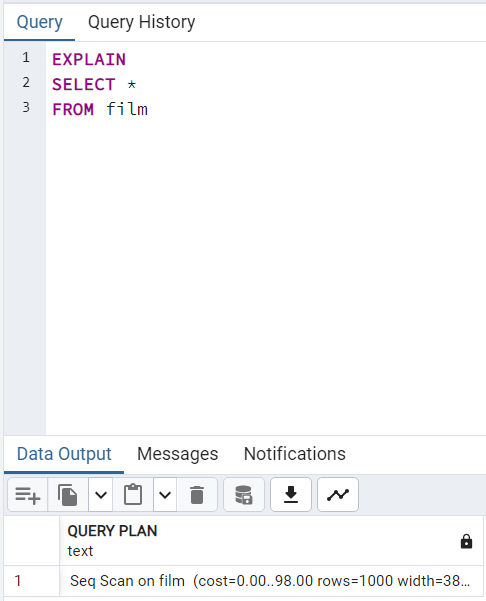
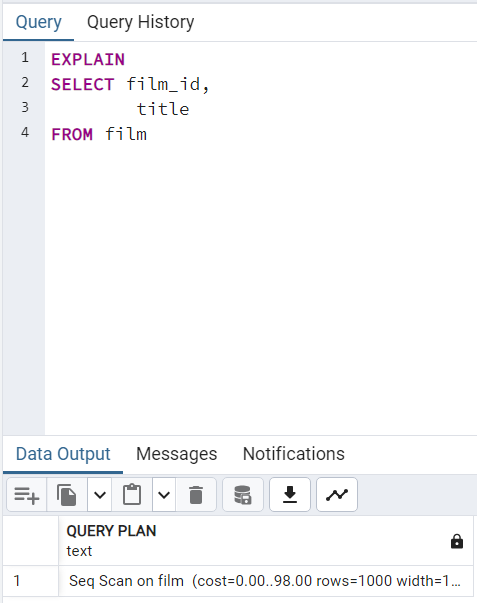
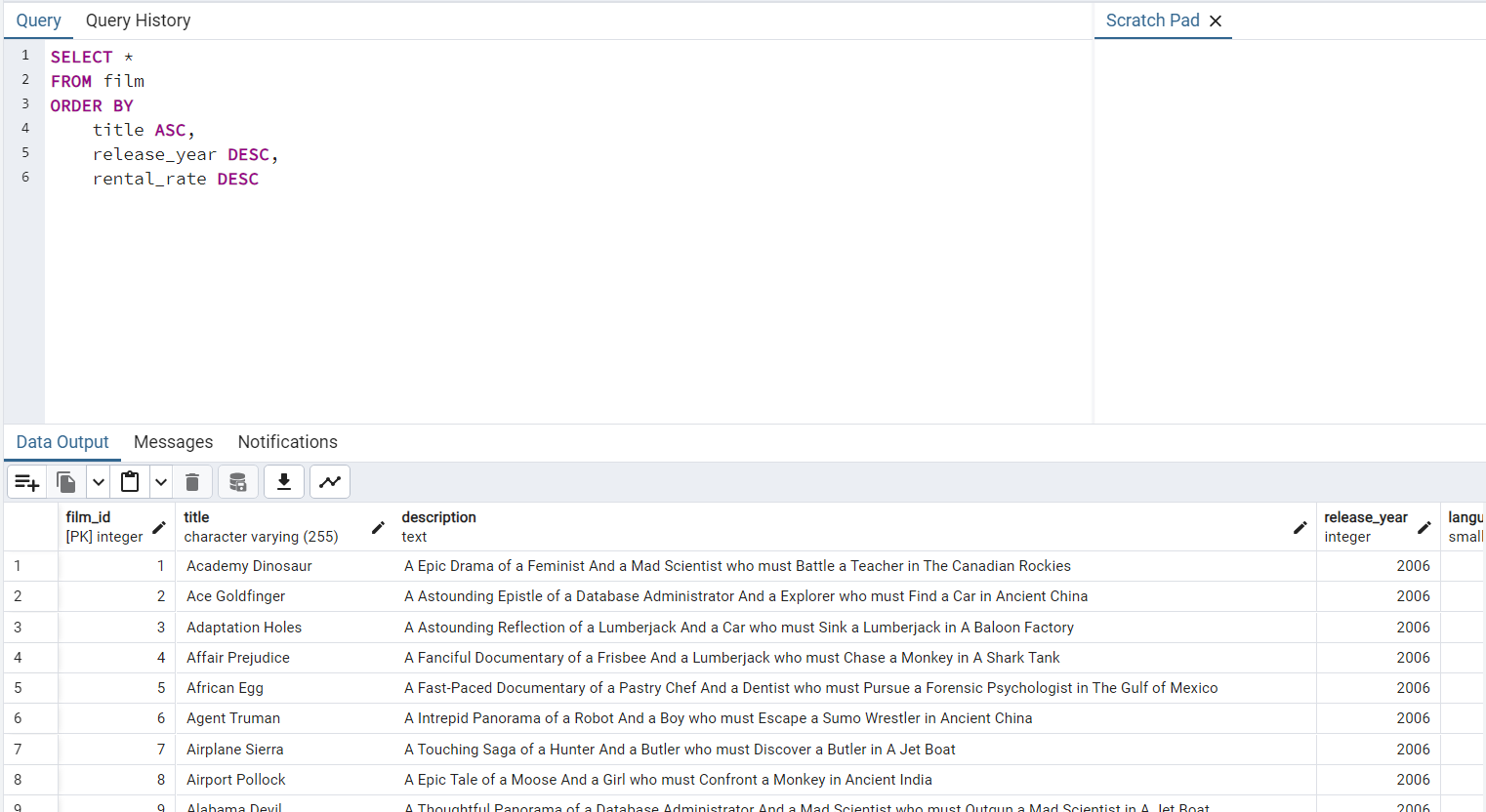
**Exercise 3.4**

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.**

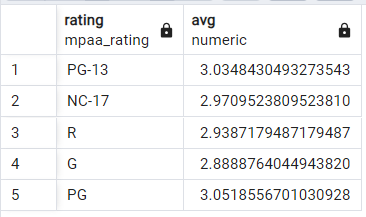
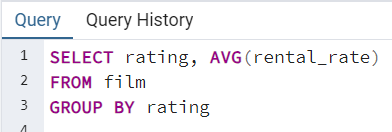
 

* + **Compare the cost of the original query and the revised query and write a few sentences explaining the comparison. Can you suggest any ways to optimize this query?**
    - The cost of the original and revised query are the same. It might be possible to optimize the query by decreasing the amount of data the query needs to go through, since in both cases here the query still must go through the entire table to return the desired result.

1. **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.**



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?**

A screenshot of a computer

Description automatically generated

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?**

Procedure:

In order to migrate the data, the ETL (extract, transform, and load) process would be followed:

1. Extract: collect all necessary and available data (data engineers or whoever is in charge of the app data)
2. Transform: convert all data into the format of the data warehouse so it can be imported correctly. (data engineers or data analysts)
3. Load: load the data into the data warehouse (data engineers)
   * **What problems do you foresee if you start analyzing the data before it’s been loaded into the data warehouse?**
     + If you start trying to analyze the data before it’s loaded into the warehouse, you are working with incomplete and raw data. As such, you would likely encounter a plethora of problems such as missing values, inconsistency, duplicates, and confusing structure. The data needs to go through a cleaning process to both condense and organize it into a format that is easier to analyze and without errors so that the resulting analysis is as accurate as possible.