

# NETFLIX

## Movies Data Analysis

SQL project  
-Ritoja Kundu

# Project Overview

- Analyze a Netflix movies dataset to uncover insights about content trends, types, ratings, and more .
- Query Tool used : PostgreSQL
- Dataset used : csv Kaggle dataset
- Why SQL?
  - Efficient querying of structured data
  - Real-world practice with filtering, aggregation, text search, and conditional logic
  - Great foundation for data analytics and BI dashboards

# Project Objectives

- Rank movies and TV shows by rating.
- Identify countries having the most content uploaded on Netflix.
- Count the number of movies released per year.
- Group movies based on the Genre.
- Find the actors having starred in the most number of movies.
- Categorize content based on certain words in the description.
- Find the average number of content released in India.

# Creating the Table in SQL

```
CREATE TABLE netflix
(
    show_id      VARCHAR(5),
    show_type     VARCHAR(10),
    title        VARCHAR(250),
    director     VARCHAR(550),
    show_cast    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)
);
```

Table headings created.

Data imported from csv file.

Dataset contains 8807 rows, 12 columns.

# Questions

1. Count the Number of Movies vs TV Shows
2. Find the Most Common Rating for Movies and TV Shows
3. List All Movies Released in a Specific Year
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 Movies/TV Shows by Director 'Rajiv Chilaka'
8. List All TV Shows with More Than 5 Seasons
9. Count the Number of Content Items in Each Genre
10. Find each year and the average numbers of content release in India on netflix.
11. List All Movies that are Documentaries
12. Find All Content Without a Director
13. Find How Many Movies Actor 'Salman Khan' Appeared in the Last 10 Years
14. Find the Top 10 Actors Who Have Appeared in the Highest Number of Movies Produced in India
15. Categorize Content Based on the Presence of 'Kill' and 'Violence' Keywords

# Solutions

```
--Count the Number of Movies vs TV Shows
SELECT show_type,
       COUNT(*) AS count_type
FROM netflix
GROUP BY show_type;
```

```
--List All Movies Released in a Specific Year
WITH movie_list AS (SELECT title, release_year
                     FROM netflix
                     WHERE show_type LIKE 'Movie')
SELECT COUNT(title), release_year
      FROM movie_list
      GROUP BY release_year
      ORDER BY COUNT(title) DESC
```

```
--Find the Most Common Rating for Movies and TV Shows
SELECT show_type, rating, count_rating
FROM (
    SELECT show_type, rating, COUNT(*) AS count_rating
          RANK() OVER (PARTITION BY show_type
                        ORDER BY COUNT(*) DESC) AS rnk
     FROM netflix
     GROUP BY show_type, rating
) ranked
WHERE rnk = 1;
```

```
--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;
```

--Identify the Longest Movie

```
SELECT title, duration FROM (SELECT title, duration
FROM netflix
WHERE show_type = 'Movie')
WHERE duration IS NOT null
ORDER BY SPLIT_PART(duration, ' ', 1)::INT DESC
LIMIT 1;
```

--List All TV Shows with More Than 5 Seasons

```
SELECT title, duration FROM (SELECT * FROM netflix
WHERE show_type = 'TV Show') as TV_shows
WHERE SPLIT_PART(duration, ' ', 1)::INT > 5;
```

--Find All Movies/TV Shows by Director 'Rajiv Chilaka'

```
SELECT title FROM netflix
WHERE director = 'Rajiv Chilaka';
```

--Count the Number of Content Items in Each Genre

```
SELECT UNNEST(STRING_TO_ARRAY(listed_in, ',')) AS genre,
      COUNT(*)
FROM netflix
GROUP BY genre
```

--Find Content Added in the Last 5 Years

```
SELECT * FROM netflix
SELECT *
FROM netflix
WHERE TO_DATE(date_added, 'Month DD, YYYY') >= CURRENT_DATE - INTERVAL '5 years';
```

```
--Find each year and the average numbers of content release in India on netflix.  
SELECT release_year,  
    ROUND(AVG(content_count),2) FROM  
        (SELECT release_year,  
        UNNEST(STRING_TO_ARRAY(country, ',')) AS country,  
        COUNT(*) AS content_count  
    FROM netflix  
    WHERE country = 'India'  
    GROUP BY country, release_year)  
GROUP BY country, release_year  
ORDER BY AVG(content_count) DESC;
```

```
--Find All Content Without a Director  
SELECT * FROM netflix  
WHERE director IS NULL
```

```
--List All Movies that are Documentaries  
SELECT COUNT(title) FROM  
    (SELECT title, UNNEST(STRING_TO_ARRAY(listed_in, ',')) AS genre  
    FROM netflix  
    WHERE show_type = 'Movie')  
WHERE genre LIKE '%Documentaries%';
```

```
--Find How Many Movies Actor 'Salman Khan' Appeared in the Last 10 Years  
SELECT title FROM  
    (SELECT title,  
    UNNEST(STRING_TO_ARRAY(show_cast, ',')) AS casts,  
    release_year  
    FROM netflix  
    WHERE release_year >= EXTRACT(YEAR FROM CURRENT_DATE) - 10)  
WHERE casts LIKE '%Salman Khan%';
```

```
--Find the Top 10 Actors Who Have Appeared in the Highest Number of Movies Produced in India
SELECT casts, movie_count FROM
  (SELECT UNNEST(STRING_TO_ARRAY(show_cast, ',')) AS casts,
         UNNEST(STRING_TO_ARRAY(country, ',')) AS country,
         COUNT(*) AS movie_count
   FROM netflix
  WHERE show_type = 'Movie'
    GROUP BY casts, country
   HAVING country = 'India')
ORDER BY movie_count DESC
LIMIT 10;
```

```
--Categorize Content Based on the Presence of 'Kill' and 'Violence' Keywords
SELECT category, COUNT(*) FROM (
  SELECT title,
  CASE
    WHEN description ILIKE '%kill%' OR description ILIKE '%violence%' THEN 'Violent'
    ELSE 'Non-violent'
  END AS category
  FROM netflix)
GROUP BY category;
```

# Key Insights

There are more number of movies on Netflix than there are TV shows.

The maximum number of movies got released in the years 2017 and 2018.

United States, India, United Kingdom and Canada have contributed to the most content on Netflix till date.

The lengthiest movie is of 312 minutes.

"Drama" is the most common genre.

99 shows have over 5 seasons.

THANK YOU