PERFORMANCE ASSESSMENT

KAILI HAMILTON

Masters of Science in Data Analytics, Western Governors University

Course: D205 Data Acquisition

Instructor: Dr. David Gagner

Program Mentor: Krissy Bryant

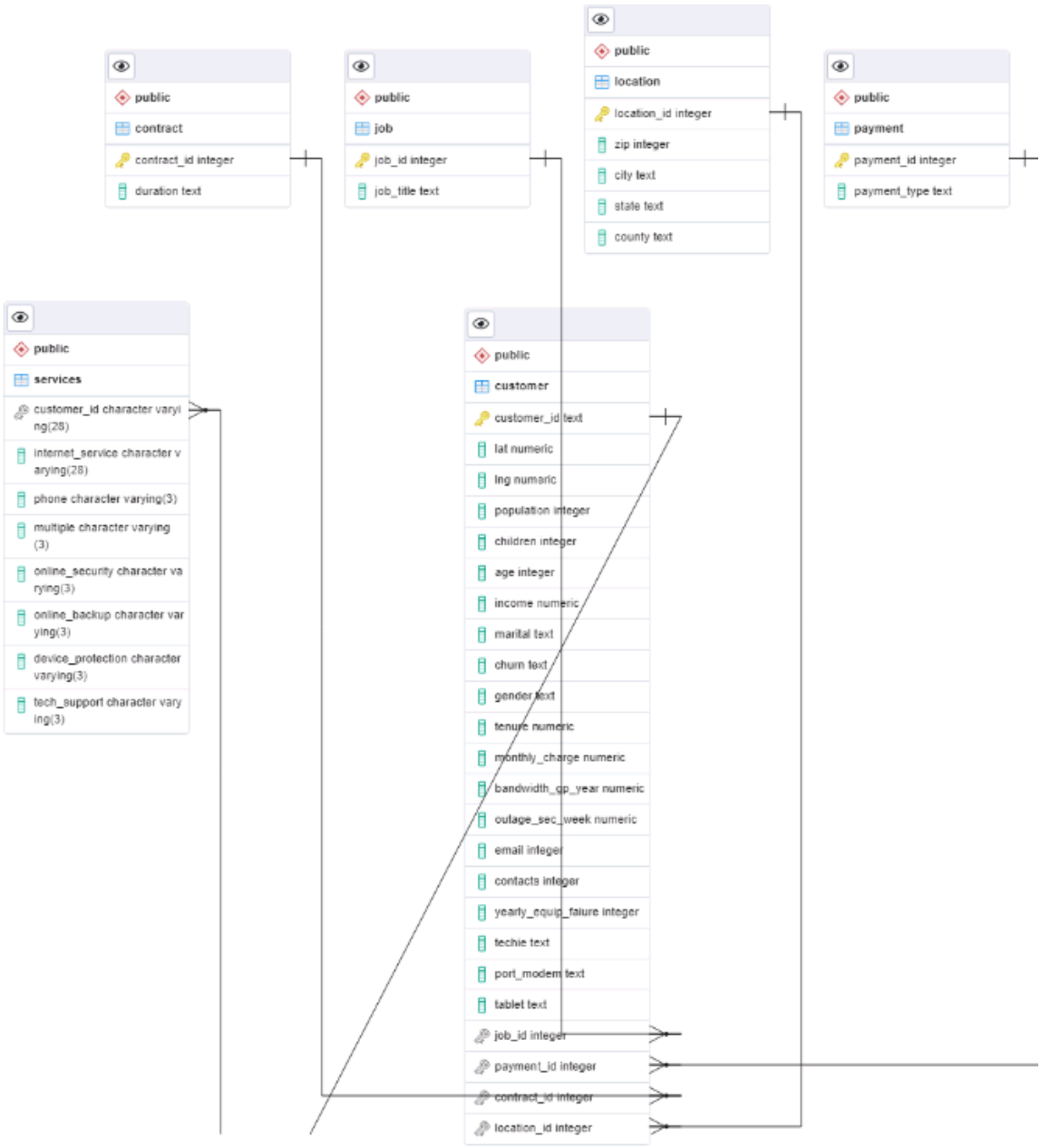
April, 2023

SECTION A

A	What are the top three services chosen by senior citizens, ages 65 and older?
A1	This research question uses data from the original database, namely the customer table filtered by age, and the new table, services, created using the services CSV file. We will join the two tables using the customer_id field. In this way we will have customers ages 65 and older (from the customer table) and the services they subscribe to (from the services table). The fields used from the customer table are: customer_id and age. The customer_id field is the primary key (or unique identifier) for the customer table and has a data type text. We use customer_id to join to the services table, referenced in the following paragraph. Age is a necessary variable to answer the research question because we are specifically filtering the data where the customer's age is 65 and older. Age has data type integer. The fields and their data types used from the services table are customer_id (varchar(28)) and the seven types of services the company offers: internet_service (varchar(28)), phone (varchar(3)), multiple (varchar(3)), online_security (varchar(3)), online_backup (varchar(3)), device_protectrion (varchar(3)), tech_support (varchar(3)). Customer_id is the primary key for the services table and is the foreign key to the customer table. It is the field that is used to join the two tables together to answer the research question. To designate if a customer subscribes to a service, the entries in the seven services fields (listed in the previous paragraph) are a "Yes" or "No", except in the case of internet_service, where the entries are "DSL", "Fiber Optic", or "None". To answer the research question, we will count the "Yes" values in these services fields, or the not "None" values in internet_service, and then assess which of these are the top three services chosen by senior citizens.

SECTION B

B	The logical model, or Entity Relationship Diagram (ERD) is displayed below and emphasizes the relational constraints. The services table, represented in the leftmost box, was added to the database using SQL code presented in part B1. Primary and foreign keys were created – again, see part B1.


public
contract
contract_id integer
duration text
public
job
job_id integer
job_title text
public
location
location_id integer
zip integer
city text
state text
county text
public
payment
payment_id integer
payment_type text
public
services
customer_id character varying(28)
internet_service character varying(28)
phone character varying(3)
multiple character varying(3)
online_security character varying(3)
online_backup character varying(3)
device_protection character varying(3)
tech_support character varying(3)
public
customer
customer_id text
lat numeric
lng numeric
population integer
children integer
age integer
income numeric
marital text
churn text
gender text
tenure numeric
monthly_charge numeric
bandwidth_gb_year numeric
outage_sec_week numeric
email integer
contacts integer
yearly_equip_failure integer
techie text
port_modem text
tablet text
job_id integer
payment_id integer
contract_id integer
location_id integer

B1

Below is the SQL code that creates the services table and includes the column names and data types for each column. Then, I checked if the table was indeed added into the database.

```
-- create a table for the services data
CREATE TABLE services (
    customer_id VARCHAR(28),
    internet_service VARCHAR(28),
    phone VARCHAR(3),
    multiple VARCHAR(3),
    online_security VARCHAR(3),
    online_backup VARCHAR (3),
    device_protection VARCHAR (3),
    tech_support VARCHAR (3)
);

-- check if table was indeed created
SELECT *
FROM information_schema.columns
WHERE table_schema = 'public'
;
```

Now that the new table is created and data is imported, I assigned primary and foreign keys to the table. The primary key for the services table is the field customer_id. It is the primary key because it uniquely identifies each entry in the services table. Customer_id is also a foreign key, pointing to the customer table. This is a one-to-many relationship because one customer could subscribe to many services (Datacamp, n.d.).

```
-- create primary and foreign keys for services table
/* customer_id = pk, created in the ERD
    customer_id = fk to customer table, created below */

ALTER TABLE services
    ADD CONSTRAINT services_fkey FOREIGN KEY (customer_id) REFERENCES customer (customer_id)
;
```

B2

The SQL code that loads the data from the services CSV file into the services table is as follows (Command Prompt, Inc., 2023):

“ ”\\copy public.services (customer_id, internet_service, phone, multiple, online_security, online_backup, device_protection, tech_support) FROM ‘C:/Users/LapUser/Desktop/Services.csv’ DELIMITER ‘,’ CSV HEADER QUOTE ‘\’” ESCAPE “”;””

```
--command " "\\copy public.services (customer_id, internet_service, phone, multiple, online_security, online_backup,
device_protection, tech_support) FROM 'C:/Users/LabUser/Desktop/Services.csv' DELIMITER ',' CSV HEADER QUOTE '\'"
ESCAPE "";""
```

SECTION C

C

The research question is, what are the top three services that are used by senior citizens, ages 65 and older. To answer this I prepared and transformed the data as follows:

I left joined the customers table to the services table and used all seven services fields from the services table and selected the age of the customer from the customer table. I also filtered by age (Datacamp n.d.).

```
SELECT c.age,
    internet_service,
    phone,
    multiple,
    online_security,
    online_backup,
    device_protection,
    tech_support
FROM services AS s
LEFT JOIN customer AS c
    ON s.customer_id = c.customer_id
WHERE c.age >= 65
;
```

Next, I used CASE statements to count the number of subscribers in each of the seven services column from the view queried above. The CASE statements are how we can get multiple COUNTs with multiple WHERE clauses (Stack Overflow, n.d.).

```
SELECT
    sum(CASE WHEN phone = 'Yes' THEN 1 ELSE 0 END) AS phone,
    sum(CASE WHEN internet_service != 'None' THEN 1 ELSE 0 END) AS internet_service,
    sum(CASE WHEN multiple = 'Yes' THEN 1 ELSE 0 END) AS multiple,
    sum(CASE WHEN online_backup = 'Yes' THEN 1 ELSE 0 END) AS online_backup,
    sum(CASE WHEN device_protection = 'Yes' THEN 1 ELSE 0 END) AS device_protection,
    sum(CASE WHEN tech_support = 'Yes' THEN 1 ELSE 0 END) AS tech_support,
    sum(CASE WHEN online_security = 'Yes' THEN 1 ELSE 0 END) AS online_security
FROM seniors_services
;
```

This final query allowed me to answer my research question, “What are the top three services chosen by senior citizens (ages 65 and older)?” The top three services are phone, internet, and multiple services. If we exclude the multiple field, which indicates that a senior citizen is subscribing to multiple services, the third most subscribed to service is online backup.

Data Output | Explain | Messages | Notifications

	phone bigint	internet_service bigint	multiple bigint	online_backup bigint	device_protection bigint	tech_support bigint	online_security bigint
1	3084	2640	1566	1528	1519	1282	1170

C1

The CSV file is included in the submission of this performance assessment.

SECTION D

D	The add-on file should be acquired and refreshed in the database daily or weekly. Likely, new subscribers are being added every day and the add-one file should be updated to follow that so that. In this way, the business will have the most up to date information to inform their marketing, sales, and other business strategies.

SECTION E

E	The SQL code that performs the process of loading the add-on data is as follows (Command Prompt, Inc., 2023): “ ”\\copy public.services (customer_id, internet_service, phone, multiple, online_security, online_backup, device_protection, tech_support) FROM ‘C:/Users/LapUser/Desktop/Services.csv’ DELIMITER ‘,’ CSV HEADER QUOTE ‘\’” ESCAPE “”;””

```
--command " "\\copy public.services (customer_id, internet_service, phone, multiple, online_security, online_backup,
device_protection, tech_support) FROM 'C:/Users/LabUser/Desktop/Services.csv' DELIMITER ',' CSV HEADER QUOTE '\"
ESCAPE '"';""
```

SECTION F

F	Here is the link for my Panopto video recording demonstrating the functionality of the code used for the analysis and a summary of the programing environment. https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=36c70247-feee-460c-807e-afe8000b26bc

SECTION G

G	Stack Overflow. (n.d.). How can I get multiple counts with one SQL query? Stack Overflow. https://stackoverflow.com/questions/12789396/how-can-i-get-multiple-counts-with-one-sql-query Command Prompt, Inc. (2000-2023). How to Import or Export CSVs to PostgreSQL Using pgAdmin. Command Prompt, Inc. https://www.commandprompt.com/education/how-to-import-or-export-csvs-to-postgresql-using-pgadmin/#:~:text=To%20do%20that%2C%20right%2Dclick,CSVs%20to%20Postgres%20via%20pgAdmin.

SECTION H

H	Datacamp. (n.d.). SQL for Joining Data [Online course]. Datacamp. https://app.datacamp.com/learn/courses/sql-for-joining-data Datacamp. (n.d.). Introduction to Relational Databases in SQL [Online course]. Datacamp. https://app.datacamp.com/learn/courses/introduction-to-relational-databases-in-sql

SECTION I

I	Professional communication is demonstrated throughout this performance assessment.