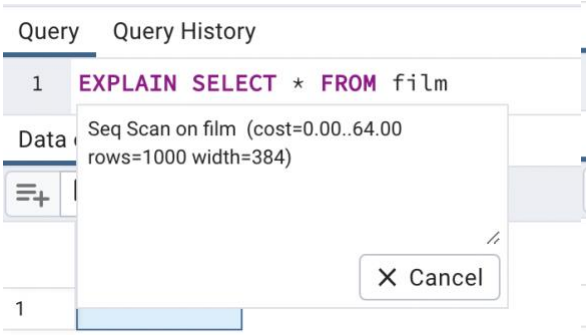
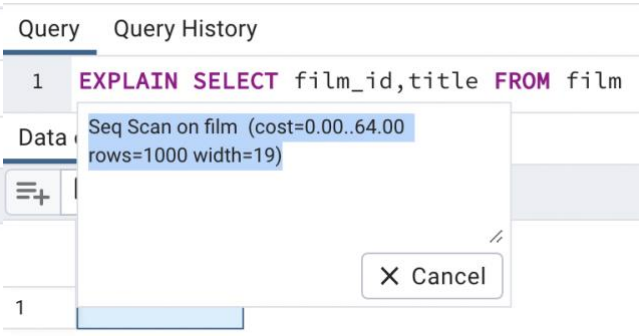
Answers 3.4

Step 1





Total rows: 1 of 1 | Query complete 00:00:00.090

Total rows: 1 of 1 | Query complete 00:00:00.053

Both queries in theory have the same cost 0.00..64.00, however in practice, the original query, which fetched more data than needed took 90 milliseconds, whilst the improved query which on fetched what needed took 53 milliseconds. The improvement results in faster run time, which on a large scale can be highly beneficial for a business, and present a cleaner result, with only the necessary information.

Step 2

Query | Query History

```
1  SELECT title, release_year, rental_rate
2  FROM film
3  ORDER BY title, release_year, 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
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
16	Alley Evolution	2006	2.99
17	Alone Trip	2006	0.99

Films Sorted CSV File

Step 3

Query | Query History

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

Data output | Messages | Notifications

	rating mpaa_rating	avg_rental_rate numeric
1	R	2.9387179487179
2	PG	3.0518556701030
3	NC-17	2.9709523809523
4	PG-13	3.0348430493273
5	G	2.8888764044943

Average Rental Rate CSV File

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
```

Data output | Messages | Notifications

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

Minimum and Maximum Rental Durations CSV File

Step 4

Can you outline the procedure for migrating the data and who will be responsible for it?
Ideally, data engineers will be responsible for migrating data into the Rockbuster database warehouse. They will need to follow the ETL process, where it is extracted from the source of the app data, transformed into a more palatable format that aligns with Rockbuster's, and then loaded into the warehouse database.

What problems do you foresee if you start analyzing the data before it's been loaded into the data warehouse?
The format may not be aligned with ours, and we will not have a useful comparison point from our own data sources of which to analyze the data. Breaking the procedural order will only waste time.

Bonus Task

Query | Query History

```
1  SELECT rating,
2  MIN(replacement_cost) AS min_replacement_cost,
3  MAX(replacement_cost) AS max_replacement_cost
4  FROM film
5  GROUP BY rating
6  ORDER BY CASE WHEN rating = 'G' THEN 1
7  WHEN rating = 'PG' THEN 2
8  WHEN rating = 'PG-13' THEN 3
9  WHEN rating = 'R' THEN 4
10 WHEN rating = 'NC-17' THEN 5
11 ELSE 6
12 END
```

Data output | Messages | Notifications

	rating mpaa_rating	min_replacement_cost numeric	max_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