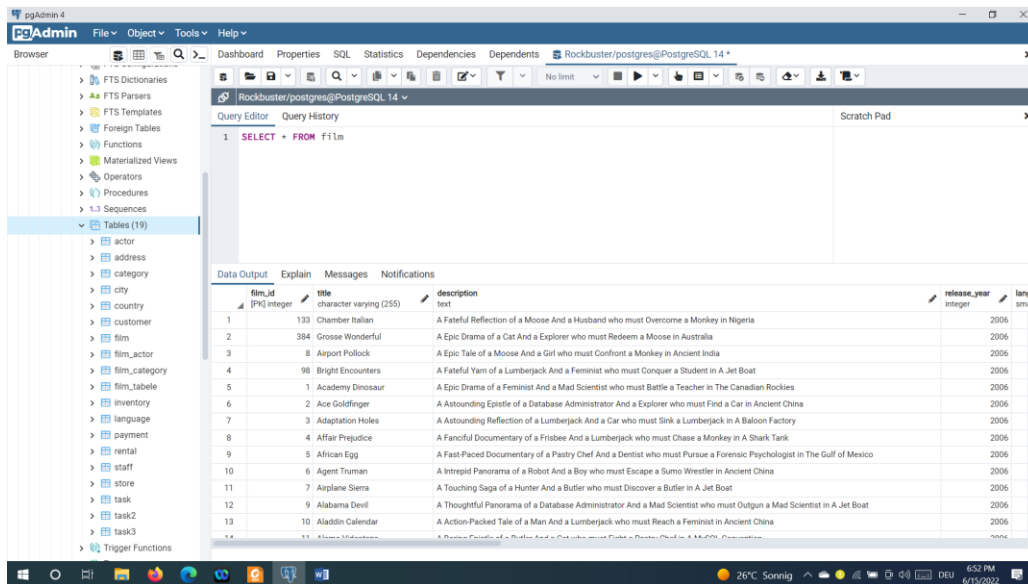
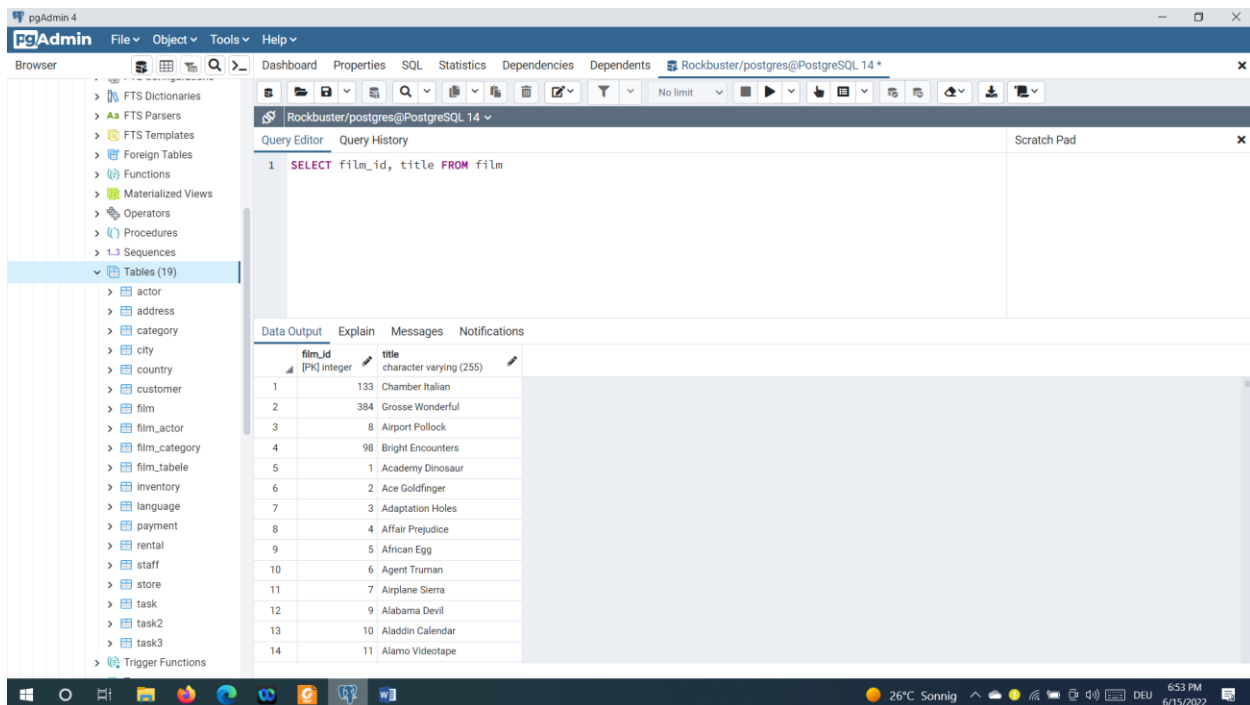


## Answers 3.4

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

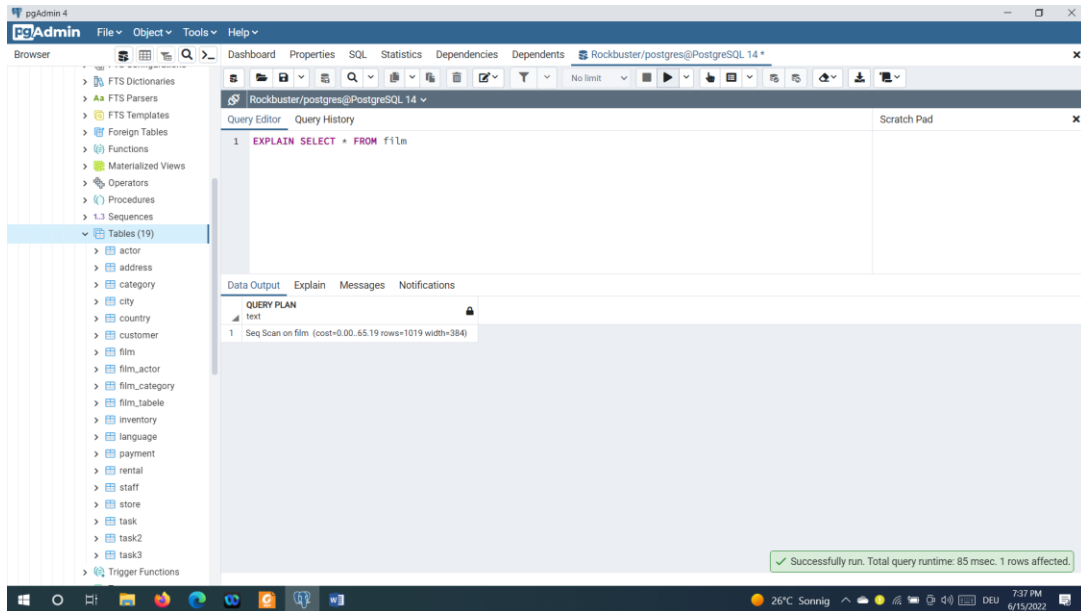


- You realize that only the “film\_id” and “title” columns are needed. Write a new query that selects only those 2 columns.

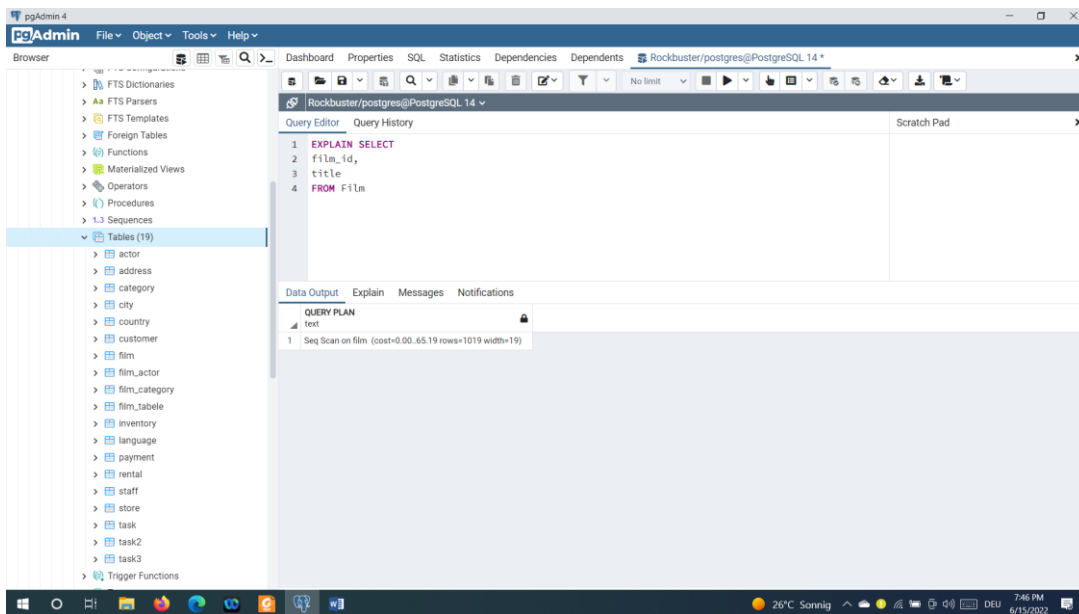


- Compare the cost of the original query and the revised query, and write a few sentences explaining the comparison. Can you suggest any ways to optimize this query?

## RAW Query



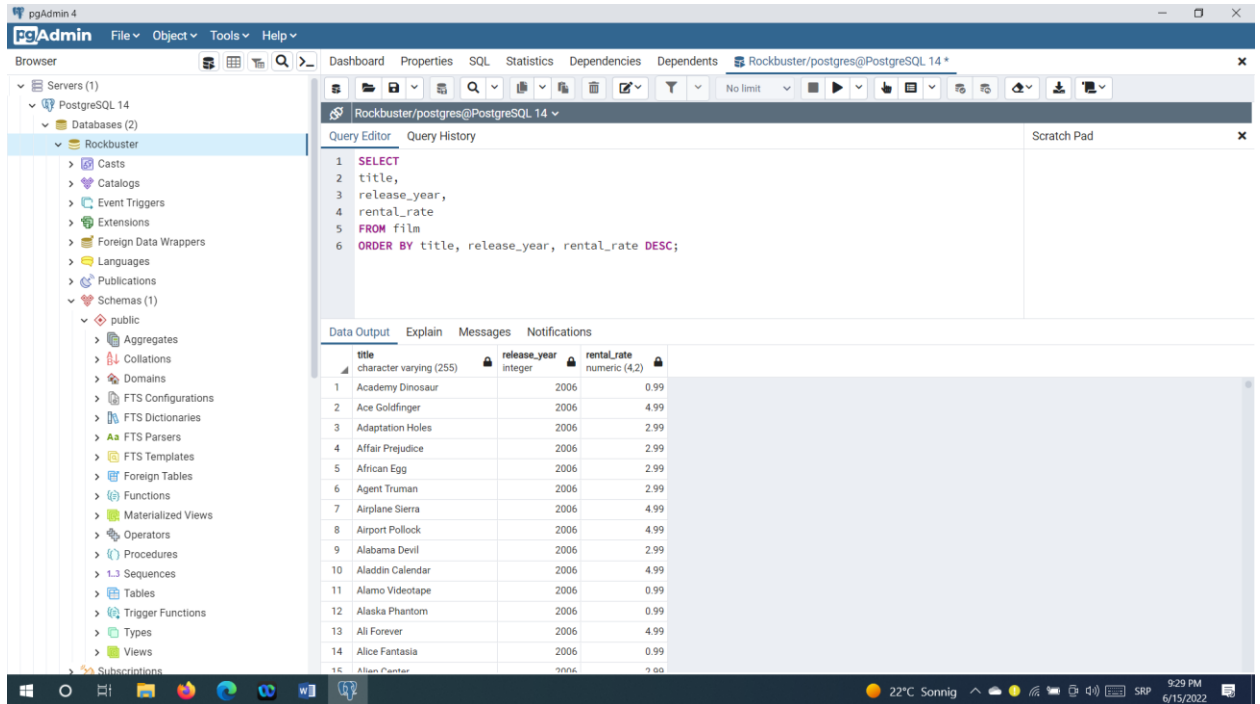
## Refined Query



Both queries have the same cost of 0.00-64.00 however, there was time saved in running the second query. The first query took 188 msec to retrieve the data and the second took 88 msec.

## 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. (You may need to explore how to save your output as a csv file in the Query Tool.)

## 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?

I have used following command:

```
SELECT rating,
```

```
AVG (rental_rate) AS avg_rental_rate
```

```
FROM film
```

```
GROUP BY rating
```

pgAdmin 4

File Object Tools Help

Browser Dashboard Properties SQL Statistics Dependencies Dependents Rockbuster/postgres@PostgreSQL 14 \*

Rockbuster/postgres@PostgreSQL 14

Query Editor Query History Scratch Pad

```
1 SELECT rating,
2 AVG (rental_rate) AS avg_rental_rate
3 FROM film
4 GROUP BY rating
5
6
7
8
9
```

Data Output Explain Messages Notifications

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

18°C Sonntag 6/16/2022 10:22 AM

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

pgAdmin 4

File Object Tools Help

Browser Dashboard Properties SQL Statistics Dependencies Dependents Rockbuster/postgres@PostgreSQL 14 \*

Rockbuster/postgres@PostgreSQL 14

Query Editor Query History Scratch Pad

```
1 SELECT rating,
2 MIN (rental_duration) AS min_rental_duration,
3 MAX (rental_duration) AS max_rental_duration
4 FROM film
5 GROUP BY rating
6
7
8
9
```

Data Output Explain Messages Notifications

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

18°C Sonntag 6/16/2022 10:33 AM

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 process of migrating data follows a procedure called ETL, Extract, Transform and Load. The job of a Data Engineer would be sourcing the data, transform it into a appropriate format and perform any aggregations and then load it into the database. However, an awareness of the basic concepts is important for data analysis, especially in startups where data analysts may be involved in aspects of database migration.

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

Working with the data before they were extracted and transformed we would be in high risk with working with incorrect information.