A: The Question

“Does a customer’s income impact the likelihood they will purchase an Online Security package?”

Exploring the relationship between income and optional extras could alert the business to affordability concerns with the offered Online Security product. Should such a relationship emerge, the business could look into more competitive pricing models, discounts, and income-based vouchers, driving an increase in usage of the service.

Within the existing database, I will be looking at the customer table. We will be using the numeric field customer.income. From the Services.csv file, we will be using the OnlineBackup column, represented as services.online\_backup using a text value.

Both tables will be joined on the customer\_id field, which is a text field shared both by the existing table and the table generated from Services.csv. In addition to these existing columns, the report will generate three new columns. The first will be income\_bracket, a text field that will be generated by separating records by customer.income according to three pre-defined thresholds. The second column will be total\_customers, a numeric field. This column will be a count of each record falling into the income\_bracket groups from before. The third column is also numeric, percentage\_with\_online\_security. This will be the count of records where services.online\_backup = true divided by total\_customers and multiplied by 100 for a result expressed as a percentage.

B: Entity Relationship Diagram

A screenshot of a computer

Description automatically generated

Above is the ERD for the two tables in question. They are joined by a shared customer\_id. This column is unique in each table and serves as the primary key. Each record from `customer` corresponds to exactly one record from `services` establishing a one-to-one relationship between the tables.

Adding the services table to the database:

CREATE TABLE services (

customer\_id TEXT PRIMARY KEY,

internet\_service TEXT,

phone TEXT,

multiple TEXT,

online\_security TEXT,

online\_backup TEXT,

device\_protection TEXT,

tech\_support TEXT,

FOREIGN KEY (customer\_id) REFERENCES customer(customer\_id)

);

Populating the services table with data from the csv:

COPY services(customer\_id, internet\_service, phone, multiple, online\_security, online\_backup, device\_protection, tech\_support) FROM 'C:\LabFiles\Services.csv' DELIMITER ',' CSV HEADER;

C: SQL

The intention is to use aggregate functions to determine the percentage of customers have Online Security packages within three income brackets: Below 20,000, between 20,000 and 60,000, and over 60,000.

This is the statement:  
  
SELECT income\_bracket, COUNT(\*) AS total\_customers, COUNT(CASE WHEN s.online\_security = 'Yes' THEN 1 END) \* 100.0 / COUNT(\*) AS percentage\_with\_online\_security FROM ( SELECT c.customer\_id, c.income, CASE WHEN c.income < 20000 THEN 'Below 20,000' WHEN c.income BETWEEN 20000 AND 60000 THEN 'Between 20,000 and 60,000' ELSE 'Above 60,000' END AS income\_bracket FROM customer c ) AS income\_groups JOIN services s ON income\_groups.customer\_id = s.customer\_id GROUP BY income\_bracket;

Output:

The results of the query are included in the D205\_Results.csv file.

D: Refresh Timing

In order to keep the report timely, the data should be refreshed annually. While income tends to increase for the average person over time, general shifts in income are not that dramatic within shorter intervals.

After times of larger change, it would help for the business to refresh the data and take another look at the numbers to see if changes in income over time makes a dramatic difference, allowing them a further insight into subscription levels as they apply to customer income.

E: Panopto Video

Below is a link to the Panopto video for the task.

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=a818fd2c-10d0-4028-86f3-b1a1013e9776