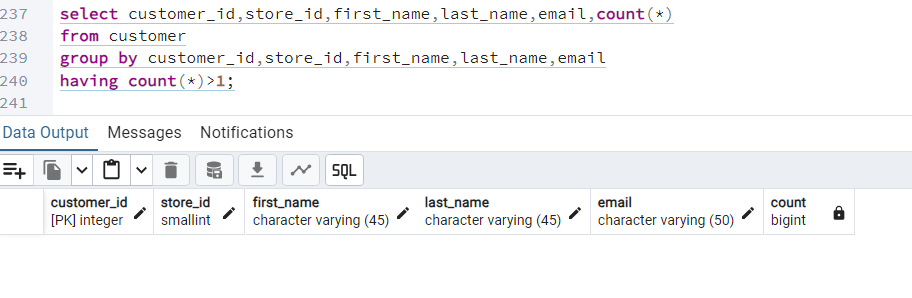
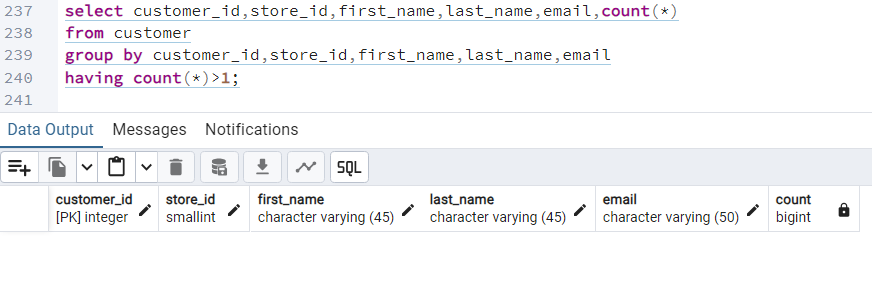
**Task 3.6**

**Check for clean and dirty data:**

Duplicate data

Film Table 

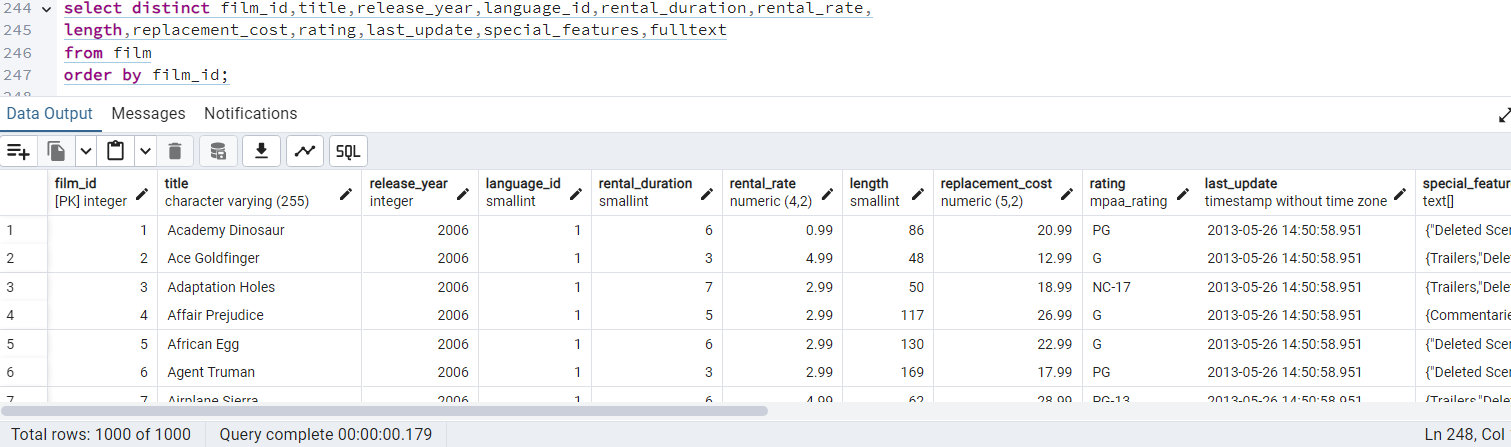
Customer Table

****

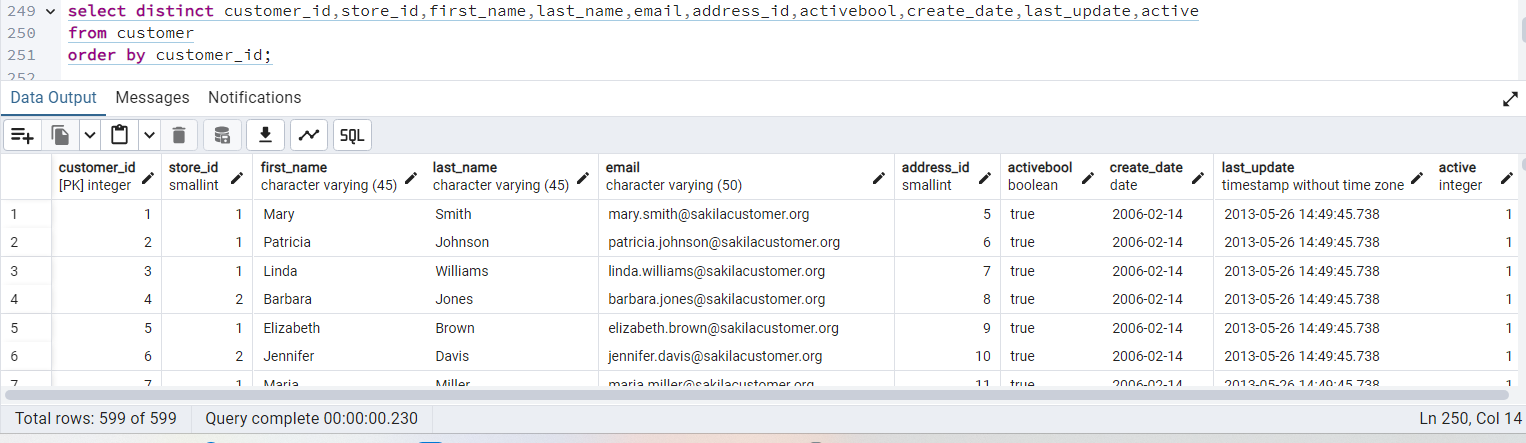
I found no duplicates in the film table or the customer table. Had there been any duplicates, I would have created a “view table” where I would exclude any duplicates. Second, any duplicates found could have been deleted from the table if granted access to the Date base. Most likely updating and deleting in databases are not permitted, so a “view table” would be the way.

**Non-Uniform Data**

Film Table

****

CustomerTable

****

I found no Nonuniform data in the film or customer table. If there had been nonuniform data, I would have corrected it with the following function:

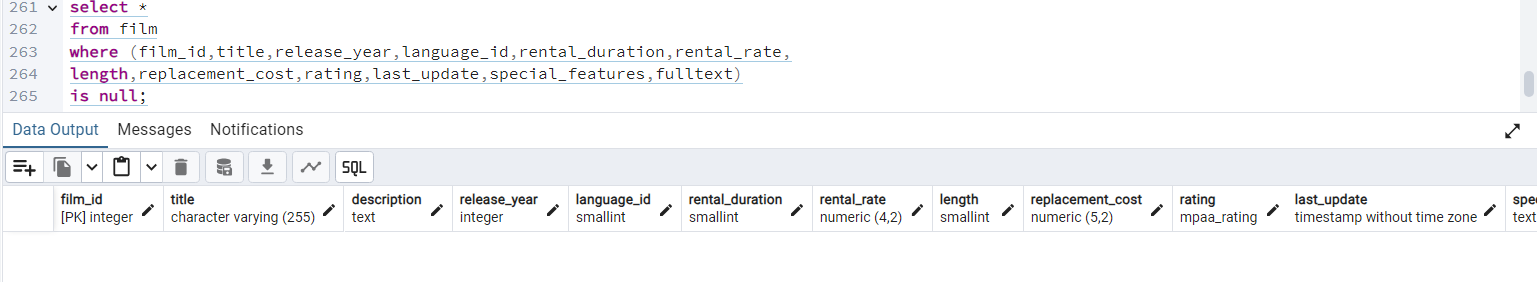
**UPDATE**

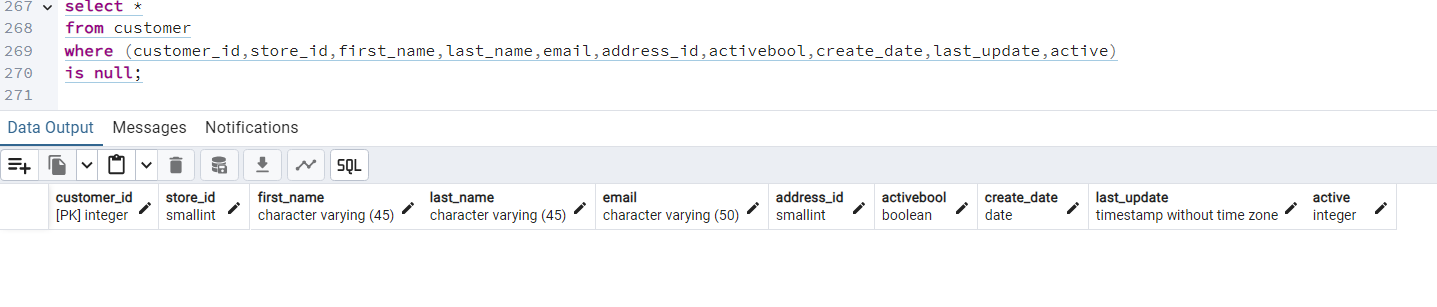
**SET**

**WHERE**

**Missing values:**

Film table

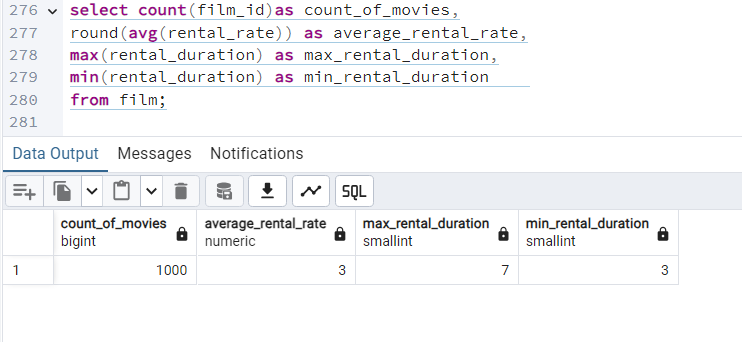


Customer Table

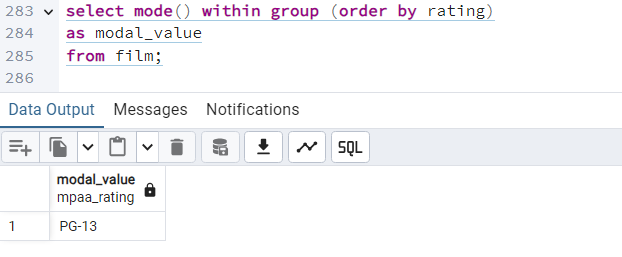
I found no missing values. If there had been any missing values, depending on what and how much is missing; I would either ignore the columns that are not important, or input values if needed depending on the number of missing values in the represented columns by using descriptive statistical method.

**Summarize your data:**

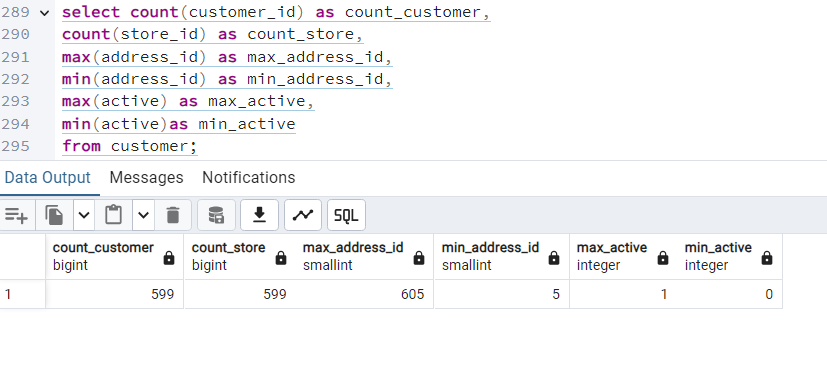
Film Table -Numerical Data



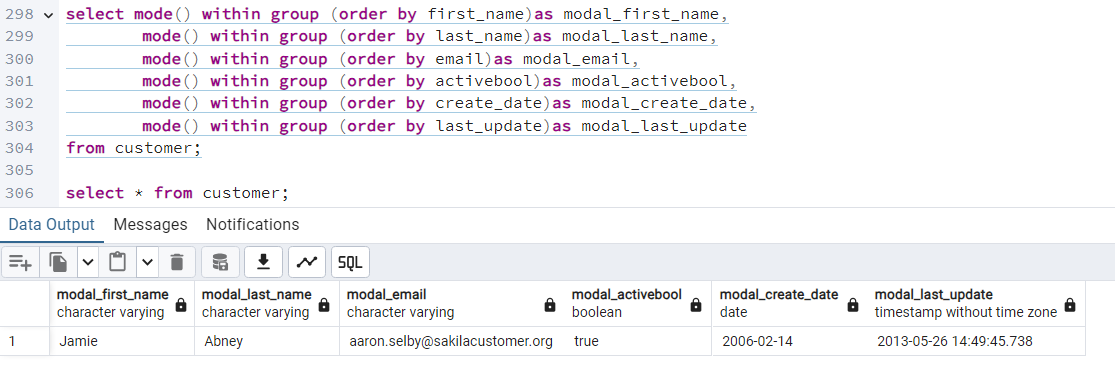
Film\_table -Non-numerical data



Customer table -Numeric Data



Customer table-Non-numeric data



**3. Reflection**

When it comes to data profiling both SQL and Excel have their respective strengths and weaknesses. Excel provides a user-friendly interface and facilitates quick data exploration, catering to users with diverse technical backgrounds. However, it struggles with efficiently handling large datasets and often requires manual processing for data profiling tasks. Conversely, SQL databases offer scalability, automation, and robust data integrity features, enabling efficient management of substantial data volumes and time-efficient analysis. Nonetheless, proficiency in SQL query language is essential, presenting a learning curve for individuals unfamiliar with SQL syntax and concepts; making it less user-friendly than Excel. I find myself leaning towards SQL for data profiling due to its capacity for automating queries and handling multiple columns simultaneously. Nevertheless, I also appreciate the visual element of examining the entire table in Exce