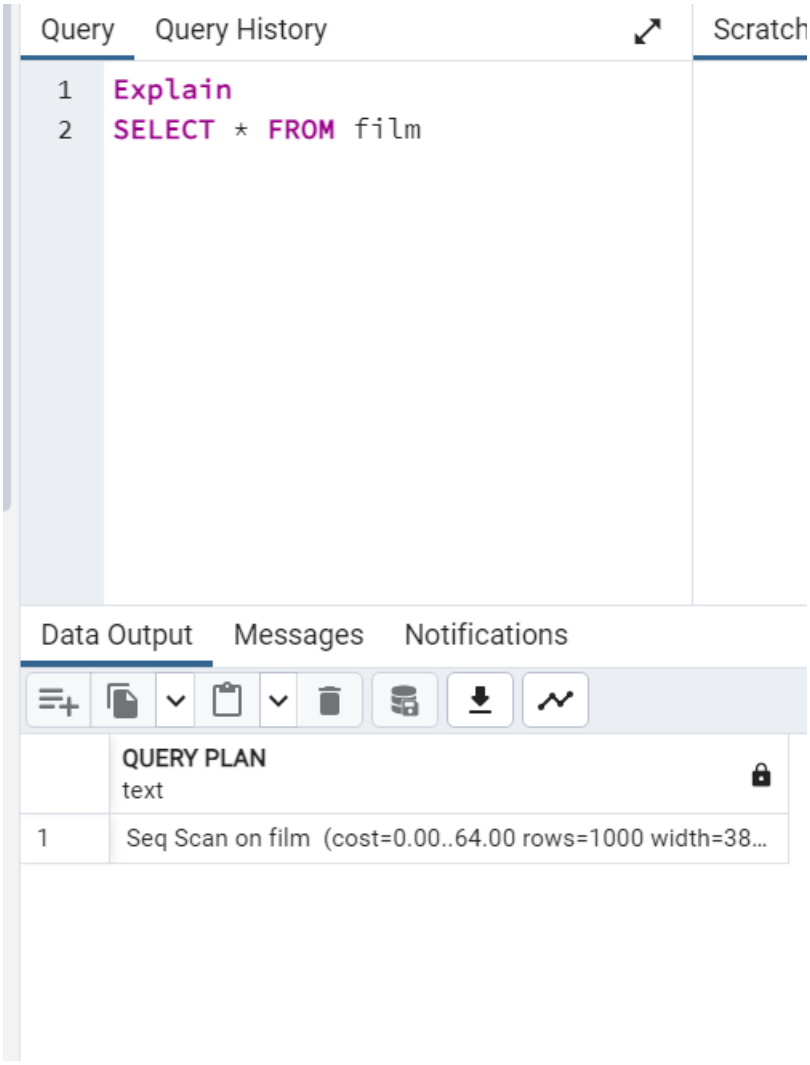
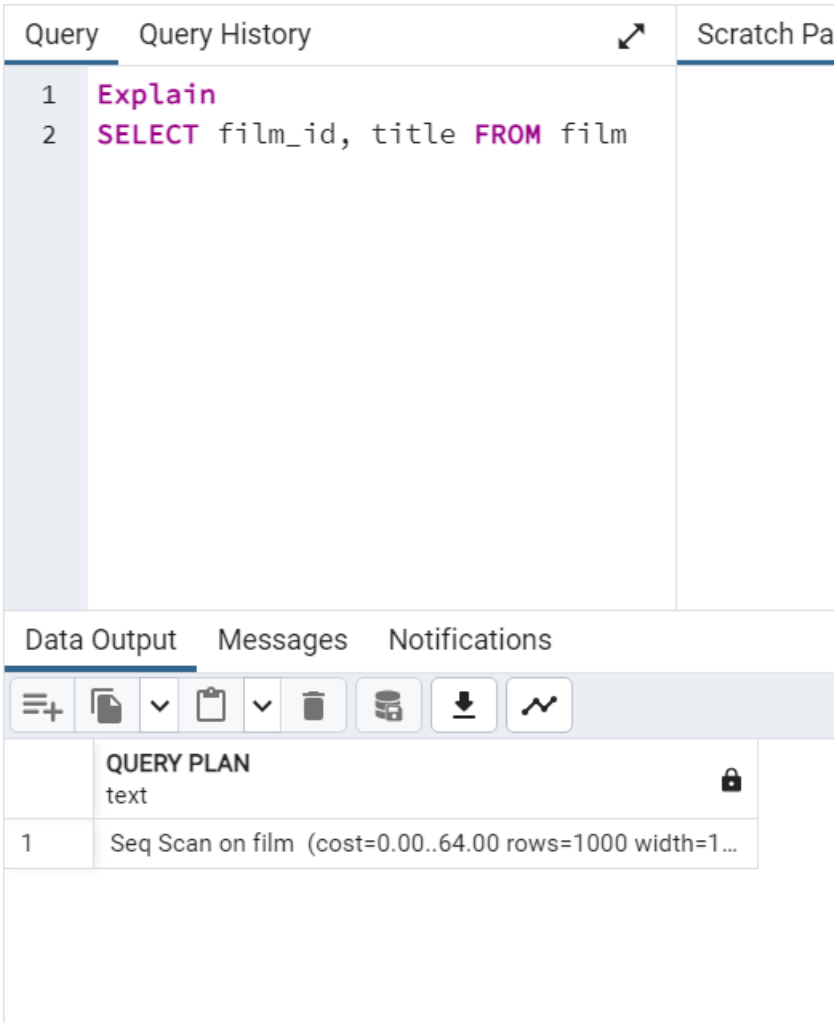
Step 1:


Query
Query History
Scratch
1 Explain
2 SELECT * FROM film
Data Output
Messages
Notifications
QUERY PLAN
text
1
Seq Scan on film (cost=0.00..64.00 rows=1000 width=38...



The cost for both Queries is the same even though the columns returned are less. To optimize this we should limit the number of rows that are returned based on aggregation to ensure we are pulling in the necessary data.

Step 2:

Query | Query History

```
1  SELECT title, release_year, rental_rate
2  FROM film
3  ORDER BY title, 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
9	Alabama Devil	2006	2.99

Step 3:

Average Rental Rate

Query | Query History

```
1  SELECT rating, AVG(rental_rate) AS average_rental_rate
2  FROM film
3  GROUP BY rating
```

Data Output | Messages | Notifications

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

Min and Max Rental Duration

Query Query History

```
1  SELECT rating, MIN(rental_duration) AS min_rental_duration, MAX(rental_duration) AS max_rental_duration
2  FROM film
3  GROUP BY rating
```

Data Output Messages Notifications

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

Step 4:

A. Data Engineers would oversee mitigating the data from an external source. They would begin by extracting the data from the android apps which store user data that then can be mitigated and organized into usable data. The usable data would then be used by the data analyst to analyze.

B. Since the data is live it would be continuously updated based on user data and this can cause issues with data integrity for the analysis. By integrity I mean there could be incorrect or even missing data since the data will have to be transferred into the data base from an external source.

Bonus Task

Query | Query History

```
SELECT rating, MIN(replacement_cost) AS minimum_replacement_cost, MAX(replacement_cost) AS
maximum_replacement_cost
FROM film
GROUP BY rating
ORDER BY CASE WHEN rating = 'G' THEN '1'
              WHEN rating = 'PG' THEN '2'
              WHEN rating = 'PG-13' THEN '3'
              WHEN rating = 'R' THEN '4'
              WHEN rating = 'NC-17' THEN '5'
END
```

Data Output | Messages | Notifications

	rating mpaa_rating	minimum_replacement_cost numeric	maximum_replacement_cost numeric
1	G	9.99	29.99
2	PG	9.99	29.99
3	PG-13	9.99	29.99
4	R	9.99	29.99
5	NC-17	9.99	29.99