**Adam Eyerman**

**Student ID #010426715**

**D205: Data Acquisition**

A.  Summarize a research question that can be answered using both the original database and the add-on CSV data. The question should require data from both these data sources.

1. Identify which data from the original data set and the add-on CSV file are needed to answer the research question.

**Are customers who self-report as a techie more likely to have Internet Service, and how satisfied are they with the reliability of their service?**

1. **Data needed:**

* **(Original) Customer Table – customer\_id column**
* **(CSV) Services Table – customer\_id column and internetservice column**
* **(CSV) Survey Responses Table – customer\_id column and reliability column**

B.  Create a logical data model for the add-on CSV file by evaluating the data contained in the file and emphasizing the relational constraints.

1. Write SQL code that creates a table that accommodates the extension of the logical data model to a physical data model by specifying the field types and relevant keys.

2. Write SQL code that loads the data from the add-on CSV file into the table created in part B1.

**The logical data model connects the Customer Table to the Services Table and Survey Responses Table with the primary key being customer\_id. They have a one-to-one relationship as each customer has one record for each connecting table. The primary key is the customer\_id column. The foreign keys are techie – customer table, internet service – services table, and reliability – survey responses table.**



1. **SQL Code for creating the Survey Responses and Services Tables:**

Graphical user interface, text, application, email

Description automatically generatedText

Description automatically generated

1. **SQL Code for Loading Add-On data to newly created tables:**

Graphical user interface, text

Description automatically generated with medium confidence

Text

Description automatically generated with medium confidence

C.  Write SQL statement(s) for a query or queries that inform the research question summarized in part A.

1. Provide a CSV file or files that capture the results from the query or queries.

**SQL Code for finding all non-techie customers with internet service and their reliability responses:**

Graphical user interface, text

Description automatically generated with medium confidence

**Non-techie customers had a 78.6% rate of having internet service and an average reliability score of 3.49.**

**SQL Code for finding all techie customers with internet service and their reliability responses:**

Graphical user interface, text, application

Description automatically generated

**Techie customers had a 79.2% rate of having internet service and an average reliability score of 3.52.**

**Data shows no statistically significant difference between groups in either metric.**

1. **CSV files provided in separate attachments.**

D.  Determine how often the add-on file should be acquired and refreshed in the database for the data to remain relevant to the business and the research question.

**The add-on files should be acquired and refreshed weekly to remain relevant to the business and the research question. This rate would ensure an accurate analysis of the customer profile and if any changes have occurred.**

E.  Create an SQL script that performs the process of loading the add-on data.

**Format for loading add-on data to newly created tables:**

Text

Description automatically generated with medium confidence

**According to Layne (2021) this code is used for loading add-on data to a newly created table in PostgreSQL**

**SQL Code for both Services and Survey Responses add-on data:**

Text

Description automatically generated

Text

Description automatically generated with medium confidence

F.  Provide a Panopto video recording that includes a demonstration of the functionality of the code used for the analysis and a summary of the programming environment.

[**Tue Jun 14 2022 1:29:29 PM (panopto.com)**](https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=0423d520-065b-4927-9376-aeb401204052)

[**https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=0423d520-065b-4927-9376-aeb401204052**](https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=0423d520-065b-4927-9376-aeb401204052)

G.  Record the web sources used to acquire data or segments of third-party code to support the application. Be sure the web sources are reliable.

**SQL Code used for add-on data upload to newly created tables:**

[<https://dataschool.com/learn-sql/importing-data-from-csv-in-postgresql/>](https://dataschool.com/learn-sql/importing-data-from-csv-in-postgresql/)

H.  Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.

**Layne, M. (2021). *Importing data from CSV in PostgreSQL*. The Data School. Retrieved June 13, 2022, from https://dataschool.com/learn-sql/importing-data-from-csv-in-postgresql/**

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