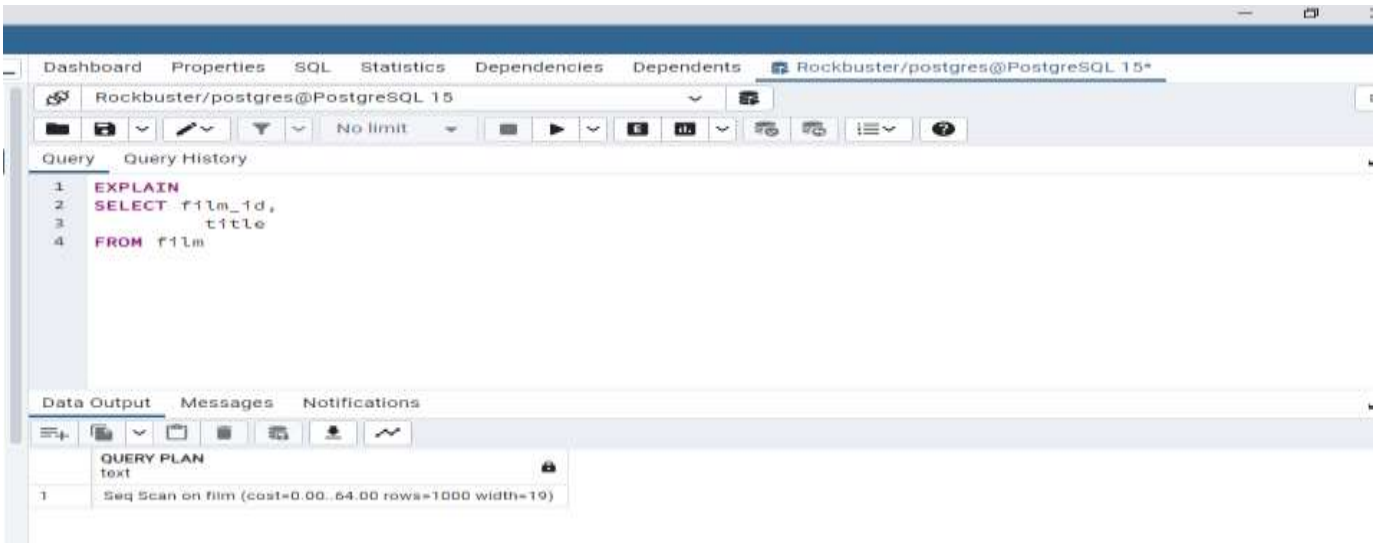
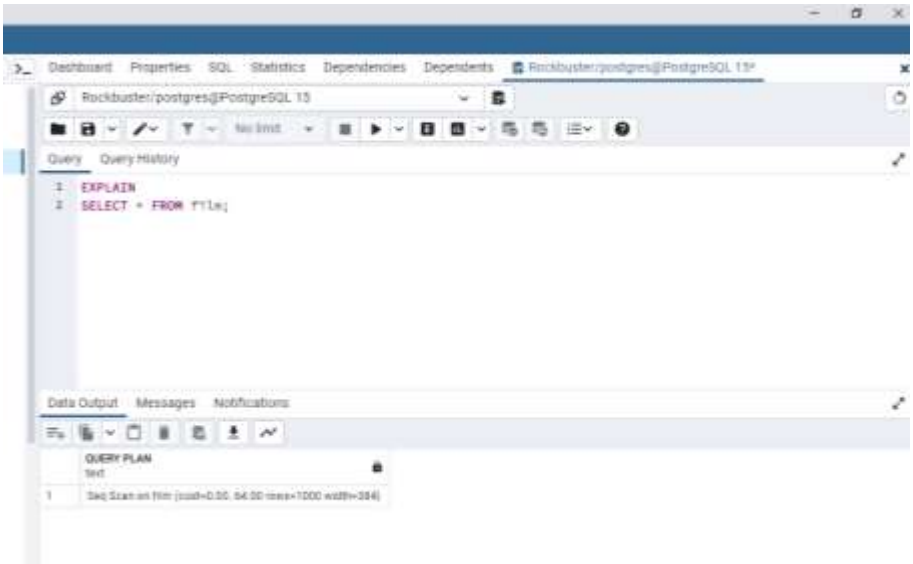
Refining Your Query: You need to get some data from the "film" table and decide to use the query SELECT * FROM film.

- You realize that only the "film_id" and "title" columns are needed. Write a new query that selects only those 2 columns.

Compare the cost of the original query and the revised query, and write a few sentences explaining the comparison. Can you suggest any ways to optimize this query?

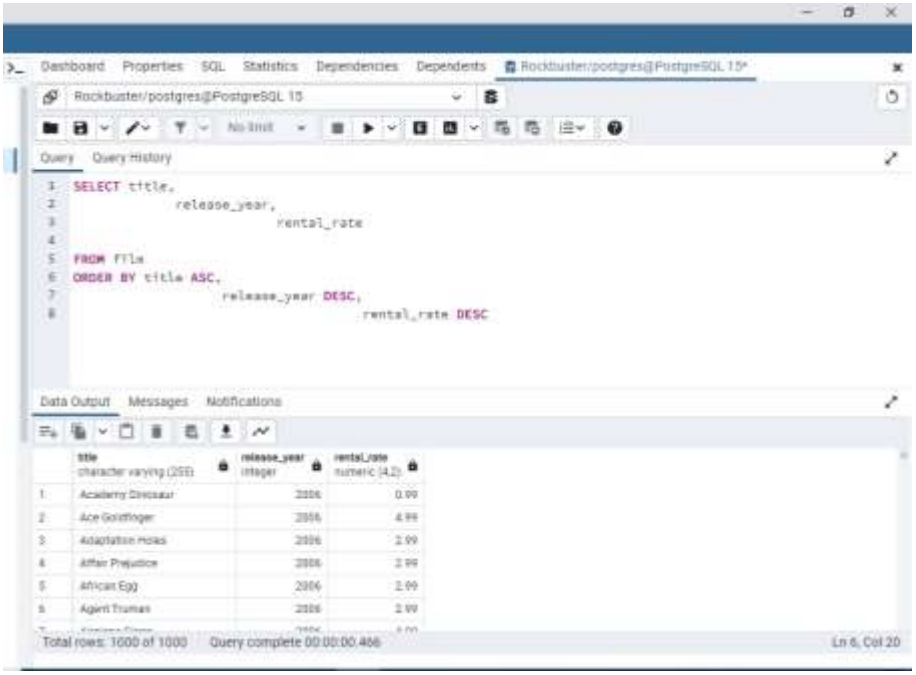




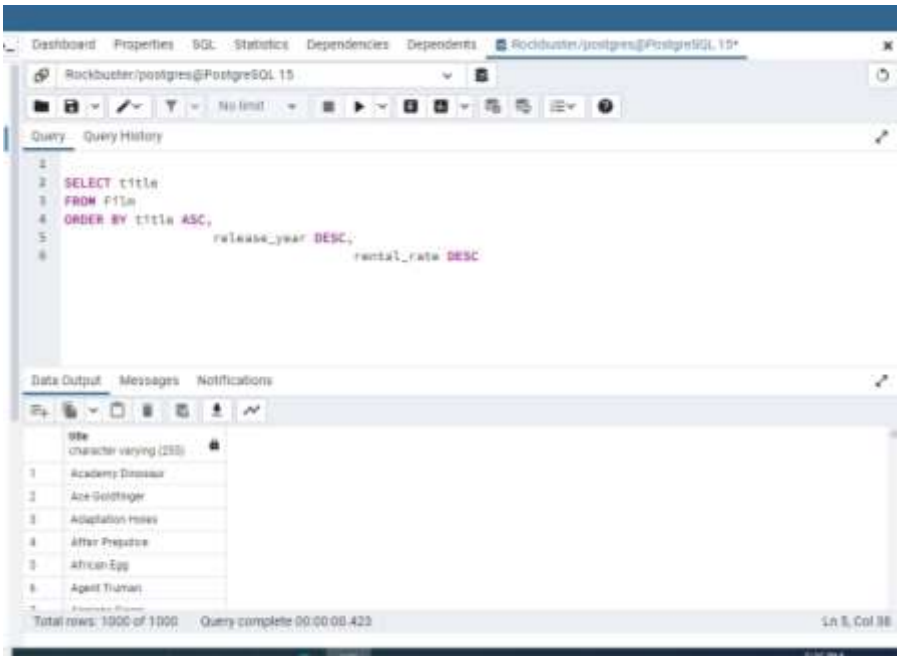
- The cost first query is more expensive than the second one. Because the first one takes longer time, checking all the 64 rows and having a width value 384. The second one only targets the intended columns, hence cheaper and faster in processing the result. Such kind of differences is more visible especially when we have a big data, overall, no need to go into all the records in a table – specific queries are efficient and cheaper.

Ordering the Data:

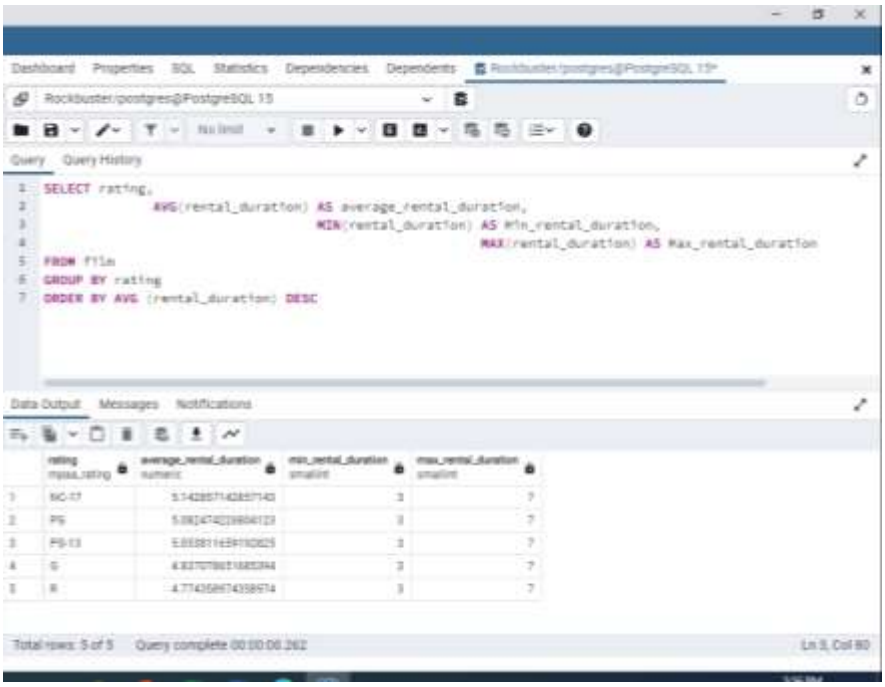
- In the pgAdmin Query Tool, run a query that selects every film from the “film” table, with the movies sorted by title from A to Z, then by most recent release year, and then by highest to lowest rental rate.
- Extract the data output of your query into a CSV file for the film collection department to analyze in Excel. To do this, click the button “Save results to file”:



- Without release year and rental rate



1. **Grouping Data:** The strategy department has asked you the questions below. Write a SQL query to retrieve the correct answers, then extract your results as a CSV file.
    - What is the average rental rate for each rating category?
    - What are the minimum and maximum rental durations for each rating category?



1. **Database Migration:** Your team has decided to use an external tool to collect data on user behavior in the new Rockbuster Android app. Data collected from this new source will need to be loaded into the data warehouse before you can analyze it.
    - Can you outline the procedure for migrating the data and who will be responsible for it?
    - What problems do you foresee if you start analyzing the data before it's been loaded into the data warehouse?

The data must be collected from Android APP data base, transformed (to a suitable format which will be needed in a data warehouse) and finally loaded to data warehouse. Process is owned by data engineers with additional support of data analysts for finding errors in migration process (pipelines).

There are risks of different format in Android APP Data base in compare to the data warehouse, which itself creates more workload like adding different variables in the script etc. In order to receive values that needed to be analyzed. This itself creates extra costs due to longer code and more functions in it which could be avoided if data would be already loaded into data warehouse