Data Acquisition – D205 1

Data Acquisition – D205

Jeffrey Williams

Western Governors University

Student #: 001173968

Date: 01/11/2023

Mentor: Jared Knepp

Data Acquisition – D205 2

**Table of Contents**

A. Research Question....................................................................................................................3

A1. Data set requirements..............................................................................................................3

B. Logical Model.............................................................................................................................3

B1. SQL Code – Creation of table...................................................................................................4

B2. SQL Code – To load data..........................................................................................................5

C. SQL Statements Supporting Research Question.....................................................................6

C1. Query results............................................................................................................................6

D. Frequency of data refresh........................................................................................................6

E. SQL Script for data load............................................................................................................7

Data Acquisition – D205 3

1. Research Question

Finding the exact number of customers with a timely response rating of 4 or higher, by city, state, service, and type of internet used.

A1. Data set requirements

The research question requires analysis of customer and location tables from the original fields given. Services and survey\_responses tables from add-on CSV data sets to answer this question. The customer table has customer information, and the location table has customer location information such as city, state, and zip code. Services table has the different type of internet services the customer used. The survey\_responses table has various ratings of timely\_fixes, responses\_given, and timely\_responses.

1. Logical Model

Shown below is the logical model with entities, attributes, keys, and relationships.

Diagram

Description automatically generated

Data Acquisition – D205 4

B1. SQL Code – creation of table

pgAdmin’s create table dialogue is used to create services, survey\_responses tables and generated CREATE script listed below.

Graphical user interface, application

Description automatically generated

CREATE TABLE public.services

(

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

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

phone character varying(4) COLLATE pg\_catalog."default",

multiple character varying(4) COLLATE pg\_catalog."default",

onlinesecurity character varying(4) COLLATE pg\_catalog."default",

onlinebackup character varying(4) COLLATE pg\_catalog."default",

deviceprotection character varying(4) COLLATE pg\_catalog."default",

techsupport character varying(4) COLLATE pg\_catalog."default",

CONSTRAINT services\_pkey PRIMARY KEY (customer\_id)

)

CREATE TABLE public.survey\_responses

(

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

timely\_response integer,

timely\_fixes integer,

timely\_replacement integer,

reliability integer,

options integer,

respectful integer,

courteous integer,

active\_listening integer,

CONSTRAINT survey\_pkey PRIMARY KEY (customer\_id)

Data Acquisition – D205 5

B2. SQL Code – To load data.

The following screen shots are captured after importing services, survey response CSV files in respective tables.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Data Acquisition – D205 6

1. SQL Statements Supporting Research Question

-- find total number of customers by service, city, state

-- and timely response rating is 4 or greater

-- order the total number of customers desc from higher to lower

SELECT count(cust.customer\_id) as total\_customers, ser.internetservice, loc.state, loc.city from customer cust

JOIN location loc

ON cust.location\_id = loc.location\_id

JOIN services ser

ON cust.customer\_id = ser.customer\_id

JOIN survey\_responses sur

on cust.customer\_id = sur.customer\_id

WHERE sur.timely\_response = >4

GROUP BY 2, 3, 4

ORDER BY 1 DESC

C.1 Query results



1. Frequency of data refresh

In order to keep the above data current the telecom company should refresh the database any time the two external sources are updated to reflect current services and survey responses.

Data Acquisition – D205 7

1. SQL Script for data load.

pgAdmin’s CSV import utility imports data from the survey responses file when new survey responses are captured.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated