



**NETFLIX**

# SQL PROJECT ON NETFLIX CONTENT

**Presented by Firdous Rahmani**

## **Project Overview**

This project focuses on analyzing Netflix's content library using SQL to extract meaningful insights. By examining movies and TV shows, the objective is to understand content distribution, genre popularity, country-wise production trends, and actor/director contributions.

### **Objectives**

- Analyze the distribution of Movies vs. TV Shows
- Identify the most common ratings and top content-producing countries
- Determine the longest movies and TV shows on Netflix
- Examine trends in content added over the last five years
- Categorize content based on specific keywords
- Gain insights into popular actors, directors, and genres



# COUNT THE NUMBER OF MOVIES VS TV SHOWS.

```
select type,  
       count(*) as total_count  
  from netflix  
 group by type;
```

|   | type<br>character varying (10)  | total_count<br>bigint  |
|---|--|---|
| 1 | Movie  | 6131  |
| 2 | TV Show  | 2676  |



# FIND THE MOST COMMON RATING FOR MOVIES AND TV SHOWS?

```
select
    type,
    mode() within group (order by rating) as most_common_rating
    from netflix
    group by type;
```

|   | type<br>character varying (10)  | most_common_rating<br>character varying  |
|---|--|---|
| 1 | Movie  | TV-MA   |
| 2 | TV Show  | TV-MA   |



# LIST ALL MOVIES RELEASED IN A SPECIFIC YEAR (E.G. 2020)

```
select title, release_year  
from netflix  
where type = 'Movie'  
and  
release_year = 2020;
```

|   | title   | release_year |
|---|---|--------------|
|   | character varying (150)                             | integer      |
| 1 | Dick Johnson Is Dead                                | 2020         |
| 2 | Europe's Most Dangerous Man: Otto Skorzeny in Spain | 2020         |
| 3 | Tughlaq Durbar                                      | 2020         |
| 4 | Omo Ghetto: the Saga                                | 2020         |
| 5 | Shadow Parties                                      | 2020         |
| 6 | Here and There                                      | 2020         |
| 7 | Shikara   | 2020         |



# FIND THE TOP 5 COUNTRIES WITH THE MOST CONTENT ON NETFLIX.

```
with count_of_country as (
  select
    trim(unnest(string_to_array(country, ','))) as countries
    from netflix
)
select countries, count(*) as total_content
from count_of_country
group by countries
order by total_content desc
limit 5;
```

|   | countries      | total_content |
|---|----------------|---------------|
|   | text           | bigint        |
| 1 | United States  | 3690          |
| 2 | India          | 1046          |
| 3 | United Kingdom | 806           |
| 4 | Canada         | 445           |
| 5 | France         | 393           |



# IDENTIFY THE LONGEST MOVIE

```
select title, duration from netflix
where
    type = 'Movie'
    and
    duration = (select MAX(duration) from netflix)
order by duration desc
limit 1;
```

|   | <b>title</b><br>character varying (150)                   |  <b>duration</b><br>character varying (15)  |
|---|---|---|
| 1 | InuYasha the Movie 2: The Castle Beyond the Looking Glass | 99 min  |



# FIND CONTENT ADDED IN THE LAST 5 YEARS.

```
select * from netflix
where
TO_DATE(date_added, 'Month DD, YYYY') >= current_date - Interval '5 years';
```

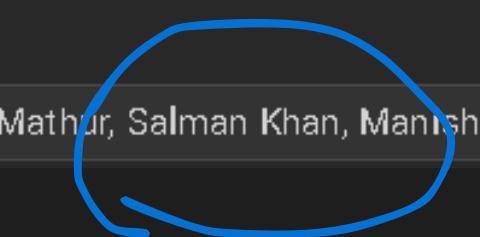
|   | show_id<br>character varying (6)  | type<br>character varying (10)  | title<br>character varying (150)  | director<br>character varying (208) |
|---|---|--|--|-------------------------------------|
| 1 | s1  | Movie  | Dick Johnson Is Dead   | Kirsten Johnson                     |
| 2 | s2  | TV Show  | Blood & Water  | [null]                              |
| 3 | s3  | TV Show  | Ganglands  | Julien Leclercq                     |
| 4 | s4  | TV Show  | Jailbirds New Orleans  | [null]                              |
| 5 | s5  | TV Show  | Kota Factory   | [null]                              |
| 6 | s6  | TV Show  | Midnight Mass  | Mike Flanagan                       |
| 7 | s7  | Movie  | My Little Pony: A New Generation   | Robert Cullen, José Luis Ucha       |
| 8 | s8  | Movie  | Sankofa  | Haile Gerima                        |



NETFLIX

# FIND HOW MANY MOVIES ACTOR 'SALMAN KHAN' APPEARED IN THE LAST 10 YEARS

```
select title, casts from netflix
where
    casts ilike '%Salman Khan%'
AND
    release_year > extract(YEAR FROM CURRENT_DATE) - 10;
```

|   | title<br>character varying (150)  | casts<br>character varying (1000)   |  |
|---|--|---|---|
| 1 | Paharganj  | Lorena Franco, Bijesh Jayarajan, Neet Chowdhary, Rajeev Goursingh, Karran Jeet, Rajesh Sharma, Gautam Kurup, Poonam Mathur, <b>Salman Khan</b> , Manish Devkush |  |



# LIST ALL TV SHOWS WITH MORE THAN 5 SEASONS

```
select type, title, duration  
from netflix  
where type = 'TV Show'  
and  
Split_part(duration, ' ', 1)::numeric > 5;
```

|   | type<br>character varying (10)  | title<br>character varying (150)  | duration<br>character varying (15)  |
|---|--|--|--|
| 1 | TV Show  | The Great British Baking Show  | 9 Seasons  |
| 2 | TV Show  | Nailed It  | 6 Seasons  |
| 3 | TV Show  | Numberblocks   | 6 Seasons  |
| 4 | TV Show  | Saved by the Bell  | 9 Seasons  |
| 5 | TV Show  | Lucifer  | 6 Seasons  |
| 6 | TV Show  | Grace and Frankie  | 7 Seasons  |
| 7 | TV Show  | On Books   | 7 Seasons  |



# FIND EACH YEAR AND THE AVERAGE NUMBER OF CONTENT RELEASED IN INDIA ON NETFLIX.

## -- RETURN TOP 5 YEARS WITH HIGHEST AVERAGE CONTENT RELEASE

```
select
    EXTRACT(YEAR from to_date(date_added, 'Month DD, YYYY')) as Year,
    count(*) as yearly_count ,
    Round(
        count()*:numeric/(select count(*) from netflix where country = 'India')::numeric * 100, 2)
    as avg_content_per_year

from netflix
where country ILIKE '%India%'
group by Year
ORDER BY avg_content_per_year desc
limit 5;
```

|   | year<br>numeric  | yearly_count<br>bigint  | avg_content_per_year<br>numeric  |
|---|---|--|---|
| 1 | 2018  | 349  | 35.91   |
| 2 | 2019  | 218  | 22.43   |
| 3 | 2020  | 199  | 20.47   |
| 4 | 2017  | 162  | 16.67   |
| 5 | 2021  | 105  | 10.80   |



# FIND THE TOP 10 ACTORS WHO HAVE APPEARED IN THE HIGHEST NUMBER OF MOVIES PRODUCED IN INDIA.

```
select
```

```
unnest(string_to_array(casts, ',')) as top_actors,  
count(*) as film_done  
from netflix  
where country ILIKE '%India%'  
group by 1  
order by film_done desc  
limit 10;
```

|    | top_actors        | film_done |
|----|-------------------|-----------|
| 1  | Anupam Kher       | 36        |
| 2  | Om Puri           | 26        |
| 3  | Boman Irani       | 25        |
| 4  | Paresh Rawal      | 25        |
| 5  | Shah Rukh Khan    | 25        |
| 6  | Akshay Kumar      | 23        |
| 7  | Naseeruddin Sh... | 20        |
| 8  | Amitabh Bachch... | 20        |
| 9  | Kareena Kapoor    | 20        |
| 10 | Asrani            | 17        |



**CATEGORIES 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.**

```
With new_table
AS
(select *,
    case
    when
        description ilike '%kill%'
    or
        description ilike '%violence%' THEN 'Bad_Content'
    else 'Good_content'
    End category
From Netflix
)
SELECT
    category,
    COUNT(*) as total_content
from new_table
Group by 1
```

|   | category     | total_content |
|---|--------------|---------------|
| 1 | Good_content | 8465          |
| 2 | Bad_Content  | 342           |



NETFLIX



# THANK YOU FOR YOUR ATTENTION

[My LinkedIn](#)

[My Github](#)