Performance Assessment

D205 - Data Acquisition

Doug Haunsperger

MSDA student, Western Governors University

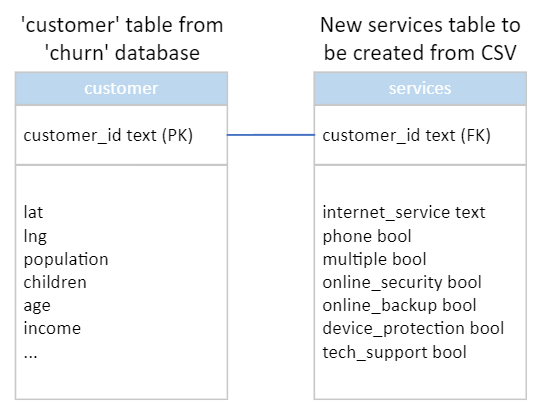
July 15, 2023

A. Research Question: “What is the proportion of customers in each income quintile that purchase the Device Protection service?” The answer to this question could lead to insights about the perceived value of the Device Protection service and could inform how the business will market this service to its customers in the future.

A1. Identifying Data: To answer this question, we will need to query the customer table in the churn database, which contains income data per customer. We will also need to import the Services CSV file to a new services table and query which customers have a ‘Yes’ in the Device Protection column.

B. Logical Data Model

Below is a diagram showing the logical data model of the data in the ‘Service.csv’ file, and how it connects to the existing ‘customer’ table in the ‘churn’ database. As it was not asked for in the requirement, the rest of the tables in the ‘churn’ database are not shown, but the diagram could easily be extended to encompass them.



B1. Physical Data Model Code

To instantiate the above logical model in PostgreSQL, one would use the following code:

CREATE TABLE services

(

customer\_id text REFERENCES customer,

internet\_service text,

/\* While the remainder of the fields are ‘Yes’/’No’, these do not cleanly import as boolean without additional preprocessing. Therefore, they are declared as text for this exercise. \*/

phone text,

multiple text,

online\_security text,

online\_backup text,

device\_protection text,

tech\_support text

);

B2. Loading CSV Data

To load the data from the ‘Services.csv’ file, one would use the following code:

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 Query

We create a view to separate the income values into quintiles. Information on income quintiles used to partition the data table was sourced from taxpolicycenter.org using U.S. Census data.

CREATE VIEW income\_view AS

SELECT c.customer\_id,

s.device\_protection,

c.income,

/\* Income quintiles as of 2021 sourced from US Census Bureau data via taxpolicycenter.org \*/

CASE WHEN c.income > 149131 THEN 'Top quintile'

WHEN c.income BETWEEN 89744 AND 149131 THEN 'Fourth quintile'

WHEN c.income BETWEEN 55000 AND 89744 THEN 'Middle quintile'

WHEN c.income BETWEEN 28007 AND 55000 THEN 'Second quintile'

ELSE 'Bottom quintile' END AS income\_quintile

FROM customer AS c

INNER JOIN services AS s

ON c.customer\_id = s.customer\_id

ORDER BY income DESC;

We then query the view to inform our research question:

SELECT income\_quintile,

-- count only those customers that have purchased Device Protection

SUM(CASE WHEN device\_protection = 'Yes'

THEN 1

ELSE 0

END) AS cust\_with\_dp,

count(customer\_id) AS total\_cust

FROM income\_view

GROUP BY income\_quintile

ORDER BY income\_quintile

C1. CSV file: The file ‘data\_out.csv’ has been submitted separately.

D. Add-On File Refresh

Since the purpose of the research question is to monitor customer behavior at different income levels, perhaps in response to a marketing campaign, the add-on data should be refreshed on a monthly or quarterly basis. Anything more frequent than monthly would be unnecessary since a change in aggregate customer behavior would not have time to show up. If the data is refreshed less frequently than quarterly, then the business could be spending on an ineffective campaign for a long period of time.

E. SQL Script

To generate a script for SQL to load the add-on file, one could save the code from section B2 of this document to a script file.

F. Panopto Video Link: <https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=8d57fcd3-9391-40c0-8070-b0430034c3fe>

G. Web Sources

Tax Policy Center. (n.d.) *Household Income Quintiles,* Urban Institute & Brookings Institution. <https://www.taxpolicycenter.org/statistics/household-income-quintiles>. Retrieved July 17, 2023.

H. Other Sources

No other sources have been cited.