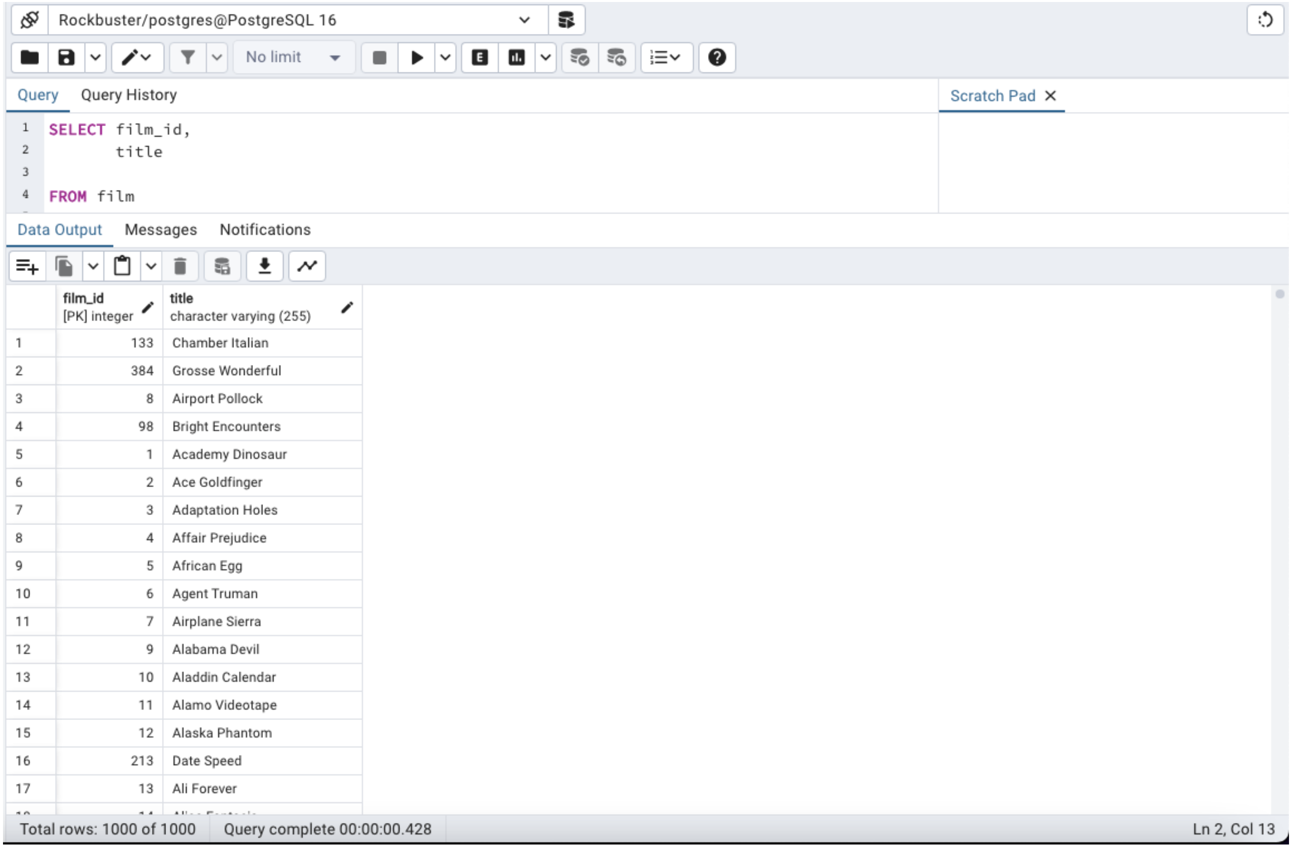
Task 3.4 (answers)

Database Querying in SQL



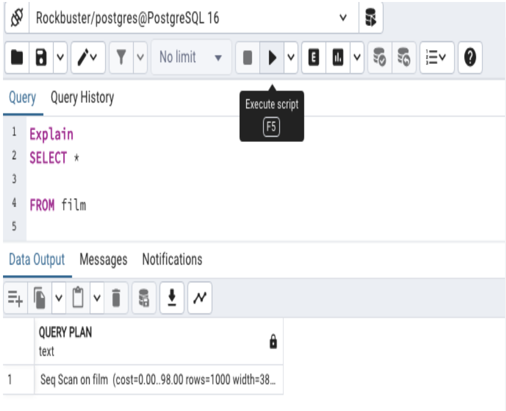
```sql
SELECT film_id,
       title

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
14	11	Alamo Videotape
15	12	Alaska Phantom
16	213	Date Speed
17	13	Ali Forever

Total rows: 1000 of 1000   Query complete 00:00:00.428   Ln 2, Col 13

COST COMPARISON

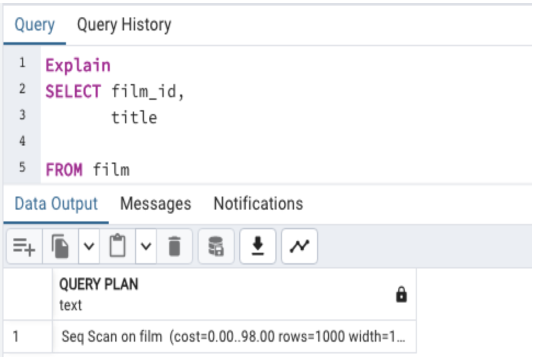
Original query



```sql
Explain
SELECT *

FROM film
```

	QUERY PLAN text
1	Seq Scan on film (cost=0.00..98.00 rows=1000 width=38...

Revised Query

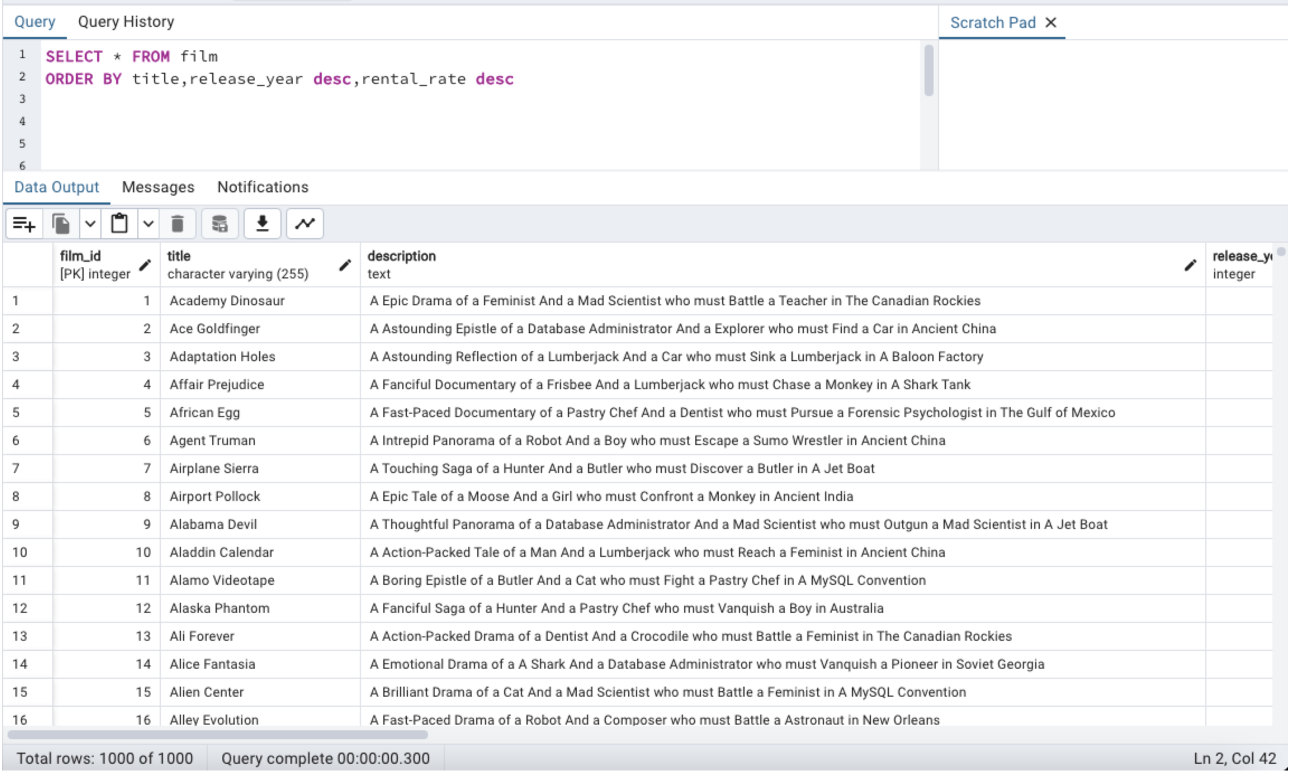


```sql
Explain
SELECT film_id,
       title

FROM film
```

	QUERY PLAN text
1	Seq Scan on film (cost=0.00..98.00 rows=1000 width=1...

The cost for both queries is identical since they both traverse the entire table and fetch the same number of records (1000 rows). To enhance query performance, limiting the number of rows returned would reduce the query's cost and increase efficiency, especially when extracting only the top few rows of records.

2:

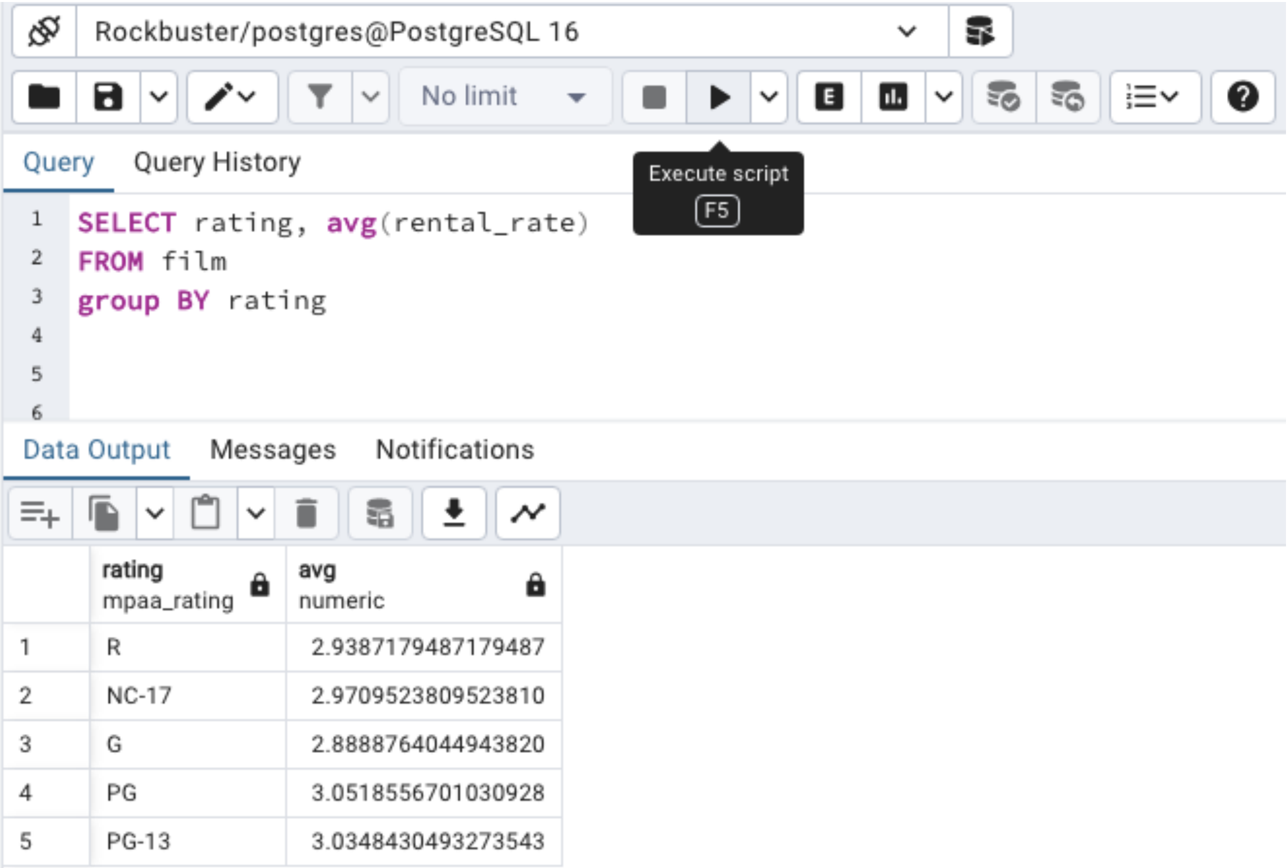


```sql
SELECT * FROM film
ORDER BY title,release_year desc,rental_rate desc
```

	film_id [PK] integer	title character varying (255)	description text	release_y integer
1	1	Academy Dinosaur	A Epic Drama of a Feminist And a Mad Scientist who must Battle a Teacher in The Canadian Rockies	
2	2	Ace Goldfinger	A Astounding Epistle of a Database Administrator And a Explorer who must Find a Car in Ancient China	
3	3	Adaptation Holes	A Astounding Reflection of a Lumberjack And a Car who must Sink a Lumberjack in A Baloon Factory	
4	4	Affair Prejudice	A Fanciful Documentary of a Frisbee And a Lumberjack who must Chase a Monkey in A Shark Tank	
5	5	African Egg	A Fast-Paced Documentary of a Pastry Chef And a Dentist who must Pursue a Forensic Psychologist in The Gulf of Mexico	
6	6	Agent Truman	A Intrepid Panorama of a Robot And a Boy who must Escape a Sumo Wrestler in Ancient China	
7	7	Airplane Sierra	A Touching Saga of a Hunter And a Butler who must Discover a Butler in A Jet Boat	
8	8	Airport Pollock	A Epic Tale of a Moose And a Girl who must Confront a Monkey in Ancient India	
9	9	Alabama Devil	A Thoughtful Panorama of a Database Administrator And a Mad Scientist who must Outgun a Mad Scientist in A Jet Boat	
10	10	Aladdin Calendar	A Action-Packed Tale of a Man And a Lumberjack who must Reach a Feminist in Ancient China	
11	11	Alamo Videotape	A Boring Epistle of a Butler And a Cat who must Fight a Pastry Chef in A MySQL Convention	
12	12	Alaska Phantom	A Fanciful Saga of a Hunter And a Pastry Chef who must Vanquish a Boy in Australia	
13	13	Ali Forever	A Action-Packed Drama of a Dentist And a Crocodile who must Battle a Feminist in The Canadian Rockies	
14	14	Alice Fantasia	A Emotional Drama of a A Shark And a Database Administrator who must Vanquish a Pioneer in Soviet Georgia	
15	15	Alien Center	A Brilliant Drama of a Cat And a Mad Scientist who must Battle a Feminist in A MySQL Convention	
16	16	Alley Evolution	A Fast-Paced Drama of a Robot And a Composer who must Battle a Astronaut in New Orleans	

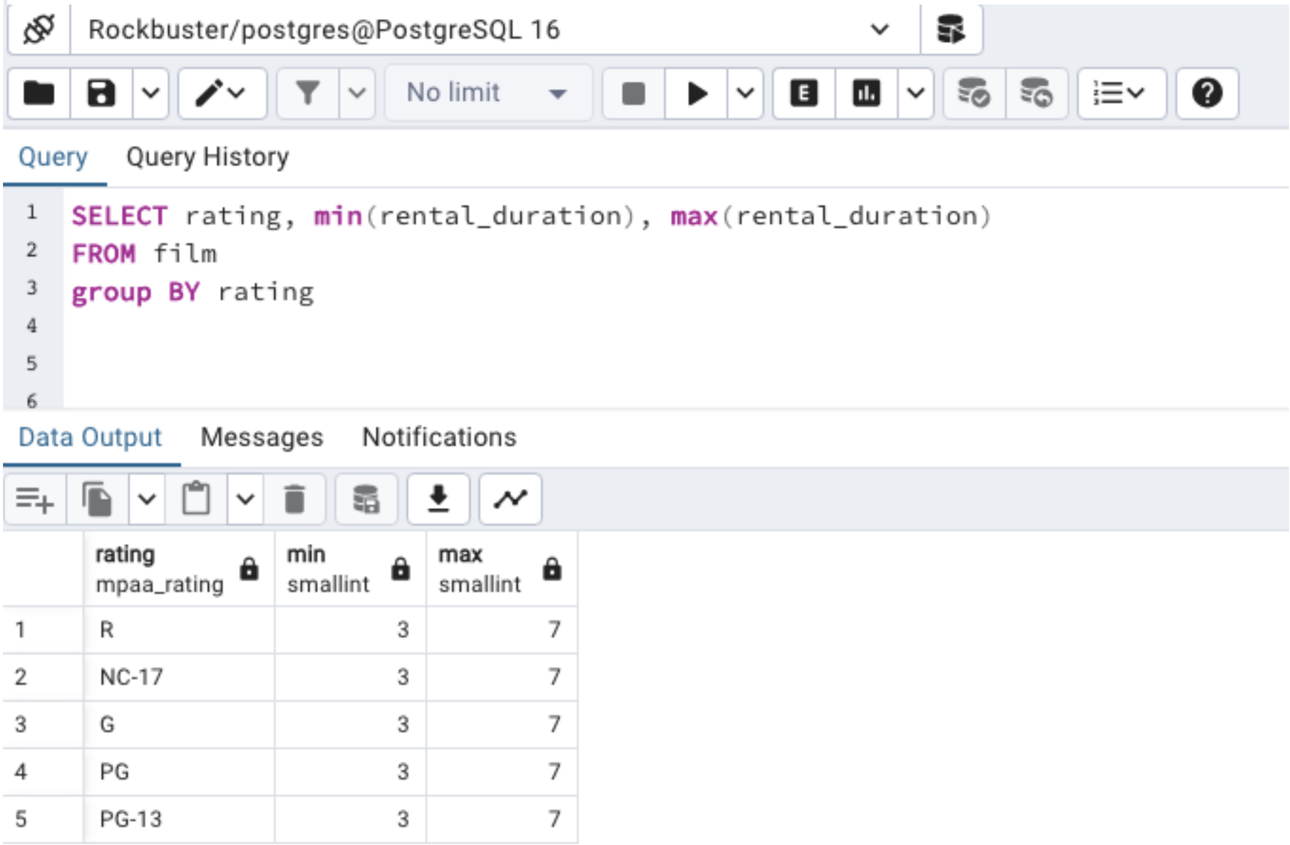
Total rows: 1000 of 1000   Query complete 00:00:00.300   Ln 2, Col 42

3:



```
SELECT rating, avg(rental_rate)
FROM film
group BY rating
```

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



```
SELECT rating, min(rental_duration), max(rental_duration)
FROM film
group BY rating
```

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

4:

Procedure for Data Migration and Responsibilities:
The steps for migrating data to the data warehouse are as follows:
1-Extracting and Collecting New Data: Data on user behavior in the Rockbuster Android app needs to be extracted and gathered.
2-Data Transformation: The extracted data must undergo transformation to align with the format and requirements for loading into the data warehouse.
3-Data Loading: The transformed data is loaded into the data warehouse.

Responsibilities for this ETL process primarily fall on data engineers. However, data analysts can assist in coordinating the migration timeline and ensuring data completeness at various stages.

Problems with Analyzing Data Before Loading into the Data Warehouse:
Analyzing data prior to loading it into the data warehouse can lead to several issues, including inconsistent, missing, or duplicate values. These issues can result in inaccurate analysis and require additional time and effort to clean the data. This may impact project timelines. Additionally, analyzing raw data takes more time compared to analyzing data from a data warehouse, where queries can be used to extract information efficiently.