**Netfix Shows Data Analysis**

**Difficulties Encountered**

While working with the Netflix Shows dataset, several challenges were encountered:

1. **Data Cleaning and Preparation**: The dataset included some missing values, especially in columns like **director** and **cast.** Handling these missing values was essential to ensure accurate analysis. This required careful inspection and cleaning of the data before performing any SQL queries.
2. **Parsing Date Fields**: The date\_added column was stored as a string in the CSV file. Converting this column to a proper date format in MySQL required additional steps. It was necessary to ensure that the dates were correctly parsed and stored to enable accurate date-based queries.
3. **Large Dataset Size**: With over 6,000 entries, performing certain operations on the entire dataset was resource-intensive and time-consuming. This required optimizing queries to ensure they ran efficiently without consuming excessive computational resources.

**Interesting Observation**

One particularly interesting observation from the dataset was the diversity and global reach of Netflix's content. The dataset included shows from a wide range of countries, highlighting Netflix's strategy to cater to a global audience.

**Interesting Fact:**

* **Global Content Diversity**: The dataset revealed that Netflix hosts shows from over 100 different countries. This global diversity is reflected in the variety of genres and languages available on the platform, indicating Netflix's commitment to offering a wide range of content to meet the preferences of a diverse, international audience.

**Queries**

**Data Fun (20 pts)**

This section showcases simple SQL queries used to explore the Netflix Shows dataset, revealing hidden facts and insights.

1. Total Number of Movies and TV Shows

**SQL Query:**

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

*FROM netflix\_titles*

*GROUP BY type;*

**Insight:**

* **Movies**: 6131 titles
* **TV Shows**: 2576 titles

2. Distribution of Ratings

**SQL Query:**

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

*FROM netflix\_titles*

*GROUP BY rating*

*ORDER BY count DESC;*

**Insight:** The most common ratings are:

* **TV-MA**: 2027 titles
* **TV-14**: 1540 titles
* **TV-PG**: 819 titles
* **R**: 796 titles
* **PG**-13: 489 titles
* **TV-Y7**: 334 titles
* **TV Y**: 307 titles
* **PG**: 286 titles
* **TV-G**: 220 titles

**Cool Fact #1: Most Common Release Year**

One interesting fact is identifying the year in which the highest number of shows were released.

**SQL Query:**

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

*FROM netflix\_titles*

*GROUP BY release\_year*

*ORDER BY count DESC*

*LIMIT 1;*

**Insight:**

* The most common release year for shows in the dataset is **2018**, with a total of **1145** shows released that year.

**Cool Fact #2: Top 5 Directors with Most Titles**

Another cool fact is identifying the top 5 directors with the most titles in the Netflix catalog.

**SQL Query:**

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

*FROM netflix\_titles*

*WHERE director IS NOT NULL*

*GROUP BY director*

*ORDER BY count DESC*

*LIMIT 5;*

**Insight:**

The top 5 directors with the most titles are:

1. **Rajiv Chilaka** - 19 titles
2. **Raúl Campos, Jan Suter** - 18 titles
3. **Suhas Kadav** - 16 titles
4. **Marcus Raboy** - 15 titles
5. **Jay Karas** - 14 titles

**Ask Away (30 pts)**

This section focuses on formulating questions about the dataset, writing SQL queries to find answers, and sharing the insights gained.

***Question 1: What is the distribution of ratings for movies released in the last 5 years?***

**SQL Query:**

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

*FROM netflix\_titles*

*WHERE type = 'Movie' AND release\_year >= YEAR(CURDATE()) - 5*

*GROUP BY rating*

*ORDER BY count DESC;*

**Insight:** By examining the distribution of ratings for movies released in the last 5 years, we can understand the trends in content rating:

* **TV-MA**: The most common rating with a significant number of titles (611 titles).
* **TV-14**: Another prevalent rating, indicating a considerable amount of content suitable for teenagers (323 titles).
* **TV-PG**: These movies come in third with a total of 128 titles.

***Question 2: Which countries have the highest number of movies directed by a single director?***

**SQL Query:**

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

***FROM netflix\_titles***

***WHERE type = 'Movie' AND director IS NOT NULL***

***GROUP BY country, director***

***ORDER BY count DESC***

***LIMIT 10;***

**Insight:** Identifying countries with the highest number of movies directed by a single director reveals:

* **United States**: Rajiv Chilaka stands out with a significant number of titles.
* **India**: Directors like Karan Johar have a notable presence.
* **United Kingdom**: Alfred Hitchcock is a prominent director.

This analysis highlights key directors in different countries and their contribution to Netflix's movie catalog.