**Netflix Movies and TV Shows Data Analysis using SQL**



**Overview**

This project involves a comprehensive analysis of Netflix's movies and TV shows data using SQL. The goal is to extract valuable insights and answer various business questions based on the dataset. The following README provides a detailed account of the project's objectives, business problems, solutions, findings, and conclusions.

**Objectives**

* Analyze the distribution of content types (movies vs TV shows).
* Identify the most common ratings for movies and TV shows.
* List and analyze content based on release years, countries, and durations.
* Explore and categorize content based on specific criteria and keywords.

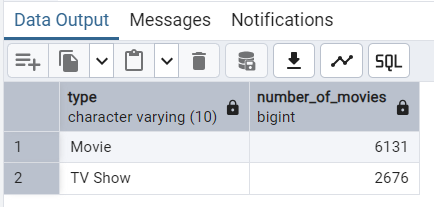
**1. Count the number of Movies vs TV Shows**

select

type , count(show\_id) as number\_of\_movies

from netflix

group by 1



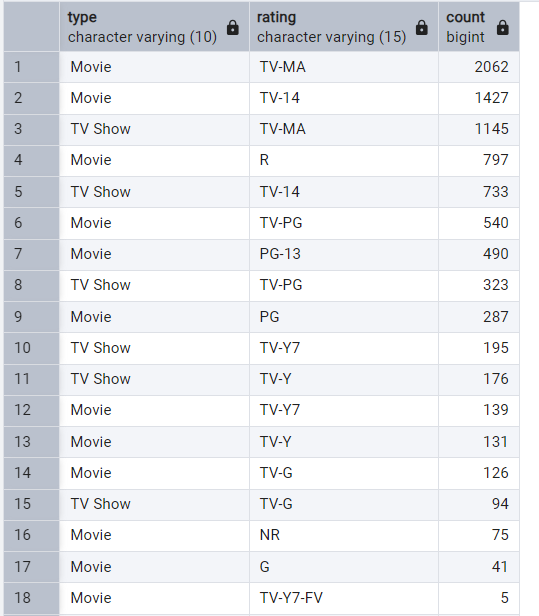
**2. Find the most common rating for movies and TV shows**

select type ,rating , count(rating)

from netflix

group by 1,2

order by 3 desc



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

select release\_year, title

from netflix

where release\_year = 2020



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

select unnest(string\_to\_array(country, ',')) as contry\_name,

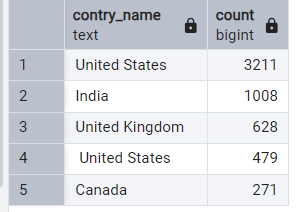
count(show\_id)

from netflix

group by 1

order by 2 desc

limit 5



**5. Identify the longest movie**

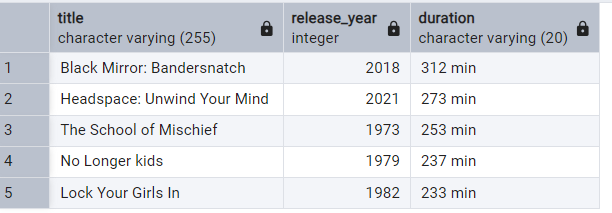
select title, release\_year, duration

from netflix

where type = 'Movie' and duration is not null

order by split\_part(duration,' ',1)::int desc

limit 5

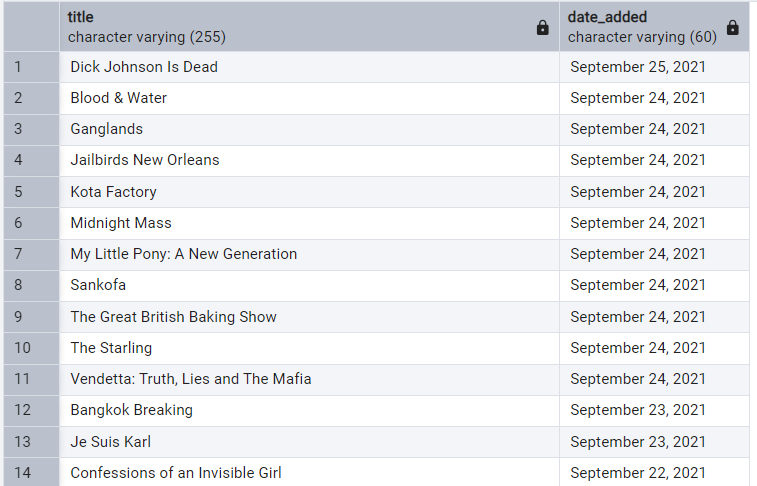


**6. Find content added in the last 5 years**

select title , date\_added

from netflix

where to\_date(date\_added ,'month dd, yyyy') >= current\_date - interval '5 years'



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

1st Method

select \*

from (select title, type,director,

unnest(string\_to\_array(director,',')) as director\_name

from netflix)

where director\_name = 'Rajiv Chilaka'

--------------------------------------------------------------

2nd method

select title, type, director

from netflix

where director like '%Rajiv Chilaka%'



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

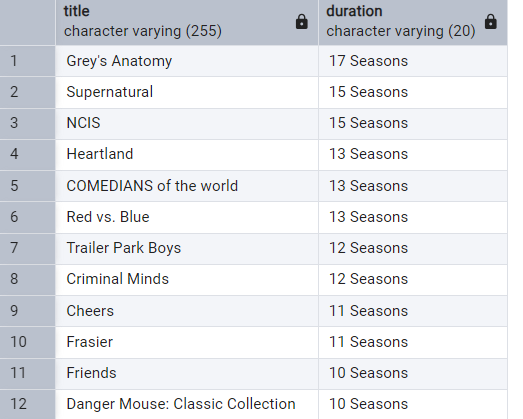
select title, duration

from netflix

where type = 'TV Show'

and split\_part(duration,' ',1)::int > 5

order by split\_part(duration,' ',1)::int desc



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

select unnest(string\_to\_array(listed\_in,', ')) as genre, count(show\_id)

from netflix

group by 1

order by 2 desc



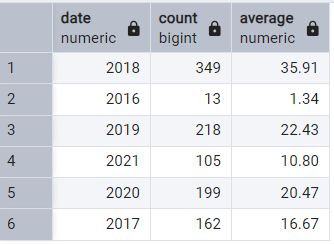
**10.Find each year and the average numbers of content release in India on netflix.**

**return top 5 year with highest avg content release.**

select extract(year from to\_date(date\_added, 'month dd, yyyy')) as date,

count(\*),

round(count(\*)::numeric/(select count(\*) from netflix where country = 'India'):: numeric \*100,2) as average

from netflix

where country like '%India%'

group by 1

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

select \* from

(select title, unnest(string\_to\_array(listed\_in,', ')) as genre

from netflix)

where genre = 'Documentaries'

------------------------------------------------------------

select title , listed\_in

from netflix

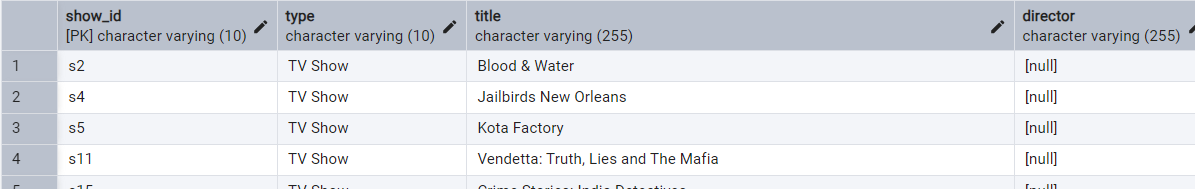
where listed\_in like '%Documentaries%'



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

select \* from netflix

where director is null



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

select title, to\_date(date\_added,'month dd, yyyy'),casts

from netflix

where casts like '%Salman Khan%'

and to\_date(date\_added,'month dd, yyyy') >= current\_date - interval '10 years'



**14. Find the top 10 actors who have appeared in the highest number of movies produced in India.**

select unnest(string\_to\_array(casts,', ')) as actors,

Count(\*)

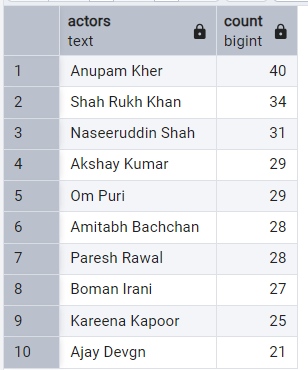
from netflix

where country like '%India%'

group by 1

order by 2 desc

limit 10



**15. Categorize the content based on the presence of the keywords 'kill' and 'violence' in the description field. Label content containing these keywords as 'Bad' and all other content as 'Good'. Count how many items fall into each category.**

select

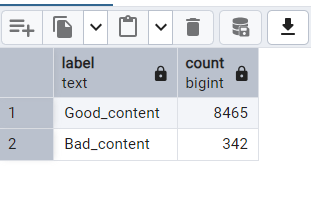
case when description Ilike '%violence%' then 'Bad\_content'

when description ilike '%kill%' then 'Bad\_content'

else 'Good\_content'

end as label, count(\*)

from netflix

group by 1