**Performance Assessment: D205**

**A. Research Question**

The research question that I will be answering is as follows: are customers who are married and have children, as in more people within the same household, more likely to purchase fiber optic internet over DSL? These two types of internet services are very different from one another, and it is well established that fiber is the fastest type of internet available today. DSL internet is the most affordable option but download speeds are within a range of 5-35 megabits per second while fiber optic is more expensive but has download speeds ranging from 250-1000 megabits per second, which makes it a better and more stable service in households that have more people (Cooper 2022).

To answer this research question, I am using the churn database “customers” table in PGAdmin4 alongside the added-on “services” CSV file. The data that will be needed from each set is the “Customer ID” and “Internet Service” data from the services csv file and the “customer ID,” “children,” and “marital” data from the customer table in the churn database.

**B2/E. Loading the Add-On Data**

![Graphical user interface, text, application, email

Description automatically generated]()

The image displayed above showcases the successful query generated to create the table that will later contain the data from the add-on CSV file. The customer\_id data has been defined to be not null and has also been created as the primary key for the table.

Graphical user interface, application, Teams

Description automatically generatedTable

Description automatically generated

These next two screenshots demonstrate the command entered in order to import the services CSV file into the newly created services table as well as a simple select all query to show that the data was imported correctly into the table.

**B1. Logical Data Model**

A picture containing diagram

Description automatically generated

The screenshot above showcases the logical data model/ERD. As only two tables were used in answering the research question, there is a one-to-one relationship between them with the customer\_id field being the relational constraint as it is the only field binding the two tables together.

**C. SQL Statements to Inform Research Question**

In order to determine if bigger households bought more fiber optic internet services, I firstly did an inner join to join the services and customer table, as is demonstrated here.

Table

Description automatically generated

The next step was to use different *WHERE* statements while using a *COUNT* command in order to match different criteria to determine if more fiber optic was being bought. These different criteria are: never married or widowed with no children, never married or widowed with children, married with children, and married without children. The SQL queries were made for each case and the count determines the number of the total 10,000 different customers who bought fiber optic in each case, which can then be divided into a percentage. The results of each are show below, respectively.

Graphical user interface, text, application, email

Description automatically generated

Figure 1: 9.94%

Graphical user interface, text, application, email

Description automatically generated

Figure 2: 29:89

Graphical user interface, text, application, email

Description automatically generated

Figure 3: 4.97%

Graphical user interface, text, application, email

Description automatically generated

Figure 4: 14.14

In none of these queries was fiber optic service the most popular product. DSL was always the primary seller. From the results of these queries within this dataset, there is no evidence that bigger households are paying for faster internet speeds. ­

**D. Refreshing Add-On CSV File**

In order for the data to remain relevant in regard to the research question, the services CSV file needs to be refreshed each time a new customer is added into the file. That way, the company will be able to accurately measure if fiber optic internet service is being purchased in larger households or not. As the previous queries have demonstrated, fiber optic is not necessarily more popular in bigger households, so it is very possible that customers do not know the difference or that the provider is charging too much for it. In either case, it would be recommended that this provider market their fiber optic service more efficiently in order to maximize their profit off of it.

**F. Web Sources**

The cited code below was used to provide a basic understanding to the reader of the differences between DSL and Fiber Optic internet in order to understand the research question. No other web sources or 3rd-party codes was used for this assessment.

Cooper, Tyler. “DSL vs Cable vs Fiber: Comparing Internet Options.” BroadbandNow, 15 Mar. 2022, https://broadbandnow.com/guides/dsl-vs-cable-vs-fiber.