Answers 3.4

Directions

As you've done for previous tasks, create a new text document for your answers and call it “Answers 3.4.”

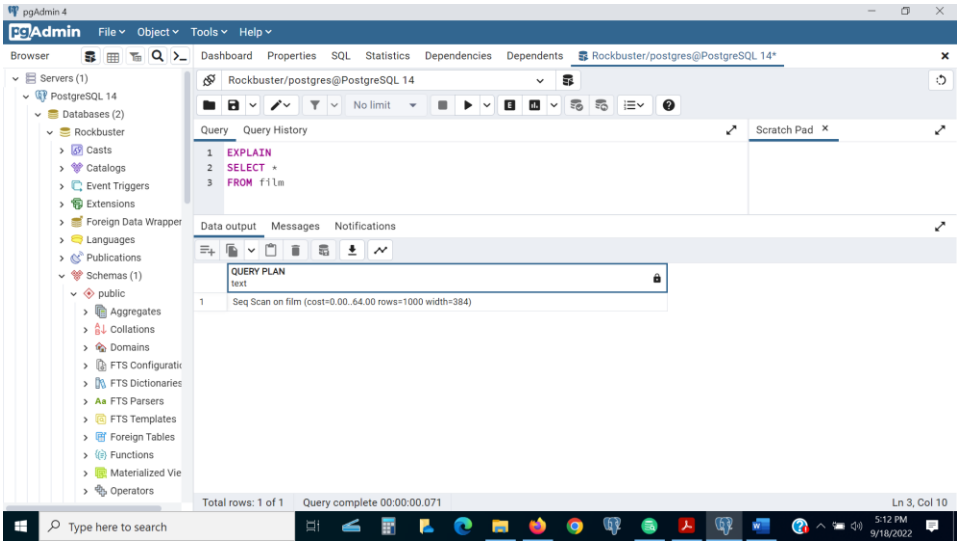
1. **Refining Your Query:** You need to get some data from the “film” table and decide to use the query SELECT * FROM film.

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

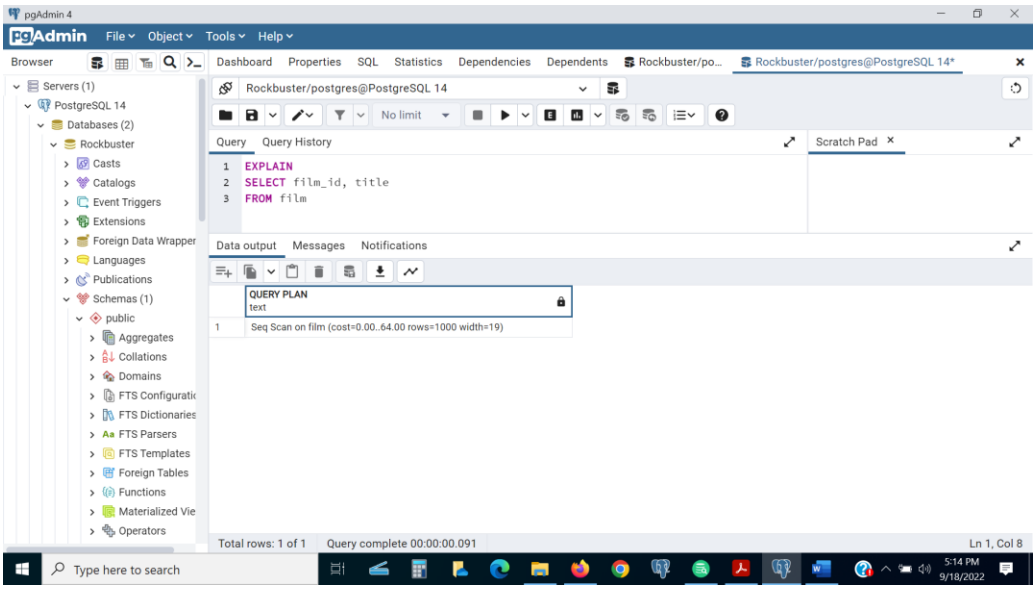
   SELECT film_id, title FROM film

   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 screenshots per EXPLAIN show that the original query and revised query costs are the same (cost=0.0..64.00). Ways to optimize query would be create and run a script.

   Cost of original query:

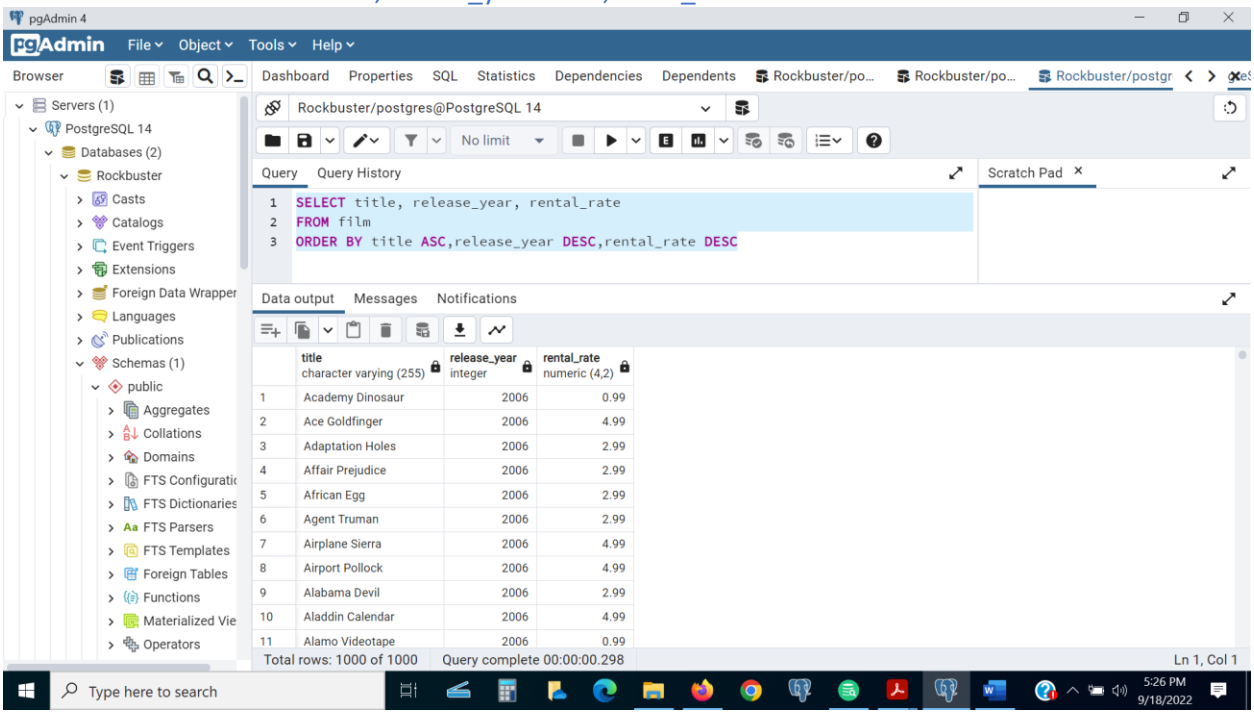


   Cost of revised query:

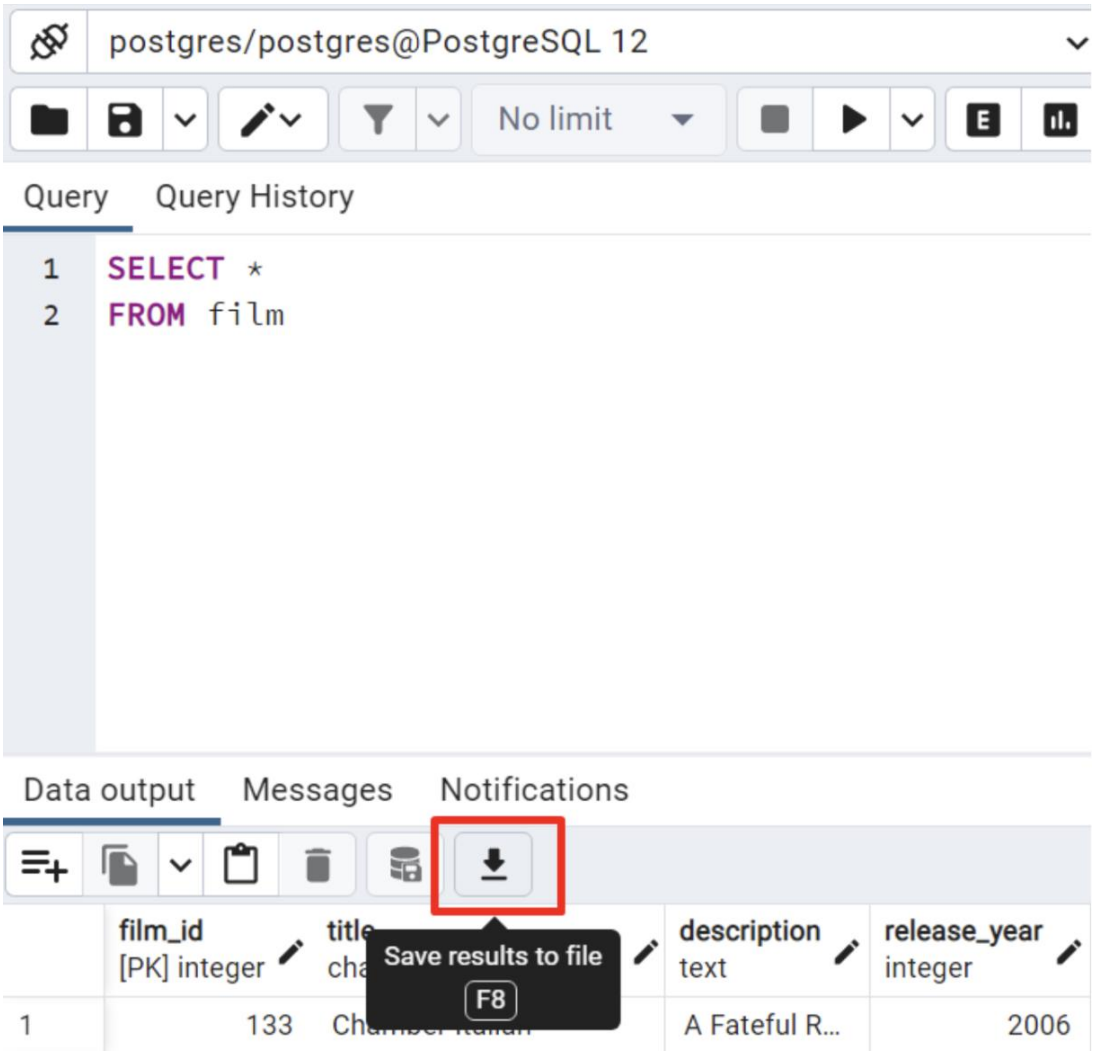


2. **Ordering the Data:**
   - In the pgAdmin Query Tool, run a query that selects every film from the "film" table, with the movies sorted by title from A to Z, then by most recent release year, and then by highest to lowest rental rate.

```
SELECT title, release_year, rental_rate
FROM film
ORDER BY title ASC,release_year DESC,rental_rate DESC
```

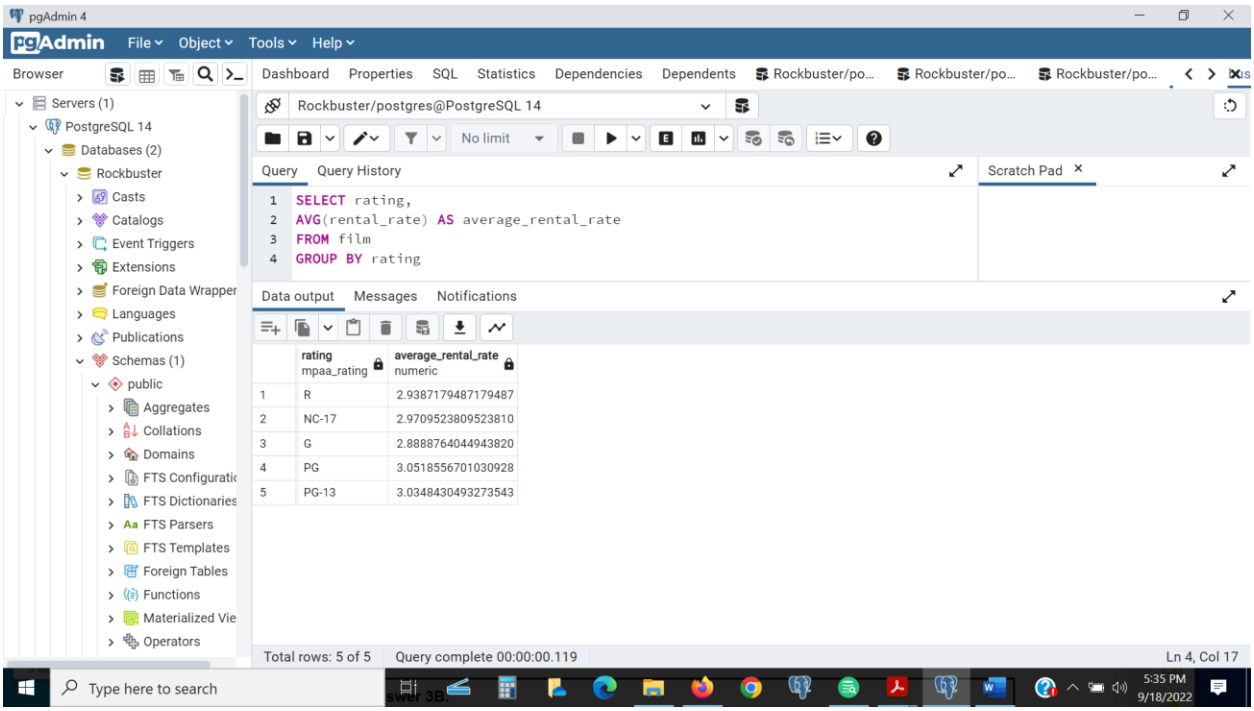


Extract the data output of your query into a csv file for the film collection department to analyze in Excel. To do this, click the button "Save results to file": see file q2 ordering data.csv

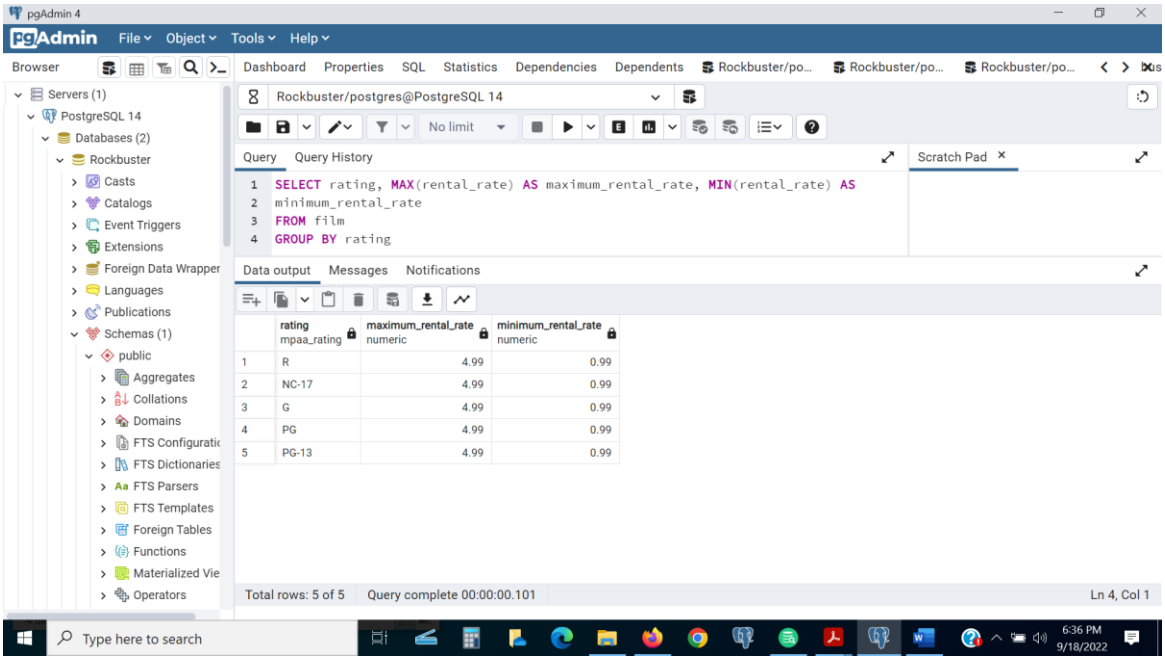


3. **Grouping Data:** The strategy department has asked you the questions below. Write a SQL query to retrieve the correct answers, then extract your results as a csv file.

    - What is the average rental rate for each rating category? See screen shot below and q2 grouping data.csv



- What are the minimum and maximum rental durations for each rating category? See screen shot below and q3 Max Min.csv file



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?
    - Procedure for migrating data is done using ETL procedures (Extract, Transform, Load). Data engineers are typically responsible for this procedure.
        1. Extract – first step involves collecting data from multiple data sources
        2. Transform - second step, extracted data is converted into another format.
        3. Load – final step, transformed data is inserted or loaded into new database.
- What problems do you foresee if you start analyzing the data before it's been loaded into the data warehouse?
    - Problems that I foresee, if start to analyze data before loaded into the data warehouse could be formatting issues and misinterpretation of the data for it has not been loaded into proper tables.

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.