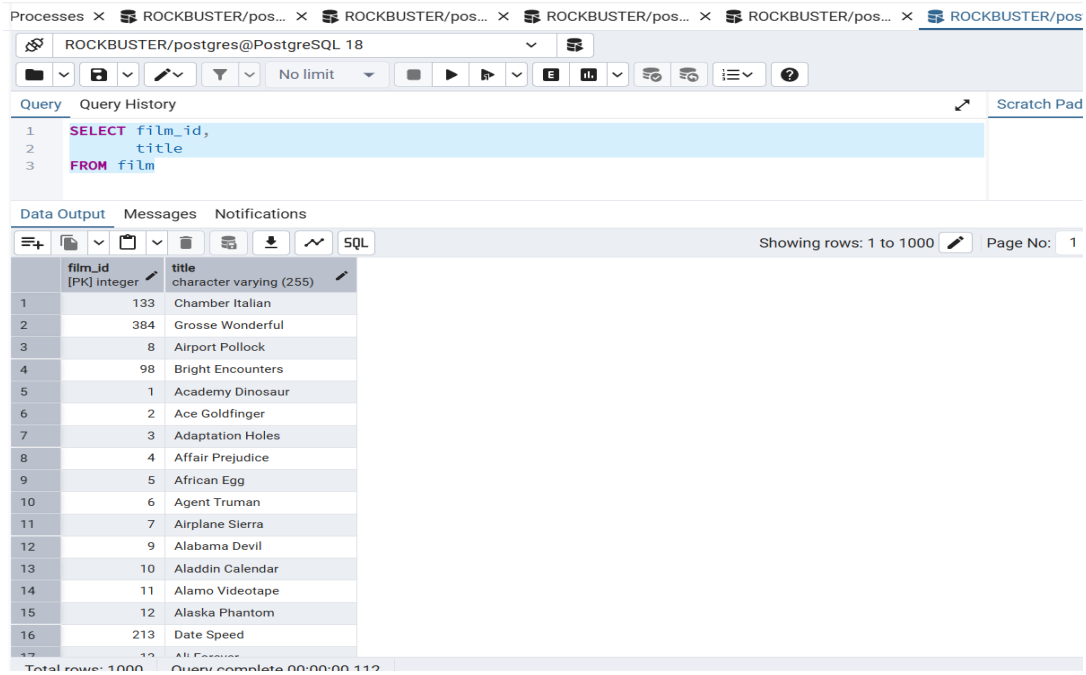
Databases & SQL for Analysts

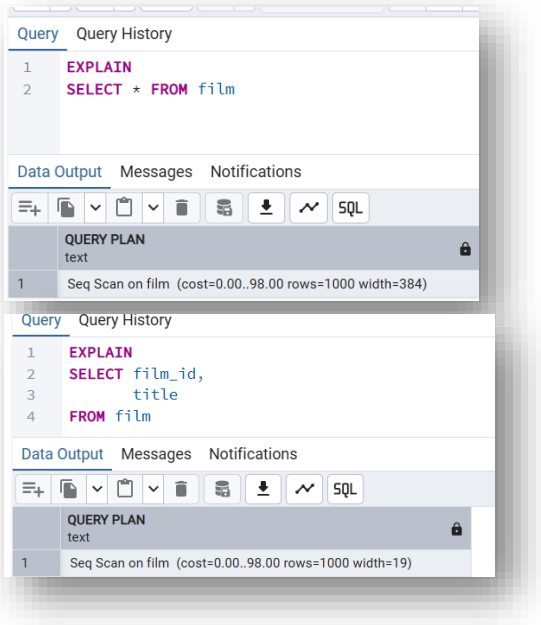
3.4: Database Querying in SQL

Refining Your Query:

1/



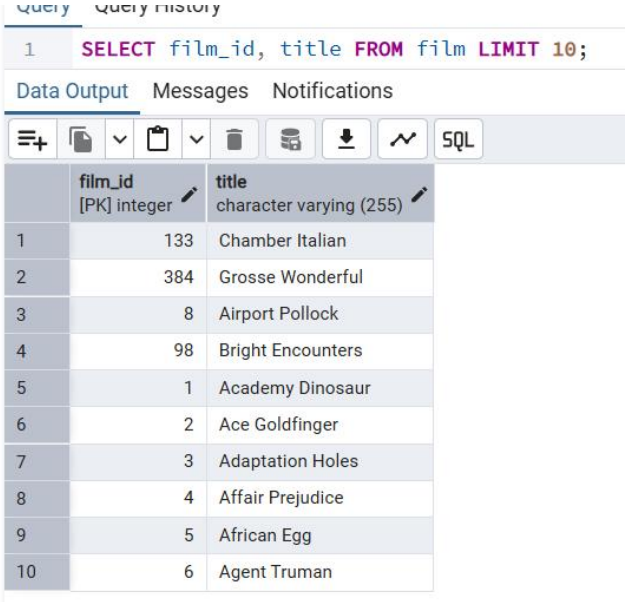
2/



The only difference is in the width: reducing the width from 384 bytes to 19 bytes.

We can say that there is no big difference but the revised query is more efficient in practice because it transfers less data from disk to memory.

Can you suggest any ways to optimize this query?



	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

Ordering the Data:

run a query that selects every film from the “film” table, with the movies sorted by title from A to Z ; by most recent release year, by highest to lowest rental rate.

Query | Query History

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

Data Output | Messages | Notifications

	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

Grouping Data:

1/

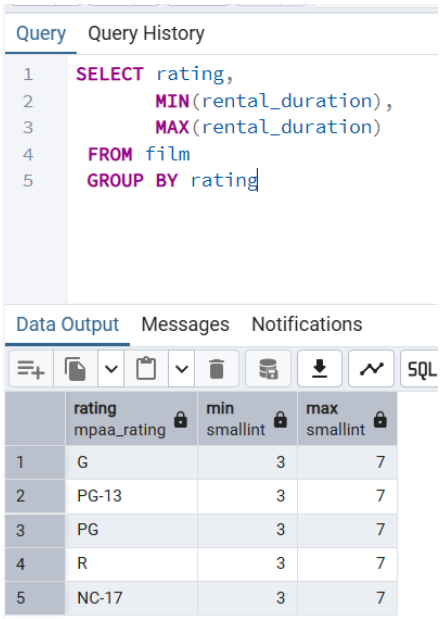
Query | Query History

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

Data Output | Messages | Notifications

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

2/



Database Migration:

Can you outline the procedure for migrating the data?

- **Extract:** The first step involves collecting the data from multiple data sources.
- **Transform:** During 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:** At this point the transformed data is inserted or loaded into the new database.

who will be responsible for i?

ETL is a data engineer's job. However, an awareness of the basic concepts is important for data analysis

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

Analyzing data before it is fully loaded into the data warehouse leads to biased results, as some extractions or transformations may be incomplete and provide a partial view. Furthermore, raw data often contains duplicates, null values, or inconsistent units, which compromises the reliability of the insights.