Noah Cina

D205 – Data Acquisition

03/05/2024

*A.  Provide a question that can be answered using structured query language (SQL) to acquire data from****both****the original database and the add-on CSV file data. The question should require data from****both****data sources.*

“What is the average age of customers with fiber optic internet service?”

*1.  Justify how your question provided in part A will be answered.*

The relevant data from the Services.csv file will be added to a new table in the churn database. A query will be run using an aggregate function that calculates the average age of all customers in the customer table whose InternetService column from the Services.csv file has the label “Fiber Optic”.

*2.  Identify which data from****both****the original dataset and the add-on CSV file are needed to answer the question provided in part A, including all tables, columns, and data types.*

The relevant fields required from the database are:  
-the customer\_id (string) column  
-the age (integer) column  
from the customer table in the churn database.   
The relevant fields required from one of the add-on CSV files are:  
-the customer\_id (string) column  
-the InternetService (string) column  
from the Services.csv file

*B.  Create an entity relationship diagram (ERD) for the add-on CSV file and any other tables and columns used to answer the question from part A by evaluating the data contained in the file and identifying the 1:1, 1:M, or M:N relationships and relational constraints.*

A screenshot of a computer

Description automatically generated

*1. Describe the relationship between the existing table and the table created in part B and discuss any issues with the relationship in the ERD.*

The relationship between the two tables is 1:1, as the new services table only adds more information for each existing customer entry in the customer table. From a data integrity perspective, this introduces redundancy and could be restructured, but works as-is.

*2.  Write a SQL statement, in text format, that creates a table for the add-on CSV based on the ERD and specifies the columns and relevant keys.*

CREATE TABLE services (

customer\_id TEXT,

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)

);

*3.  Write a SQL statement, in text format, that loads the data from****one****of the add-on CSV files into the table created in part B.*

COPY services

FROM ‘C:\LabFiles\Services.csv’

DELIMITER ‘,’

CSV HEADER;

*C.  Write****one or more****SQL statements in text format for a query or queries that answer the question provided in part A.*

SELECT ROUND(AVG(age), 2) AS average\_age, services.internet\_service

FROM customer

LEFT JOIN services USING(customer\_id)

WHERE internet\_service = ‘Fiber Optic’

GROUP BY internet\_service;

*1.  Provide data files that capture the results from the query or queries.*

Included in submission as NCinaResults.csv

*D.  Identify the specific time period for how often the add-on file should be acquired and refreshed in the database for the data to remain relevant to the business activities and the question from part A.*

Monthly

*1.  Explain why the time period identified in part D is relevant to the specific business activities.*

The average age of customers who use fiber optic cable can be used in advertising targeting the age range of the most customers who require high speed internet connections for services such as online gaming or HQ video streaming. The file can afford to be updated monthly as advertising campaigns generally last at least as long. Updating the file more often provides little benefit for this case.

*E.  Provide a Panopto video recording that can be used in a preproduction code review.*

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=c26bee80-c7b5-49fb-a135-b12b003dbf96>

*F.  Acknowledge****two****web sources for any code used to support the application. Be sure the web sources are reliable.*

[*https://www.postgresqltutorial.com/postgresql-tutorial/import-csv-file-into-posgresql-table/*](https://www.postgresqltutorial.com/postgresql-tutorial/import-csv-file-into-posgresql-table/)

[*https://www.pgadmin.org/docs/pgadmin4/development/erd\_tool.html*](https://www.pgadmin.org/docs/pgadmin4/development/erd_tool.html)