

# Netflix Data Analysis Using SQL

## INTRODUCTION:

This project focuses on analyzing Netflix's extensive content library, including Movies and TV Shows, using SQL. The dataset provides detailed information such as titles, directors, cast, release years, ratings, genres, and descriptions. By addressing key business questions, the project uncovers patterns in content distribution, popular genres, and regional trends. The insights gained from this analysis aim to support strategic decisions for enhancing content curation and audience engagement.

## SQL Queries

```
CREATE TABLE netflix
(
  show_id varchar(10),
  type varchar(10),
  title varchar(200),
  director varchar(250),
  casts varchar(1000),
  country varchar(200),
  date_added varchar(150),
  release_year INT,
  rating varchar(10),
  duration varchar(20),
  listed_in varchar(300),
  description varchar(600)
);
```

```
SELECT * FROM netflix;
```

```
SELECT COUNT(*) as total_count
FROM netflix;
```

```
SELECT DISTINCT TYPE
FROM netflix;
```

## Business Problems and Solutions

### Q1. Count the number of Movies vs TV Shows

SELECT type, COUNT(\*) AS count

FROM netflix

GROUP BY type;

	type character varying (10)	count bigint
1	Movie	6131
2	TV Show	2676

### Q2. 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

FROM RankedRatings

WHERE rank = 1;

	type character varying (10)	most_frequent_rating character varying (10)
1	Movie	TV-MA
2	TV Show	TV-MA

### Q3. List all movies released in a specific year (e.g., 2020)

SELECT \* FROM netflix

WHERE release\_year = 2020;

	country character varying (200)	date_added character varying (150)	release_year integer	rating character varying (10)
1	United States	September 25, 2021	2020	PG-13
2	[null]	September 22, 2021	2020	TV-MA
3	Mexico	September 22, 2021	2020	TV-MA
4	United Kingdom	September 17, 2021	2020	TV-MA
5	[null]	September 17, 2021	2020	TV-Y7
6	[null]	September 16, 2021	2020	TV-MA
7	[null]	September 11, 2021	2020	TV-14
8	Nigeria	September 10, 2021	2020	TV-MA

**Q4. 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;
```

	country text	total_content bigint
1	United States	3211
2	India	1008
3	United Kingd...	628
4	United States	479
5	Canada	271

**Q5. Identify the longest movie**

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

	date_added character varying (150)	release_year integer	rating character varying (10)	duration character varying (20)	listed_in character varying (300)
1	August 15, 2016	2015	66 min	[null]	Movies
2	September 16, 2016	2010	84 min	[null]	Movies
3	April 4, 2017	2017	74 min	[null]	Movies
4	December 28, 2018	2018	TV-MA	312 min	Dramas, International Movies, Sci-Fi & Fantasy
5	June 15, 2021	2021	TV-G	273 min	Documentaries
6	May 21, 2020	1973	TV-14	253 min	Comedies, Dramas, International Movies
7	May 21, 2020	1979	TV-14	237 min	Comedies, Dramas, International Movies
8	May 21, 2020	1982	TV-PG	233 min	Comedies, International Movies, Romantic Movies
9	May 21, 2020	1984	TV-14	230 min	Comedies, Dramas, International Movies
10	September 1, 2021	1984	R	229 min	Classic Movies, Dramas

### Q6. Find content added in the last 5 years

```
SELECT * FROM netflix
```

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

	country character varying (200)	date_added character varying (150)	release_year integer	rating character varying (10)
1	United States	September 25, 2021	2020	PG-13
2	South Africa	September 24, 2021	2021	TV-MA
3	[null]	September 24, 2021	2021	TV-MA
4	[null]	September 24, 2021	2021	TV-MA
5	India	September 24, 2021	2021	TV-MA
6	[null]	September 24, 2021	2021	TV-MA
7	[null]	September 24, 2021	2021	PG
8	United States, Ghana, Burkina Faso, United Kingdom, Germany, Ethio...	September 24, 2021	1993	TV-MA
9	United Kingdom	September 24, 2021	2021	TV-14

### Q7. 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';
```

	show_id character varying (10)	type character varying (10)	title character varying (200)	director character varying (250)	casts character varying
1	s407	Movie	Chhota Bheem - Neeli Pahaadi	Rajiv Chilaka	Vatsal Dubey, Juli
2	s408	Movie	Chhota Bheem & Ganesh	Rajiv Chilaka	Vatsal Dubey, Juli
3	s409	Movie	Chhota Bheem & Krishna: Mayanagari	Rajiv Chilaka	Vatsal Dubey, Juli
4	s410	Movie	Chhota Bheem & Krishna: Pataliputra- City of the Dead	Rajiv Chilaka	Vatsal Dubey, Juli
5	s411	Movie	Chhota Bheem And The Broken Amulet	Rajiv Chilaka	Vatsal Dubey, Juli
6	s412	Movie	Chhota Bheem And The Crown of Valhalla	Rajiv Chilaka	Vatsal Dubey, Juli
7	s413	Movie	Chhota Bheem and the Incan Adventure	Rajiv Chilaka	Vatsal Dubey, Juli
8	s414	Movie	Chhota Bheem and The ShiNobi Secret	Rajiv Chilaka	Vatsal Dubey, Juli
9	s415	Movie	Chhota Bheem Aur Hanuman	Rajiv Chilaka	Vatsal Dubey, Juli
10	s416	Movie	Chhota Bheem aur Krishna	Rajiv Chilaka	Vatsal Dubey, Juli

### Q8. List all TV shows with more than 5 seasons

```
SELECT *FROM netflix
```

```
WHERE type = 'TV Show'
```

```
AND SPLIT_PART(duration, ' ', 1)::INT > 5;
```

	country character varying (200)	date_added character varying (150)	release_year integer	rating character varying (10)	duration character varying (20)	listed_in character varying (300)
1	United Kingdom	September 24, 2021	2021	TV-14	9 Seasons	British TV Shows, Reality TV
2	United States	September 15, 2021	2021	TV-PG	6 Seasons	Reality TV
3	United Kingdom	September 15, 2021	2021	TV-Y	6 Seasons	Kids' TV
4	United States	September 15, 2021	1994	TV-PG	9 Seasons	Kids' TV, TV Comedies
5	United States	September 10, 2021	2021	TV-14	6 Seasons	Crime TV Shows, TV Comedi
6	United States	August 13, 2021	2021	TV-MA	7 Seasons	TV Comedies, TV Dramas
7	United States	August 1, 2021	2012	TV-14	7 Seasons	TV Comedies
8	Japan	August 1, 2021	2014	TV-14	6 Seasons	Anime Series, International T
9	United States	July 28, 2021	2021	TV-14	7 Seasons	Crime TV Shows, TV Action &
10	United States	July 26, 2021	2019	TV-MA	10 Seasons	Classic & Cult TV, TV Action I

**Q9. 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;
```

	genre text	total_content bigint
1	International TV Shows	774
2	Stand-Up Comedy	9
3	Spanish-Language TV Shows	172
4	Romantic Movies	613
5	Anime Features	21
6	TV Horror	11
7	Stand-Up Comedy & Talk Shows	34
8	Kids' TV	63
9	Docuseries	221

**Q10. Find each year and the average numbers of content release in 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, release_year
ORDER BY avg_release DESC
LIMIT 5;
```

	country character varying (200)	release_year integer	total_release bigint	avg_release numeric
1	India	2017	101	10.39
2	India	2018	94	9.67
3	India	2019	87	8.95
4	India	2020	75	7.72
5	India	2016	73	7.51

### Q11. List all movies that are documentaries

```
SELECT * FROM netflix
WHERE listed_in LIKE '%Documentaries';
```

	date_added character varying (150)	release_year integer	rating character varying (10)	duration character varying (20)	listed_in character varying (300)	description character varying (1000)
1	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father
2	September 16, 2021	2021	PG	23 min	Documentaries	Robin Wilts
3	September 2, 2021	2021	PG-13	94 min	Documentaries	This document
4	July 1, 2019	2017	TV-PG	54 min	Documentaries	The chariot
5	August 25, 2021	2021	TV-14	93 min	Documentaries	Bob Ross b
6	August 5, 2021	2021	TV-14	97 min	Documentaries	'Queen of S
7	July 1, 2019	2017	TV-PG	53 min	Documentaries	Drilling into
8	July 28, 2021	2019	TV-14	81 min	Documentaries	Delve into t

### Q12. Find all content without a director

```
SELECT * FROM netflix
WHERE director IS NULL;
```

	type character varying (10)	title character varying (200)	director character varying (250)	casts character varying (1000)
1	TV Show	Blood & Water	[null]	Ama Qamata, Khosi
2	TV Show	Jailbirds New Orleans	[null]	[null]
3	TV Show	Kota Factory	[null]	Mayur More, Jitendi
4	TV Show	Vendetta: Truth, Lies and The Mafia	[null]	[null]
5	TV Show	Crime Stories: India Detectives	[null]	[null]
6	TV Show	Dear White People	[null]	Logan Browning, Br
7	TV Show	Khawatir	[null]	[null]
8	TV Show	Falsa identidad	[null]	Luis Ernesto Franco

### Q13. 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;
```

	director character varying (250)	casts character varying (1000)	country character varying (100)
1	Sooraj R. Barjatya	Salman Khan, Sonam Kapoor, Anupam Kher, Neil Nitin Mukesh, Sanjay Mishra, Swara Bhaskar, Deepak Dobriyal, Arman Kohli, Manoj Joshi, Samaira Rao, Karuna Pandey, Sucheta Kh...	India
2	Rakesh Ranjan Kumar	Lorena Franco, Bijesh Jayarajan, Neet Chowdhary, Rajeev Gaursingh, Karran Jeet, Rajesh Sharma, Gautam Kurup, Poonam Mathur, Salman Khan, Manish Devkush	India

**Q14. 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 actor
ORDER BY COUNT(*) DESC
LIMIT 10;
```

	actor text	count bigint
1	Anupam Kher	36
2	Paresh Rawal	24
3	Shah Rukh Khan	24
4	Om Puri	23
5	Akshay Kumar	22
6	Boman Irani	21
7	Amitabh Bachchan	20
8	Kareena Kapoor	19
9	Naseeruddin Shah	19
10	Asrani	17

**Q15. 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;
```

	category text	content_count bigint
1	Good	8465
2	Bad	342

## **Conclusion**

The analysis highlights that Netflix's content library is predominantly composed of Movies, with "TV-MA" being the most common rating for both Movies and TV Shows. Regional trends show a significant contribution from countries like India, while genres such as Documentaries and Dramas are popular. Additionally, the analysis identifies gaps in metadata, like missing director details, which can be improved. These findings provide actionable insights for optimizing Netflix's content strategy and user satisfaction.