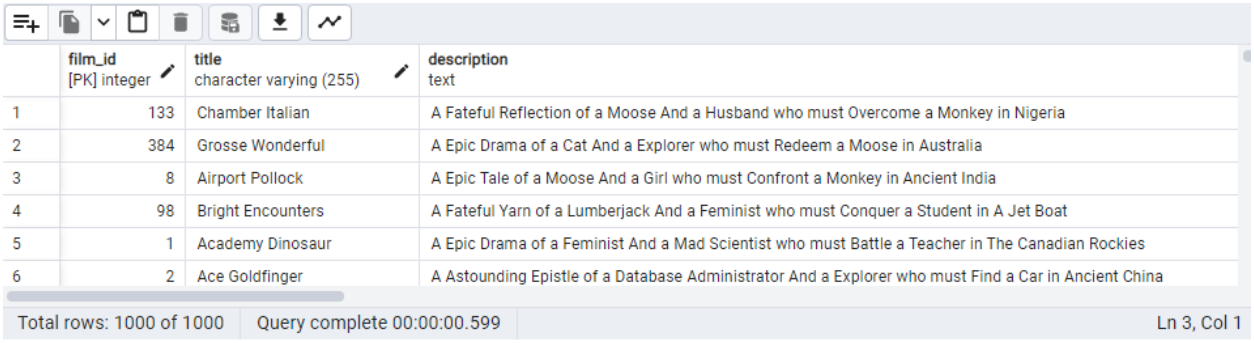
Task 3.4

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?

Original Query

Query | Query History

```
1  SELECT *
2  FROM film
3
```

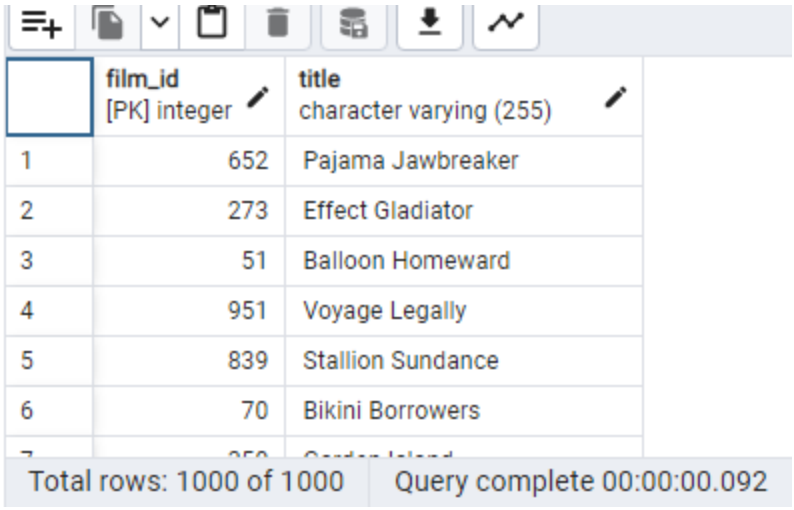


	film_id [PK] integer	title character varying (255)	description text
1	133	Chamber Italian	A Fateful Reflection of a Moose And a Husband who must Overcome a Monkey in Nigeria
2	384	Grosse Wonderful	A Epic Drama of a Cat And a Explorer who must Redeem a Moose in Australia
3	8	Airport Pollock	A Epic Tale of a Moose And a Girl who must Confront a Monkey in Ancient India
4	98	Bright Encounters	A Fateful Yarn of a Lumberjack And a Feminist who must Conquer a Student in A Jet Boat
5	1	Academy Dinosaur	A Epic Drama of a Feminist And a Mad Scientist who must Battle a Teacher in The Canadian Rockies
6	2	Ace Goldfinger	A Astounding Epistle of a Database Administrator And a Explorer who must Find a Car in Ancient China

Total rows: 1000 of 1000 | Query complete 00:00:00.599 | Ln 3, Col 1

Revised Query

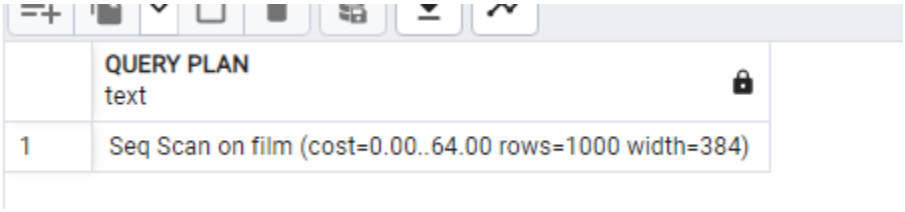
```
SELECT film_id,
title
FROM film
GROUP BY film_id,
title
```



	film_id [PK] integer	title character varying (255)
1	652	Pajama Jawbreaker
2	273	Effect Gladiator
3	51	Balloon Homeward
4	951	Voyage Legally
5	839	Stallion Sundance
6	70	Bikini Borrowers

Total rows: 1000 of 1000 Query complete 00:00:00.092

Cost/Explain for Unrevised Query



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

Cost/Explain for Revised Query

	QUERY PLAN text
1	HashAggregate (cost=66.50..76.50 rows=1000 width=19)
2	Group Key: film_id
3	-> Seq Scan on film (cost=0.00..64.00 rows=1000 width=19)

Optimizing queries saves time and money. Both the revised and unrevised can take up to a cost of 64.

Although this is just an estimate it still shows that a company can choose to run either query.

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. To do this, click the button “Save results to file”.

```sql
SELECT title,
rental_rate,
release_year
FROM film
ORDER BY title,
release_year DESC,
rental_rate DESC

```

	title character varying (255)	rental_rate numeric (4,2)	release_year integer
1	Academy Dinosaur	0.99	2006
2	Ace Goldfinger	4.99	2006
3	Adaptation Holes	2.99	2006
4	Affair Prejudice	2.99	2006
5	African Egg	2.99	2006
6	Agent Truman	2.99	2006

Total rows: 1000 of 1000 Query complete 00:00:00.053

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

```
10  SELECT rating,
11  AVG(rental_rate)
12  FROM film
13  GROUP BY rating
```

Data Output | Messages | Notifications

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

```
SELECT rating,
MAX(rental_duration)
FROM film
GROUP BY rating
```

Output | Messages | Notificatior

rating mpaa_rating	max smallint
G	7
PG-13	7
PG	7
R	7
NC-17	7

```
15  SELECT rating,
16  MIN(rental_duration)
17  FROM film
18  GROUP BY rating
```

Data Output | Messages | Notifications

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

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

The procedure is extracting the data, transforming the data, and loading the data. Data migration is usually done by data engineers, but data analysts should also understand the process. The engineer would extract the data needed for Rockbuster from an external tool. Then the data would be formatted to match Rockbuster's database. Then the data would be loaded into the database.

Analyzing the data before it has been loaded into the warehouse will be difficult. The engineer or analyst first needs to understand the data and establish connections in order to confidently manipulate the data to produce the results intended.