



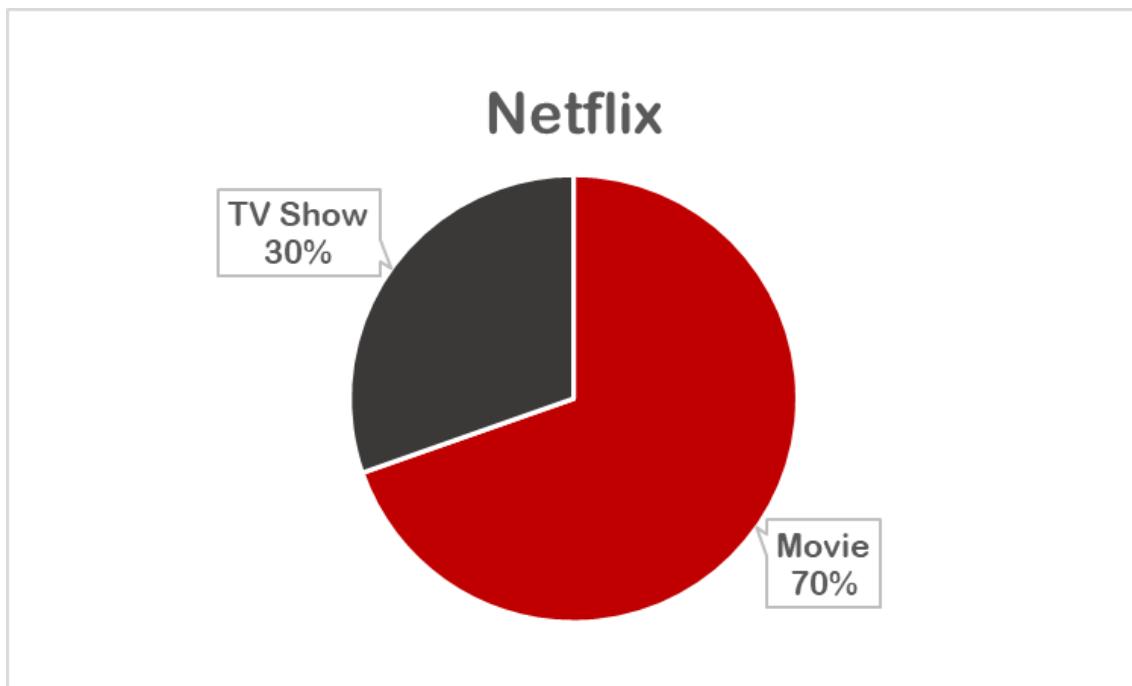
Netflix - SQL Project

The database schema includes columns for movie types and titles, countries, release years, ratings, etc.

Exploring data

```
SELECT type, COUNT(type)
FROM netflix_titles
GROUP BY type;
```

The dataset comprises a total of 8,807 rows, with 6,131 movies and 2,676 TV shows.

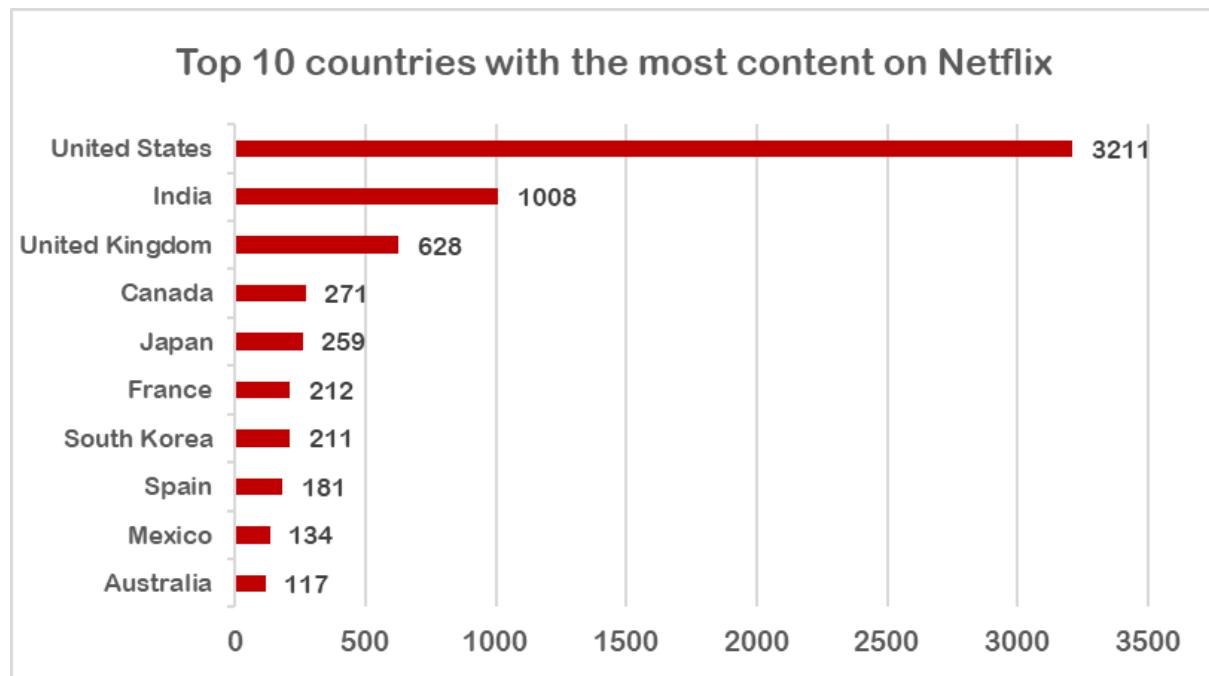


[1] Understanding which country has the most content

Aggregate the number of movies and shows by country. In cases where the data contains a list of countries, choose the first country in the list to represent the country from which the content originates. This analysis will help determine which country has the most content.

In the query, we need to handle two situations because the 'country' column sometimes contains lists of countries and sometimes just a single country. To manage this, we use a `CASE Expression`. When the countries has a list, we extract the first country by finding the word before the first comma using `substr()` and `instr()` functions.

```
SELECT
CASE
    WHEN instr(country, ',') > 0
    THEN substr(country, 1, instr(country, ',') -1)
    ELSE country
END AS new_country,
COUNT(*) as Amount
FROM netflix_titles
group by new_country
order by Amount desc;
```



Most content on Netflix is produced by the United States, India, and the United Kingdom.

[2] Type of content that Netflix has more focus on in recent years

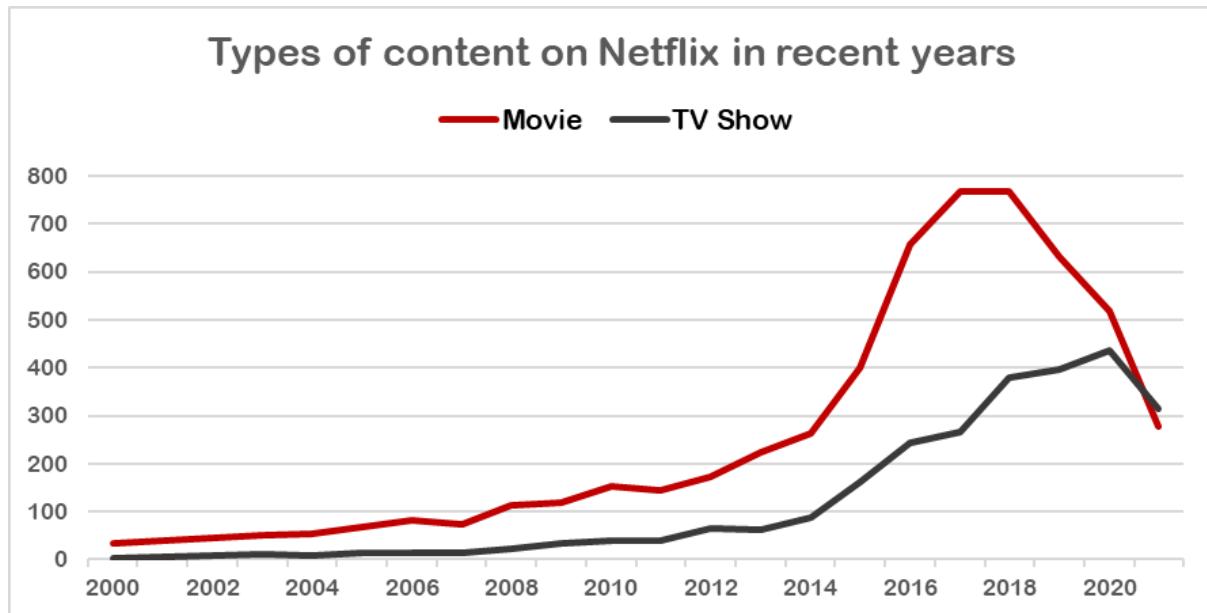
Aggregate the number of movies and TV shows released each year. Then, compare the difference between the two to see which type Netflix has focused on more in recent years.

In the query, we use `subquery` to count each category , movies and TV shows, by `release year`. Then, we employ `CASE Expression` with `Aggregate functions` to create columns for categorizing data and filter released year since 2000.

```

SELECT release_year,
       MAX(CASE WHEN type = 'Movie' THEN value ELSE 0 END) AS "Movie",
       MAX(CASE WHEN type = 'TV Show' THEN value ELSE 0 END) AS "TV Show"
  FROM (SELECT release_year, type, COUNT(*) as value
        FROM netflix_titles
       GROUP by release_year, type)
 WHERE release_year >= 2000
 GROUP BY release_year;

```



Since 2000, Netflix focused more on movies than TV shows. But in 2021, they started to focus more on TV shows. Movies peaked in 2018 but are now decreasing, while TV shows keep growing

[3] Rating and Genres Distribution

[3.1] Rating Distribution

While exploring the ratings, we found that the most common ratings on Netflix are TV-MA, TV-14, and TV-PG. Additionally, there are some titles without ratings, and some titles have incorrect data such as duration in the rating field.

```

SELECT rating AS Rating,
       count(*) AS N
  FROM netflix_titles
 GROUP BY rating
 ORDER BY N;

```

In the query, we decided to exclude incorrect data by using a `where clause` to filter it out.

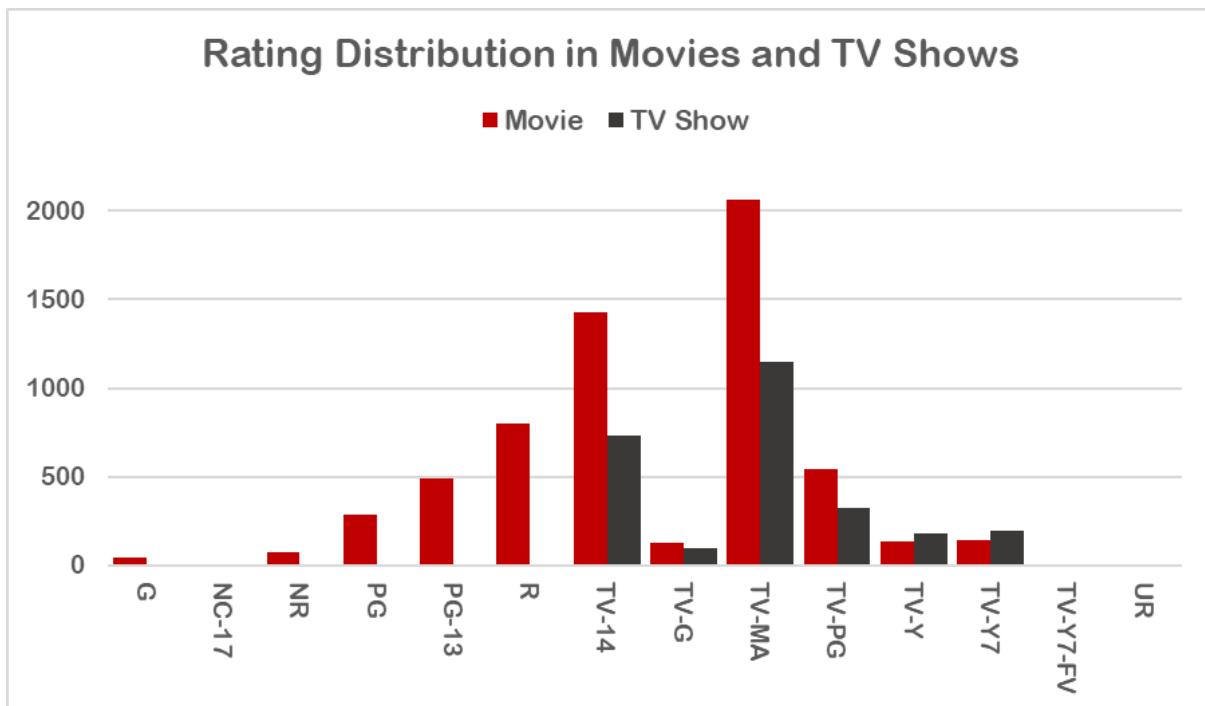
Rating	N
66 min	1
74 min	1
84 min	1
NC-17	3
UR	3
	4
TV-Y7-FV	6
G	41
NR	80
TV-G	220
PG	287
TV-Y	307

```

SELECT type,
       rating,
       count(*) AS N
  FROM netflix_titles
 WHERE rating not IN ('66 min', '74 min', '84 min')
 GROUP BY type, rating

```

Rating	N
TV-Y7	334
PG-13	490
R	799
TV-PG	863
TV-14	2160
TV-MA	3207



When we analyze movies and TV shows separately, we discover that TV-MA and TV-14 ratings are the top two for both categories. However, there's a contrast in the third position, with 'R' rating for movies and 'TV-PG' for TV shows. Next, we'll take a closer look at the genres in each category.

[3.2] Genres Distribution

We can determine genres for each title by choose the first genres in column 'listed_in' to reveals the popular common genres in each category with the top three ratings.

For Movies ⇒ TV-MA, TV-14, and R

For TV shows ⇒ TV-MA, TV-14, and TV-PG

In the query, we will utilize a `CASE Expression` along with `substr()` and `instr()` functions similar to our previous approach, to determine the genres.

Movie Genres

```

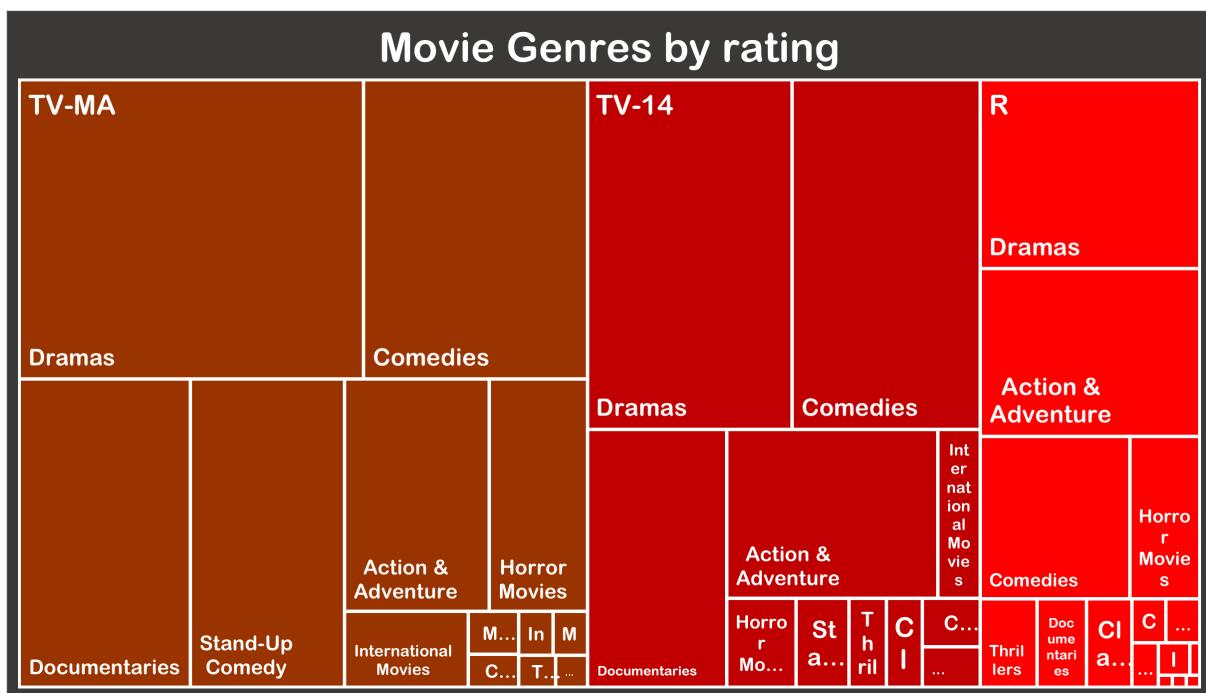
SELECT
       rating,
       CASE
           WHEN instr(listed_in, ',') > 0
           THEN substr(listed_in, 1, instr(listed_in, ',' ) -1)

```

```

        ELSE listed_in
    END AS genres,
    count(*) as N
FROM netflix_titles
WHERE type LIKE 'Movie' and
      rating IN ('TV-MA', 'TV-14', 'R')
group by rating, genres;

```



In movies, we found that '**Dramas**' is the most common genre across all of them.

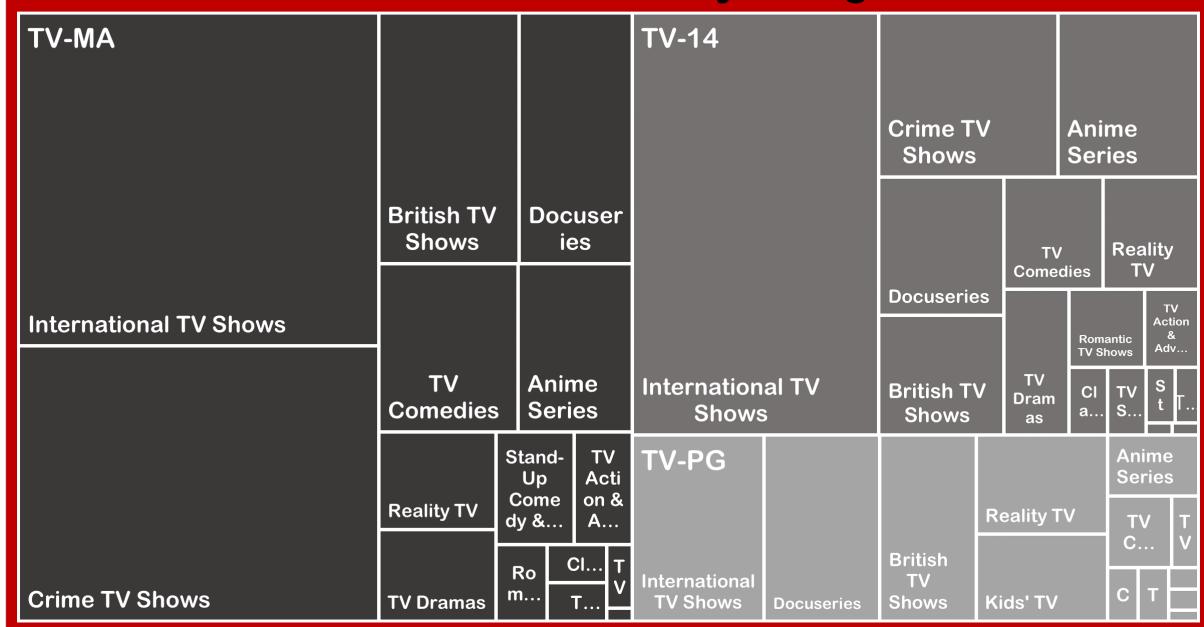
TV Show Genres

```

SELECT
    rating,
    CASE
        WHEN instr(listed_in, ',') > 0
        THEN substr(listed_in, 1, instr(listed_in, ',') -1)
        ELSE listed_in
    END AS genres,
    count(*) as N
FROM netflix_titles
WHERE type LIKE 'TV Show' and
      rating IN ('TV-MA', 'TV-14', 'TV-PG')
group by rating, genres

```

TV Show Genres by rating



In TV show, we found that '***International TV Shows***' is the most common genre across all of them.

From the analysis, we observed differences between genres in each category. This could be attributed to the expectation that in movies, audiences anticipate actors performing within engaging storylines, while for TV shows, they expect a more relaxed and entertaining atmosphere.