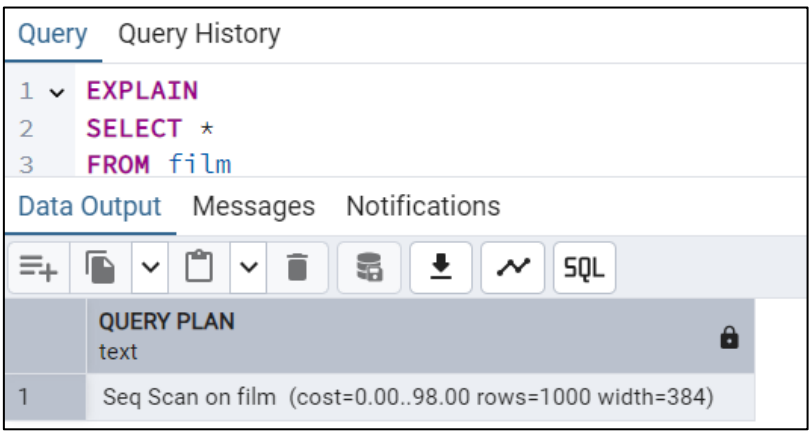
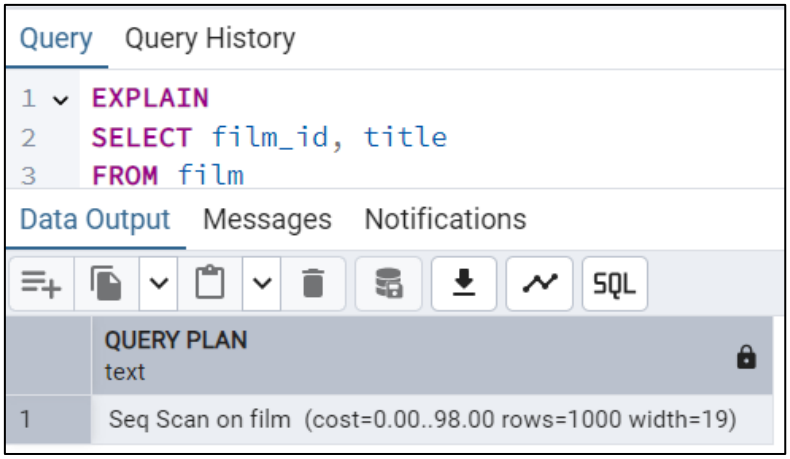
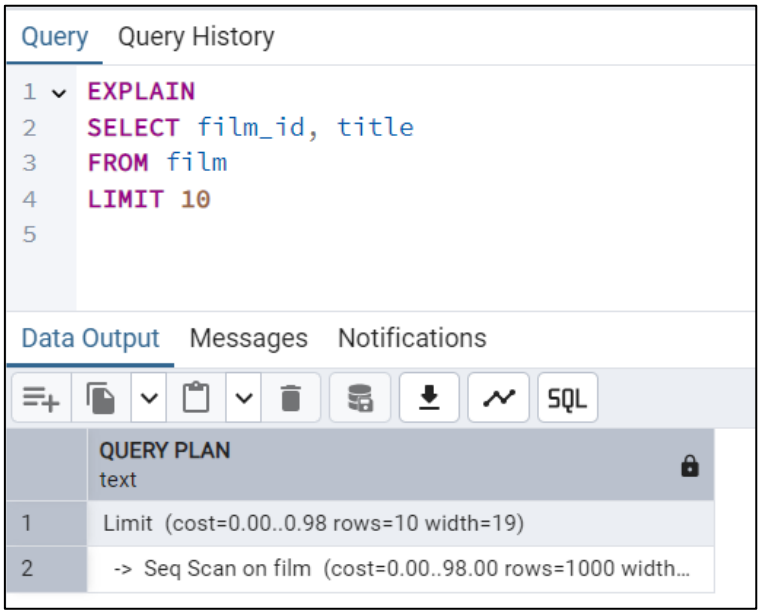
STEP 1

Compare the cost of the original query and the revised query.





Both have the same cost because the same search was conducted throughout the entire table.



Query | Query History

```
1  EXPLAIN
2  SELECT film_id, title
3  FROM film
4  Order by title desc
5  LIMIT 1
6
```

Data Output | Messages | Notifications

	QUERY PLAN text
1	Limit (cost=0.28..0.41 rows=1 width=19)
2	-> Index Scan Backward using idx_title on film (cost=0.28..135.33 rows=1000 width...

An attempt was made to minimize costs using the ORDER BY and LIMIT functions, but even so, the result increased the cost.

STEP 2

- **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.**

Query | Query History

```
1  SELECT title, release_year, rental_rate
2  FROM film
3  ORDER BY title asc, release_year desc, rental_rate desc
4
```

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
9	Alabama Devil	2006	2.99
10	Aladdin Calendar	2006	4.99
11	Alamo Videotape	2006	0.99
12	Alaska Phantom	2006	0.99
13	Ali Forever	2006	4.99
14	Alice Fantasia	2006	0.99
15	Alien Center	2006	2.99

Total rows: 1000 | Query complete 00:00:00.232

STEP 3

What is the average rental rate for each rating category?

Query Query History

```
1 SELECT rating,
2 AVG(rental_rate)
3 FROM film
4 GROUP BY rating
5
```

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

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

Query Query History

```
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
```

Data Output Messages Notifications

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

STEP 4:

Procedure for Migrating Data from the Rockbuster Android App to the Data Warehouse

Data Collection & Storage

Responsible: Mobile App Development Team

The app collects user behavior data using an external tool.

Data Extraction

Responsible: ETL (Extract, Transform, Load) Engineers

Extract raw data from the external tool's API, database, or storage system.

Define a schedule for data extraction (real-time streaming or batch processing).

Data Transformation & Cleaning

Responsible: Data Engineering Team

Standardize formats (e.g., timestamps, user IDs).

Handle missing or duplicate data.

Map the extracted data to the correct schema used in the data warehouse.

Data Loading into the Data Warehouse

Responsible: ETL Engineers & Database Administrators

Load transformed data into the data warehouse.

Implement indexing and partitioning for efficient querying.

Validation & Quality Assurance

Responsible: Data Analysts & QA Team

Run queries to verify data integrity and completeness.

Cross-check with raw data to ensure accuracy.

Data Analysis & Reporting

Responsible: Data Analysts & Business Intelligence Team

Once validated, data is used for analytics, dashboards, and reporting.

Analysts can now derive insights from user behavior in the Rockbuster Android app.

Problems If Analysis Starts Before Data is Loaded into the Data Warehouse

Incomplete Data

If the data hasn't been fully migrated, reports will be inaccurate or misleading.

Inconsistent Formats

Raw data may have inconsistencies (different date formats, missing values), leading to errors in analysis.

Performance Issues

Running queries on raw or scattered data sources may be slow and inefficient compared to querying a structured data warehouse.

Duplicate or Redundant Data

Without proper ETL processes, duplicate records might distort analysis results.

Security & Compliance Risks

Analyzing unverified data could expose sensitive information without proper access controls.

Wrong Business Decisions

If key stakeholders act on unprocessed or incorrect data, it could lead to faulty strategies and decisions.