**Task 1: Data Acquisition**

Alex Rivera

Western Governors University

D205 - Data Acquisition

June 6, 2023

**A: Research Question**

The research question for this project is, “How many techie and non-techie internet service customers churned last month?” The result of this query can be used to gain insight into the cause of Internet service customer churn and help in setting a strategy to mitigate customer churn within this population of customers. Higher retention can be accomplished through promotional targeting, educating customers, and offering services to internet service customers.

**A1: Identifying Data**

Two data sources will be used to answer the research question; the churn dataset and the Services CSV file provided in the virtual machine environment. The churn dataset will utilize the customer table with the fields: “Customer\_id,” “Techie” and “Churn.” These three fields will provide a unique customer id, whether they consider themselves a techie based on a questionnaire, and if they discontinued service in the last month. The services CSV will utilize the “Customer\_id” and “InternetService” fields to pull whether the customer had either of the following internet services: DSL, fiber optic, or None.

**B: Logical Data Model**

A screenshot of a computer

Description automatically generated with medium confidence

Above is an illustration of the logic data model of the add-on services CSV file and its relationship to the customer table in the churn database. For the add-on CSV, a table will be created to hold the eight text fields that will not accept NULL inputs into the table. Services\_CSV table once created will have a foreign key on “customer\_id” used to build a relationship with the primary key “customer\_id” on the customer table. A cockroach labs blog post went into detail about the impact of this relationship, “The existence of a foreign key column establishes a foreign key constraint – a database rule that ensures that a value can be added or updated in column\_a only if the same value already exists in column\_b” (Custer, 2023, What is a foreign key?).

**B1: Code for The Physical Data Model**

A screenshot of a computer program

Description automatically generated with low confidence

The SQL code above was created with the assistance of the pgAdmin graphical interface in the virtual machine environment. Within the churn database tables, pgAdmin allows you to create a table by adding the following: Table name, Columns, Data type, Not NULL, Primary key, and constraints. This interface allowed me to input the necessary fields that were needed to hold the data from the Services.csv add-on file. The primary and foreign keys created allowed for the relationships between the Services\_CSV and Customer tables. All SQL code is available once the table is created, and the final output published above.

**B2: Loading CSV Data**

A black text on a white background

Description automatically generated with low confidence

PgAdmin allows importing from a file through the query tool interface, but it does not show the output of the SQL code used to import into the table. Because of this limitation, I used the learnsql website for a detailed walkthrough of the code needed to copy into a table from a file that I referenced for the SQL code above (Romanowski, 2021, How to Import CSVs to PostgreSQL). The SQL code copies the contents from the Services CSV in the Labfiles folder provided in the virtual machine environment into the Services\_CSV table; additional options of using a comma delimiter and skipping the CSV header.

**C: SQL Query**

A screenshot of a computer code

Description automatically generated with low confidence

The SQL statement needed to acquire the necessary data for the research question is quite straightforward only needing two tables and four fields. The left join relies on the relationship between the customer and Services\_CSV tables on the customer\_id field created in section B1. Filtering of the two tables is done by looking at customers that have churned in the past month equal to “Yes” and the internet service field not equal to “None.” Next, the field needed is internet service, techie, and a count of techie as techie\_count. Group and order by statements help summarize the data in a manner that allows us to gain insight into the research question: how many techie and non-techie internet service customers churned last month? The output of the provided query above states the following churn: DSL techie 895, DSL non-techie 219, Fiber Optic techie 800, and Fiber Optic non-techie 240.

**C1: CSV File**

A screenshot of a computer screen

Description automatically generated with low confidence

A snapshot of the query results is provided above, and an export of the query results was submitted with the report.

**D: Add-On File**

For the query to remain relevant to the research question the add-on file will need to be updated at least once a month to match the frequency of the churn field discontinues update of yes or no. A monthly update cadence will also provide the most up-to-date services the customers are subscribed to, this will be relevant to the business to allow targeted promotions and increase revenue.

**E: SQL Script**

A picture containing text, screenshot, font

Description automatically generated

To refresh the data from the add-on file, this will follow a similar approach to section B2 to load the data. The main difference in the process is that data already resides with the Services\_CSV table which will need to be truncated before loading the data back into the table. For the refresh to work like in section B2 the flat file used for the data import will need to reside in C:\LabFiles\Services.csv as done previously.

**F: Panopto Video**

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=22ca4b3a-9a10-4d23-8860-b01c0112611c

**G: Web Sources**

Custer, C. (2023, May 4). *What is a foreign key? (with SQL examples)*. Cockroach Labs. https://www.cockroachlabs.com/blog/what-is-a-foreign-key/

Romanowski, J. (2021, February 5). *How to import CSVS to PostgreSQL using PgAdmin*. LearnSQL.com. https://learnsql.com/blog/how-to-import-csv-to-postgresql/