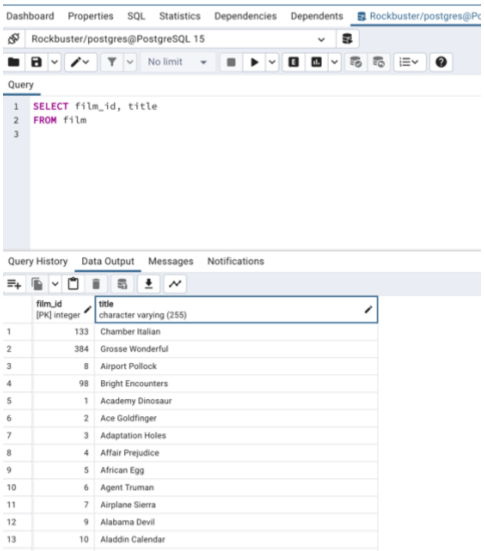
3.4- Database Querying in SQL

Answers 3.4.

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

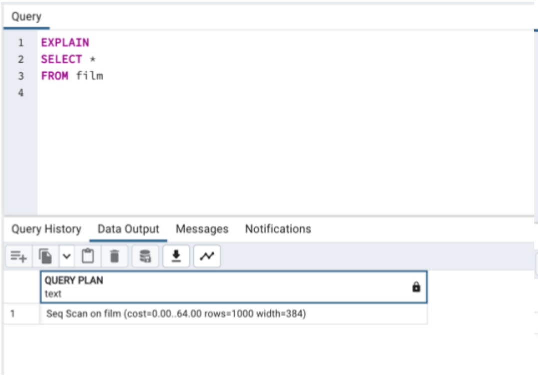
   ○ You realize that only the "film_id" and "title" columns are needed. Write a new query that selects only those two columns.



```
SELECT film_id, title
FROM film
```

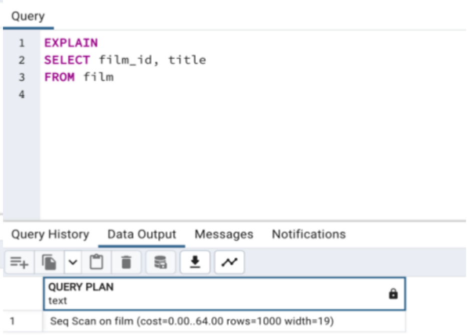
	film_id [PK] integer	title character varying (255)
1	133	Chamber Italian
2	384	Grosse Wonderful
3	8	Airport Pollock
4	98	Bright Encounters
5	1	Academy Dinosaur
6	2	Ace Goldfinger
7	3	Adaptation Holes
8	4	Affair Prejudice
9	5	African Egg
10	6	Agent Truman
11	7	Airplane Sierra
12	9	Alabama Devil
13	10	Aladdin Calendar

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



```
EXPLAIN
SELECT *
FROM film
```

	QUERY PLAN text
1	Seq Scan on film (cost=0.00..64.00 rows=1000 width=384)



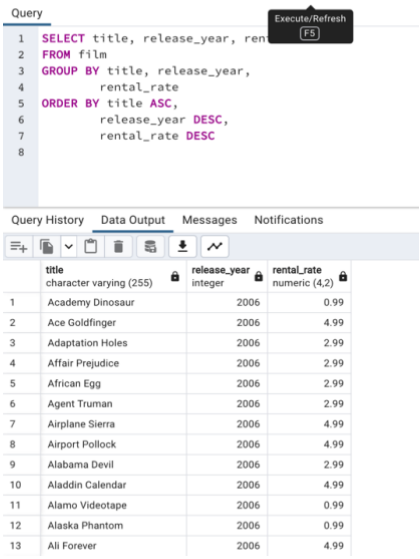
```
EXPLAIN
SELECT film_id, title
FROM film
```

	QUERY PLAN text
1	Seq Scan on film (cost=0.00..64.00 rows=1000 width=19)

The costs are the same in both queries (cost=0.00..64.00). The important thing about this example is that creating a more specific query can save time and costs when selecting the desired information at once. That is why to optimize it as much as possible before execution. After all, faster scripts mean lower prices.

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, ren
FROM film
GROUP BY title, release_year,
         rental_rate
ORDER BY title ASC,
         release_year DESC,
         rental_rate DESC
```

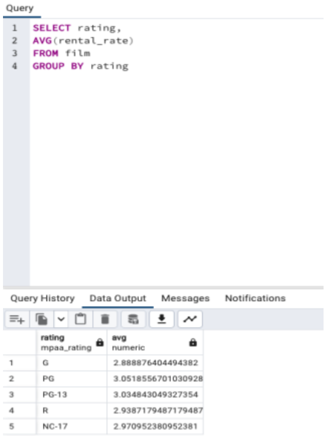
	title character varying (255)	release_year integer	rental_rate numeric (4,2)
1	Academy Dinosaur	2006	0.99
2	Ace Goldfinger	2006	4.99
3	Adaptation Holes	2006	2.99
4	Affair Prejudice	2006	2.99
5	African Egg	2006	2.99
6	Agent Truman	2006	2.99
7	Airplane Sierra	2006	4.99
8	Airport Pollock	2006	4.99
9	Alabama Devil	2006	2.99
10	Aladdin Calendar	2006	4.99
11	Alamo Videotape	2006	0.99
12	Alaska Phantom	2006	0.99
13	Ali Forever	2006	4.99

- **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.**

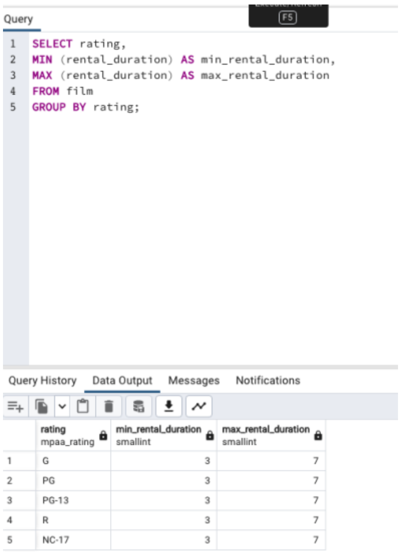
   ○ **What is the average rental rate for each rating category?**



```
SELECT rating,
AVG(rental_rate)
FROM film
GROUP BY rating
```

	rating mpaa_rating	avg numeric
1	G	2.888876404494382
2	PG	3.0518556701030928
3	PG-13	3.034843049327354
4	R	2.9387179487179487
5	NC-17	2.970952380952381

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



```sql
SELECT rating,
MIN (rental_duration) AS min_rental_duration,
MAX (rental_duration) AS max_rental_duration
FROM film
GROUP BY rating;
```

	rating mpaa_rating	min_rental_duration smallint	max_rental_duration smallint
1	G	3	7
2	PG	3	7
3	PG-13	3	7
4	R	3	7
5	NC-17	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?**

The migration will be done via ETL (Extract, Transform, Load). Data engineers carry this out.

The steps; Extract: This is the first step, and it involves the collection of data from various data sources Transform: In 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: In this step, the transformed data is inserted or 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?**

Getting data from different sources could always be problematic. They might be a problem with the data, such as formatting issues or unrelated data, to mention a few.