

Netflix Movies and TV Shows Data Analysis Using SQL





Overview:

Overview:

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. The following README provides a detailed account of the project's objectives, business problems, solutions, findings, and conclusions.



Objective:

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.



Questions:

1. How many total shows (Movies + TV Shows) are there?

Query:

```
SELECT COUNT(*) AS total_shows FROM netflix1;
```

Output:

	total_shows
1	8807

2. How many movies and TV shows are there separately?

Query:

```
SELECT type, COUNT(*) AS total FROM netflix1 GROUP BY type;
```

Output:

	type	total
1	Movie	6131
2	TV Show	2676

3. What are the top 10 countries with the most Netflix shows?

Query:

```
SELECT country, COUNT(*) AS total FROM netflix1  
GROUP BY country ORDER BY total DESC LIMIT 10;
```

Output:

	country
1	United States
2	India

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

Query:

```
SELECT *  
FROM  
(  
    SELECT  
        -- country,  
        UNNEST(STRING_TO_ARRAY(country, ',')) as country,  
        COUNT(*) as total_content  
    FROM netflix1  
    GROUP BY 1  
)as t1  
WHERE country IS NOT NULL  
ORDER BY total_content DESC  
LIMIT 5
```

Output:

	country	total_content
1	United States	3211
2	India	1008
3	United King...	628
4	United Stat...	479
5	Canada	271

5. Find the oldest show available on Netflix.

Query:

```
SELECT * FROM netflix1 ORDER BY release_year ASC LIMIT 1;
```

Output:

	show_id
1	s4251

6. Find the most recent show added to Netflix.

Query:

```
SELECT * FROM netflix1 ORDER BY date_added DESC LIMIT 1;
```

Output:

	show_id
1	s6067

7. List all shows directed by 'Rajkumar Hirani'.

Query:

```
SELECT * FROM netflix1 WHERE director = 'Rajkumar Hirani';
```

Output:

	show_id
1	s1115
2	s4508
3	s4661

8. How many shows have a rating of 'TV-MA'?

Query:

```
SELECT COUNT(*) AS total FROM netflix1 WHERE rating = 'TV-MA';
```

Output:

	total
1	3207

9. Find all the shows where 'Shah Rukh Khan' is in the casts.

Query:

```
SELECT * FROM netflix1 WHERE casts LIKE '%Shah Rukh Khan%';
```

Output:

show_id	type
s115	Movie
s302	Movie

10. Find each year and the average numbers of content release by India on netflix.

Query:

```
SELECT
    country,
    release_year,
    COUNT(show_id) as total_release,
    ROUND(
        COUNT(show_id)::numeric/
        (SELECT COUNT(show_id) FROM netflix1 WHERE country = 'India')::numeric * 100
     ,2
    )
    as avg_release
FROM netflix1
WHERE country = 'India'
GROUP BY country, 2
ORDER BY avg_release DESC
LIMIT 5
```

Output:

	country
1	India
2	India
3	India
4	India
5	India

11. List the top 10 directors with the most shows.

Query:

```
SELECT director, COUNT(*) AS total FROM netflix1  
WHERE director IS NOT NULL  
GROUP BY director ORDER BY total DESC LIMIT 10;
```

Output:

	director
1	Rajiv Chilaka
2	Raúl Campos, Jan Suter
3	Suhas Kadav

12. List the shows that belong to the 'Drama' category.

Query:

```
SELECT * FROM netflix1 WHERE listed_in LIKE '%Drama%';
```

Output:

	show_id
1	s2
2	s6
3	s8
4	s10
5	s13

13. Count of shows released each year.

Query:

```
SELECT release_year, COUNT(*) AS total FROM netflix1  
GROUP BY release_year ORDER BY release_year;
```

Output:

	release_year	total
	integer	bigint
1	1925	1
2	1942	2

14. Find shows with no director mentioned.

Query:

```
SELECT * FROM netflix1 WHERE director IS NULL
```

Output:

	show_id
	character varying (6)
1	s2
2	s4
3	s5

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:

```
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 netflix1
) AS categorized_content
GROUP BY 1,2
ORDER BY 2
```

Output:

	category	type
1	Bad	Movie
2	Good	Movie
3	Bad	TV Show
4	Good	TV Show



Conclusion:

Conclusion:

Content Distribution: The dataset contains a diverse range of movies and TV shows with varying ratings and genres.

Common Ratings: Insights into the most common ratings provide an understanding of the content's target audience.

Geographical Insights: The top countries and the average content releases by India highlight regional content distribution.

Content Categorization: Categorizing content based on specific keywords helps in understanding the nature of content available on Netflix.



NETFLIX

Thank You