**Netflix Data Analysis Using SQL**

**INTRODUCTION:**

This project focuses on analyzing Netflix's extensive content library, including Movies and TV Shows, using SQL. The dataset provides detailed information such as titles, directors, cast, release years, ratings, genres, and descriptions. By addressing key business questions, the project uncovers patterns in content distribution, popular genres, and regional trends. The insights gained from this analysis aim to support strategic decisions for enhancing content curation and audience engagement.

**SQL Queries**

CREATE TABLE netflix

(

show\_id varchar(10),

type varchar(10),

title varchar(200),

director varchar(250),

casts varchar(1000),

country varchar(200),

date\_added varchar(150),

release\_year INT,

rating varchar(10),

duration varchar(20),

listed\_in varchar(300),

description varchar(600)

);

SELECT \* FROM netflix;

SELECT COUNT(\*) as total\_count

FROM netflix;

SELECT DISTINCT TYPE

FROM netflix;

**Business Problems and Solutions**

**Q1. Count the number o-f Movies vs TV Shows**

SELECT type, COUNT(\*) AS count

FROM netflix

GROUP BY type;



**Q2. Find the most common rating for movies and TV shows**

WITH RatingCounts AS (

SELECT type, rating, COUNT(\*) AS rating\_count

FROM netflix

GROUP BY type, rating

),

RankedRatings AS (

SELECT type, rating, rating\_count,

RANK() OVER (PARTITION BY type ORDER BY rating\_count DESC) AS rank

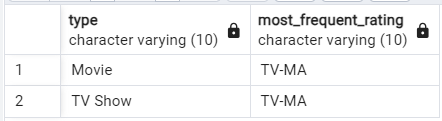
FROM RatingCounts

)

SELECT type, rating AS most\_frequent\_rating

FROM RankedRatings

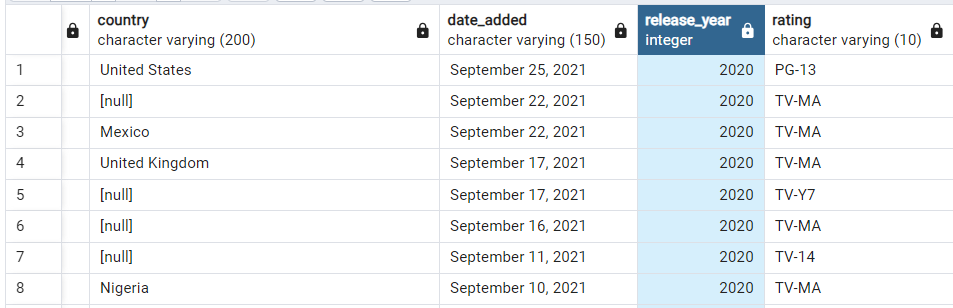
WHERE rank = 1;



**Q3. List all movies released in a specific year (e.g., 2020)**

SELECT \* FROM netflix

WHERE release\_year = 2020;



**Q4. Find the top 5 countries with the most content on Netflix**

SELECT \* FROM

(

SELECT UNNEST(STRING\_TO\_ARRAY(country, ',')) AS country,

COUNT(\*) AS total\_content

FROM netflix

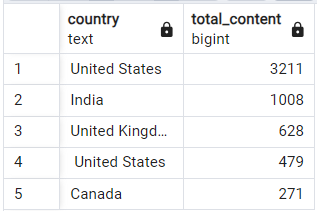
GROUP BY 1

) AS t1

WHERE country IS NOT NULL

ORDER BY total\_content DESC

LIMIT 5;

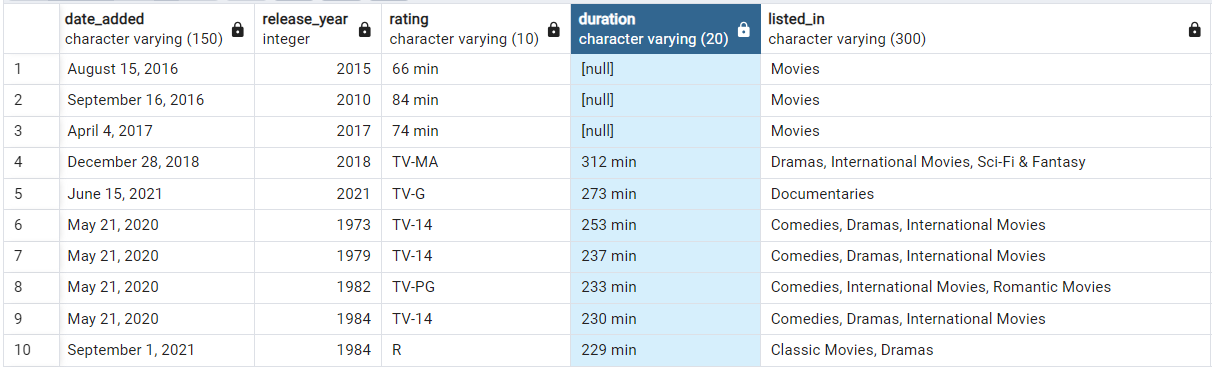


**Q5. Identify the longest movie**

SELECT \* FROM netflix

WHERE type = 'Movie'

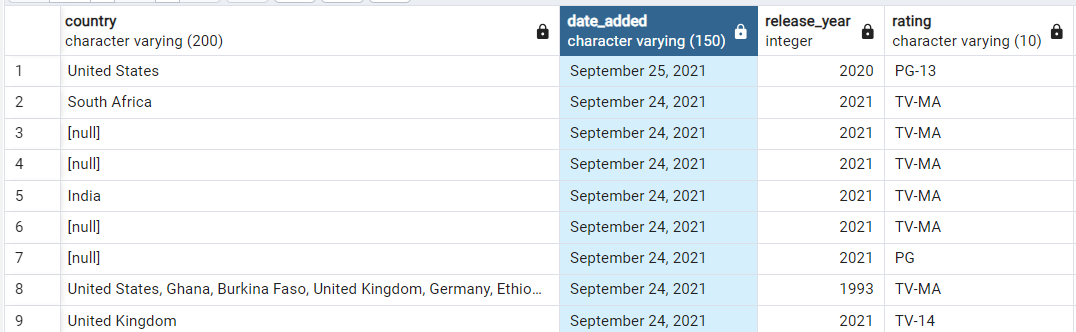
ORDER BY SPLIT\_PART(duration, ' ', 1)::INT DESC;



**Q6. Find content added in the last 5 years**

SELECT \* FROM netflix

WHERE TO\_DATE(date\_added, 'Month DD, YYYY') >= CURRENT\_DATE - INTERVAL '5 years';



**Q7. Find all the movies/TV shows by director 'Rajiv Chilaka'**

SELECT \*FROM (

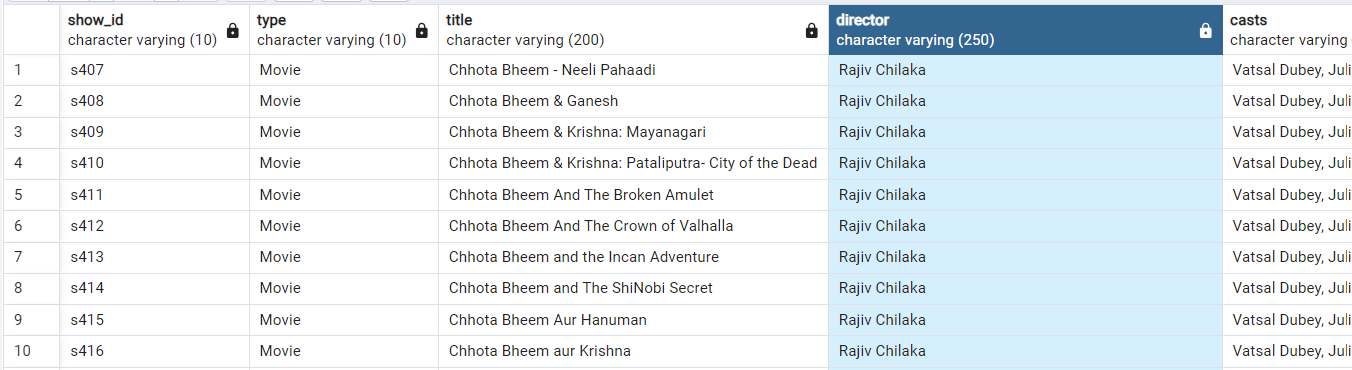
SELECT \*,

UNNEST(STRING\_TO\_ARRAY(director, ',')) AS director\_name

FROM netflix

)

WHERE director\_name = 'Rajiv Chilaka';

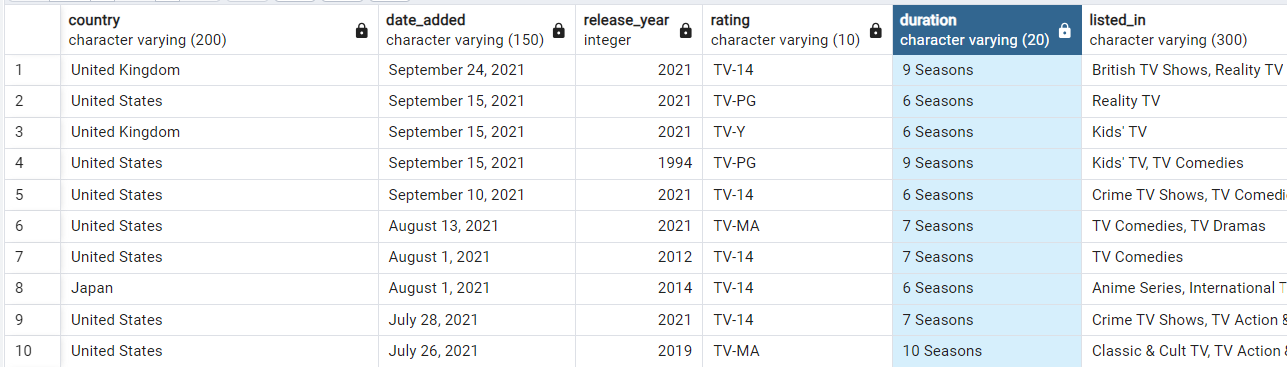


**Q8. List all TV shows with more than 5 seasons**

SELECT \*FROM netflix

WHERE type = 'TV Show'

AND SPLIT\_PART(duration, ' ', 1)::INT > 5;



**Q9. Count the number of content items in each genre**

SELECT

UNNEST(STRING\_TO\_ARRAY(listed\_in, ',')) AS genre,

COUNT(\*) AS total\_content

FROM netflix

GROUP BY 1;



**Q10. Find each year and the average numbers of content release in India on netflix.   
return top 5 year with highest avg content release!**

SELECT country, release\_year,

COUNT(show\_id) AS total\_release,

ROUND(

COUNT(show\_id)::numeric /

(SELECT COUNT(show\_id) FROM netflix WHERE country = 'India')::numeric \* 100, 2

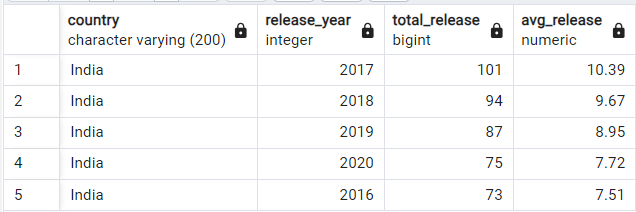
) AS avg\_release FROM netflix

WHERE country = 'India'

GROUP BY country, release\_year

ORDER BY avg\_release DESC

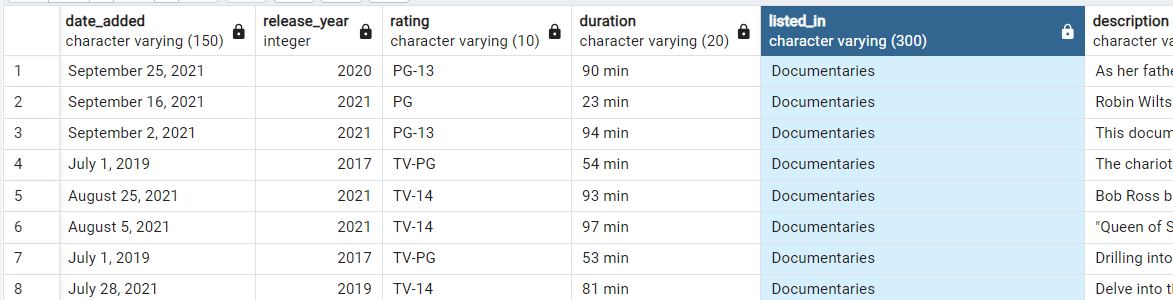
LIMIT 5;



**Q11. List all movies that are documentaries**

SELECT \* FROM netflix

WHERE listed\_in LIKE '%Documentaries';



**Q12. Find all content without a director**

SELECT \* FROM netflix

WHERE director IS NULL;



**Q13. Find how many movies actor 'Salman Khan' appeared in last 10 years!**

SELECT \* FROM netflix

WHERE casts LIKE '%Salman Khan%'

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



**Q14. Find 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(\*)

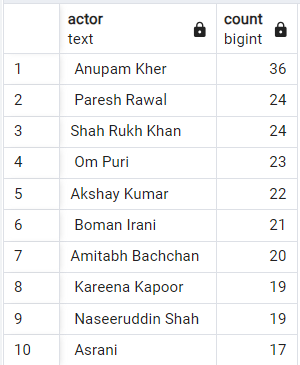
FROM netflix

WHERE country = 'India'

GROUP BY actor

ORDER BY COUNT(\*) DESC

LIMIT 10;



**Q15. Categorize Content Based on the Presence of 'Kill' and 'Violence' Keywords**

SELECT category, COUNT(\*) AS content\_count

FROM (

SELECT

CASE

WHEN description ILIKE '%kill%' OR description ILIKE '%violence%' THEN 'Bad'

ELSE 'Good'

END AS category

FROM netflix

) AS categorized\_content

GROUP BY category;



### ****Conclusion****

The analysis highlights that Netflix's content library is predominantly composed of Movies, with "TV-MA" being the most common rating for both Movies and TV Shows. Regional trends show a significant contribution from countries like India, while genres such as Documentaries and Dramas are popular. Additionally, the analysis identifies gaps in metadata, like missing director details, which can be improved. These findings provide actionable insights for optimizing Netflix's content strategy and user satisfaction.