Andrew Smith

Career Foundry Data Immersion

January 14, 2022

3.4: Database Querying in SQL (Answers 3.4)

1)

a)

SELECT film\_id,

title

FROM film

Graphical user interface, application, Word

Description automatically generated

b)

**Original Query**

EXPLAIN

SELECT \* FROM film

"Seq Scan on film (cost=0.00..64.00 rows=1000 width=384)"

Graphical user interface, application

Description automatically generated

**Revised Query**

EXPLAIN

SELECT film\_id,

title

FROM film

"Seq Scan on film (cost=0.00..64.00 rows=1000 width=19)"

Graphical user interface, application

Description automatically generated

The cost for both queries is the same with 64 and the rows are both 1000. However, with the revised query, the width is significantly lowered from 384 to 19. I’m unsure of any ways to optimize this query more since it is a simple query and only needs data from two columns.

2)

-Movies by title from A to Z

SELECT title,

film\_id

FROM film

ORDER BY title ASC

Graphical user interface, application, table

Description automatically generated

-Movies by most recent release year

SELECT title,

release\_year

FROM film

ORDER BY release\_year DESC

Graphical user interface, application

Description automatically generated

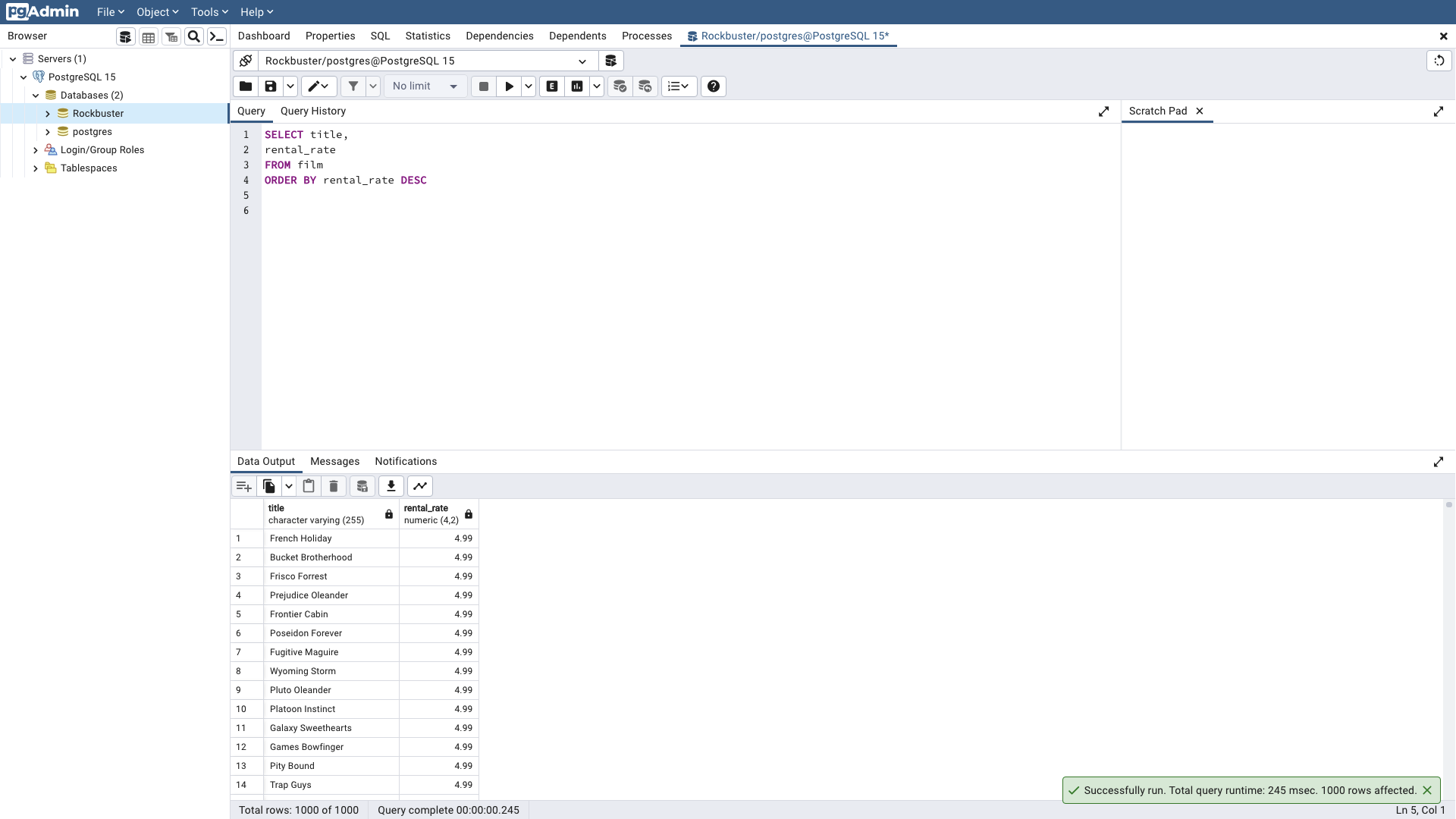
-Movies by highest to lowest rental rate

SELECT title,

rental\_rate

FROM film

ORDER BY rental\_rate DESC



3)

a)

SELECT rating,

avg(rental\_rate)

FROM film

GROUP BY rating

**Rating** **Average Rental Rate per Rating**

"R" 2.9387179487179487

"NC-17" 2.970952380952381

"PG" 3.0518556701030928

"PG-13" 3.034843049327354

"G" 2.888876404494382

Graphical user interface, application

Description automatically generated

b)

SELECT rating,

min(rental\_duration)

FROM film

GROUP BY rating

**Rating Minimum Rental Duration per Rating**

"R" 3

"NC-17" 3

"PG" 3

"PG-13" 3

"G" 3

Graphical user interface, application

Description automatically generated

SELECT rating,

max(rental\_duration)

FROM film

GROUP BY rating

**Rating Maximum Rental Duration per Rating**

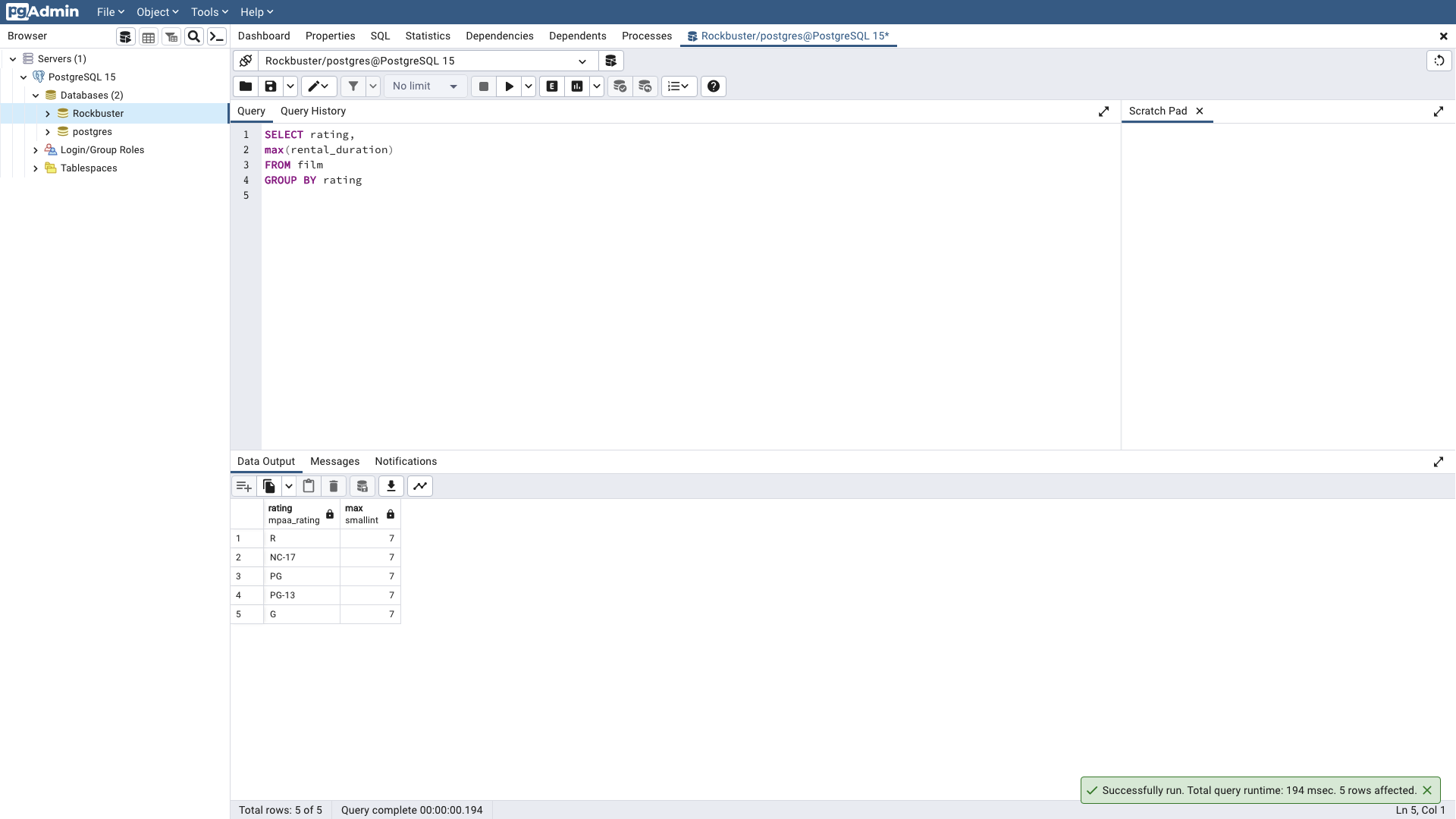
"R" 7

"NC-17" 7

"PG" 7

"PG-13" 7

"G" 7



4)

a)When migrating data from the Rockbuster Android app to the data warehouse, there is a procedure with three steps called Extract, Transform, and Load (ETL). Extract means collecting data from the app. Then the extracted data needs to be transformed or converted into a format that fits the data warehouse. Finally load the transformed data into the data warehouse. This procedure is completed by data engineers which means they are responsible for the data migration.

b)If you analyze data before it’s been loaded into the data warehouse, then there is likely going to be different results since the data is in a different format than the data warehouse. In addition, this can affect the analysis because you are not finding results along with other data in the data warehouse.