

NETFLIX_SQL Project



By Shivam Maheshwari

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. This PPT provides a detailed account of the project's objectives, business problems, solutions, findings, and conclusions.

Project Overview

Project Objective

- 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.

Database Schema

```
CREATE TABLE netflix
(
    show_id      VARCHAR(5),
    type         VARCHAR(10),
    title        VARCHAR(250),
    director     VARCHAR(550),
    casts        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)
);
SELECT * FROM netflix;
```

Solved 15 real life Business Problems

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 (e.g., 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 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 Number of Content Releases 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. 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

Count the Number of Movies vs TV Shows

```
SELECT  
    type,  
    COUNT(*) as total_content  
FROM netflix  
GROUP BY type
```


Find the Most Common Rating for Movies and TV Shows

```
SELECT
    type,
    rating
FROM

(
    SELECT
        type,
        rating,
        COUNT(*),
        RANK() OVER(PARTITION BY type ORDER BY COUNT(*)) as ranking
    FROM netflix
    GROUP BY 1, 2
) as t1
WHERE
    ranking = 1
```

List All Movies Released in a Specific Year (e.g., 2020)

```
SELECT * FROM netflix
WHERE
    type = 'Movie'
    AND
    release_year = 2020
```


Find the Top 5 Countries with the Most Content on Netflix

```
SELECT
    UNNEST (STRING_TO_ARRAY (country, ',')) as new_country,
    COUNT (show_id) as total_content
FROM netflix
GROUP BY 1
ORDER BY 2 DESC
LIMIT 5
```

Identify the Longest Movie

```
SELECT * FROM netflix
WHERE
    type = 'Movie'
AND
    duration = (SELECT MAX (duration) FROM netflix)
```

Find Content Added in the Last 5 Years

```
SELECT *  
FROM netflix  
WHERE  
    date_added ~ '^[0-9]{2}-[A-Za-z]{3}-[0-9]{2}$'  
    AND TO_DATE(date_added, 'DD-Mon-YY') >= CURRENT_DATE - INTERVAL '5 years'
```

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

```
SELECT * FROM netflix  
WHERE Director ILIKE '%Rajiv Chilaka%'
```

List All TV Shows with More Than 5 Seasons

```
SELECT * FROM netflix
WHERE
    type = 'TV Show'
    AND
    SPLIT_PART(duration, ' ', 1)::numeric > 5
```

Count the Number of Content Items in Each Genre

```
SELECT  
UNNEST (STRING_TO_ARRAY (listed_in, ',')) as genre,  
COUNT (show_id) as total_content  
FROM netflix  
GROUP BY 1
```

Find Each Year and the Average Number of Content Releases in India on Netflix

```
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
```


List All Movies that are Documentaries

```
SELECT * FROM netflix
WHERE
listed_in ILIKE '%documentaries%'
```

Find All Content Without a Director

```
SELECT * FROM netflix  
WHERE  
    director is null
```

Find How Many Movies Actor 'Salman Khan' Appeared in the Last 10 Years

```
SELECT * FROM netflix
WHERE
    casts ILIKE '%salman khan%'
    AND
    release_year > EXTRACT(YEAR FROM CURRENT_DATE) - 10
```

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
```

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

- 1. The dataset contains a diverse range of movies and TV shows with varying ratings and genres.**
- 2. Insights into the most common ratings provide an understanding of the content's target audience.**
- 3. The top countries and the average content releases by India highlight regional content distribution.**
- 4. Categorizing content based on specific keywords helps in understanding the nature of content available on Netflix.**
- 5. This analysis provides a comprehensive view of Netflix's content and can help inform content strategy and decision-making.**