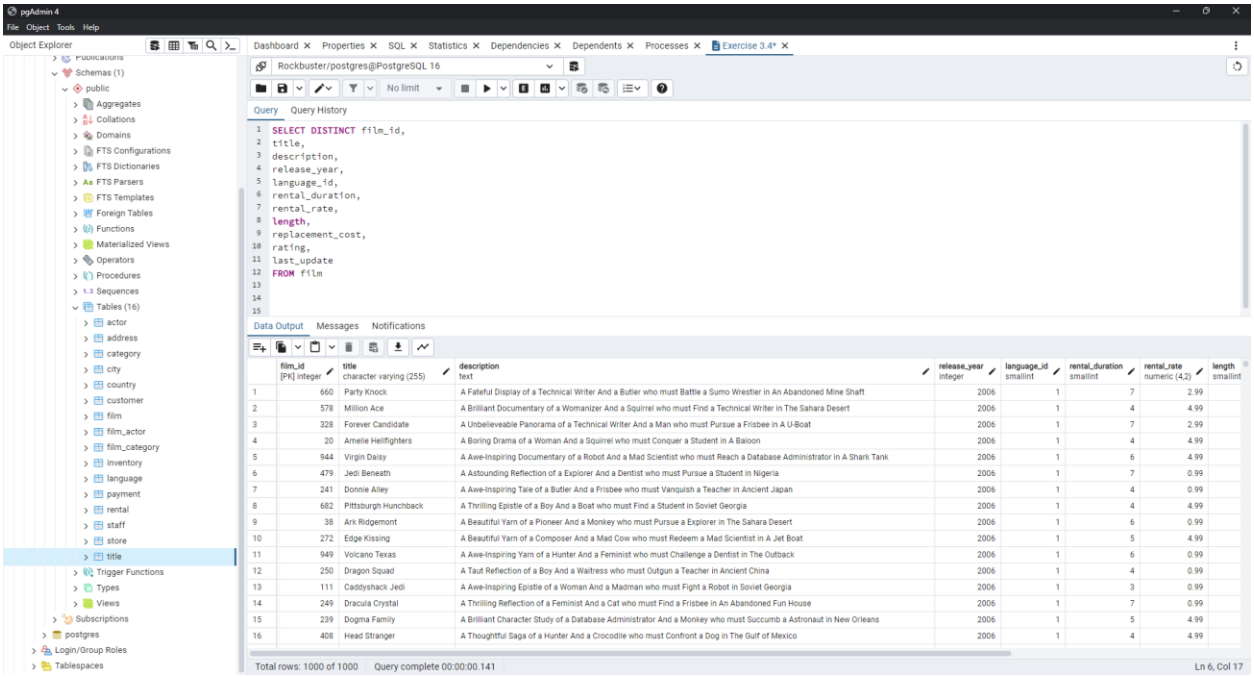
Mandy Yeung
Data Immersion 3.6

Summarizing & Cleaning Data in SQL

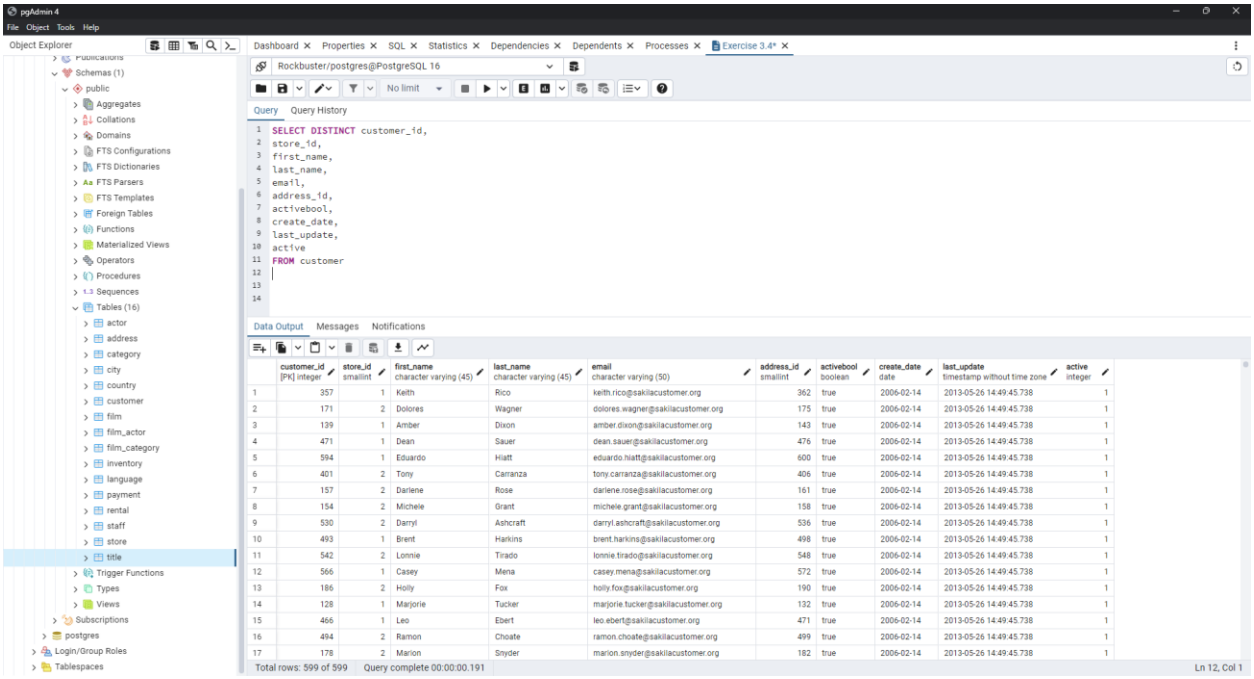
1. Check for and clean dirty data:

Nonuniform Data

FILM



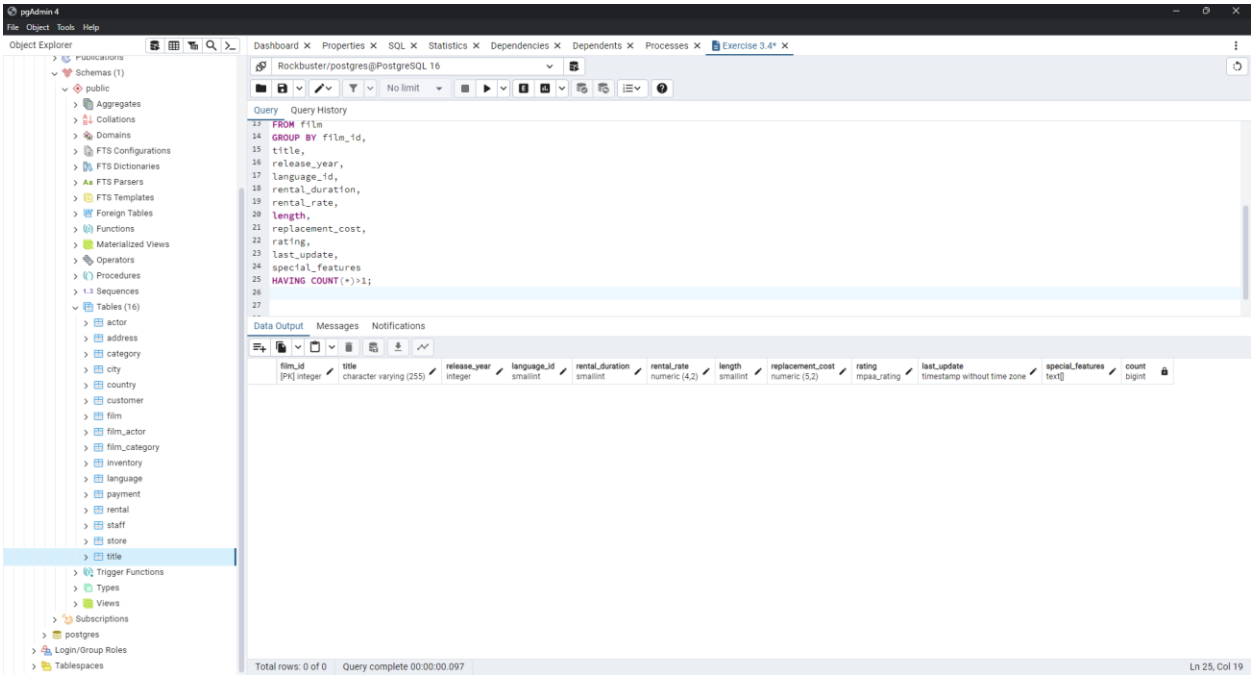
CUSTOMER



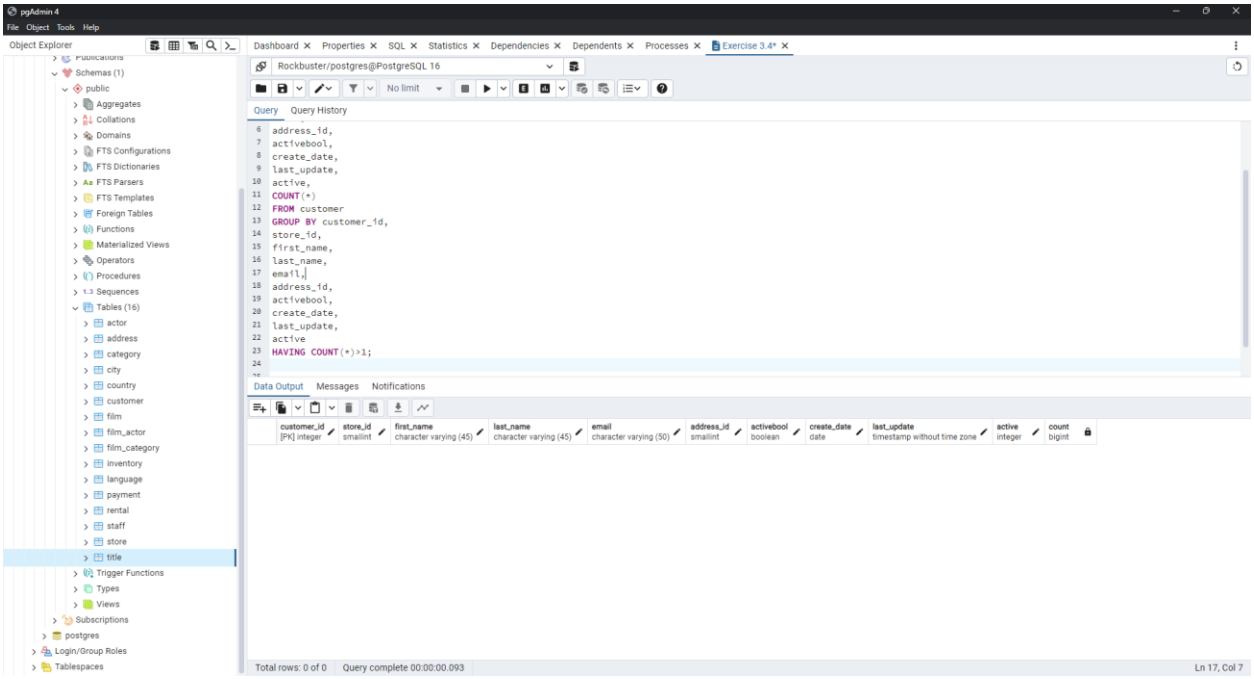
If there were non-uniform for the tables above, I would use UPDATE and SET commands.

Duplicate Date

FILM



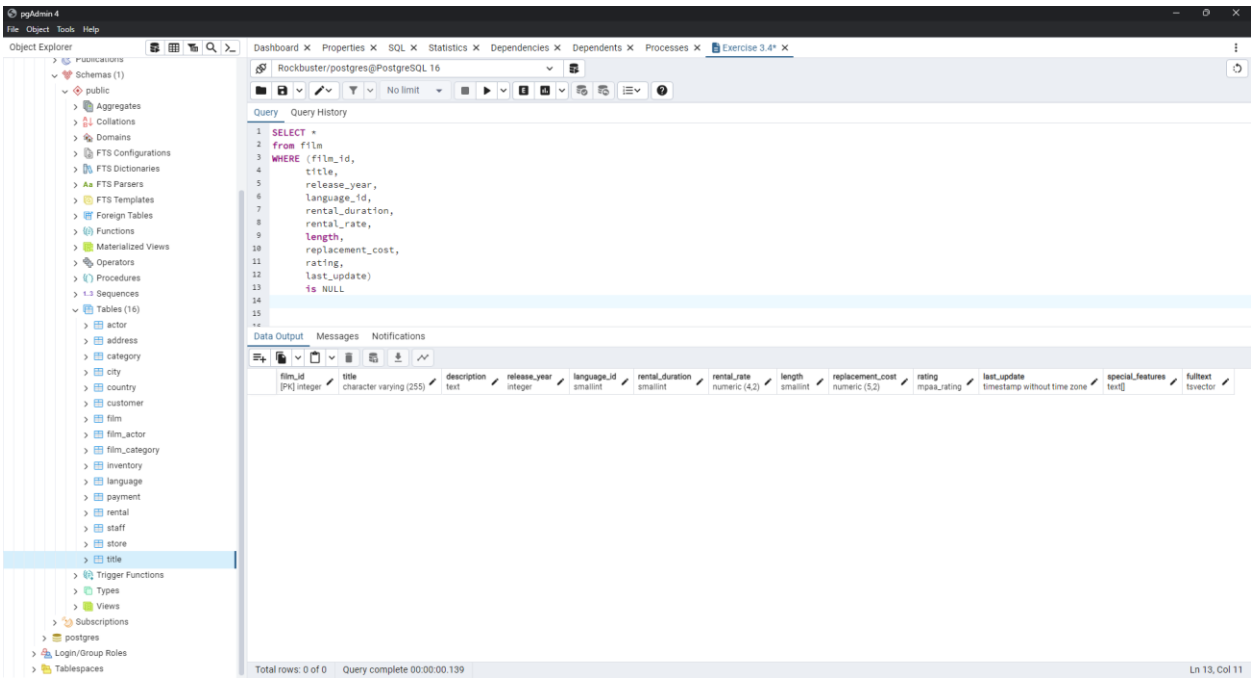
CUSTOMER



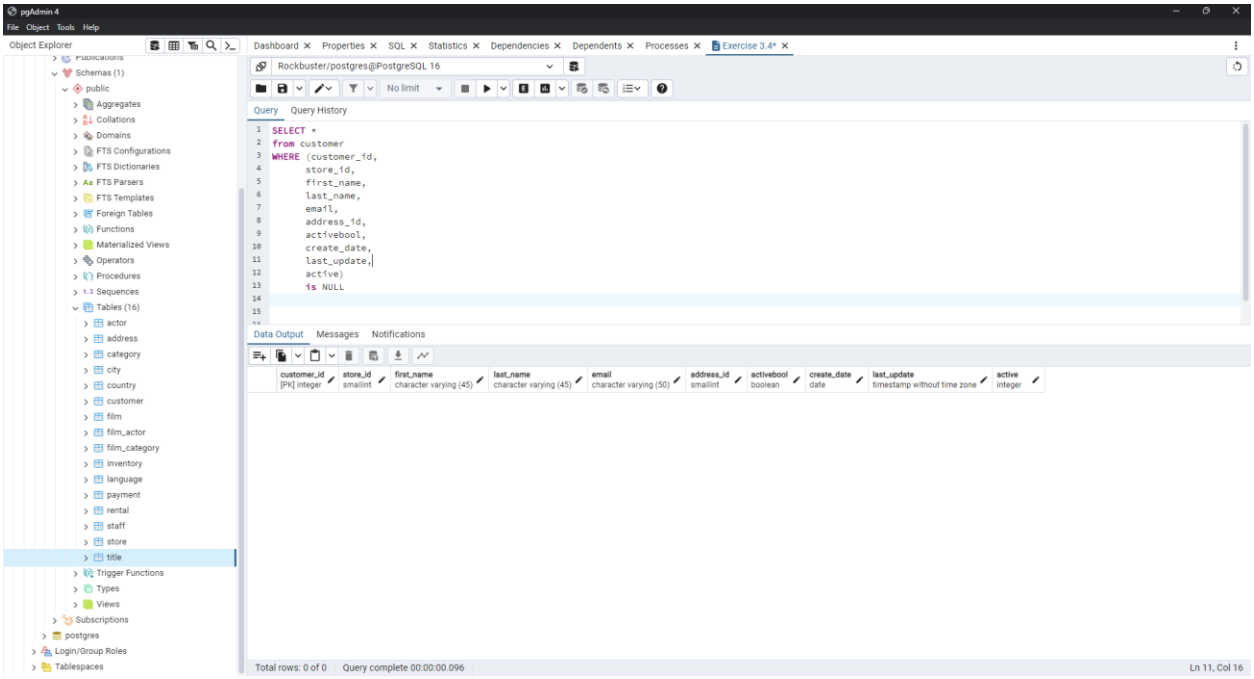
There are no duplicate values in both film and customer. If there were, I would fix it by deleting the duplicate values or creating a VIEW table displaying only the duplicate values.

Missing Data

FILM



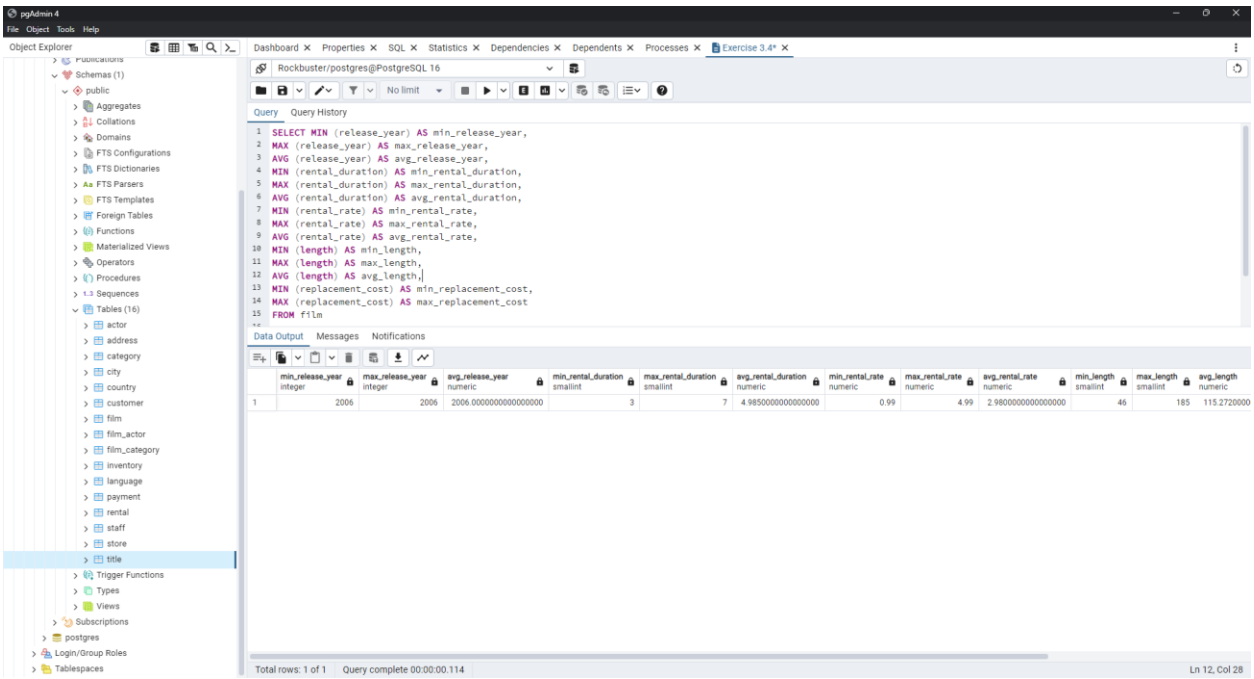
CUSTOMER



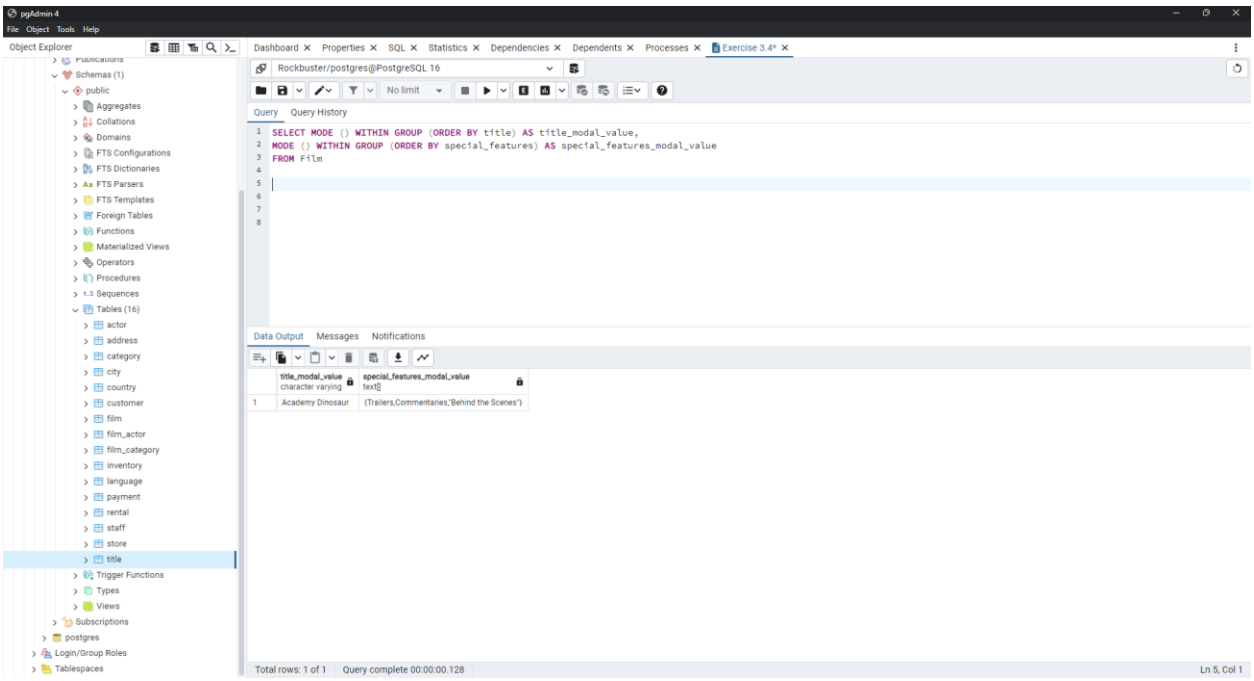
There are no missing values but if there were, I could impute values there aren't a lot of missing data.

2. **Summarize your data:**

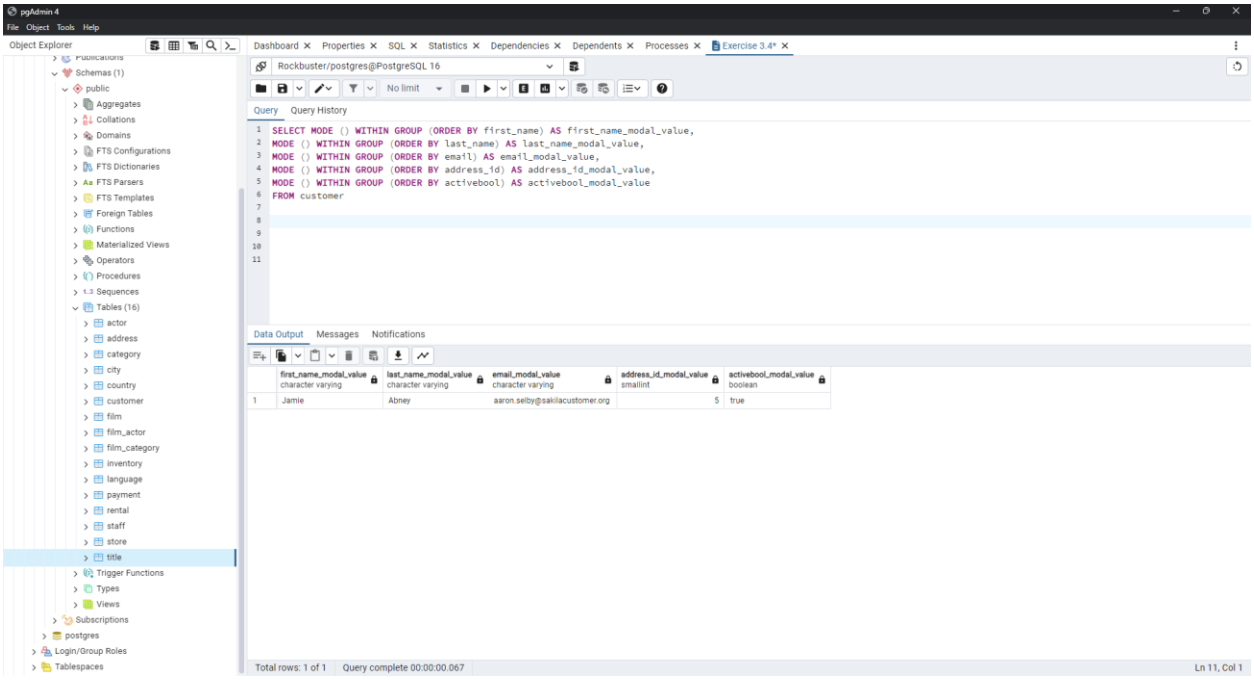
Numerical – Film



Non-numerical – Film



Non-numerical – CUSTOMER



3. **Reflect on your work:**

SQL is very efficient for advanced data tasks and handling large volumes of data. One thing I love about SQL is it prevents you from accidentally deleting or editing data and it tells you when your query is incorrect. But it does require a lot of practice and learning. I find that

although it’s great, I do struggle with remembering the syntax or knowing which and when to use them.
Excel on the other hand is user friendly but it’s geared towards basic data tasks and handling smaller volumes of data. I think both tool is great depending on what you’re using it for and I do really enjoy learning SQL and pretty fascinated on what it can do.