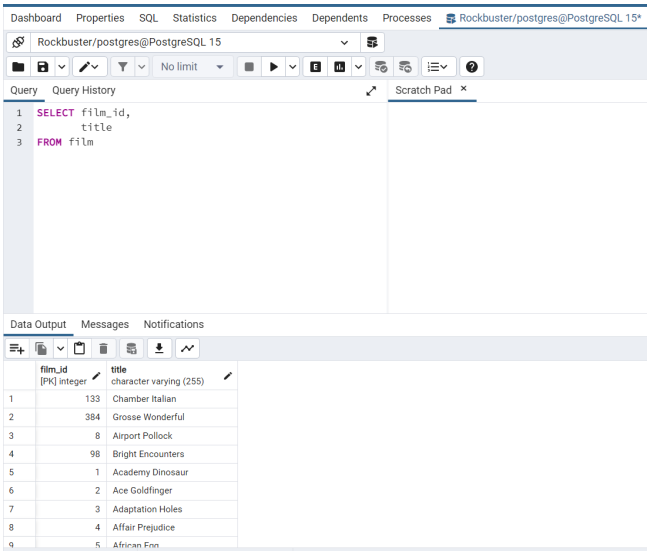
1.) **<u>Refining Your Query: You need to get some data from the “film” table and decide to use the query SELECT * FROM film.</u>**

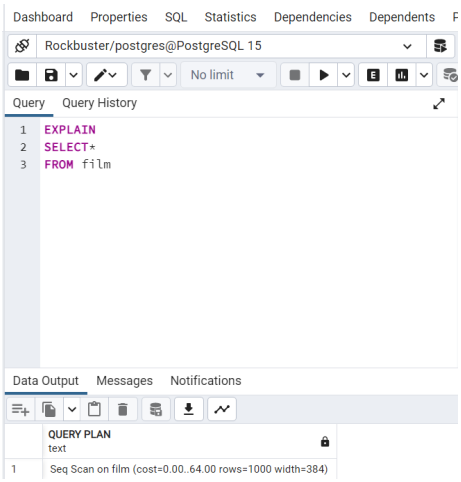
- **You realize that only the “film_id” and “title” columns are needed. Write a new query that selects only those 2 columns.**



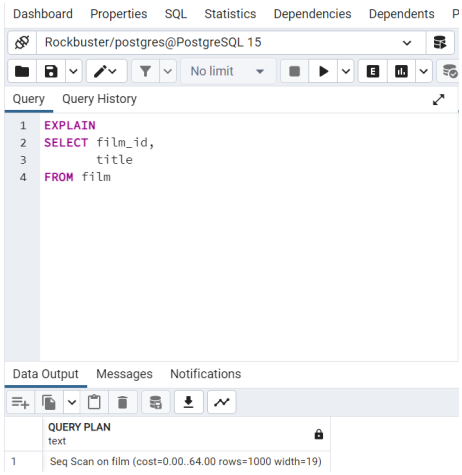
- **Compare the cost of the original query and the revised query, and write a few sentences explaining the comparison. Can you suggest any ways to optimize this query?**

  The costs of the respective queries are the same. The respective widths differ, though, with the revised query returning a narrower (i.e. less wide) result. Depending on what type of information is being sought, these queries could be optimized by limiting the number of rows returned with LIMIT or WHERE, to restrict returns based on certain specifications.

<u>Query 1</u>

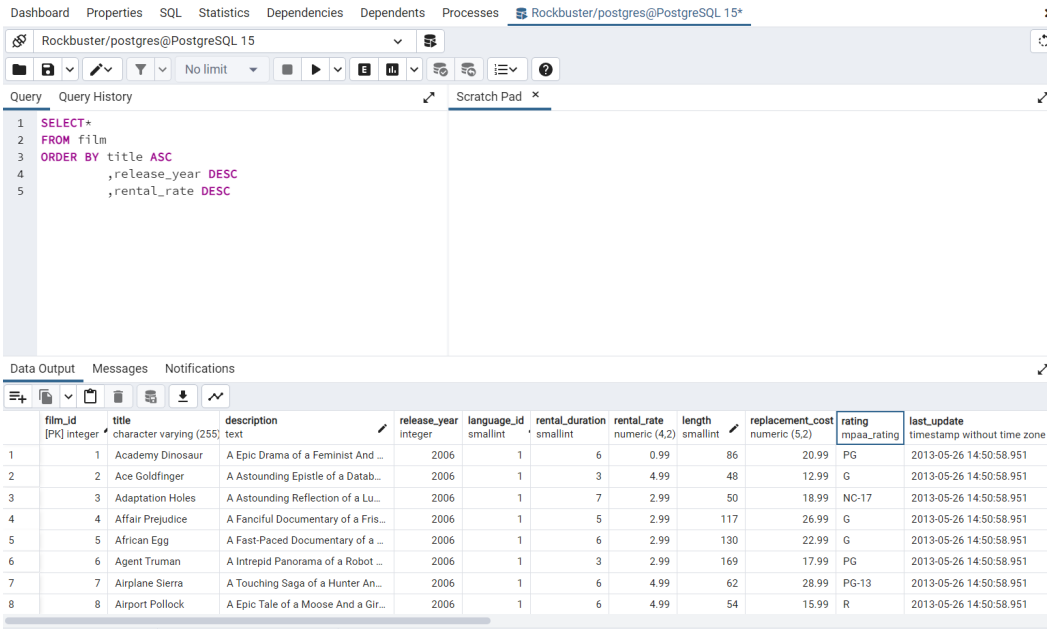


<u>Query 2</u>



2.) **<u>Ordering the Data:</u>**

- **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.**

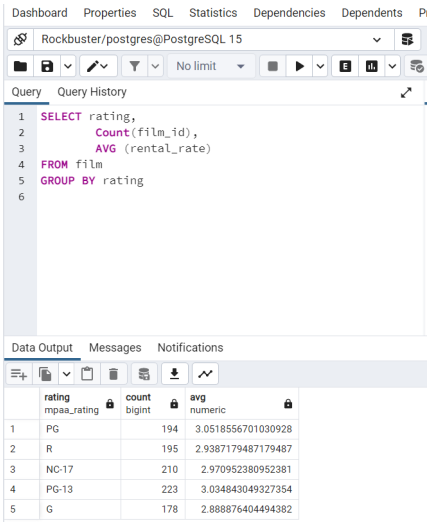


	film_id [PK] integer	title character varying (255)	description text	release_year integer	language_id smallint	rental_duration smallint	rental_rate numeric (4,2)	length smallint	replacement_cost numeric (5,2)	rating mpaa_rating	last_update timestamp without time zone
1	1	Academy Dinosaur	A Epic Drama of a Feminist And ...	2006	1	6	0.99	86	20.99	PG	2013-05-26 14:50:58.951
2	2	Ace Goldfinger	A Astounding Epistle of a Datab...	2006	1	3	4.99	48	12.99	G	2013-05-26 14:50:58.951
3	3	Adaptation Holes	A Astounding Reflection of a Lu...	2006	1	7	2.99	50	18.99	NC-17	2013-05-26 14:50:58.951
4	4	Affair Prejudice	A Fanciful Documentary of a Fris...	2006	1	5	2.99	117	26.99	G	2013-05-26 14:50:58.951
5	5	African Egg	A Fast-Paced Documentary of a ...	2006	1	6	2.99	130	22.99	G	2013-05-26 14:50:58.951
6	6	Agent Truman	A Intrepid Panorama of a Robot ...	2006	1	3	2.99	169	17.99	PG	2013-05-26 14:50:58.951
7	7	Airplane Sierra	A Touching Saga of a Hunter An...	2006	1	6	4.99	62	28.99	PG-13	2013-05-26 14:50:58.951
8	8	Airport Pollock	A Epic Tale of a Moose And a Gir...	2006	1	6	4.99	54	15.99	R	2013-05-26 14:50:58.951

- **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"**

3.) **<u>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.</u>**

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



	rating mpaa_rating	count bigint	avg numeric
1	PG	194	3.0518556701030928
2	R	195	2.9387179487179487
3	NC-17	210	2.970952380952381
4	PG-13	223	3.034843049327354
5	G	178	2.888876404494382

PG = 3.0518556701030928
R= 2.9387179487179487
NC-17= 2.970952380952381
PG-13= 3.034843049327354
G= 2.888876404494382

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

Rockbuster/postgres@PostgreSQL 15

No limit

Query | Query History

```sql
SELECT rating,
        Count(film_id),
        AVG (rental_rate),
        MAX (rental_duration),
        MIN (rental_duration)
FROM film
GROUP BY rating
```

Data Output | Messages | Notifications

	rating mpaa_rating	count bigint	avg numeric	max smallint	min smallint
1	PG	194	3.0518556701030928	7	3
2	R	195	2.9387179487179487	7	3
3	NC-17	210	2.970952380952381	7	3
4	PG-13	223	3.034843049327354	7	3
5	G	178	2.888876404494382	7	3

Rating	Maximum	Minimum
PG	7	3
R	7	3
NC-17	7	3
PG-13	7	3
G	7	3

4.) **<u>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.</u>**

- **Can you outline the procedure for migrating the data and who will be responsible for it?**
  This process is commonly referred to as Extract, Transform, and Load in which the data is downloaded from the application, subsequently transformed into a different format, and then saved in the data warehouse. Typically, this is the domain of the data engineer but analysts should be aware of the general process.

- **What problems do you foresee if you start analyzing the data before it's been loaded into the data warehouse?**
  Inconsistent formatting could lead to unintentional analysis (i.e. analyzing the wrong fields or analyzing data that is not yet in the intended format.) There could also be consistency issues and, if uploading data from multiple sources, faulty association of records.