

PERFORMANCE ASSESSMENT

Data Acquisition - D205

Western Governors University, College of IT

Natalie Toler
June 23, 2023

A. Research Question

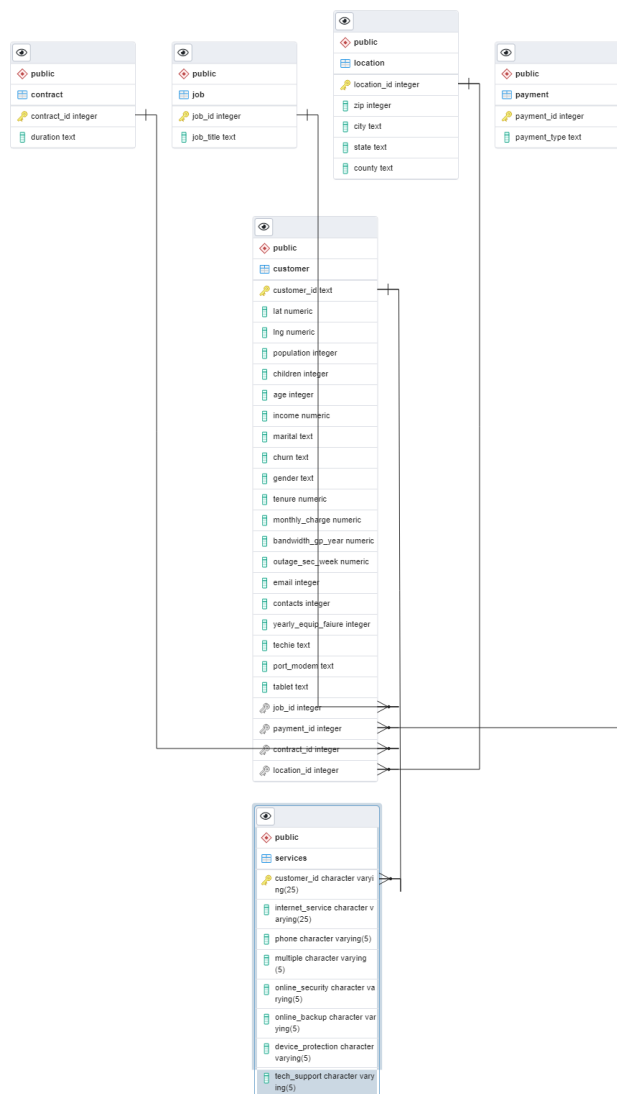
Telecommunications companies are interested in keeping new customers and lowering their rate of churn and raising customer retention. To best advise a company it is important to look at the existing variables that affect customer turnover. As such I will be looking into one variable to see if there is a correlation with customer churn. I will be looking at customers that have opted for the Technical Support service along with their other services to see if access to Tech Support correlates to higher retention of services. In other words, *does having access to tech support from the company affect customer retention?*

1. Identifying Data

In the provided customer table, I will be using the “customer_id” column as the common key between datasets and the column for “churn” which is a yes/no binary, with yes meaning that the customer has discontinued service within the last month. In the supplementary dataset titled Services, I will use the “customer_id” to link the tables and the column for “TechSupport” which is also a yes/no binary, where yes indicates that the customer has the tech support service option.

To effectively answer the SQL question: does having access to tech support from the company affect customer retention? I will create a query that joins the two tables and counts the number of customers in four columns: *does the customer have tech support and did they stay, does the customer have tech support and did they leave, does the customer not have tech support and did they stay, and finally does the customer not have tech support and did they leave.*

B. Logical Data Model



have any nulls.

Using the additional CSV file “services.csv” I will create a data table. I will only be using two columns from the additional CSV file; however the other columns will be useful for other SQL queries into customer churn rates and therefore I will create the table with all the provided columns. Creating a table titled “services” I will use the **customer_id** as the primary key since customer ids are unique to each customer. Since the **customer_id** column is also shared with the other tables it will also serve as the foreign key. The **customer_id** column should be the same throughout the tables as a customer would need an id to have their data stored and services available.

In the existing tables the **customer_id** column is a “text” column which is fine, however I am going to create my columns as “varchar” for the sake of practicing. Since the columns are filled with “yes” and “no” options, as well as the type of internet service or “none” I will import the data with a “not null” constraint since this data shouldn’t

1. Code for the physical data model

```
Query Editor  Query History
1  -- create table for services.csv data
2  CREATE TABLE services(
3      customer_id varchar(10) NOT NULL,
4      internet_service varchar(20) NOT NULL,
5      phone varchar(5) NOT NULL,
6      multiple varchar(5) NOT NULL,
7      online_security varchar(5) NOT NULL,
8      online_backup varchar(5) NOT NULL,
9      device_protection varchar(5) NOT NULL,
10     tech_support varchar(5) NOT NULL,
11     PRIMARY KEY (customer_id), -- Use customer_id as primary key
12     CONSTRAINT customer_id_fkey FOREIGN KEY (customer_id)
13         REFERENCES customer (customer_id)
14         -- Use customer_id as foreign key connected to customer table
15 );
```

As stated above I used the varchar column type to set up my columns and set the constraints as not null. I double checked the format of the constraints as well as my set up of customer_id with the postgresSQL documentation for constraints (The PostgreSQL Global Development Group, 5.4. *constraints* 2023). The primary key and foreign key are both the customer_id column and I made sure that they customer_id column references the customer table.

2. Loading CSV data

```
1  -- populate the blank table with data from services.csv
2  COPY services
3  FROM 'C:\LabFiles\Services.csv'
4  DELIMITER ','
5  csv header;
```

To upload the csv file, I used the postgresSQL documentation for the copy function (The PostgreSQL Global Development Group, *Copy* 2023). Since I had already labeled the columns with lowercase headers, I used the “csv header” to remove the original headers of the data.

C. SQL Query

```
1  SELECT s.tech_support, c.churn, COUNT(*) as customer
2  FROM customer AS c
3  INNER JOIN services AS s ON c.customer_id = s.customer_id
4  GROUP BY s.tech_support, c.churn
5  ORDER BY customer DESC;
```

To gather the data that would answer the SQL question: Does having access to technical support affect customer retention I wanted to see the tech_support column from the services table, the churn column from the customer table, and the count of customers by whether they have tech support first and second if they discontinued service. In order to best visualize the results, I ordered the rows by the number of customers with the highest customer count at top. Using the postgresSQL style guide to ensure that they query was readable (Holywell, SQL style guide 2022).

1. CSV File

query results

tech_support	churn	customer
No	No	4634
Yes	No	2716
No	Yes	1616
Yes	Yes	1034

These are the results of the SQL query. We can see the four columns:

1. No tech support and the customer kept the services with 4634 customers.
2. Tech support and the customer kept the services with 2716 customers.
3. No tech support and the customer left the company with 1616 customers.
4. Tech support and the customer left the company with 1034 customers.

D. Add-on File

The CSV file for services will be updated with new customers and the services they have signed up for or as customers change their services. However, a day-to-day update of customer service choices is not necessary within the database itself. Since the services data is relevant to churn analysis, I would advise that the add-on file be acquired and refreshed in the database either on a monthly basis to track the changes or as new queries are needed to be run.

As analysis of the variables which could affect customer retention is done and business decisions are made and implemented, it would be important to update the add-on file in the database to show any changes in customer retention.

E. SQL Script

```
1  -- Refresh add-on data as needed
2  -- clear the existing data
3  DELETE FROM services
4  -- Upload updated csv
5  COPY services
6  FROM 'C:\LabFiles\Services.csv'
7  DELIMITER ','
8  csv header;
```

Since creating a better way to update this data is not a part of the project, I believe the best way to update the add-on file is by deleting the old data and re-uploading the new. This is because all rows of the add-on file will need to be updated and cherry picking the new data to update will be tedious and time consuming. Since customers are often changing their services and discontinuing their service altogether the data will go through large changes on a monthly basis.

F. Panopto Video

Link provided in the attachments.

G. Web Sources

No web sources were used to acquire data or code in the process of this project.

H. Sources

Holywell, S. (2022, November 15). SQL style guide by Simon Holywell.

<https://www.sqlstyle.guide/>

The PostgreSQL Global Development Group. (2023a, May 11). 5.4. *constraints*.

PostgreSQL Documentation. <https://www.postgresql.org/docs/current/ddl-constraints.html>

The PostgreSQL Global Development Group. (2023b, May 11). *Copy*. PostgreSQL

Documentation. <https://www.postgresql.org/docs/15/sql-copy.html>

I. Professional Communication

Demonstrate professional communication in the content and presentation of your submission.