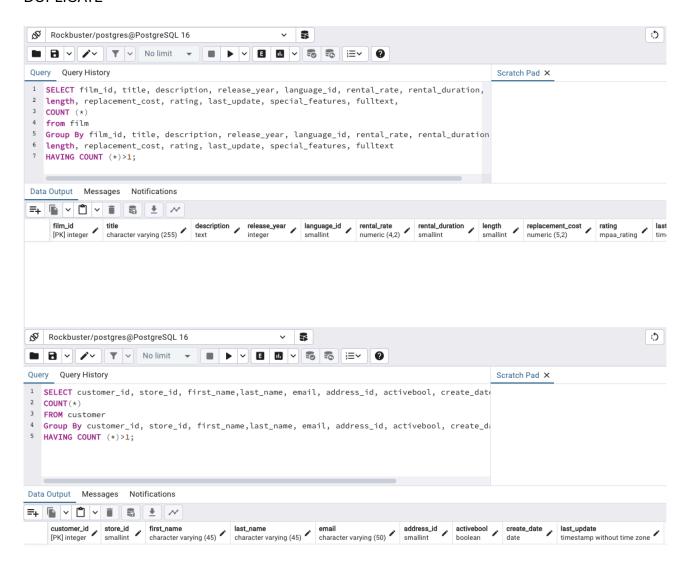
# **TASK 3.6**

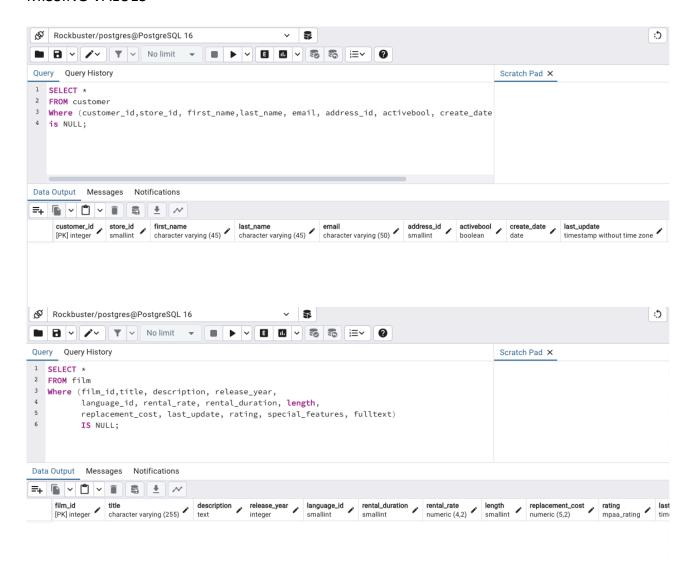
## Summarizing & Cleaning Data in SQL

### 1: DUPLICATE



I didn't come across any duplicate entries in the film\_table or customer\_table. But if there were any, I would address them by either removing the duplicate entry or creating a "view table" that selects only one instance of the duplicates.

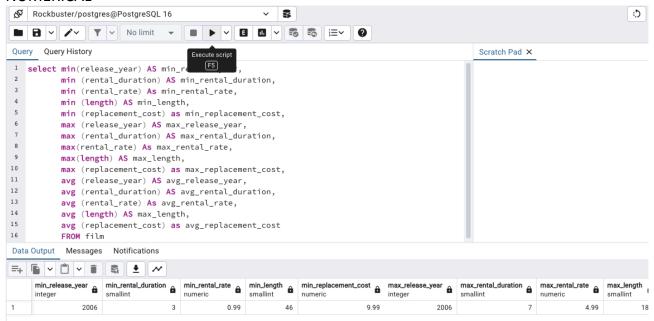
#### MISSING VALUES



I haven't encountered any missing values in the tables so far. However, if any were present, I would fill them in as needed. For instance, if the last\_name field in the customer\_table was blank, I would skip it. But if the release\_year of a film was missing, I would make sure to find the information and populate the record accordingly.

#### 2:

#### **NUMERICAL**

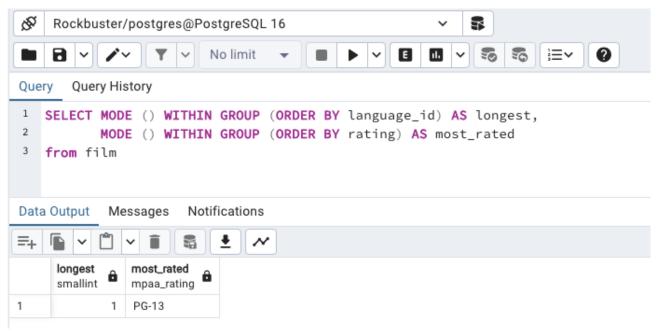


These data records offer crucial insights into several aspects:

- The overall satisfaction level of customers with the rented movies.
- The average duration customers typically keep a rented movie.
- The potential cost associated with replacing a specific movie.
- Additionally, analyzing the popularity of different movie lengths.

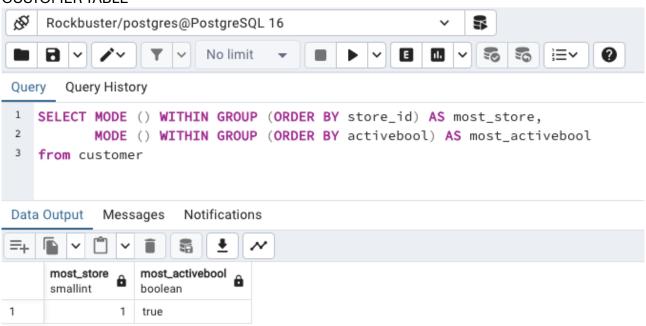
In the Customer\_table, there isn't pertinent numerical data for retrieving minimum, maximum, and average values. For instance, customer\_id, being a unique identifier for each customer, wouldn't yield meaningful insights through minimum, maximum, or average calculations.

#### NON NUMERICAL



In this analysis, it's feasible to determine the most prevalent language among the movie collection and identify which audience demographic could derive the most value from the current selection. In my perspective, querying data from text fields like description wouldn't be pertinent as the information contained within those columns wouldn't offer substantial insights.

#### **CUSTOMER TABLE**



In this case, the analysis shows that Store 1 has the most customers signed up. It also asks if there are more active customers than inactive ones. But checking things like first names or last names wouldn't really help find useful information, like the most common names in the database.

#### 3:

SQL is really handy because it gives you exactly what you ask for when you make a query. It's great for connecting data from different places and makes sure your data stays safe from accidental changes. But since I'm still learning, I sometimes find it hard to use the right words. With practice, though, I think I'll get better at it.

Excel is good too because it helps automate tasks with lots of functions. I'm more used to Excel, so I can do similar things by moving tables around and using functions. But when it comes to big amounts of data, Excel can struggle a bit compared to SQL. Even though Excel is easier to use and understand, I'm enjoying learning SQL and getting better at it. Just like learning any new language, it takes time to get the hang of it.