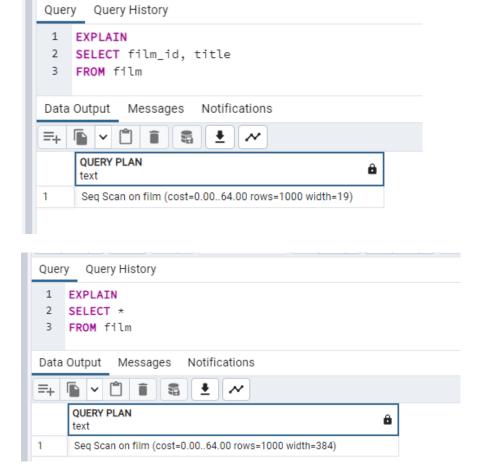
Task 3.4 – Database Querying in SQL

- 1. **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?

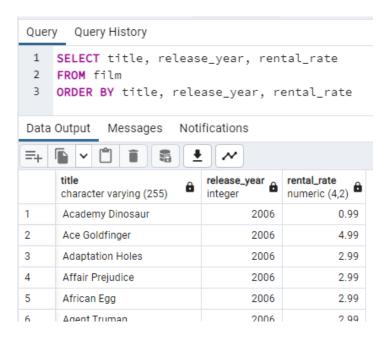
SELECT film_id, title FROM film



The 'cost' of the two queries is the same, but the widths are different. The best way to optimize the query is by using scripts to run multiple queries at once.

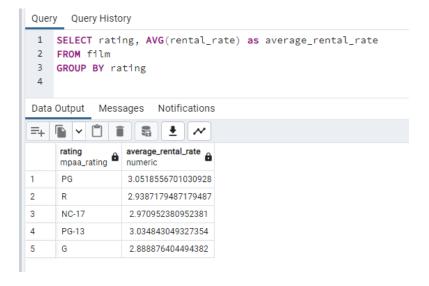
Ordering the Data:

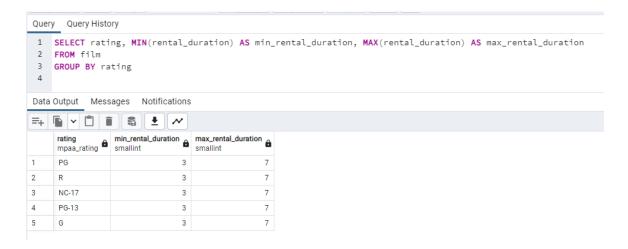
- In the pgAdmin Query Tool, run a query that selects every film from the "film" table, with the movies sorted by title from A to Z, then by most recent release year, and then by highest to lowest rental rate.
- Extract the data output of your query into a csv file for the film collection department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



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?





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?

Typically, data migration is done by data engineers via a procedure called ETL, or Extract, Transform, and Load.

- Extract The data is collected from multiple data sources.
- Transform The extracted data is converted into another format to make things uniform. For example, calculating ages from dates, or combining data points to create contact numbers.
- Load The transformed data is loaded into a new database.

The biggest problem that I can see with data being analyzed before being uploaded into a data warehouse is unclean, mismatching data and data formats. This could lead to inaccurate analyses, as well as increasing the amount of time and cost needed to do such analyses.