Netflix Movies and TV Shows Data Analysis using SQL

Overview

- Analyzed the Netflix Movies & TV Shows dataset using SQL to uncover insights into global content trends. This project highlights data cleaning, querying, and analysis skills through 20+ real-world SQL problems.

Objectives

- Import, clean, and transform Netflix data for analysis.
- Compare Movies vs. TV Shows and identify popular ratings.
- Explore top countries, directors, and actors contributing to Netflix.
- Analyze yearly & recent content release trends.
- Categorize and filter content by keywords, genres, and duration.

Schema

```sql
DROP TABLE IF EXISTS netflix;
CREATE TABLE netflix (
    show_id       VARCHAR(50),
    type          VARCHAR(50),
    title         VARCHAR(300),
    director      VARCHAR(500),
    cast          VARCHAR(2000),
    country       VARCHAR(200),
    date_added    VARCHAR(100),
    release_year  VARCHAR(10),
    rating        VARCHAR(20),
    duration      VARCHAR(50),
    listed_in     VARCHAR(500),
    description   TEXT
);
```

Business Problems

1. Count The Number of Movies Vs TV Shows

Query:

```sql
SELECT type, COUNT(*) AS Total_Count FROM netflix

GROUP BY type;
```

Output:

type	Total_Count
Movie	6131
TV Show	2676

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

Query:

```sql
SELECT type, rating AS Most_common_rating, total_rating_count
FROM (
SELECT type, rating, COUNT(*) AS total_rating_count,
RANK() OVER(PARTITION BY type ORDER BY COUNT(*) DESC) AS ranking
FROM netflix
GROUP BY type, rating) AS ranked
Where ranking = 1;
```

Output:

type	Most_common_rating	total_rating_count
Movie	TV-MA	2062
TV Show	TV-MA	1145

3. List all movies released in a specific year (e.g., 2020)

Query:

```sql
SELECT type, title, release_year FROM netflix

WHERE type = "Movie" AND release_year = 2020;
```

Output:

type	title	release_year
Movie	Dick Johnson Is Dead	2020
Movie	Europe's Most Dangerous Man: Otto Skorzeny i...	2020
Movie	Tughlaq Durbar	2020
Movie	Omo Ghetto: the Saga	2020
Movie	Shadow Parties	2020
Movie	Here and There	2020

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

Query:

```sql
WITH RECURSIVE split_cte AS (
    -- Start with the original row
    SELECT
        show_id,
        TRIM(SUBSTRING_INDEX(country, ',', 1)) AS country,
        SUBSTRING(country, LENGTH(SUBSTRING_INDEX(country, ',', 1)) + 2) AS rest
    FROM netflix
    WHERE country IS NOT NULL AND country <> ''

    UNION ALL

    -- Recursively split remaining part
    SELECT
        show_id,
        TRIM(SUBSTRING_INDEX(rest, ',', 1)) AS country,
        SUBSTRING(rest, LENGTH(SUBSTRING_INDEX(rest, ',', 1)) + 2)
    FROM split_cte
    WHERE rest <> ''
)
-- Now we will list the top 5 countries as per total_content
SELECT country, COUNT(DISTINCT show_id) AS total_content
FROM split_cte
WHERE country <> ''
GROUP BY country
ORDER BY total_content DESC
LIMIT 5;
```

Output:

country	total_content
United States	3690
India	1046
United Kingdom	806
Canada	445
France	393

5. Identify the longest movie

Query:

```sql
SELECT title, duration FROM netflix
WHERE type = 'Movie' AND duration = ( SELECT MAX(duration) FROM netflix)
LIMIT 1;
```

Output:

title	duration
InuYasha the Movie 2: The Castle Beyond the L...	99 min

6. Find content added in the last 5 years

Query:

```sql
SELECT * FROM netflix

WHERE date_added >= current_date() - INTERVAL 5 year;
```

Output:

show_id	type	title	director	cast	country	release_year	rating	duration	listed_in
s1	Movie	Dick Johnson Is Dead	Kirsten Johnson		United States	2020	PG-13	90 min	Documentaries
s2	TV Show	Blood & Water		Ama Qamata, Khosi Ngema, Gail Mabalane, Tha...	South Africa	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas
s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...		2021	TV-MA	1 Season	Crime TV Shows, International TV S
s4	TV Show	Jailbirds New Orleans				2021	TV-MA	1 Season	Docuseries, Reality TV

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

Query:

```sql
SELECT show_id, type, title, director FROM netflix

WHERE director LIKE '%Rajiv Chilaka%';
```

Output:

show_id	type	title	director
s407	Movie	Chhota Bheem - Neeli Pahaadi	Rajiv Chilaka
s408	Movie	Chhota Bheem & Ganesh	Rajiv Chilaka
s409	Movie	Chhota Bheem & Krishna: Mayanagari	Rajiv Chilaka
s410	Movie	Chhota Bheem & Krishna: Pataliputra- City of th...	Rajiv Chilaka
s411	Movie	Chhota Bheem And The Broken Amulet	Rajiv Chilaka

8. List all TV shows with more than 5 seasons

Query:

```sql
SELECT title, type, duration FROM netflix
WHERE type = 'TV Show' AND duration > 5;
```

Output:

title	type	duration
The Great British Baking Show	TV Show	9 Seasons
Nailed It	TV Show	6 Seasons
Numberblocks	TV Show	6 Seasons
Saved by the Bell	TV Show	9 Seasons
Lucifer	TV Show	6 Seasons

9. Count the number of content items in each genre

Query:

```sql
WITH RECURSIVE split_listed_in AS (
SELECT
        show_id,
        TRIM(SUBSTRING_INDEX(listed_in, ',', 1)) AS listed_in,
        SUBSTRING(listed_in, length(SUBSTRING_INDEX(listed_in, ',', 1)) +2) as rest
        FROM netflix
        WHERE listed_in IS NOT NULL AND listed_in <> ''

UNION ALL

SELECT
        show_id,
        TRIM(SUBSTRING_INDEX(rest, ',', 1)) AS listed_in,
        SUBSTRING(rest, LENGTH(SUBSTRING_INDEX(rest, ',', 1)) +2)
        from split_listed_in
        WHERE rest IS NOT NULL AND rest <> ''
)
SELECT listed_in, COUNT(*) Total_Content FROM split_listed_in
GROUP BY listed_in
ORDER BY Total_Content DESC;
```

Output:

listed_in	Total_Content
International Movies	2752
Dramas	2427
Comedies	1674
International TV Shows	1351
Documentaries	869

10. Find each year and the numbers of content release in that year in India on Netflix, return top 5 year with highest avg monthly content release.

Query:

```sql
SELECT release_year, COUNT(show_id) AS Total_content_released, ROUND(COUNT(show_id)/12, 0) AS avg_content_per_month FROM netflix
WHERE country LIKE '%India%'
GROUP BY release_year
ORDER BY avg_content_per_month DESC
LIMIT 5;
```

Output:

release_year	Total_content_released	avg_content_per_month
2017	111	9
2019	93	8
2018	101	8
2016	80	7
2015	70	6

11. List all movies that are documentaries

Query:

```sql
SELECT title, type, listed_in FROM netflix
WHERE type = 'Movie' AND listed_in LIKE '%Documentaries%';
```

Output:

title	type	listed_in
Dick Johnson Is Dead	Movie	Documentaries
Europe's Most Dangerous Man: Otto Skorzeny i...	Movie	Documentaries, International Movies
My Heroes Were Cowboys	Movie	Documentaries
Schumacher	Movie	Documentaries, International Movies, Sports M...
Blood Brothers: Malcolm X & Muhammad Ali	Movie	Documentaries, Sports Movies

12. Find all content without a director

Query:

```sql
SELECT show_id, director FROM netflix
WHERE director IS NULL OR director = '';
```

Output:

show_id	director
s2	
s4	
s5	
s11	
s15	

13. Find how many movies actor 'Salman Khan' appeared in last 10 years!

Query:

```sql
SELECT COUNT(*) AS total_movies FROM netflix
WHERE type = 'Movie' AND cast LIKE '%Salman Khan%' AND date_added >= current_date() - Interval 10 year;
```

Output:

total_movies
20

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

Query:

```sql
WITH RECURSIVE split_cast AS (
    SELECT
        type,
        country,
        TRIM(SUBSTRING_INDEX(cast, ',', 1)) AS actor_name,
        SUBSTRING(cast, LENGTH(SUBSTRING_INDEX(cast, ',', 1)) + 2) AS rest
    FROM netflix
    WHERE cast IS NOT NULL
      AND cast <> ''
      AND type = 'Movie'
      AND country LIKE '%India%'

    UNION ALL

    SELECT
        type,
        country,
        TRIM(SUBSTRING_INDEX(rest, ',', 1)) AS actor_name,
        SUBSTRING(rest, LENGTH(SUBSTRING_INDEX(rest, ',', 1)) + 2) AS rest
    FROM split_cast
    WHERE rest IS NOT NULL AND rest <> ''
)

SELECT actor_name, COUNT(*) AS total_movies
FROM split_cast
GROUP BY actor_name
ORDER BY total_movies DESC
LIMIT 10;
```

Output:

actor_name	total_movies
Anupam Kher	40
Shah Rukh Khan	34
Naseeruddin Shah	31
Om Puri	29
Akshay Kumar	29
Paresh Rawal	28
Amitabh Bachchan	28
Boman Irani	27
Kareena Kapoor	25
Ajay Devgn	21

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.*/

Query:

```sql
SELECT Label, COUNT(*) AS total_items
FROM (
    SELECT
        CASE
            WHEN LOWER(description) LIKE '%kill%'
                OR LOWER(description) LIKE '%violence%'
            THEN 'Bad'
            ELSE 'Good'
        END AS Label
    FROM netflix
) t
GROUP BY Label;
```

Output:

Label	total_items
Good	8465
Bad	342

16. Find the month with the highest number of content additions.

Query:

```sql
SELECT monthname(date_added) AS month_num, COUNT(date_added) AS Total_content FROM netflix
GROUP BY monthname(date_added)
ORDER BY Total_content DESC
LIMIT 1;
```

Output:

month_num	Total_content
July	827

17. Find the top 3 directors with the most content on Netflix.

Query:

```sql
SELECT director, COUNT(*) AS total_content FROM split_director
GROUP BY director
ORDER BY total_content DESC
LIMIT 3;
```

Output:

director	total_content
Rajiv Chilaka	22
Jan Suter	21
Raúl Campos	19

18. Find the average duration of Movies.

Query:

```sql
SELECT ROUND(AVG(duration),2)  Avg_duration_min FROM netflix
WHERE type = 'Movie';
```

Output:

Avg_duration_min
99.53

19. Which country produces the most TV Shows?

Query:

```sql
SELECT country, COUNT(*) AS total_TV_Shows
FROM split_country
WHERE country <> '' AND type = 'TV Show'
GROUP BY country
ORDER BY total_TV_Shows DESC
LIMIT 1;
```

Output:

country	total_TV_Shows
United States	938

20. List all content titles that contain the word "Love" in the title.

Query:

```sql
SELECT title FROM netflix

WHERE title LIKE '%Love%';
```

Output:

title
Love on the Spectrum
Love Don't Cost a Thing
Love in a Puff
LSD: Love, Sex Aur Dhokha
Really Love
Man in Love
Resort to Love
The Last Letter From Your Lover
Lethal Love

Thank You