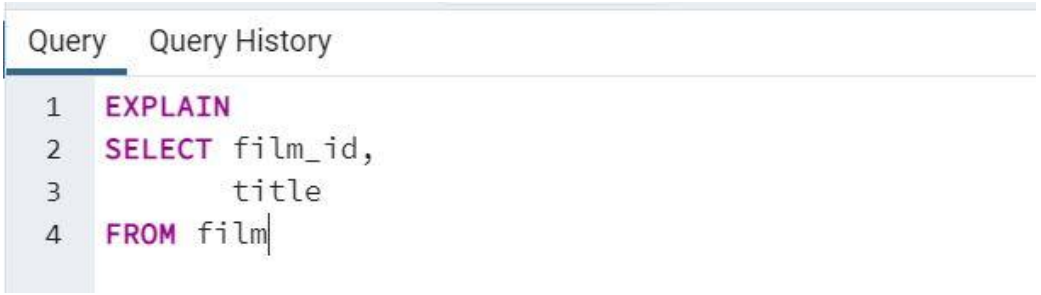
**1) Refining Your Query:**

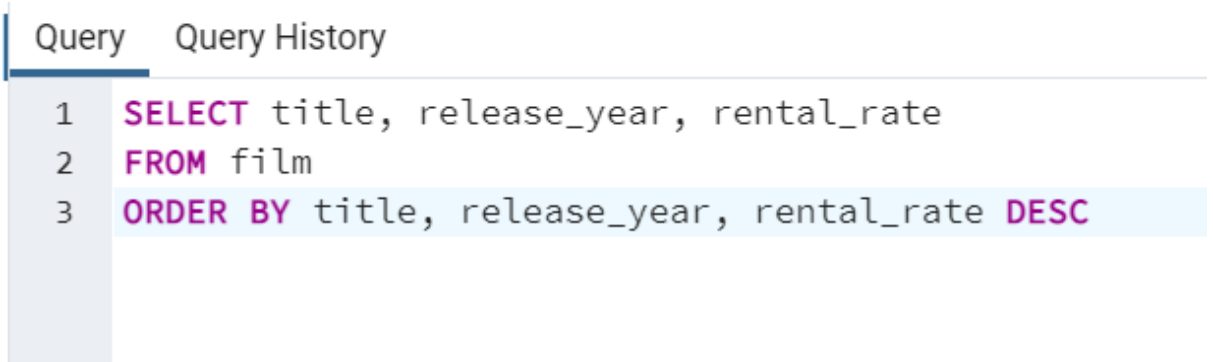
You realize that only the “film_id” and “title” columns are needed.  Write a new query that selects only those 2 columns




- 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 the queries (EXPLAIN SELECT film_id,  title FROM film and EXPLAIN SELECT * FROM film) have the same cost: "Seq Scan on film  (cost=0.00..64.00 rows=1000 width=19)"
    - If we use the actual column names in the SELECT statement rather than '*' and if we separate all words with a single space, SQL queries should typically be faster, but not in this case. We may also use "LIMIT" and "GROUP BY" to narrow down the number of results.

**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.



**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?

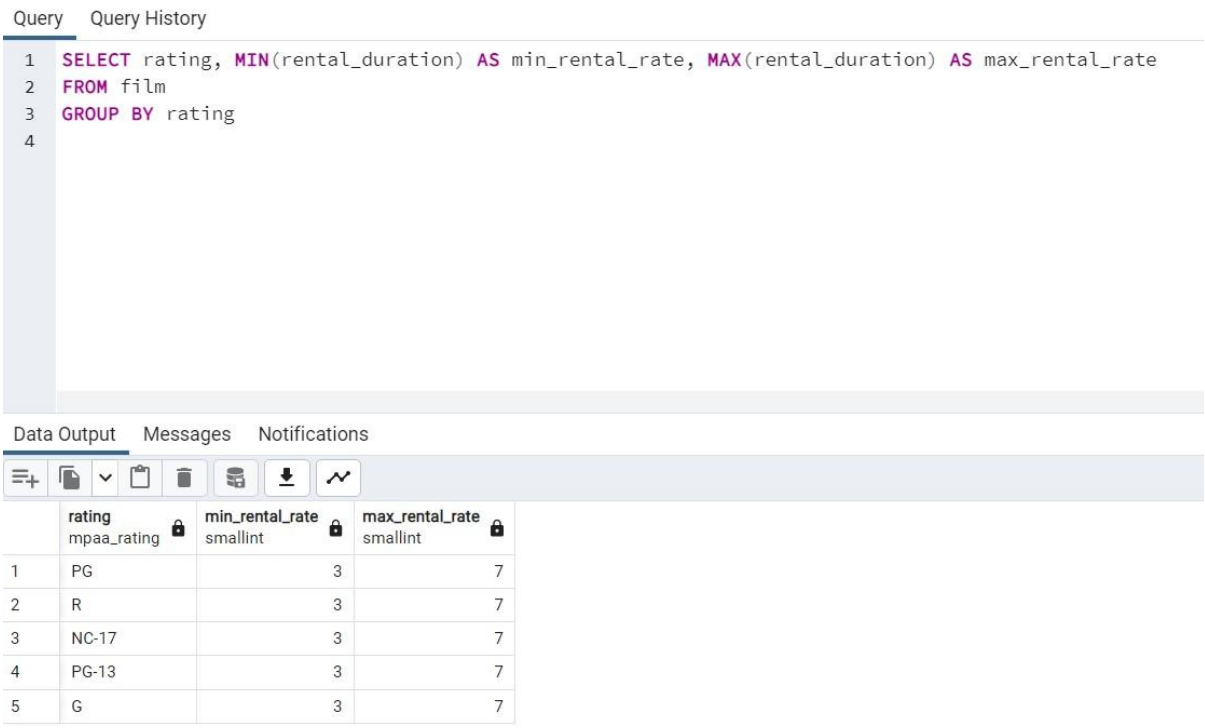
Query Query History

```
1  SELECT rating, AVG(rental_rate)  AS average_rental_duration
2  FROM film
3  GROUP BY rating
```

Data Output Messages Notifications

| | rating mpaa_rating | average_rental_duration numeric |
|---|---|---|
| 1 | PG | 3.0518556701030928 |
| 2 | R | 2.9387179487179487 |
| 3 | NC-17 | 2.9709523809523810 |
| 4 | PG-13 | 3.0348430493273543 |
| 5 | G | 2.8888764044943820 |

- What are the minimum and maximum rental durations for each rating category?



Query | Query History

```
SELECT rating, MIN(rental_duration) AS min_rental_rate, MAX(rental_duration) AS max_rental_rate
FROM film
GROUP BY rating
```

Data Output | Messages | Notifications

| | rating mpaa_rating | min_rental_rate smallint | max_rental_rate smallint |
|---|---|---|---|
| 1 | PG | 3 | 7 |
| 2 | R | 3 | 7 |
| 3 | NC-17 | 3 | 7 |
| 4 | PG-13 | 3 | 7 |
| 5 | G | 3 | 7 |

**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?
    i. Data migration is the responsibility of a data engineer, but data analysts could be involved in some stages of the process. Data should be taken from the Android app, converted into another format, and then loaded into a new database.
- What problems do you foresee if you start analyzing the data before it's been loaded into the data warehouse?

i. If we started analyzing the data before it's been loaded into the data warehouse we could find data incomplete or even in formats we are unable to open, as during the process the various formats and types are transformed in order to adhere to one consistent system. Plus we will waste a lot of time and be less productive as ETL codifies and reuses data without a need for technical skills.