Netflix Data Analysis using SQL

Author: Jyoti Rekha Dutta

Tools Used: pgAdmin 4, Excel

# PROJECT OBJECTIVE:

To analyze Netflix data using SQL by cleaning the dataset and uncovering insights related to content type, ratings, release year, country distribution, and genre trends.

# CREATE TABLE:

CREATE TABLE netflix

(

show\_id VARCHAR(5),

type VARCHAR(10),

title VARCHAR(250),

director VARCHAR(550),

casts VARCHAR(1050),

country VARCHAR(550),

date\_added VARCHAR(55),

release\_year INT,

rating VARCHAR(15),

duration VARCHAR(15),

listed\_in VARCHAR(250),

description VARCHAR(550)

);

# VIEWING ALL IMPORTED DATA:

The following query is used to view the complete Netflix dataset after importing it into the SQL Server environment:

SELECT \* FROM Netflix

# SELECT COUNT (\*) as total\_content FROM netflix

# Screenshot (84).png

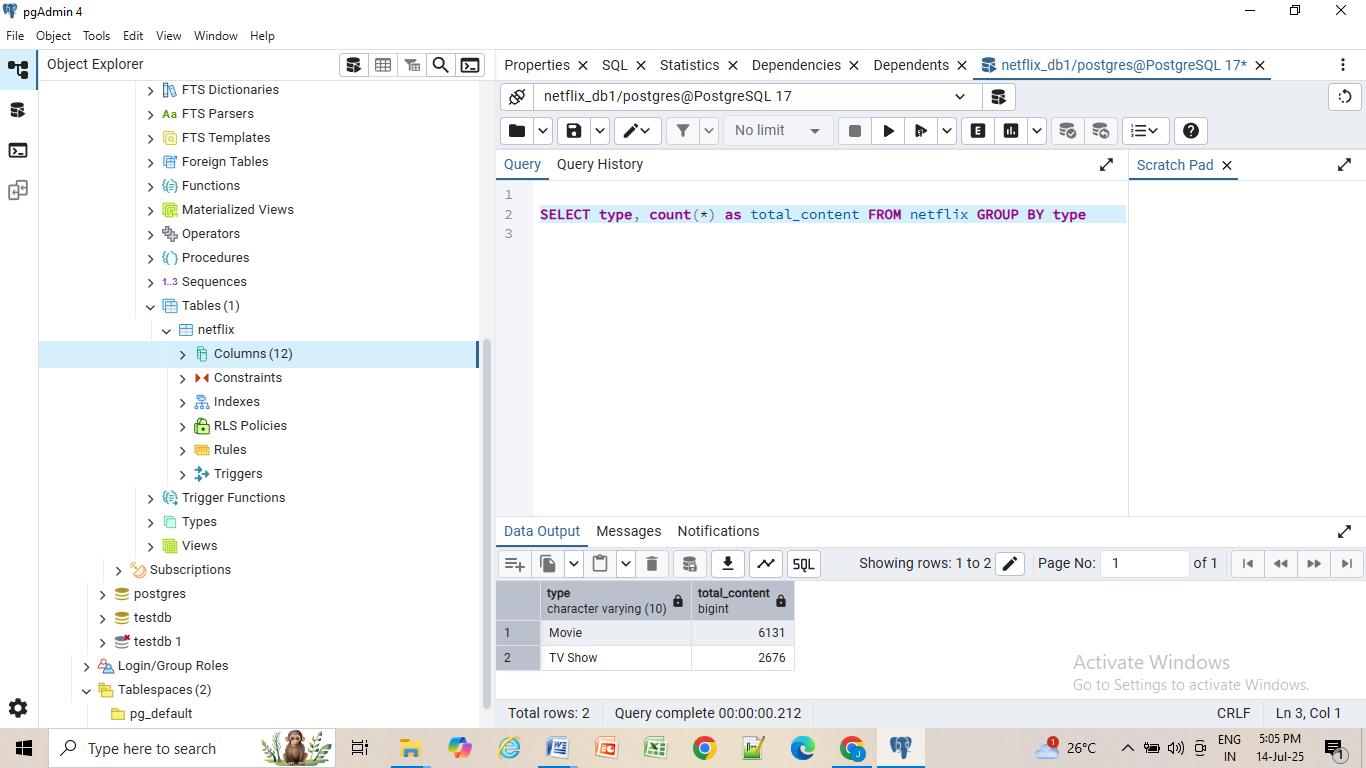
SELECT DISTINCT type FROM netflix

# Screenshot (85).png

# SOLUTIONS OF BUSINESS PROBLEMS:

# 1.Count the number of Movies vs TV Shows

SELECT type, count(\*) as total\_content FROM netflix GROUP BY type



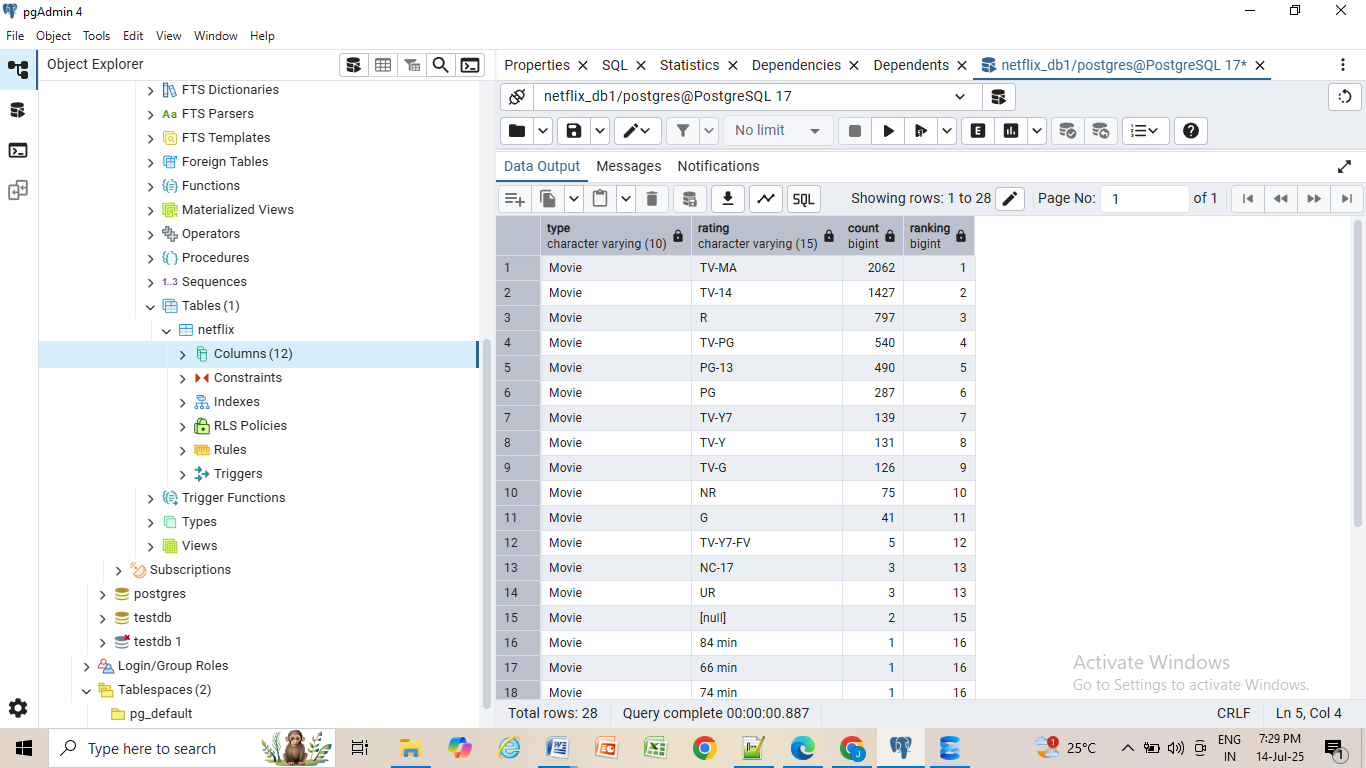
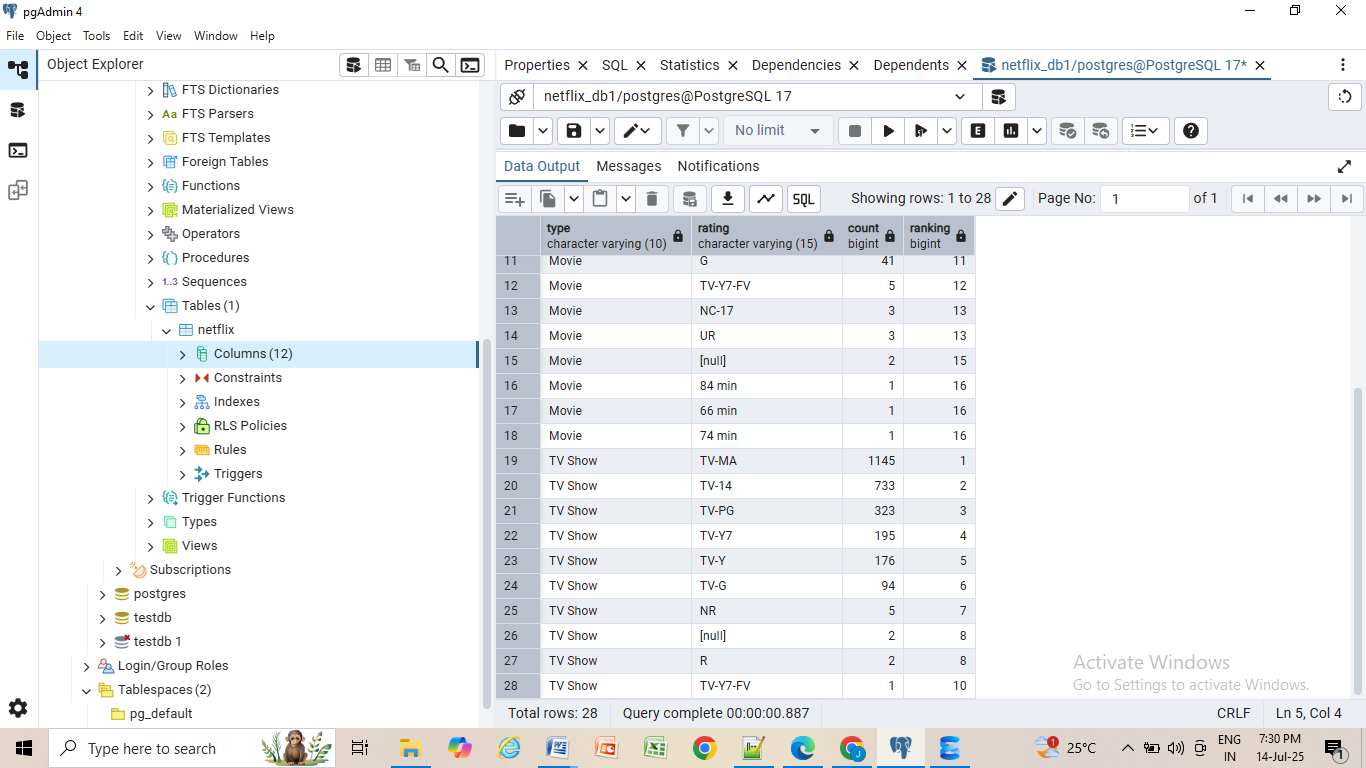
# 2. The most common rating for movies and TV shows

SELECT type,rating, count(\*),

RANK() OVER(PARTITION BY type ORDER BY count(\*) DESC) AS ranking

FROM netflix

GROUP BY 1,2



SELECT

type,

rating

FROM

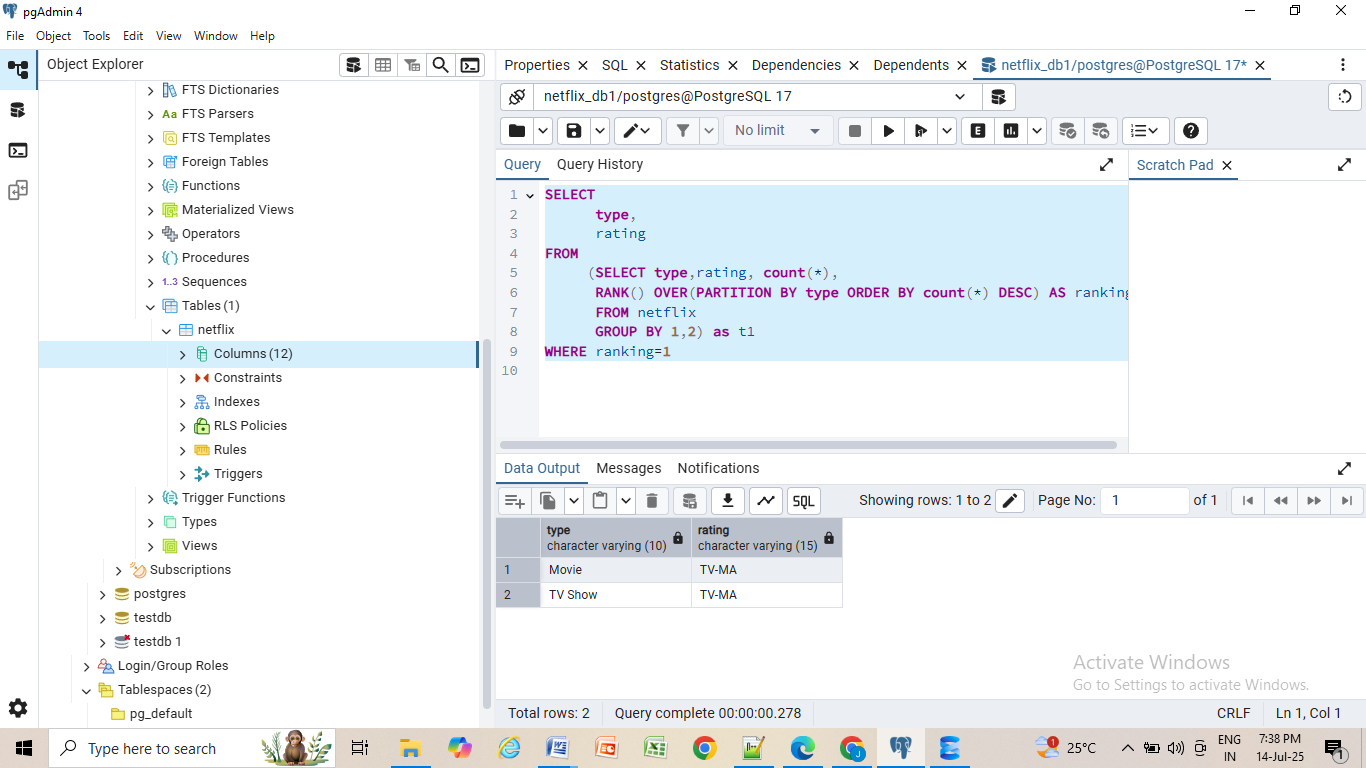
(SELECT type,rating, count(\*),

RANK() OVER(PARTITION BY type ORDER BY count(\*) DESC) AS ranking

FROM netflix

GROUP BY 1,2) as t1

WHERE ranking=1



# 3. List of all movies released in a specific year (e.g., 2020)

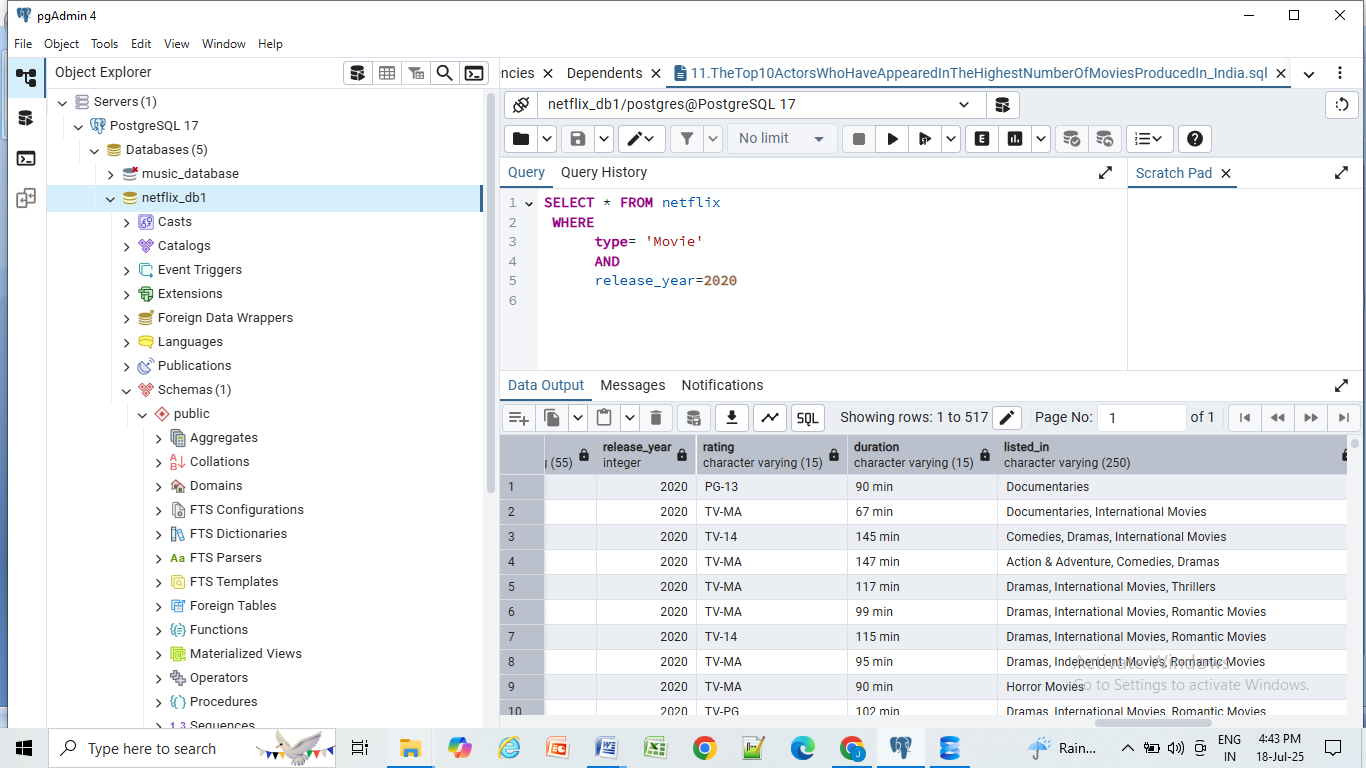
SELECT \* FROM netflix

WHERE

type= 'Movie'

AND

release\_year=2020



# 4.The top 5 countries with the most content on Netflix

SELECT

UNNEST(STRING\_TO\_ARRAY(country,',')) AS new\_country,

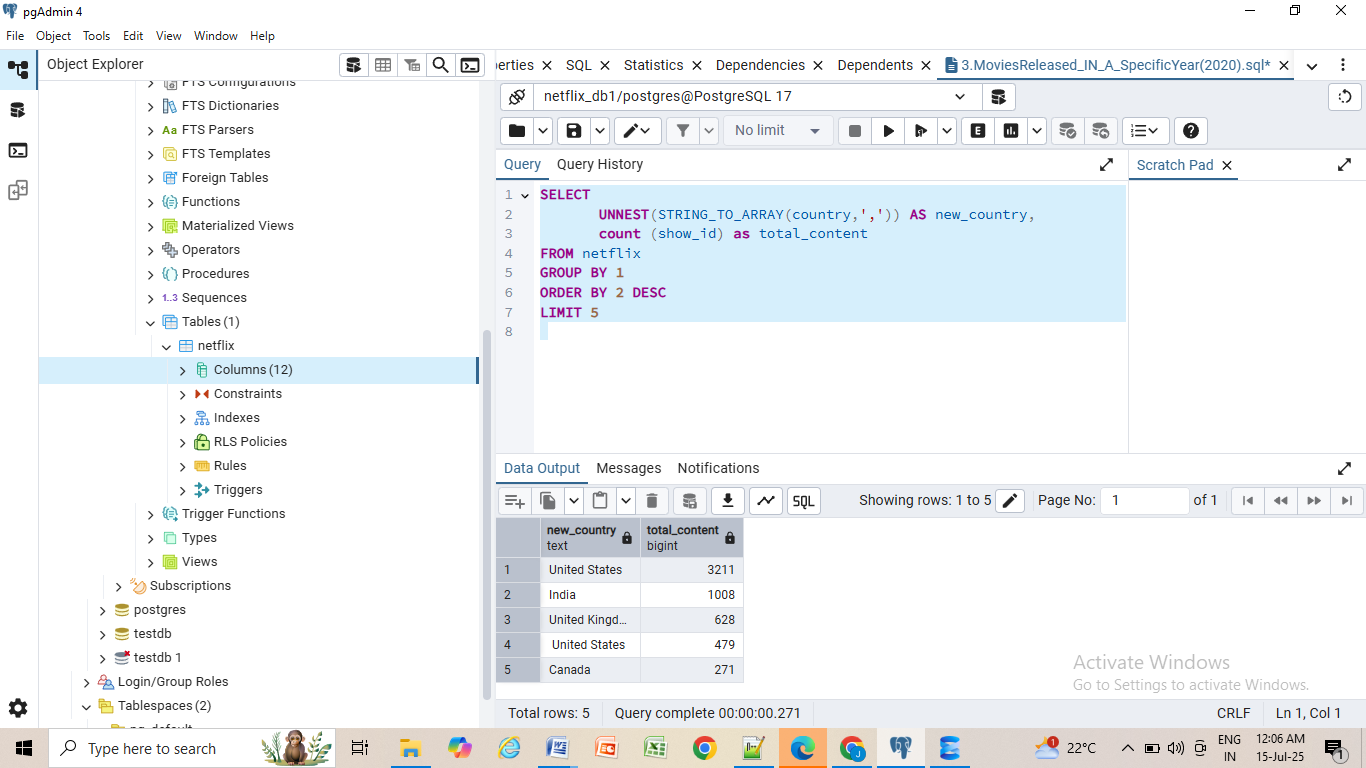
count (show\_id) as total\_content

FROM netflix

GROUP BY 1

ORDER BY 2 DESC

LIMIT 5



# 5. Identification of the longest movie

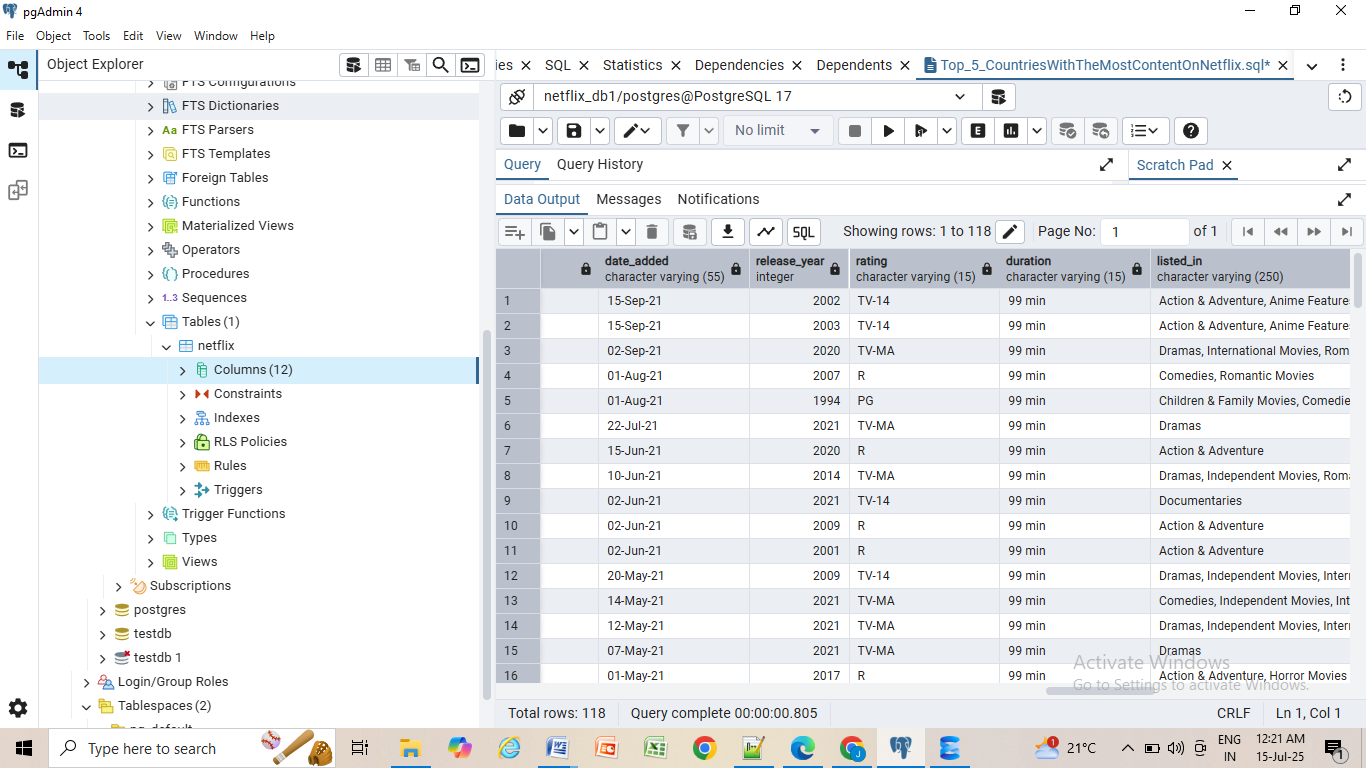
SELECT \* FROM netflix

WHERE

type='Movie'

AND

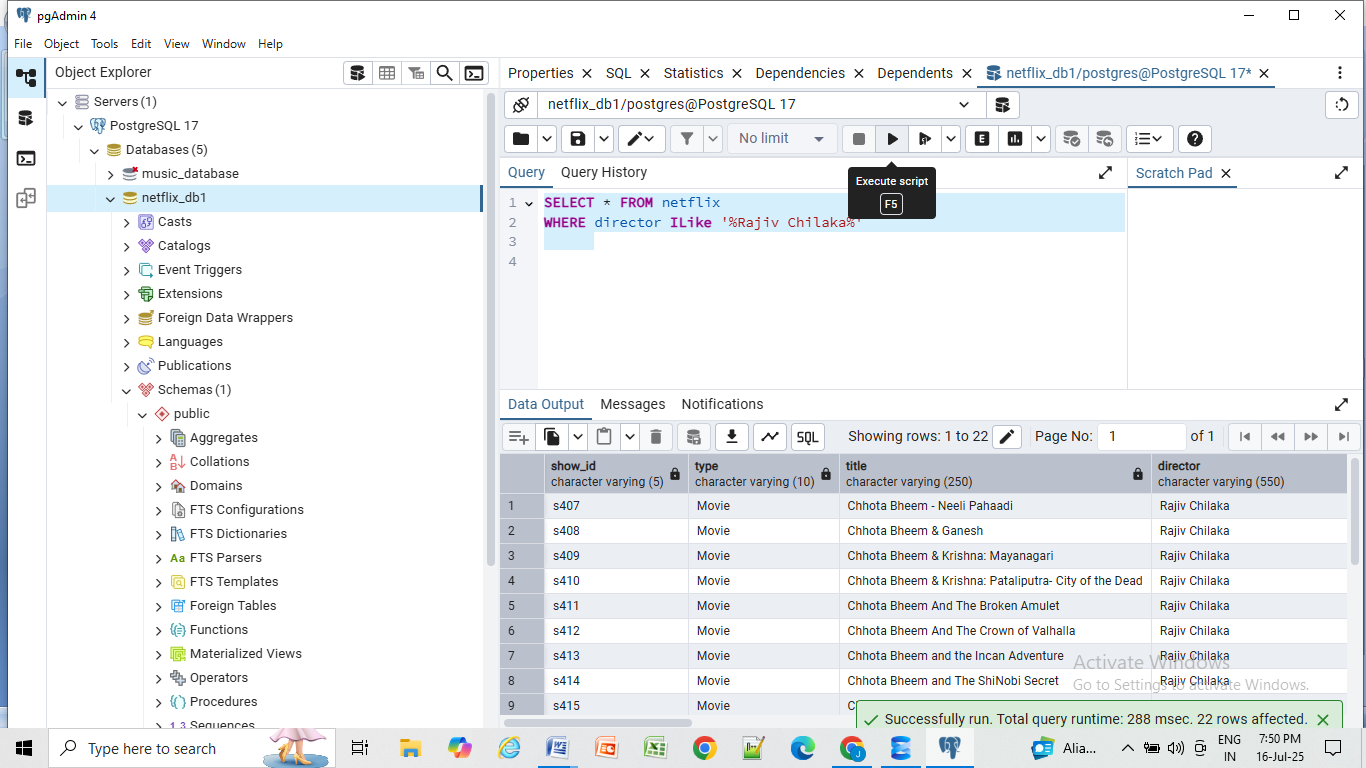
duration= (SELECT MAX(duration) FROM netflix)



# 6. All the movies/TV shows by director 'Rajiv Chilaka'

SELECT \* FROM netflix

WHERE director ILike '%Rajiv Chilaka%'



# 7. All TV shows with more than 5 seasons

SELECT \*

FROM netflix

WHERE

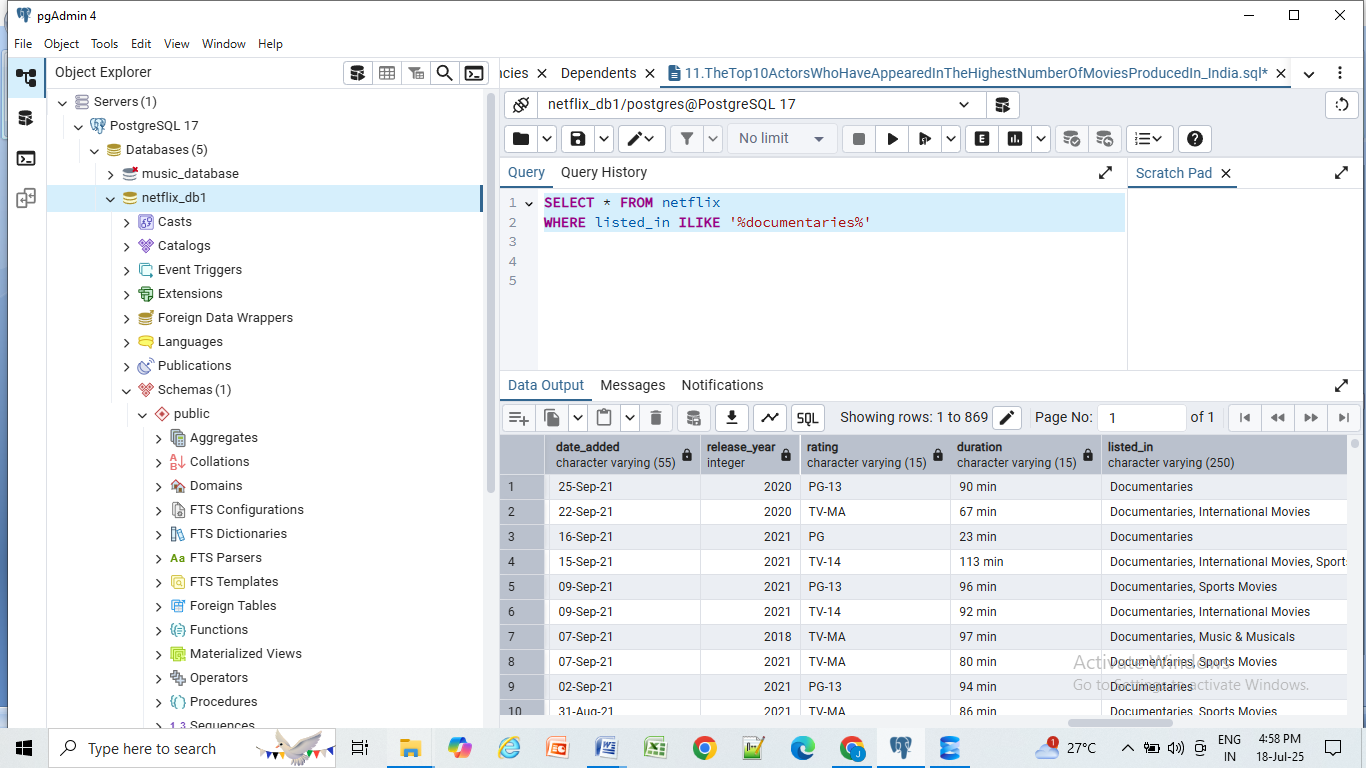
type = 'TV Show'

AND CAST(SPLIT\_PART(duration, ' ', 1) AS INTEGER) > 5;

# Screenshot (111).png8. All movies that are documentaries

SELECT \* FROM netflix

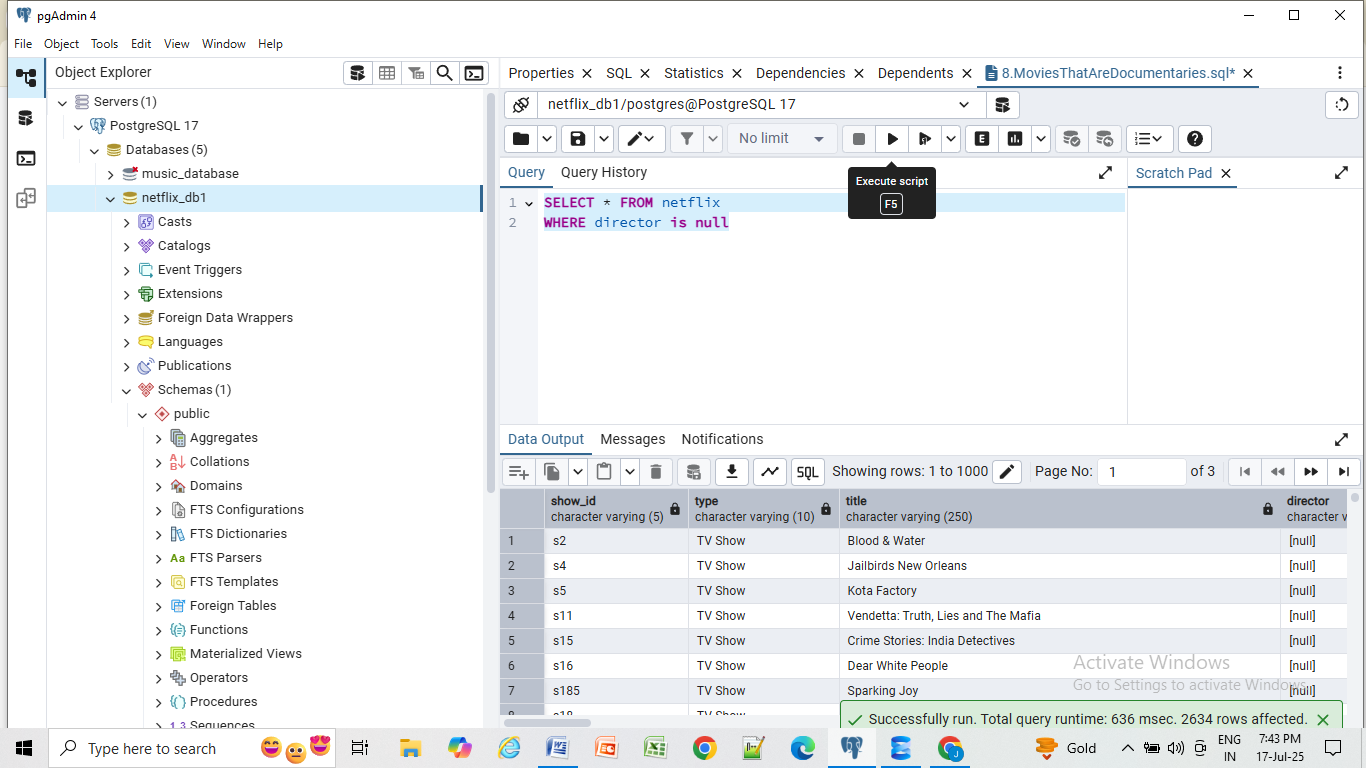
WHERE listed\_in ILIKE '%documentaries%'



# 9. All content without a director

SELECT \* FROM netflix

WHERE director is null



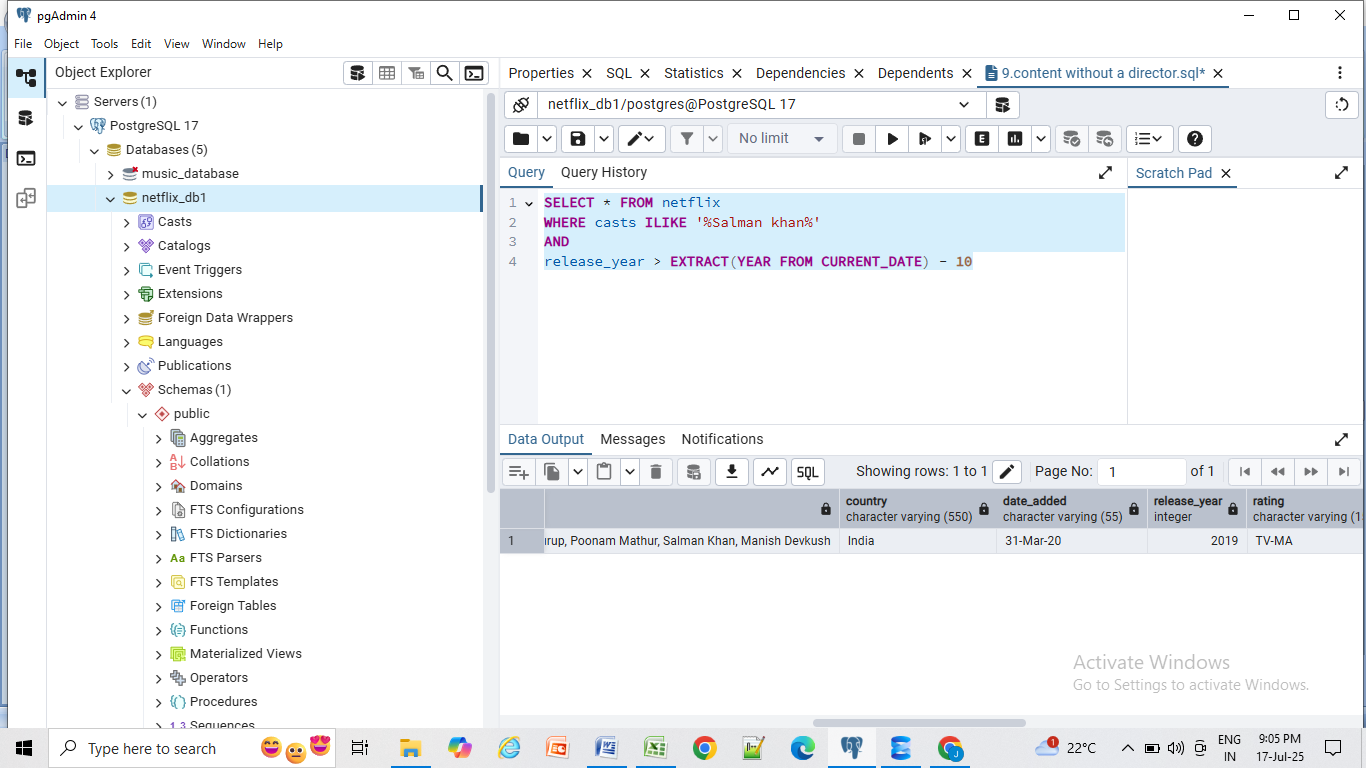
# 10. How many movies actor 'Salman Khan' appeared in last 10 years

SELECT \* FROM netflix

WHERE casts ILIKE '%Salman khan%'

AND

release\_year > EXTRACT(YEAR FROM CURRENT\_DATE) - 10



# 11. the top 10 actors who have appeared in the highest number of movies produced in India.

SELECT

UNNEST(STRING\_TO\_ARRAY(casts, ',')) as actor,

COUNT(\*)

FROM netflix

WHERE country = 'India'

GROUP BY 1

ORDER BY 2 DESC

LIMIT 10

