D205: Data Acquisition

Performance Assessment

Logan Rosemeyer

Western Governor's University

D205: Data Acquisition

Dr. David Gagner

December 11, 2022

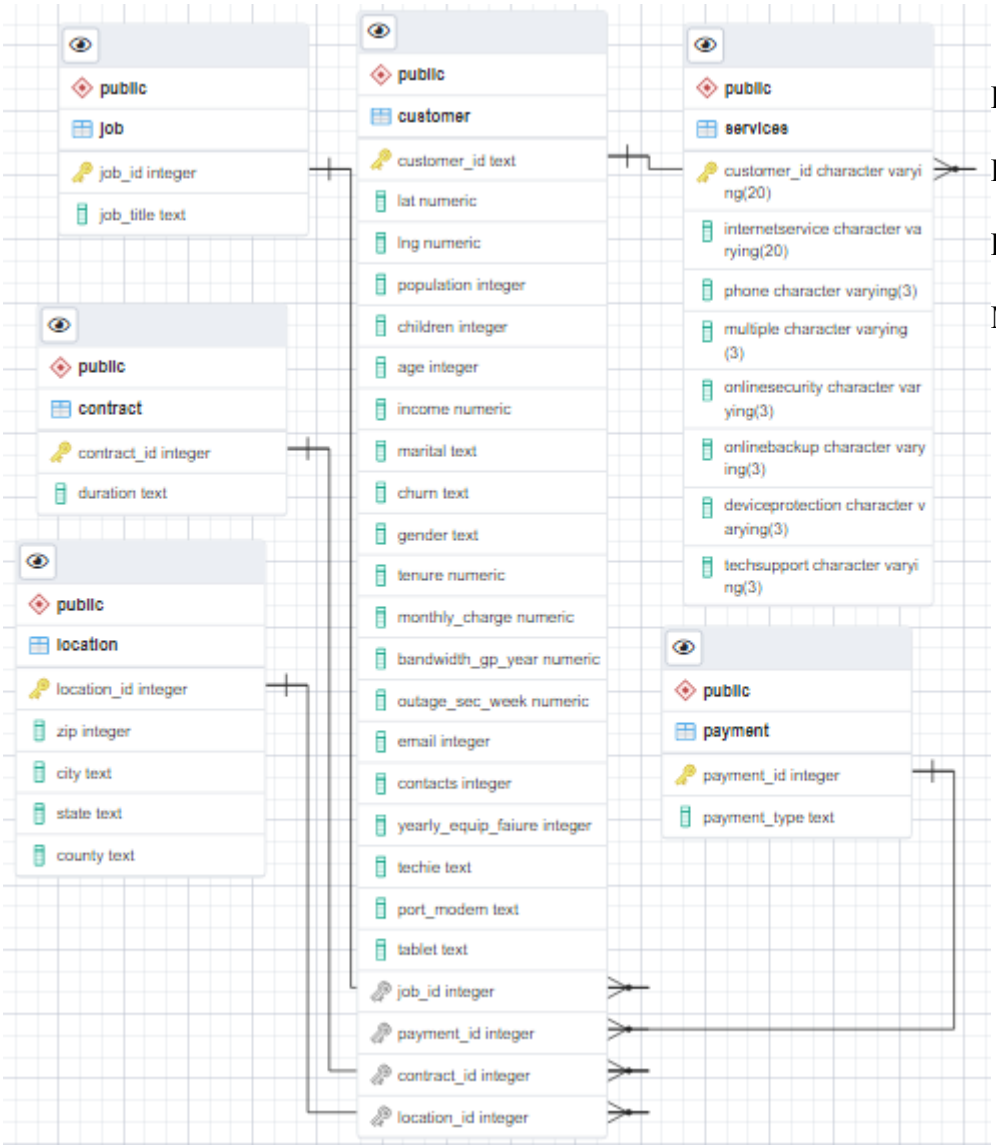
A: Research Question

What is the age distribution of customers that use tech support vs. customers that do not use tech support?

A1: Identifying Data

I will use the age data from the customers table and the tech support column from the services table. I will join the two tables using the customer_id columns in both tables.

B: Logical Data Model



B: Logical Data Model

(continued)

My ERD is shown above. It shows a primary key in all tables, as well as foreign keys in all tables referencing a column in the customer table.I was able to create this logical data model using the ERD tool in pgAdmin (“ERD Tool”, n.d.).

B1: Code for the Physical Data Model

Below is the code I used to create my services table with all columns as character varying columns. The customer_id column cannot have null values and is also the primary key.

```
create table services (
    customer_id varchar(20) not null,
    internetservice varchar(20),
    phone varchar(3),
    multiple varchar(3),
    onlinesecurity varchar(3),
    onlinebackup varchar(3),
    deviceprotection varchar(3),
    techsupport varchar(3),
    primary key(customer_id)
);
```

B2: Loading CSV Data

I was able to load the csv file into my table using the copy statement (PostgreSQL Tutorial, n.d.). I used the Services csv file from the LabFiles folder. I used a comma as the delimiter and specified that there is a header.

```
copy services
from 'C:\LabFiles\Services.csv'
delimiter ','
CSV header;
```

C: SQL Query

This query will return each customer id, the age of each customer from the customer table, and whether or not they use tech support from the services table. I joined the customer and services tables using the customer_id column from both tables.

```
select customer_id, c.age, s.techsupport
from customer c
inner join services s
using (customer_id);
```

C1: CSV File

The CSV file has been uploaded as a separate file in the submission.

D: Add-On File

The add-on file should be updated every two months. This will allow us to track how our age distribution is changing, if at all. If we decide to take action to change the distribution of customers using tech support, we can track the age distribution to see if those action steps are working.

E: SQL Script

```
copy services
from 'C:\LabFiles\Services.csv'
delimiter ','
CSV header;
```

F: Panopto Video

The Panopto video has been uploaded in the D205 folder.

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=88471f3b-e340-41cf-9a29-af680137ba33

G: Web Sources

I did not use any web sources to acquire data.

H: Sources

Import CSV file into postgresql table. PostgreSQL Tutorial. (n.d.). Retrieved December 11, 2022, from https://www.postgresqltutorial.com/postgresql-tutorial/import-csv-file-into-posgresql-table/

ERD Tool. PgAdmin 4 6.17 documentation. (n.d.). Retrieved December 11, 2022, from https://www.pgadmin.org/docs/pgadmin4/6.17/erd_tool.html