**Difficulties Encountered**

1. Data Formatting Issues:

Sometimes, CSV files contain special characters or inconsistent formatting that can cause errors during import. For example, missing values or incorrect date formats .

2.Handling Large Files:

Handling this kind of a large file has been a challenge since have never interacted with such a large dataset.

Interesting Observation about the Netflix Shows Dataset

Diversity in Content:

One interesting thing about the Netflix Shows dataset is the diversity of content available on the platform. The dataset includes a wide range of genres, countries, and ratings, highlighting Netflix's global reach and varied content library

**THE DATA FUN.**

Finding the most common genre, the most common genre is Action $ adventure.

SELECT listed\_in, COUNT(\*) AS count

FROM netflix\_titles

GROUP BY listed\_in

ORDER BY count DESC;

To find out the number of shows added per year;there are 100 shows added per year

SELECT YEAR(date\_added) AS year, COUNT(\*) AS count

FROM netflix\_titles

GROUP BY year

ORDER BY year;

**Cool Facts Discovered:**

Action and adventure is a Popular Genre: Action $ adventure is one of the most common genres on Netflix, indicating that this type of content is highly prevalent on the platform.

Steady Addition of New Content: Netflix has consistently added new shows every year, reflecting a steady growth in its content library.

**Finding the average duration of shows** ;the average duration is 100.7273

SELECT AVG(CAST(SUBSTRING\_INDEX(duration, ' ', 1) AS UNSIGNED)) AS avg\_duration\_minutes

FROM netflix\_titles

WHERE duration LIKE '%min';

**Finding the total number of shows per country;** the total number of shows is united states with 17 shows followed by japan with 13 shows

SELECT country, COUNT(\*) AS total\_shows

FROM netflix\_title

GROUP BY country

ORDER BY total\_shows DESC;

Counts of shows by rating;the highest rated show is TV-MA with 32 total shows followed by TV-14 with 7 total shows

SELECT rating, COUNT(\*) AS total\_shows

FROM netflix\_titles

GROUP BY rating

ORDER BY total\_shows DESC;

3.**Ask away**

**The following are the questions with the findings;**

**a)Which directors appear in most shows?.... T**oshiya Shinohara has directed a total of 4 shows

**SELECT director, COUNT(\*) AS count**

**FROM netflix\_titles**

**GROUP BY director**

**ORDER BY count DESC**

**LIMIT 10;**

**What is the distribution of shows by their release year?** Between 1975-1998 there has only been a single show per year,,,from 2001-2019 there has been a variation in the number of show ranging between 1and 4,,in 2020 there has been 8 shows and 2021 there has been 53 shows

The number of shows released in 2021 clearly indicated that a lot of people have fully dived in finding entertainments in neflix.

**SELECT release\_year, COUNT(\*) AS count**

**FROM shows**

**GROUP BY release\_year**

**ORDER BY release\_year;**

**Finding the total number of shows with multiple genres . the total number of shows with multiple genre is 77**

**SELECT COUNT(\*) AS multi\_genre\_count**

**FROM shows**

**WHERE listed\_in LIKE '%,%';**

**What I have learned from the dataset.**

1. Identifying the most common genre helps us understand the predominant type of content on Netflix.
2. Knowing the average duration of shows can provide insights into the typical length of content, helping to understand viewing habits and content type (e.g., movies vs. series).
3. Understanding which countries produce the most shows helps identify key content production regions and Netflix's global content strategy.
4. The distribution of shows by rating category helps understand the target audience demographics and content suitability.
5. Identifying directors with the most shows can highlight prolific creators and potentially popular content based on director fame.

These insights help understand Netflix's content strategy, regional production diversity, popular genres, and more, providing a comprehensive view of the Netflix Shows dataset.

**Graph Visualization**

