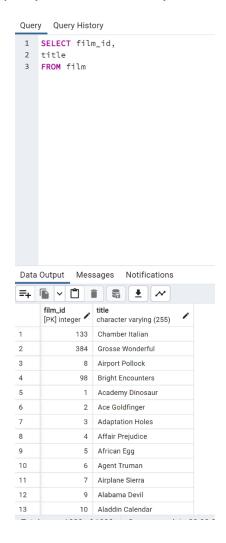
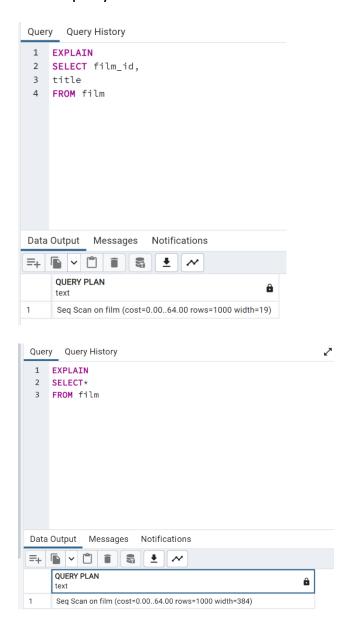
Answers 3.4.

- 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.



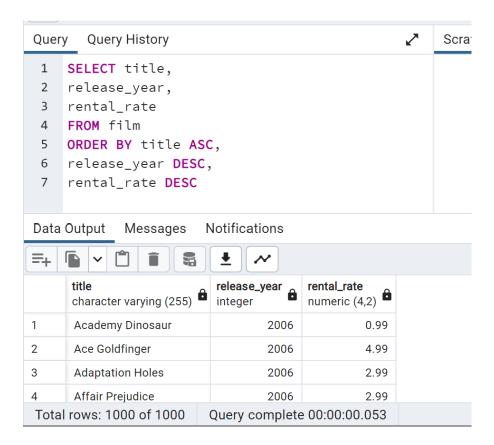
 Compare the cost of the original query and the revised query, and write a few sentences explaining the comparison. Can you suggest any ways to optimize this query?



Both have the same cost, as we can see but their runtimes are different. If we create a script we could save costs.

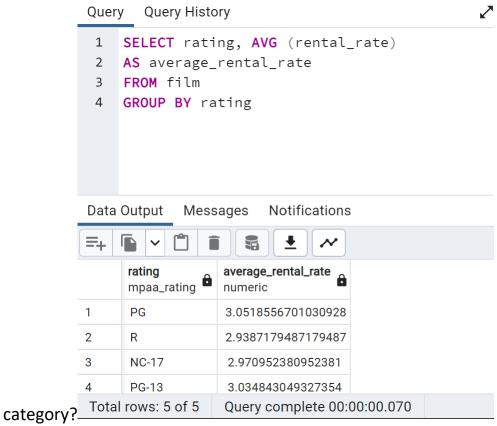
2. Ordering the Data:

- 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":



- 3. **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.
- o What is the average rental rate for each rating category?

What are the minimum and maximum rental durations for each rating



Query Query History Scratch SELECT rating, MAX (rental_rate) AS maximum_rental_rate, 4 MIN (rental_rate) 5 AS minimum_rental_rate 6 FROM film **GROUP BY** rating Data Output Messages Notifications =+ minimum_rental_rate maximum_rental_rate rating mpaa_rating numeric numeric PG 4.99 0.99 1 2 R 4.99 0.99 NC-17 3 4.99 0.99 PG-13 4.99 0.99 Total rows: 5 of 5 Query complete 00:00:00.038

- 4. **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.
 - 4.1 Can you outline the procedure for migrating the data and who will be responsible for it?

The procedure is called **Extract**, **Transform**, and **Load** (ETL).

"The ETL process is:

- Extract: The first step involves collecting the data from multiple data sources.
- Transform: During this step, the extracted data is converted into another format. This could mean calculating ages from dates of birth or combining multiple data points like area codes and telephone numbers to get a contact number, for example.
- Load: At this point, the transformed data is inserted or loaded into the new database. "
 - 4.1 What problems do you foresee if you start analyzing the data before it's been loaded into the data warehouse?

There might be a problem with incomplete data format, including irrelevant and unclear data. That will result in an inaccurate analysis, time, and cost consumption for the analysis.