Data Acquisition (D205)

Nina S. Gooden  
Western Govenors Univeristy  
Student ID: 009823504

08/25/2022

Instructor: David Gagner

## 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.

While looking at the data provided, I was drawn to the information about the different demographics that the company serves. As a marketer, I looked at these datapoints in order to better inform my marketing personas. As such, I wanted to segment this data immediately and started with the simplest and most obvious question: Are customers who identify as female more likely to be satisfied by personable customer service and does that satisfaction contribute to the purchase of multiple products?

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

In order to satisfy my research question, I need a few key pieces of information. The data I will need in order to satisfy my research question is as follows:

Original data set information necessary: customer\_id, gender

Survey Responses: customer\_id, Respectful Response, Courteous Exchange, Evidence of Active Listening

Services: customer\_id, Multiple

Table

Description automatically generated

## 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.

I created the following logical model to outline the referential integrity between the three tables I would be working in due to the fact that each database needs a primary key, I called foreign keys to get around that constraint and to make it easier—and cleaner—to join the data.

Graphical user interface, application, Teams

Description automatically generated

### 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.

The SQL code I used to create the tables that would hold the ADD-ON data as follows:

CREATE TABLE public."survey\_responses"

(

customer\_id character varying(50) COLLATE pg\_catalog."default" NOT NULL,

timely\_responses integer,

timely\_fixes integer,

timely\_replacement integer,

reliability integer,

options integer,

respectful\_response integer,

courteous\_exchange integer,

active\_listening integer,

CONSTRAINT survey\_responses\_pky PRIMARY KEY (customer\_id)

)

TABLESPACE pg\_default;

ALTER TABLE public."survey\_responses"

OWNER to postgres;

*CREATE TABLE public.services*

*(*

*customer\_id character varying(30) COLLATE pg\_catalog."default" NOT NULL,*

*internetservice character varying(25) COLLATE pg\_catalog."default",*

*phone character varying(25) COLLATE pg\_catalog."default",*

*multiple character varying(25) COLLATE pg\_catalog."default",*

*onlinesecurity character varying(25) COLLATE pg\_catalog."default",*

*onlinebackup character varying(25) COLLATE pg\_catalog."default",*

*deviceprotection character varying(25) COLLATE pg\_catalog."default",*

*techsupport character varying(25) COLLATE pg\_catalog."default",*

*CONSTRAINT services\_pky PRIMARY KEY (customer\_id)*

*)*

*TABLESPACE pg\_default;*

*ALTER TABLE public.services*

*OWNER to postgres;*

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

Rather than using the BULK INSERT code for this step, I utilized the program’s built-in import system to load the data from the CSVs to the newly created tables, because I knew that the import system uses the COPY FROM command behind the scenes. This was a quick and effective time to use an integrated tool.Graphical user interface, text, application, email

Description automatically generated

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

My first step in finding an answer to my hypothesis was to look at the relationship between all of the datapoints I had decided were important. I filtered out the entries that were not considered highly rated, as well as evaluating the number of overall Female entries. I also looked at those who had more than one service.

Table

Description automatically generated with medium confidence

After that query was complete, I was able to look at the difference between whether or not being satisfied with customer service made a difference to said gender. Unfortunately, my findings were that regardless of customer service satisfaction, there was not a considerable difference between whether or not the individuals had multiple services.



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

Attached as:   
count\_data\_multiple\_vs\_not.csv

full\_data\_female\_interaction\_multiple.csv

## 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 company should refresh the database any time either of the external sources are updated in order to keep the most accurate measurement of customer satisfaction and purchases within the chosen demographic. In order to quickly access up-to-date service information, I would recommend that the data be stored in an OLTP system that updates in real time.

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

The company should use the program’s import utility to refresh the data quarterly as more individuals make purchases and interact with customer service.

Graphical user interface, text, application, email

Description automatically generated

## 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.

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=c683bb81-3b1a-45be-b930-aefd002155e4

