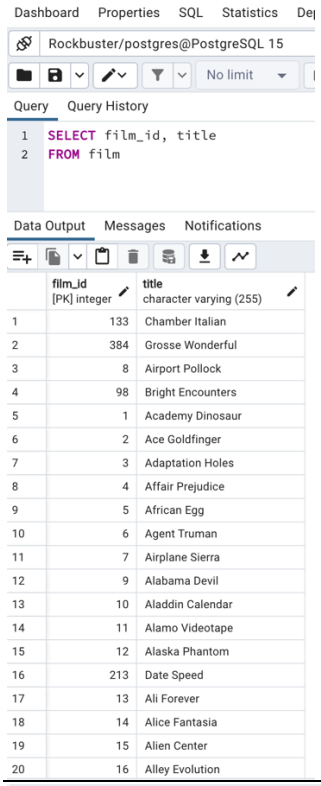
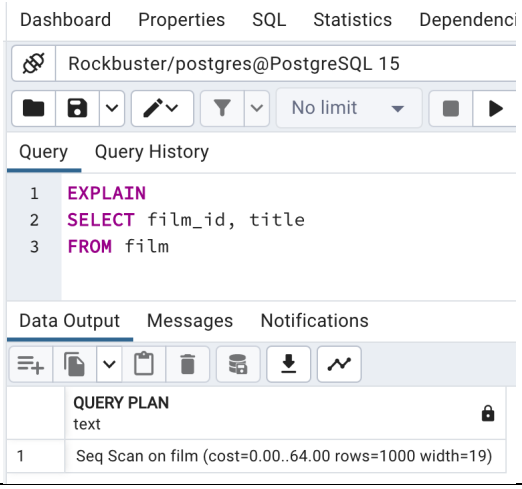
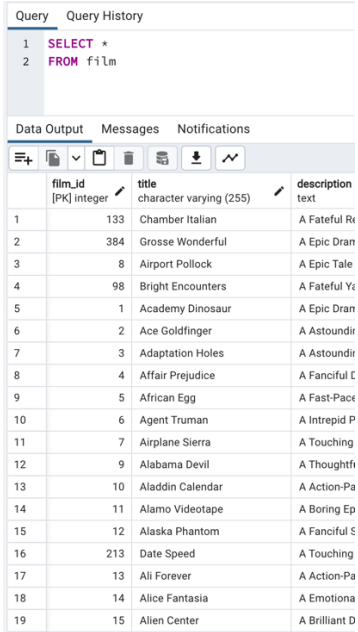
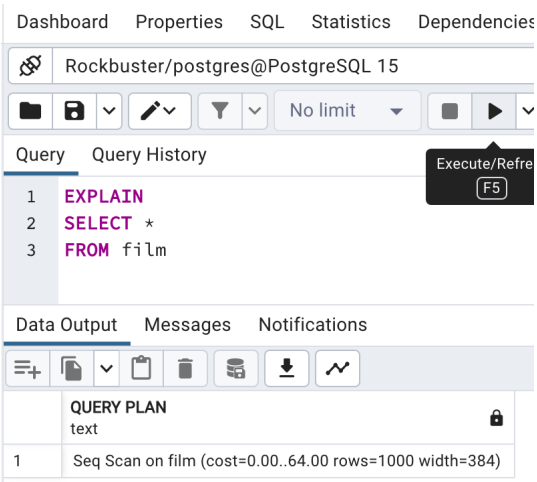
Answers Task 3.4: Database Querying in SQL

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.





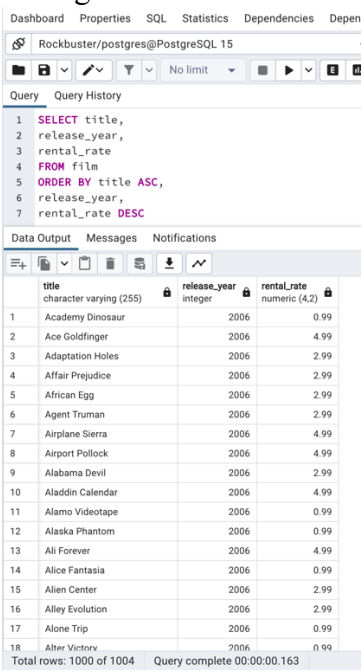


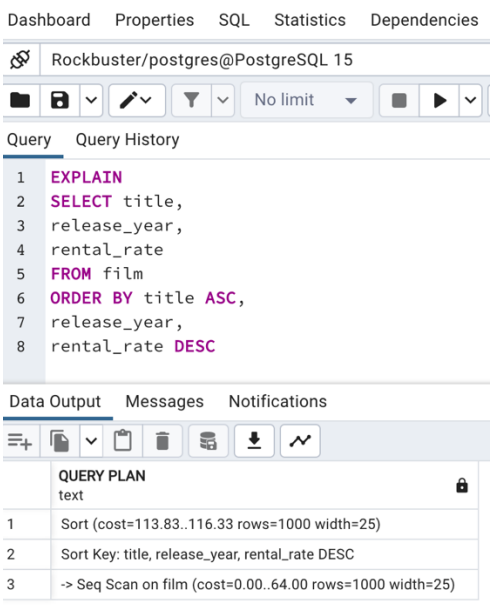


- Compare the cost of the original query and the revised query, and write a few sentences explaining the comparison. Can you suggest any ways to optimize this query?

  Data shows that the cost for both queries is 64, but we see a difference on width, the first one resulting on 19 and the second one resulting on 384. This could be optimized by adding the LIMIT query.

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.

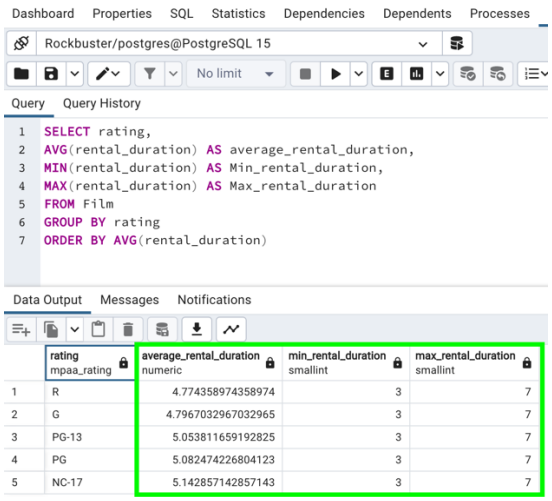




- Extract the data output of your query into a CSV file for the film collection department to analyze in Excel. To do this, click the button "Save results to file":

DONE

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.
   - What is the average rental rate for each rating category?
   - What are the minimum and maximum rental durations for each rating category?



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.

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

     The correct procedure to follow during a data migration process would be applying the method of Extraction, Transformation and Loading, better known as ETL. During extraction, this process involves collecting data from multiple sources. Followed by transform, which is when data that was extracted is converted into another format. Lastly, the transformed data in loaded into the new database.

   - What problems do you foresee if you start analyzing the data before it's been loaded into the data warehouse?

     If data isn't loaded correctly, it won't be consistent and hard to follow. This would make it difficult to analyze and retrieve information.

5. Save your "Answers 3.4" document as a pdf (with screenshots) and your CSV files as a single .xlsx Excel file and upload it here for your tutor to review.