**NETFLIX DATASET ANALYSIS**

USING POSTGRE SQL

Prepared By: Ankit Kumar Rajak

**Introduction**

This report explores Netflix’s extensive content library to uncover trends and viewers behavior. By analyzing data on popular genres, release dates, and audience engagement. It aims to provide actionable insights into viewing preferences and content strategy.

Questions

1.Count the numbers of movies and TV shows

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

3.List all movies released in a specific year (eg. 2020)

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

5.Identify the longest movie

6. Find content added in the last 5 years

7.Find all the movies/TV shows by director ‘Rajiv Chilaka’

8.List all tv shows with more than 5 seasons

9.Count the no. of content items in each genre

10.find each year and the average number of content release by india on Netflix and return top 5 year with highest average content release

11.list all movies that are documentaries

12.find all content without a director

13. find how many movies actor 'salman khan' appeared in last 10 years

14. find the top 10 actors who have appeared in the highest number of movies produced in india

15.Categorize the content based on the presence of the keywords ‘kill' and 'violence' in the description field. label content containing the keyword as 'Bad' and all other content as 'Good'. count how many items fall into these category

**1.Count the numbers of movies and TV shows**

Select type, count(\*) from netflix  
group by type

|  |  |
| --- | --- |
| type | Count |
| Movie | 6131 |
| TV Show | 2676 |

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

Select type, rating, Total\_rating

from(

select

type,

rating,

count(\*) as Total\_rating,

rank() over (partition by type order by count(\*) desc) as ranking

from netflix

group by 1, 2

)

where ranking=1

|  |  |  |
| --- | --- | --- |
| type | rating | Total\_rating |
| Movie | TV-MA | 2062 |
| TV Show | TV-MA | 1145 |

**3. List all movies released in a specific year (eg. 2020)**

Select \* from netflix

where

type= 'Movie’

and

release\_year=2020



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

select

trim(unnest(string\_to\_array(country,',') ) ) as new\_country,

count(\*)

From netflix

Group by new\_country

Order by 2 desc

Limit 5

|  |  |
| --- | --- |
| New\_country | count |
| United States | 3690 |
| India | 1046 |
| United Kingdom | 806 |
| Canada | 445 |
| France | 393 |

**5. Identify the longest movie**

select \* from netflix

where

type='Movie' and

duration is not null

order by duration desc



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

select \*, to\_date(date\_added, 'Month DD YYYY')from netflix

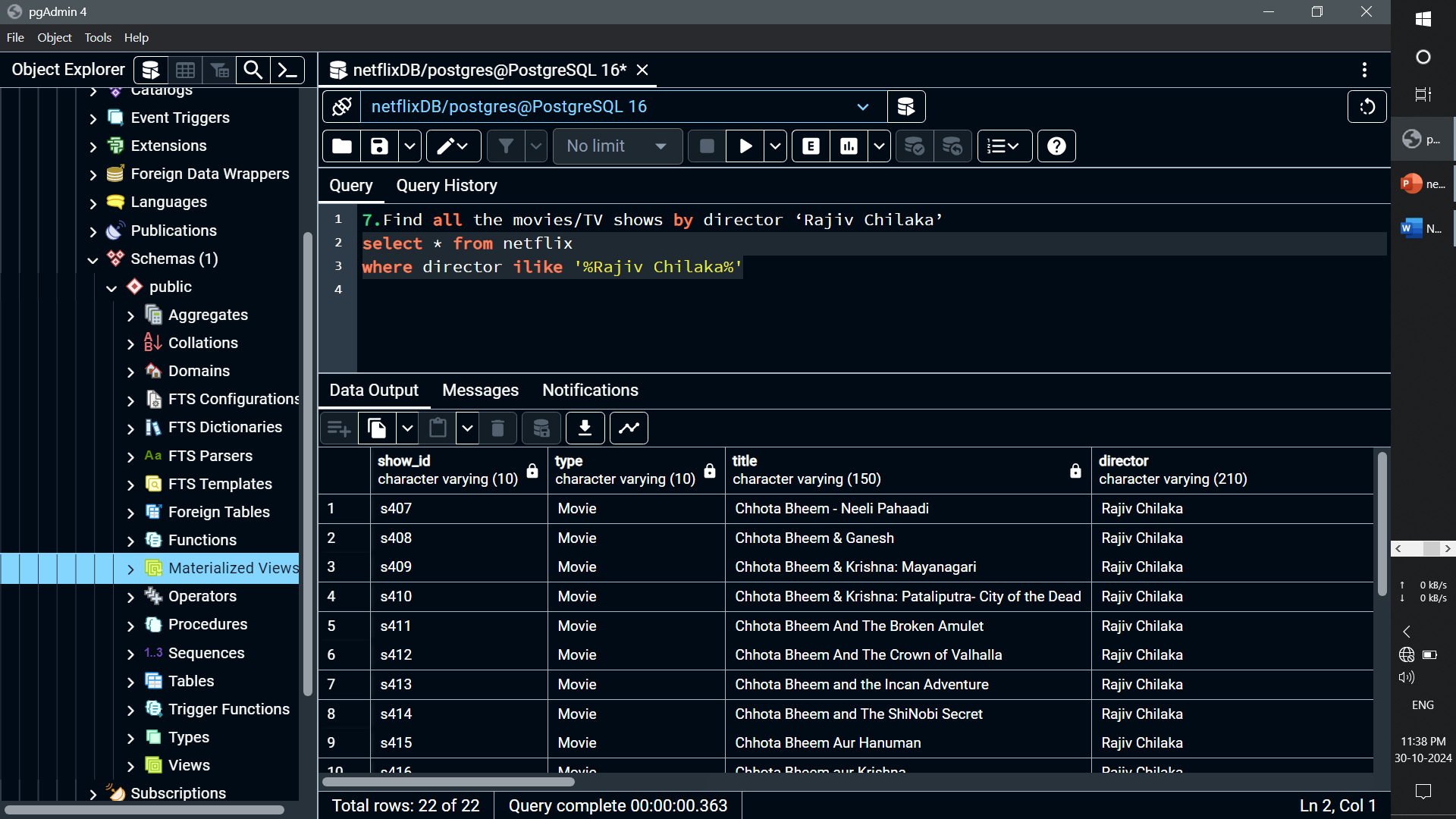
where to\_date(date\_added, 'Month DD YYYY' ) >= current\_date - interval '5year'



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

select \* from netflix

where director ilike '%Rajiv Chilaka%'



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

select \* from netflix

where

type = 'TV Show'

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



**9. Count the no. of content items in each genre**

select

trim(unnest(string\_to\_array(listed\_in,','))) as Genre,

count(\*)

from netflix

group by 1

order by 2 desc

**10.find each year and the average number of content release by india on Netflix and return top 5 year with highest average content release**

Select

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

round(count(\*)::numeric /

(select count(\*) from netflix where country='India')::numeric\*100,2)as Average\_content

from netflix

where country= 'India'

group by 1

order by 2 desc

limit 5

|  |  |  |
| --- | --- | --- |
| Years | Count | Average\_content |
| 2018 | 333 | 34.16 |
| 2019 | 203 | 20.88 |
| 2020 | 189 | 19.44 |
| 2017 | 142 | 14.61 |
| 2021 | 95 | 9.77 |

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

select \* from netflix

where listed\_in ilike '%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**

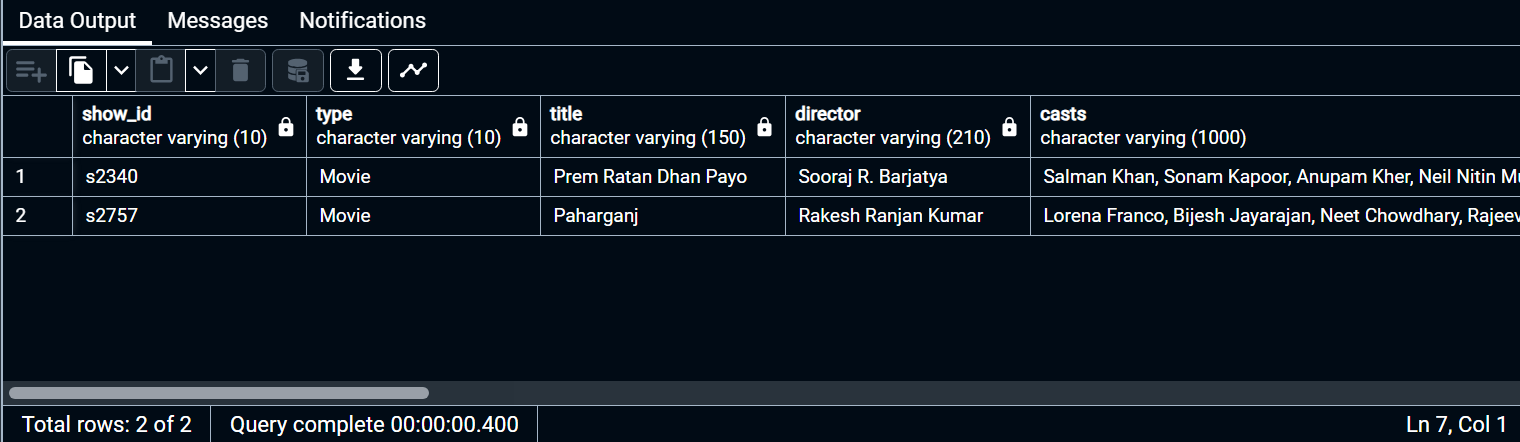
select \* from netflix

where

casts ilike '%salman khan%'

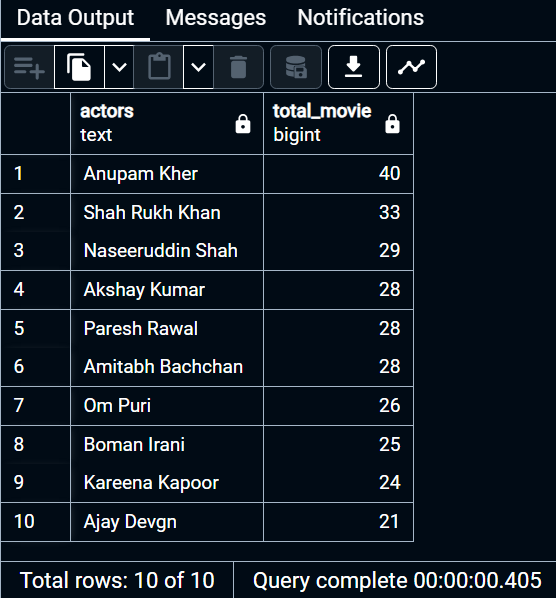
and

release\_year > extract(year from current\_date)-10



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

select

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

count(\*) as total\_movie from netflix

where country ilike '%india'

group by actors

order by total\_movie desc

limit 10

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

with cat\_table as(

select \* ,

case when

description ilike '%kill%' or description ilike '%violen%'

then 'Bad\_film'

else 'Good\_Film' end category

from netflix )

select category , count(\*) from cat\_table

group by category

|  |  |
| --- | --- |
| Category | Count |
| Bad\_film | 409 |
| Good\_Film | 8398 |

THANK YOU