**Unveiling Netflix Analytics: Insights into Movies, TV Shows, Genres, Duration, and Country**

**Introduction**

Welcome to my comprehensive analysis of Netflix data! In this exploration, I delve into various aspects of the dataset, including the distribution of movies and TV shows, genre preferences, release trends, duration analysis, and content distribution by country. Join me as I uncover key insights into Netflix's content landscape, highlighting interesting patterns and providing possible explanations for observed trends.

**Data Cleaning**

Handling Null Values, Inconsistencies, Garbage Values, and Splitting the "listed-in" Column in the Netflix Dataset.

In my Netflix data analysis project, data cleaning played a vital role in ensuring the accuracy and usability of the dataset. I employed various techniques to address specific data quality issues and enhance the dataset for further analysis.

To start, I added filters, allowing me to focus on specific columns and streamline the data cleaning process.

For the director and country columns, null values were a concern. To tackle this, I introduced new columns that examined each corresponding cell. If a null value was found in the director column, it was replaced with the label "Data Unavailable," ensuring consistent representation. Similarly, the country column was addressed in a similar manner, with null values being replaced by the label "Country Unknown." This approach enhanced the dataset's completeness and integrity.

In the "date\_added" column, inconsistencies arose due to a mixture of date and text data types. To resolve this, I created a new column that evaluated the original data type in each corresponding cell. If the data type was identified as text, it was converted to a date format. Otherwise, the original value was retained, ensuring the column's consistency, and facilitating accurate analysis.

The "rating" column contained garbage and blank values in a few rows, requiring special attention. To address this issue, I created a new column that removed these unwanted values and replaced them with the label "Ratings Unavailable." This process ensured that the rating column remained reliable and usable for analysis purposes.

Additionally, the "listed-in" column contained comma-separated values indicating the genres to which each movie belonged. To facilitate further analysis, I split this column into multiple columns, allowing for more granular genre-based analysis. This transformation expanded the dataset's capabilities and provided insights into specific genre trends and preferences.

These data cleaning techniques highlight my attention to detail and commitment to data quality. By addressing null values, inconsistencies, garbage values, and transforming the "listed-in" column, I ensured a robust dataset for meaningful analysis and interpretation.

**Analysis**

Movies and TV Shows Distribution:

In our dataset of 8,807 records, we find that movies account for 6,131 entries, representing 69.62% of the total, while TV shows make up the remaining 2,676 entries, representing 30.38%. This breakdown showcases the dominance of movies on the platform, with TV shows making up a significant portion as well.

Genre Analysis:

The genre information in the dataset is originally provided as comma-separated values, indicating that each show can belong to multiple genres. To facilitate analysis, I employed a "calculated item" approach to group similar genres, allowing me to bucket them for analysis. The findings reveal that the "international" category stands out as the most prevalent genre throughout the dataset, followed by the "romance" category. On the other hand, the "musical" category has the fewest number of shows.

Release Trends:

By analyzing the number of shows released each year over the past 20 years, I observed an overall increase in content from 2000 to 2018. There is a particularly steep increase between 2014 and 2018, indicating a period of significant growth in content production. However, a subsequent dip in content releases is evident after 2018, extending until 2021. Possible reasons for this fall in content after 2018 could include content consolidation, shifting strategies, and the impact of the COVID-19 pandemic.

Duration Analysis:

To analyze the duration of different types of content, I encountered variations between movies and TV shows. Movies have their duration information stored in minutes, while TV shows use the number of seasons as their duration metric. To address this, I employed Excel functions like IF and VLOOKUP. For TV shows, I retained the duration as the number of seasons. For movies, I used VLOOKUP calculations to assign the appropriate duration bucket based on the duration in minutes, creating intervals of every 20 minutes. This analysis provides insights into preferred content lengths for both movies and TV shows.

Content Distribution by Country:

The country column in the dataset contained comma-separated values, which I addressed by using Power Query to split and unpivot the information. This step allowed me to obtain one country value per row, enabling accurate analysis of content distribution by country. The analysis reveals that the United States has the highest number of releases, followed by India, in both the "Movies" and "TV Shows" categories. Possible reasons for this dominance include the influence of Hollywood in the US and the extensive reach and diverse content production of Bollywood in India.

**Conclusion**

The comprehensive analysis of the Netflix dataset has unveiled significant insights into the distribution of movies and TV shows, genre preferences, release trends, duration analysis, and content distribution by country. Through the utilization of various data analysis techniques, including calculated items, Excel functions, and Power Query, I have extracted valuable information from the dataset. Join me as I continue to unravel the intricacies of Netflix's content landscape, providing a deeper understanding of the platform's offerings and trends.