# Netflix Movies and TV Shows Data Analysis Report

This report provides an in-depth analysis of the Netflix content dataset stored in the netflix_data table of the netflix_db database. The data includes movies and TV shows from various countries with their associated metadata, such as title, director, cast, country, release year, and more.

We have solved 15 problems related to the dataset, covering various aspects such as content categorization, ratings, countries, directors, actors, and more. Below is a summary of the analysis along with the SQL queries used to solve each problem.

---

## 1. Count the Number of Movies vs TV Shows

We counted the number of movies and TV shows available on Netflix.

**Query:**

```sql
select
type,
COUNT(*) as counts
from netflix_data
group by type;
```

---

## 2. Find the Most Common Rating for Movies and TV Shows

We identified the most common rating for both movies and TV shows.

**Query:**

```sql
select
type ,
rating
from (
select
type,
rating,
count(*) as counts,
rank() over(partition by type order by count(*) desc)as ranking
from netflix_data
group by type ,rating
) as t1
where ranking=1;
```

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

This query lists all movies released in 2020.

**Query:**

```sql
select * from netflix_data
where type='Movie' and release_year=2020;
```

### 4. Find the Top 5 Countries with the Most Content on Netflix

We found the top 5 countries with the most content, based on the country field, which contains multiple countries for each show.

**Query:**

```sql
SELECT top 5
    trim(value) AS new_country,
    COUNT(show_id) AS total_content
FROM netflix_data
CROSS APPLY STRING_SPLIT(country, ',')
GROUP BY trim(value)
order by 2 desc ;
```

### 5. Identify the Longest Movie

This query identifies the longest movie by its duration.

```sql
SELECT
    title
FROM netflix_data
WHERE
    type = 'Movie'
    AND CAST(LEFT(duration, CHARINDEX(' ', duration + ' ') - 1) AS INT) =
        (SELECT MAX(CAST(LEFT(duration, CHARINDEX(' ', duration + ' ') - 1) AS INT))
         FROM netflix_data
         WHERE type = 'Movie');
```

**Query:**

---

## 6. Find Content Added in the Last 5 Years

We identified all content that was added to Netflix in the past 5 years.

**Query:**

```sql
SELECT *,
        CAST(date_added AS DATE) AS converted_date
FROM netflix_data
WHERE CAST(date_added AS DATE) >= DATEADD(YEAR, -5, GETDATE());
```

---

## 7. Find All Movies/TV Shows by Director 'Rajiv Chilaka'

This query lists all movies and TV shows directed by 'Rajiv Chilaka'.

**Query:**

```sql
select * from netflix_data where director LIKE '%Rajiv Chilaka%';
```

---

## 8. List All TV Shows with More Than 5 Seasons

We listed all TV shows that have more than 5 seasons.

**Query:**

```sql
SELECT
    *,
    CAST(LEFT(duration, CHARINDEX(' ', duration + ' ') - 1) AS INT) AS sessions
FROM netflix_data
WHERE type = 'TV Show'
  AND CAST(LEFT(duration, CHARINDEX(' ', duration + ' ') - 1) AS INT) > 5;
```

## 9. Count the Number of Content Items in Each Genre

This query counts the number of content items in each genre, using the listed_in field.

**Query:**

```sql
select
COUNT(show_id) AS total_content,
trim(value) as genre
from netflix_data cross apply string_split(listed_in,',')
group by TRIM(value)
order by count(show_id) desc
```

## 10. Find Each Year and the Average Number of Content Released in India on Netflix

We calculated the average number of content items released in India per year on Netflix, sorted by the top 5 years with the highest average.

**Query:**

```sql
SELECT
    YEAR(CONVERT(DATE, date_added, 101)) AS year,   -- Convert to date and extract the year
    COUNT(*) AS yearly_content,
    ROUND(
        CAST(COUNT(*) AS FLOAT) /
        (SELECT COUNT(*) FROM netflix_data WHERE country = 'India') * 100, 2
    ) AS avg_content_per_year
FROM netflix_data
WHERE country = 'India'
GROUP BY YEAR(CONVERT(DATE, date_added, 101))
order by avg content per year desc;
```

## 11. List All Movies That Are Documentaries

We listed all movies in the genre "Documentaries".

**Query:**

```sql
with table1 as(
select trim(value) as genre from netflix_data cross apply string_split(listed_in,',')
  group by TRIM(value)
  )
  select * from netflix_data n,table1 t1 where t1.genre='documentaries' and n.type='Movie'
```

## 12. Find All Content Without a Director

This query lists all content without a director.

**Query:**

```
select * from netflix_data where director is null
```

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

We counted how many movies actor 'Salman Khan' appeared in during the last 10 years.

**Query:**

```
select
count(show_id) as no_of_movies
from netflix_data
where cast like'%Salman Khan%'
and cast like '%salman khan%'
and release_year> year(GETDATE())-10
```

## 14. Find the Top 10 Actors Who Have Appeared in the Highest Number of Movies Produced in India

This query lists the top 10 actors who have appeared in the most movies produced in India.

```
select top 10 trim(value)as actors , count(*) as total_content
from netflix_data cross apply string_split(cast,',')
where country like'%India%' or country like'%india%'
group by TRIM(value)
order by total_content desc , actors asc;
```
**Query:**

## 15. Categorize Content Based on the Presence of 'Kill' and 'Violence' in the Description

We categorized content as "Bad" if the description contains the keywords "kill" or "violence", and "Good" otherwise.

**Query:**

```sql
SELECT
    CASE
        WHEN description LIKE '%kill%' OR description LIKE '%violence%' THEN 'Bad'
        ELSE 'Good'
    END AS content_category,
    COUNT(*) AS content_count
FROM netflix_data
GROUP BY
    CASE
        WHEN description LIKE '%kill%' OR description LIKE '%violence%' THEN 'Bad'
        ELSE 'Good'
    END
```