

In Summary;

I analyzed Netflix genre data by integrating Excel with MySQL to derive valuable insights. First, I created a table in MySQL and successfully loaded the Excel dataset into the database. This setup allowed me to systematically organize and manage the data for analysis.

Using SQL queries, I explored the dataset to gain a general understanding of its structure, including the distribution of genres and the total number of movies in each genre. Additionally, I addressed specific questions by extracting and aggregating data, such as identifying popular genres, movie counts, and other trends. This approach provided clear and actionable insights into Netflix's genre-based content, showcasing the potential for data-driven content strategy and decision-making.

```
USE new_schema;
```

```
CREATE TABLE netflix  
(  
  show_id VARCHAR(525),  
  title VARCHAR(150),  
  director VARCHAR(208),  
  casts VARCHAR(1000),  
  country VARCHAR(150),  
  date_added VARCHAR(50),  
  release_year INT,  
  rating VARCHAR(10),  
  duration VARCHAR(15),  
  listed_in VARCHAR(205),  
  description VARCHAR(250),  
  type VARCHAR(200)  
);
```

```
DESCRIBE netflix;
```

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/netflix2.csv'  
INTO TABLE netflix  
FIELDS TERMINATED BY '%' -- Use the correct delimiter  
LINES TERMINATED BY '\r\n' -- Adjust for line endings (e.g., '\r\n' for Windows)  
IGNORE 1 ROWS; -- Skip the header row if it exists
```

```
SELECT
```

```
*
```

```
FROM
```

```
netflix;
```

```
SELECT
```

```
COUNT(*) as total_content
```

```
FROM
```

```
netflix;
```

```
SELECT
```

```
DISTINCT type
```

```
FROM
```

```
netflix;
```

```
SELECT
```

```
*
```

```
FROM
```

```
netflix
```

```
WHERE
```

```
type = 'Movie';
```

```
UPDATE netflix
```

```
SET type = 'Movie'
```

```
WHERE type = 'Movie';
```

```
UPDATE netflix
```

```
SET type = 'TV Show'
```

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

-- Count the number of movies and tv shows.

```
SELECT
type,
COUNT(*) as total_content
FROM
netflix
GROUP BY type;
```

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

```
SELECT
type, rating
FROM
(
SELECT
type, rating, COUNT(*),
RANK () OVER (PARTITION BY type ORDER BY COUNT(*) DESC) as ranking
FROM
netflix
GROUP BY 1,2) as t1
WHERE
ranking = 1;
```

-- List all movies released in a specific year(2015).

```
SELECT
*
FROM
netflix
WHERE
type = "Movie" AND release_year = 2015;
```

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

```
SELECT
    new_country,
    COUNT(show_id) AS total_content
FROM (
    SELECT
        show_id,
        TRIM(JSON_UNQUOTE(json_each.value)) AS new_country
    FROM netflix,
    JSON_TABLE(
        CONCAT('["', REPLACE(country, ',', "'"), '"'], ''),
        '$[*]' COLUMNS (value VARCHAR(100) PATH '$')
    ) AS json_each
) AS exploded
GROUP BY new_country
ORDER BY total_content DESC
LIMIT 5;
```

-- Identify the longest movie or Tv shows duration.

```
SELECT
*
FROM
netflix
WHERE
type = 'Movie' AND duration = (SELECT MAX(duration) FROM netflix);
```

-- Find content added in the last 5 years.

```
SELECT
    *
FROM
    netflix
WHERE
    release_year >= YEAR(curdate()) - 5;
```

-- Find all the movies/Tv shows by director 'Rajiv Chilaka'.

```
SELECT
    *
FROM
    netflix
WHERE
    director LIKE '%Rajiv Chilaka%';
```

-- List all Tv shows with more than 5 seasons.

```
SELECT
    *
FROM
    netflix
WHERE
    type = 'TV Show' AND
    SUBSTRING_INDEX(duration, ' ', 1) > 5 ;
```

-- Count the number of content items in each genre.

```
WITH RECURSIVE split_listed_in AS (  
  SELECT  
    show_id,  
    SUBSTRING_INDEX(listed_in, ',', 1) AS genre,  
    SUBSTRING(listed_in FROM LOCATE(',', listed_in) + 1) AS remaining  
  FROM netflix  
  WHERE listed_in LIKE '%,%'  
UNION ALL  
  SELECT  
    show_id,  
    SUBSTRING_INDEX(remaining, ',', 1) AS genre,  
    SUBSTRING(remaining FROM LOCATE(',', remaining) + 1)  
  FROM split_listed_in  
  WHERE remaining LIKE '%,%'  
UNION ALL  
  SELECT  
    show_id,  
    remaining AS genre,  
    NULL AS remaining  
  FROM split_listed_in  
  WHERE remaining NOT LIKE '%,%'  
)
```

```
SELECT genre, COUNT(*) AS count  
FROM split_listed_in  
GROUP BY genre  
ORDER BY count DESC;
```

-- Find the average release year for content produced in a specific country.

```
SELECT
    EXTRACT(YEAR FROM DATE_ADD('1899-12-30', INTERVAL date_added DAY)) AS date,
    COUNT(*) as yearly_content,
    ROUND(COUNT(*) / (SELECT COUNT(*) FROM netflix WHERE country = 'India')*100
,2) as avg_content_per_year
FROM
    netflix
WHERE
    country = 'India'
GROUP BY 1;
```

-- List all movies that are documentaries.

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

-- Find all content without a director.

```
SELECT
    *
FROM
    netflix
WHERE
    director = 'N/A';
```


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

SELECT

*

FROM

netflix

WHERE

casts LIKE '%Salman Khan%' AND

release_year > 2010;