

NETFLIX MOVIES AND TV SHOWS DATA ANALYSIS USING SQL

NETFLIX

Solutions to 15 business problems:

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

```
SELECT
  type AS content_type,
  COUNT(*) AS total_titles
FROM netflix
GROUP BY type
ORDER BY total_titles DESC;
```

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

```
WITH RatingCounts AS (
  SELECT
    type,
    rating,
    COUNT(*) AS rating_count
  FROM netflix
  GROUP BY type, rating
),
RankedRatings AS (
  SELECT
    type,
    rating,
    rating_count,
    RANK() OVER (PARTITION BY type ORDER BY rating_count DESC) AS rank
  FROM RatingCounts
)
SELECT
```

```
type,  
rating AS most_frequent_rating,  
rating_count  
FROM RankedRatings  
WHERE rank = 1;
```

```
--3. List all movies released in a specific year (e.g., 2020)  
SELECT *  
FROM netflix  
WHERE release_year = 2020
```

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

```
SELECT *  
FROM  
(  
    SELECT  
        -- country,  
        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
```

```
-- 5. Identify the longest movie
```

```
SELECT  
    *  
FROM netflix  
WHERE type = 'Movie'  
ORDER BY SPLIT_PART(duration, ' ', 1)::INT DESC
```

```
-- 6. Find content added in the last 5 years
```

```
SELECT  
    *  
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'
```

```
SELECT *
```

```

FROM
(
SELECT
    *,
    UNNEST(STRING_TO_ARRAY(director, ',')) as director_name
FROM
netflix
)
WHERE
    director_name = 'Rajiv Chilaka'

```

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

```

SELECT *
FROM netflix
WHERE
    TYPE = 'TV Show'
    AND
    SPLIT_PART(duration, ' ', 1)::INT > 5

```

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

```

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

```

-- 10. Find each year and the average numbers of content release by India on netflix.
-- return top 5 year with highest avg content release !

```

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, 2
ORDER BY avg_release DESC
LIMIT 5
```

-- 11. List all movies that are documentaries

```
SELECT * 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 * FROM netflix
WHERE
    casts LIKE '%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
    UNNEST(STRING_TO_ARRAY(casts, ',')) as actor,
    COUNT(*)
FROM netflix
WHERE country = 'India'
GROUP BY 1
ORDER BY 2 DESC
LIMIT 10
```

/*

Question 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
```

```
category,
    TYPE,
    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 1,2
ORDER BY 2

-- End of reports
```