_support TEXT);
```

B2. SQL Code to Load Data from the Add-On CSV

Below is the SQL code that loads data from the add-on file, Services.csv.

```
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 ‘”’;””
```

C. The SQL Statement to answer the Research Question

Below is the SQL code used to generate the contingency table that informed the research question. The SQL query serves to join the customer database with the services database on the customer_id key with a one-to-one relationship. The query returns the count of both outcomes of churn, either yes or no, utilizing a CASE WHEN to filter for each. Finally, the query groups by online_backup, to complete the contingency table, with the ORDER BY statement serving to reorient the table such that (1, 1) = (yes, yes).

```
SELECT online_backup,
        COUNT(CASE WHEN c.churn = 'Yes'
                THEN 1 END) AS churn_yes,
        COUNT(CASE WHEN c.churn = 'No'
                THEN 1 END) AS churn_no
FROM customer AS c
LEFT JOIN services AS s
   ON c.customer_id = s.customer_id
GROUP BY online_backup
ORDER BY online_backup DESC;
```

C1. Data File with Results of this Query

The results of this query can be found in the submission entitled D205_PA_Mendez_Drew.csv.

D. Refresh Time Period

The data from the add-on file should be refreshed in the database monthly to remain relevant.

D1. Time Period Relevance

According to the churn data dictionary, churn is defined to be whether the customer discontinued service within the last month. This implies that the data must be updated at least once per month to stay relevant to the business needs.

E. Panopto Recording

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=686c80ca-32d0-4bcd-97d9-b12701759fa8

E1. Vocalized Demonstration

A vocalized demonstration describing the programs used to complete this task is provided in the Panopto video linked above.

F. Acknowledgement of Web Sources

No web sources were used to acquire data, nor were segments of third-party code used.

G. Acknowledgement of Sources

Grossenbacher, T., Weber, I., Ludolf, J., et al (n.d.). *Introduction to Relational Databases in SQL* [MOOC]. DataCamp. https://app.datacamp.com/learn/courses/introduction-to-relational-databases-in-sql

H. Demonstration of Professional Communication