WESTERN GOVENORS UNIVERSITY

D205 PERFORMANCE ASSESSMENT

Date: 19 AUGUST 2023

Ann Hagan
AHAGA22@WGU.EDU

T<small>ABLE OF</small> C<small>ONTENTS</small>

A. SQL QUESTION

The data acquisition performance process focuses on a telecommunications business that needs to analyze customer churn data or the turnover rate of customers. To analyze churn data, the SQL query performed focuses on answering the business question, "How many internet customers using DSL or Fiber with at least one equipment failure in the past year, disconnected service in the past month?". The data acquisition process utilized PostgreSQL and PgAdmin4. The SQL query retrieves the necessary data from the customer and services tables. The customer table provided the churn data and the number of equipment failures for each customer. The services table provided the internet service type, DSL or Fiber Optic, procured by each customer. Utilizing data from both tables allows analysts to provide insight into determining if internet service equipment failures are related to the rate of customer turnover each month.

A1. IDENTIFIED DATA FROM SOURCES

The original database contained five tables, one of which is the customer table. Three columns from the customer table are used to create the final query results. The three columns and their respective data types include 'customer_id' with the data type text, 'yearly_equip_faiure' with the data type integer, and 'churn' with the data type text. The 'customer_id' is the primary key of the customer table.

The new table created in the database, named services, contained service-related information for each customer. The columns in the services table that are use include 'customer_id' and 'internetservice'. The 'customer_id' column served as the primary key for the services table and foreign key reference to the customer table's 'customer_id' column. The 'internetservice' column's data answered whether each customer had used 'Fiber Optic', 'DSL', or 'None' service. All the data types for the services table were established as 'text' to support the importing of 10,000 rows from an unknown original source.

The final requested query consists of a SQL Join between customer and services table. The SELECT statement focuses on counting the number of customers using internet services of DSL or Fiber Optic that had at least one equipment failure in the past year and terminated their service from the telecommunications company.

B. ENTITY RELATIONSHIP DIAGRAM

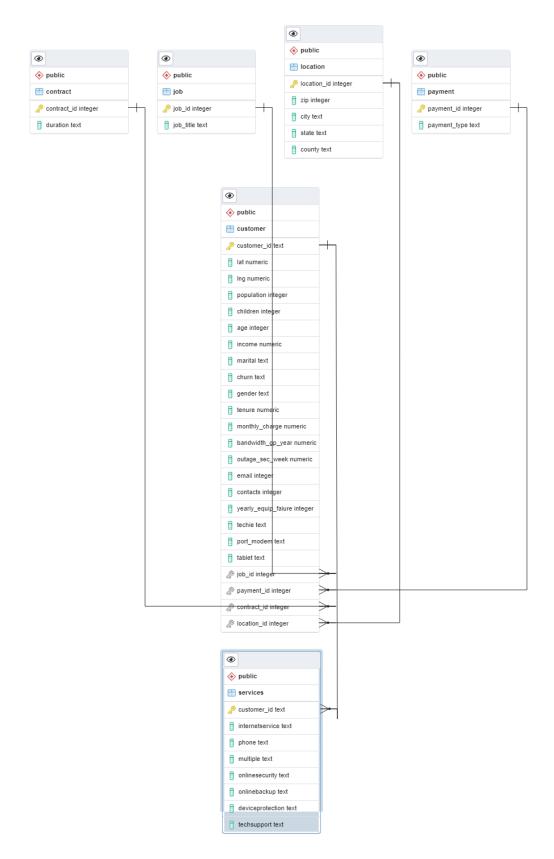Below is the Entity Relationship Diagram showing the relationships between the tables in the 'churn' database.

FIGURE 1: CHURN DATABASE ENTITY RELATIONSHIP DIAGRAM

B1. ENTITY RELATIONSHIP CODE

The SQL code that creates the Entity-Relationship Diagram is as follows:

```
BEGIN;

CREATE TABLE public.contract

(

        contract_id integer NOT NULL,

        duration text,

        PRIMARY KEY (contract_id)

);

CREATE TABLE public.customer

(

        customer_id text NOT NULL,

        lat numeric,

        lng numeric,

        population integer,

        population integer,

        children integer,

        age integer,

        Income numeric,

        marital text,

        churn text,

        tenure numeric,

        monthly_charge numeric,

        bandwidth_gp_year numeric,

        outage_sec_week numeric,
```

```sql
        email integer,

        contacts integer,

        techie text,

        job_id integer,

        payment_id integer,

        contract_id integer,

        location_id integer,

        PRIMARY KEY (customer_id)

);

CREATE TABLE public.job

(

        job_id integer NOT NULL,

        job_title text,

        PRIMARY KEY (job_id)

);

CREATE TABLE public.location

(

        location_id integer NOT NULL,

        zip integer,

        city text,

        state text,

        country text,

        PRIMARY KEY (location_id)

);

CREATE TABLE pulic.payment
```

```
(
        payment_id integer NOT NULL,

        payment_type text,

        PRIMARY KEY (payment_id)
):
CREATE TABLE public.services

(
        customer_id text NOT NULL,

        internetservice text,

        phone text,

        multiple text,

        onlinesecurity text,

        onlinebackup text,

        deviceprotection text,

        techsupport text,

        PRIMARY KEY (customer_id)
);
ALTER TABLE public.customer

        ADD FOREIGN KEY (contract_id)

        REFERENCES public.contract (contract_id)

        NOT VALID;
ALTER TABLE public.customer

        ADD FOREIGN KEY (job_id)

        REFERENCES public.job (job_id)

        NOT VALID;
```

```
ALTER TABLE public.customer

        ADD FOREIGN KEY (location_id)

        REFERENCES public.location (location_id)

        NOT VALID;

ALTER TABLE public.customer

        ADD FOREIGN KEY (payment_id)

        REFERENCES public.payment (payment_id)

        NOT VALID;

ALTER TABLE public.customer

        ADD FOREIGN KEY (customer_id)

        REFERENCES public.customer (customer_id)

        NOT VALID;

END;
```

B2. SQL IMPORT CODE

The following SQL code imported the comma-separated values file into PGAdmin (Wilbert, 2020):

--command" "\\copy public.services(customer_id, internetservice, phone, multiple, onlinesecurity, onlinebackup, deviceprotection, techsupport) FROM 'C:/LabFiles/Services.csv' DELIMITER ","CSV HEADER QUOTE'\"ESCAPE"","

The services table data types are set to 'text.' The 'customer_id' is the primary key and with the not null constraint. The not null constraint ensures only qualified data associated with a customer is considered in the analytical process. The alter table command performed post-upload, set the 'customer_id' of the services table as a foreign key reference to the customer table 'customer_id.'

C. SQL QUERIES UTILIZED

C1. COMMA SEPARATED VALUE FILES RESULTING FROM QUERY

The SQL query used intended to aggregate data from the customer and services table that met all the requirements. The requirements include the customer must have had DSL or FiberOptic internet service, have disconnected service, and have had at least one or more equipment failures in the past year. A SELECT and JOIN query with WHERE and GROUP BY clauses are chosen to accomplish this. (Malik, 2019) The final query used:

SELECT services.internetservice, COUNT(internetservice)

FROM customer

JOIN services

ON customer.customer_id = services.customer_id

WHERE churn = 'Yes' AND internetservice != 'None' AND yearly_equip_faiure > 0

GROUP BY internetservice;

The result consisted of two rows of internet services, DSL and Fiber Optic, with the total number of customers for each internet service that disconnected their service and had at least one equipment failure in the past year.

Final Data:

	internetservice	count
1	DSL	361
2	Fiber Optic	334

D. RECOMMENDATIONS FOR THE FILE REFRESH RATE
The recommendation for refreshing the file utilizing the SQL query is monthly.

D1. JUSTIFICATION FOR RECOMMENDED REFRESH FILE RATE
The recommendation for refreshing the file utilizing the SQL query is monthly. The recommendation is based on facts that a monthly report of how many users disconnected services last month and had equipment failure in the past year provides only an entry point into further analysis and research. In order to change the churn rate of customers due to equipment failure, the company would need to use the resulting data to perform a further investigation into each user's age of equipment, the average age of equipment at failure, the vendors with the most equipment failures, and which existing customers have similar equipment that has a likelihood of failing. The utilized query needs to provide sufficient data for the business to make decisions that predict

potential equipment failures. However, the company could use the query data to support further research into equipment types or equipment age to test hypotheses on whether changing equipment affects the churn rate. Realistically, the timeline for finding additional data on the equipment type, age, and vendors will take time. Employing staff to locate, find, and replace similar equipment pieces current customers use will take even more time. Therefore, the data from the first to the second month may not change. However, over a quarter, six months, or even a year, the business could implement policy changes with the equipment refresh rate and monthly run the report to determine if the changes positively affect internet customers' churn rate. The impact on churn rate would are best measured on a month-by-month basis. An alternative timeline could be quarterly, but the business risks not assessing the effectiveness of changes at a fast enough rate to reduce the risk of wasting resources on changes that do not positively impact the churn rate.

E. AUDIO VISUAL DEMONSTRATION

Audio visual demonstration was created using Panopto Application (Western Govenors University, 2023). Audio visual demonstration of the SQL query and results can be located at: https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=72946555-7c8c-4b31-8c96-b0630132b25f

F. Web Sources

Western Govenors University. (2023, 08 10). *Panopto How-To Videos*. Retrieved from WGU: https://wgu.hosted.panopto.com/Panopto/Pages/Sessions/List.aspx#folderID=%22291182 f0-cb6a-466a-929c-2cbac05f0109%22

Wilbert, A. (2020). *PostgreSQL Essential Training (2020) [Video]*. (LinkedIn Learning) Retrieved 08 2023, from LinkedIn Learning: https://www.linkedin.com/learning/postgresql-essential-training-2020/import-data-from-a-csv?resume=false

G. Sources

Malik, U. G. (2019). *SQL for data analytics: Perform fast and efficient data analysis with the power of sql.* Packt Publishing, Limited.