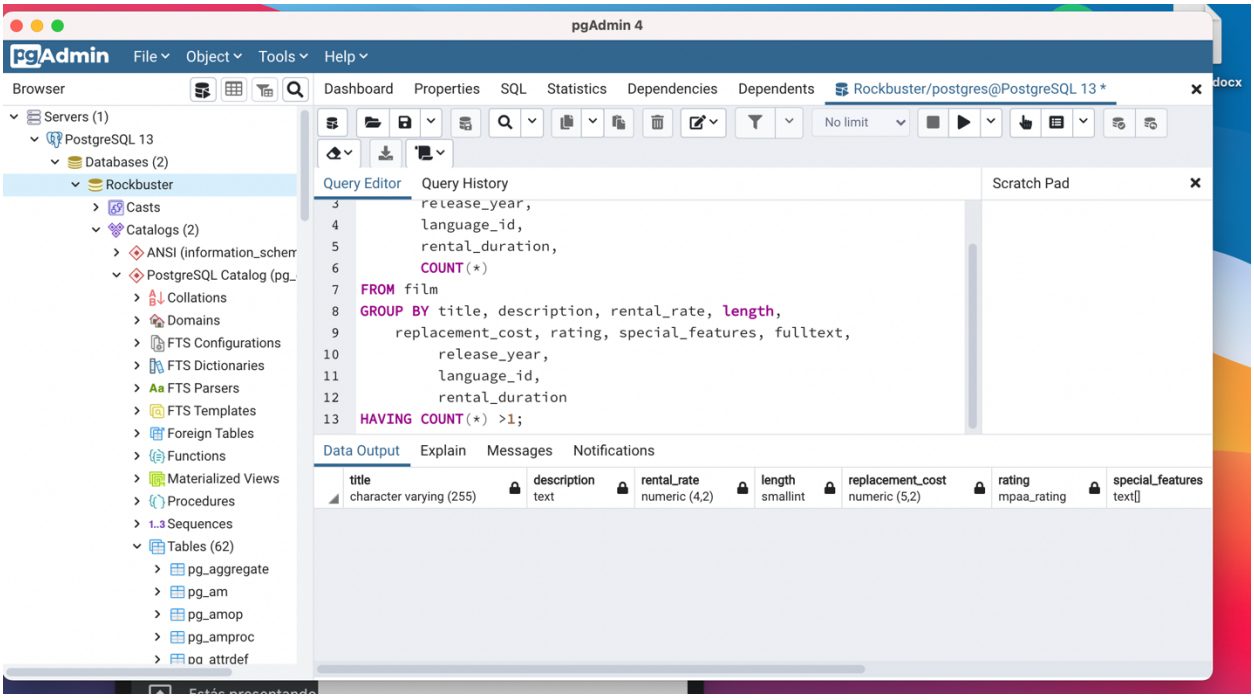
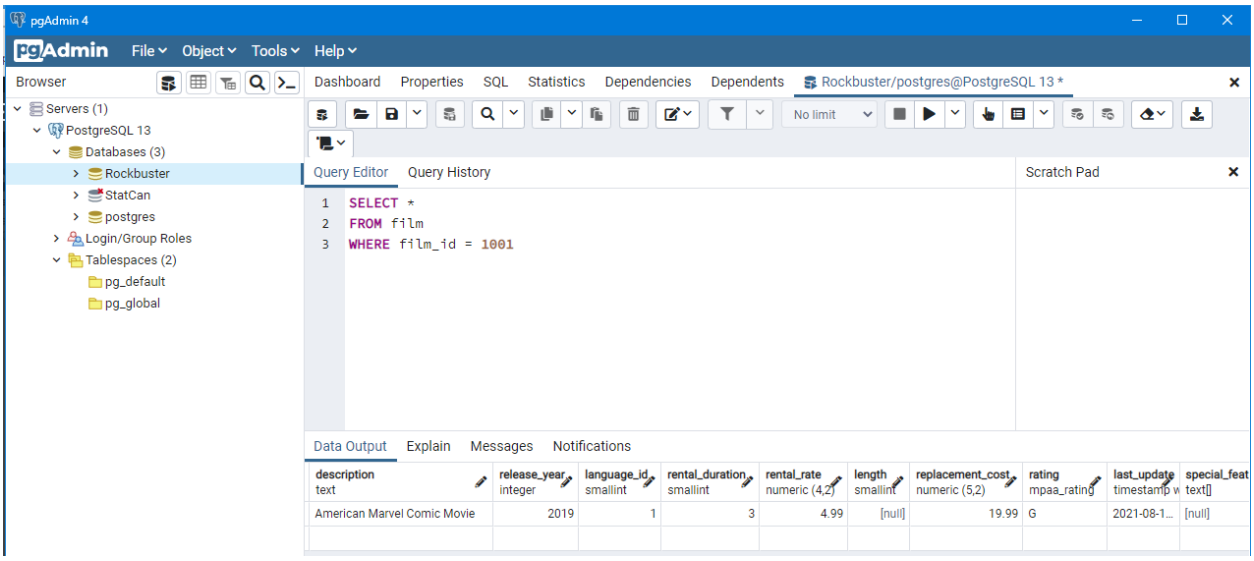
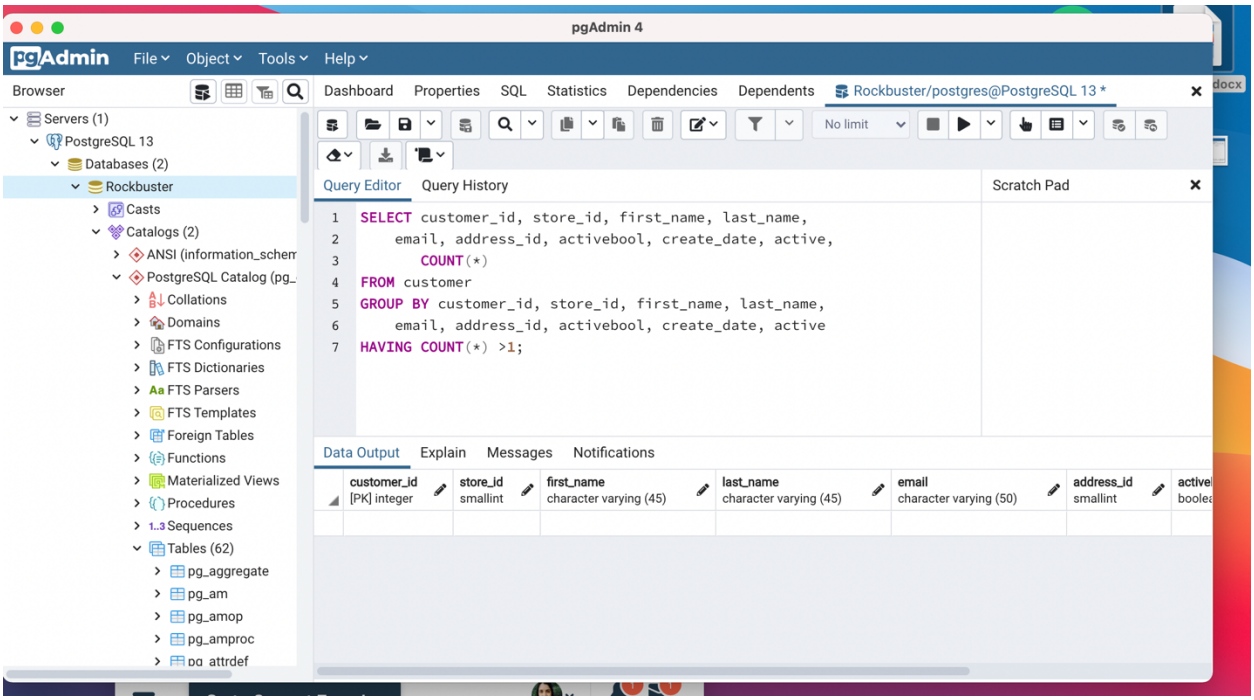
3.6 SUMMARIZING AND CLEANING DATA IN SQL TASK

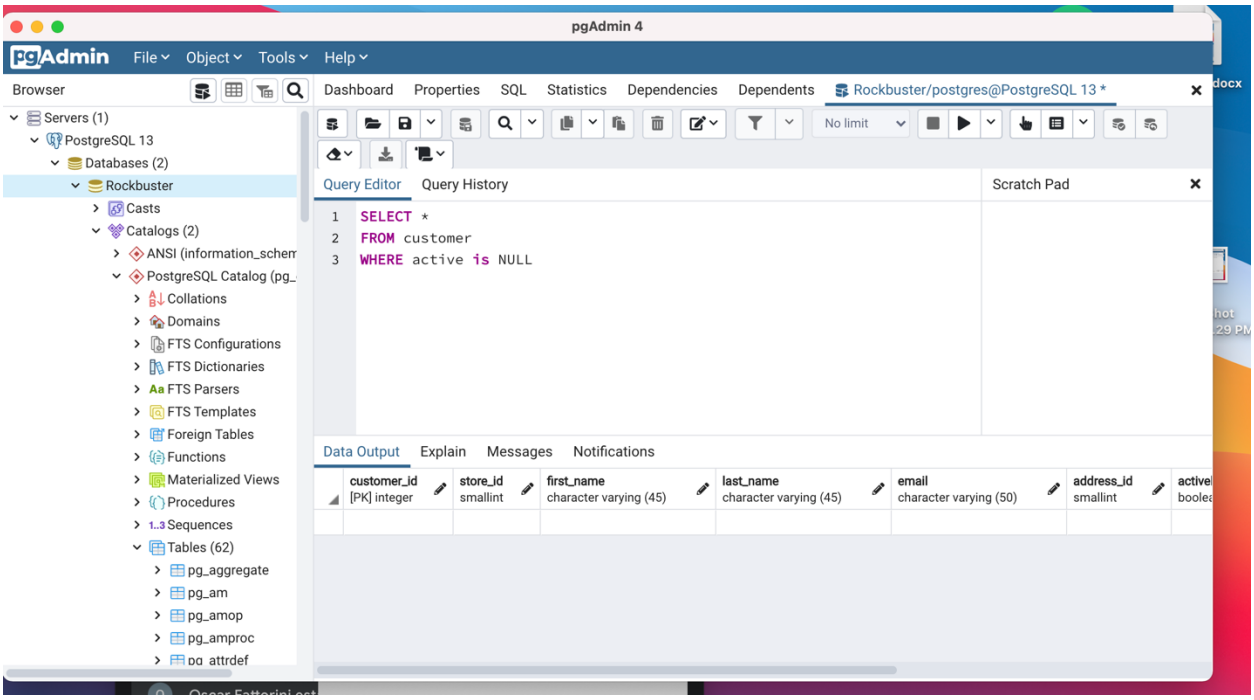
1. **Check for and clean dirty data:** Find out if the film table and the customer table contain any dirty data, specifically non-uniform or duplicate data, or missing values. Create a new "Answers 3.6" document and copy-paste your queries into it. Next to each query write 2 to 3 sentences explaining how you would clean the data (even if the data is not dirty).





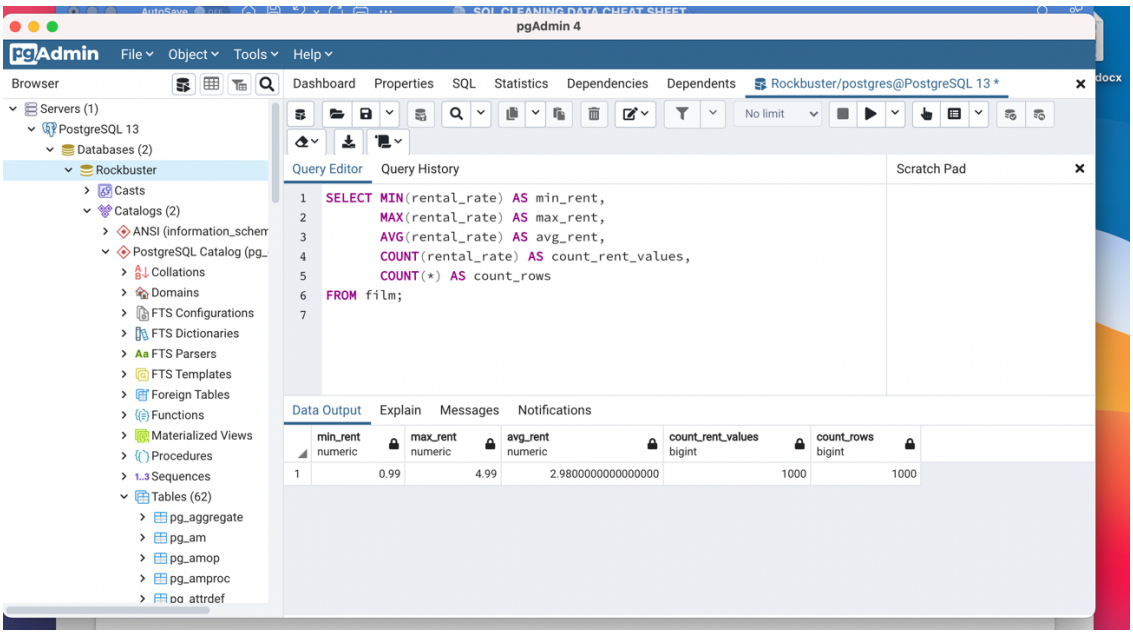
Found No duplicates in Customer



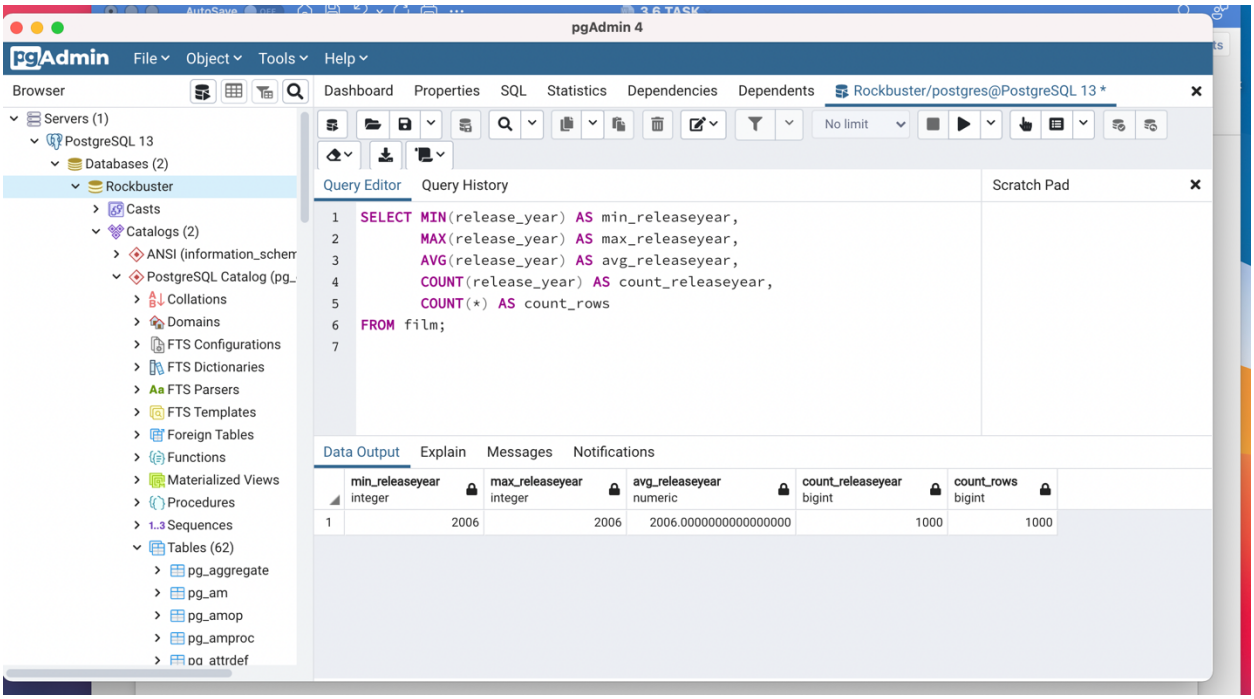


2. **Summarize your data:** Use SQL to calculate descriptive statistics for both the film table and the customer table. For numerical columns, this means finding the minimum, maximum, and average values. For non-numerical columns, calculate the mode value. Copy-paste your SQL queries and their outputs into your answers document.

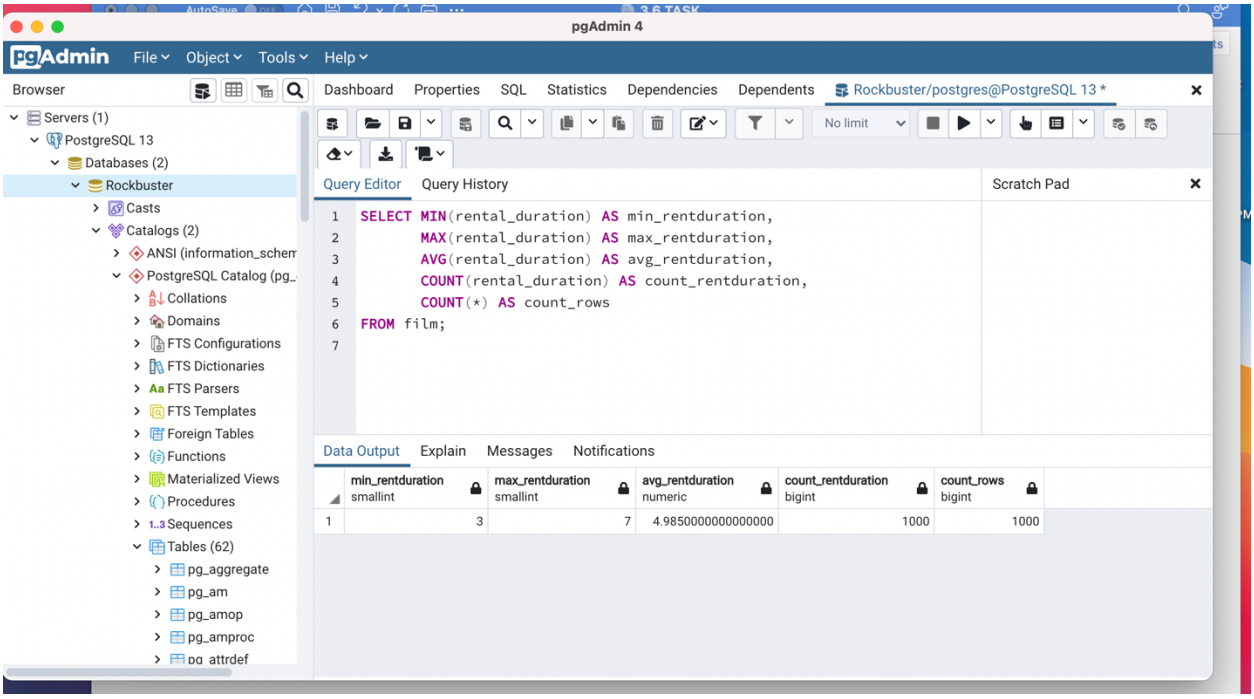
FILM TABLE



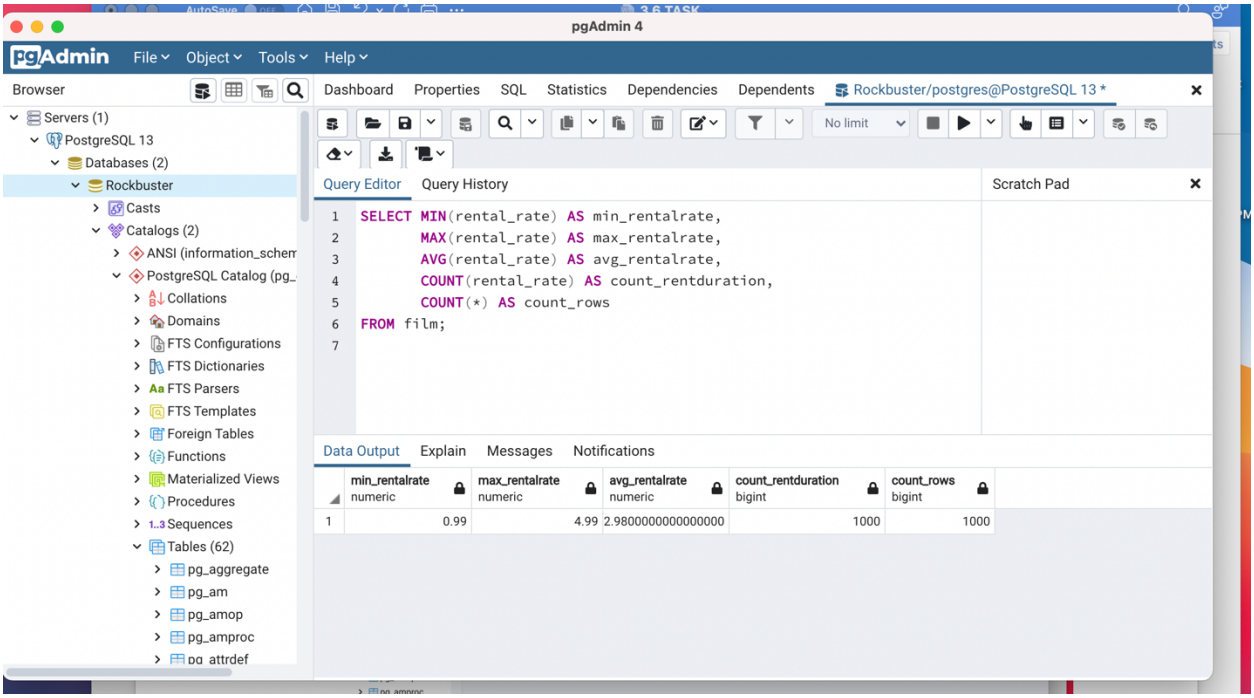
BY RELEASE YEAR



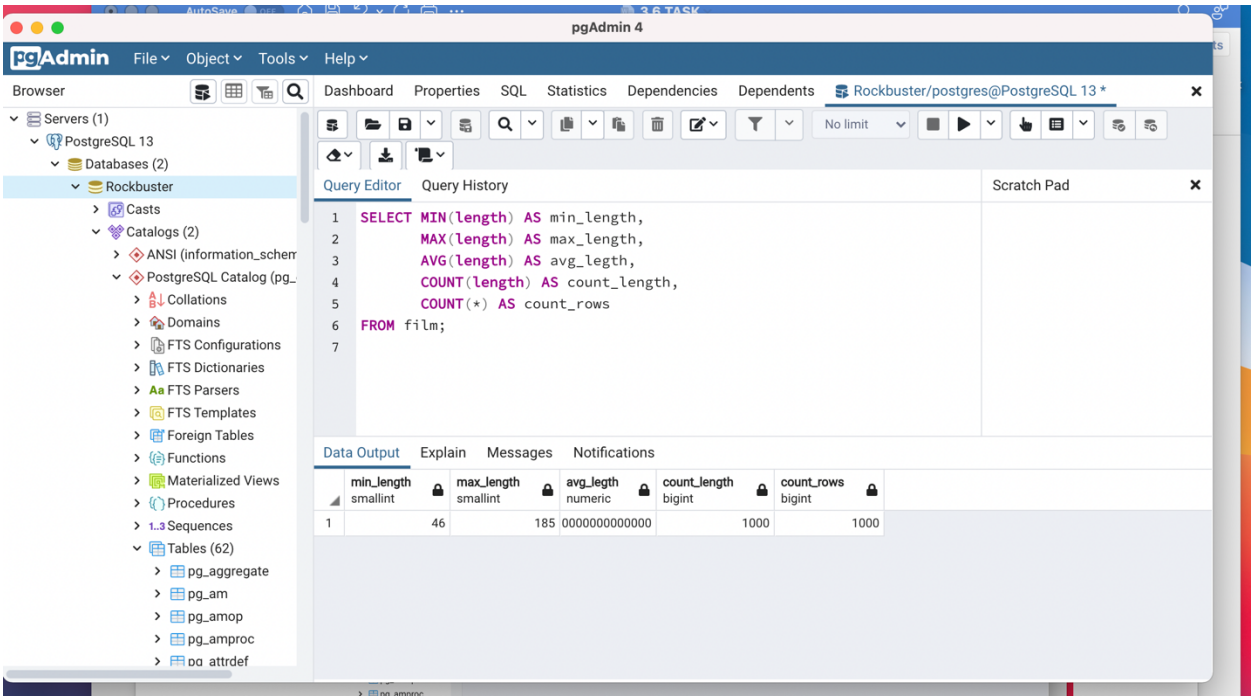
BY RENTAL DURATION



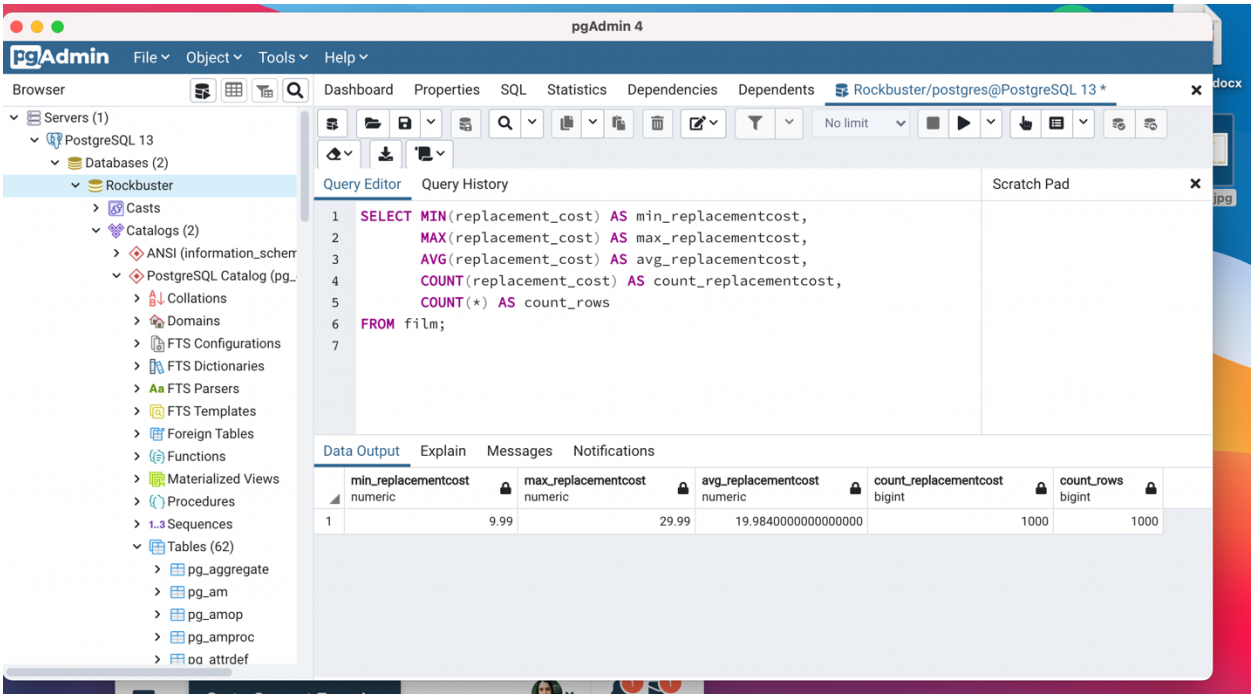
BY RENTAL RATE



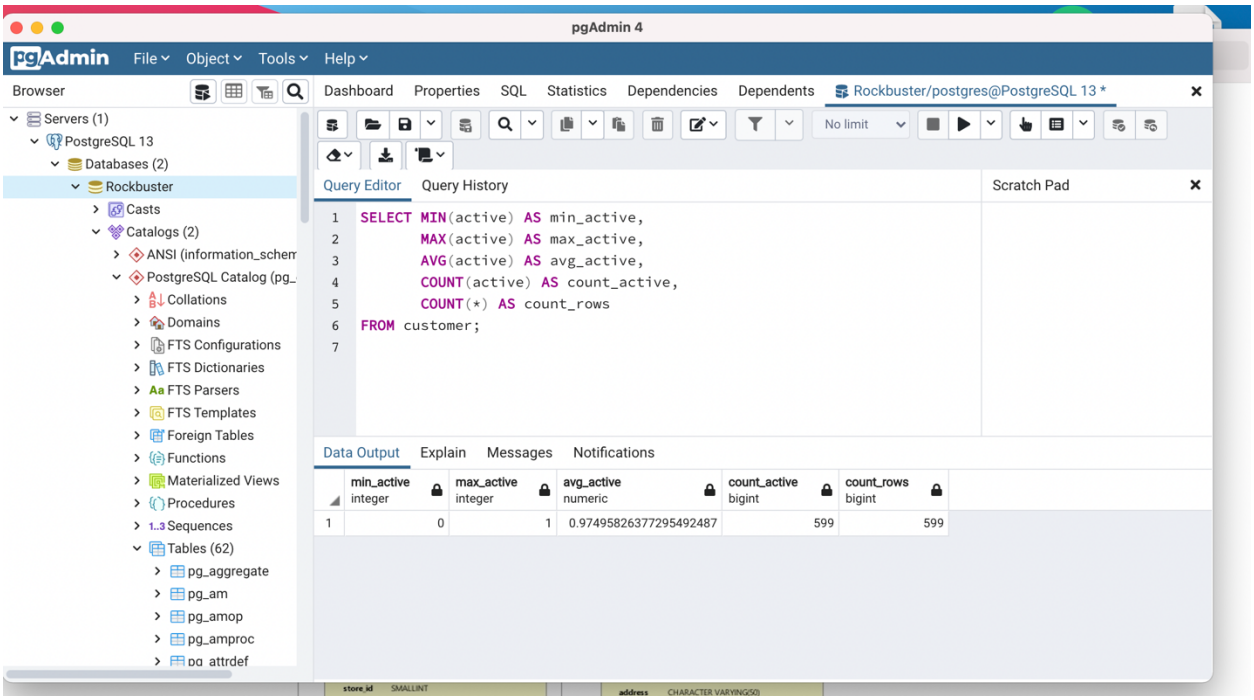
BY LENGTH



REPLACEMENT COST



CUSTOMER TABLE



3. **Reflect on your work:** Back in Achievement 1 you learned about data profiling in Excel. Based on your previous experience, which tool (Excel or

SQL) do you think is more effective for data profiling, and why? Consider their respective functions, ease of use, and speed. Write a short paragraph in the running document that you have started.

*Personally, I prefer using SQL because **is** easier, as it's more efficient and much faster than Excel. In addition, SQL gives you more options to analyze your data simultaneously.*