Bratislav Petkovic

November 7th, 2022

Data Acquisition

Dr. Sewell

Relationships Between Customer’s Age and Services Received:

Internet Company Churn Data

Before coming up with a research question, I carefully examined the data, taking a close look at data types, tables, and the way they relate to one another. Many columns were of interest to me such as churn, monthly bill, gender, and location. Researching age and technology proficiency I discovered that older groups of people are generally not as proficient with technology as their younger counterparts. A New York Times article stated that around 75% of people older than 60 required assistances with their electronic devices [1]. Therefore, I decided to formulate my question as follows:

How does a customer’s age impact the services they are likely to require from their telecommunications company?

I hypothesized that older customers – older than 65 – are more likely to receive technical support relative to middle aged – 40 to 65 years old – and younger customers – less than 40 years old. My goal was to detect any relationship between age and services which could in turn help the telecommunication company support different customer demographics better.

Table

Description automatically generated

*Figure 1: ERD of the Churn Database*

Before I could answer the question, I had to incorporate the external services dataset into the database. Primarily, I created a schema for it (‘CREATE TABLE table\_name’) and specified every field as varchar of appropriate minimal length. Next, I populated the newly created table via the ‘COPY table\_name FROM ‘local\_file\_path.csv’ and made sure the data was correctly loaded. I was able to generate the Entity Relationship Diagram – also known as a Logical Data Model – in the pgAdmin application. Figure 1 depicts the entity relationships of the 6 tables in the Churn database. In it, it is evident that the central entity is the customer table by which location, payment, job, and contract can be joined to via the foreign keys location\_id, payment\_id, job\_id and contract\_id, respectively. These foreign keys establish and preserve the relationships between the individual entities. For example, to join the payment table and the job table, it would be necessary to perform 2 joins: payment to customer and customer to job. The newly added ‘services’ entity has a direct relationship to the customer table via the customer\_id field which is the primary key of each of those tables. The relationship is one to one as each unique customer is constrained to having only a single set of services.

To answer the question I performed an inner join between the customer table and the service table. For the customer table, a new column called Age\_Band was constructed from the ‘CASE WHEN’ statement specifying the age brackets from my hypothesis. Age\_Band was split into 3 categories – ‘Senior Citizen’ , ‘Middle Aged’ and ‘Young’. To calculate the percentage of customers which received a particular service, the result was grouped by Age\_Band and the functions ROUND, AVG, COUNT, CASE WHEN were used in conjunction.

I was surprised to find no relationship between age and technology services received (Figure 2). Upon further inspection, I grouped by age, creating a scatter plot of the data but did not find any other insight (Figure 3). There was weak or no correlation between age and any of the services received.

*Figure 2 : Percentage of Customers who received the service for Each Age Group*

*Figure 3: Percentage of People Who Received Tech Support Vs their Age showing no correlation*

Because the external dataset entity – service – has the same primary key as customer entity, those two tables would ideally be updated on the same basis. This would allow every customer in the customer database to have a corresponding set of services associated with it. However, as the data showed no relationship between age and gender, to keep the add-on file relevant to the research question, the business would be mandated to refresh the services data on a yearly basis or semi-yearly basis.

[1] *nytimes.com*. (2020, March 27). https://www.nytimes.com/2020/03/27/technology/virus-older-generation-digital-divide.html

[2] Romanowski, J. (2021, February 5). *How to Import CSVs to PostgreSQL Using PgAdmin*. LearnSQL.com. https://learnsql.com/blog/how-to-import-csv-to-postgresql/