**Business Problem**

Analyze the data and generate insights that could help Streaming Platform in deciding which type of shows/movies to produce and how they can grow the business in different countries.

**Basic Analysis**

Un-nesting the columns

Un-nest the columns those have cells with multiple commas separated values by creating multiple rows

Handling null values

For categorical variables with null values, update those rows as unknown\_column\_name.

Example: Replace missing value with Unknown Actor for missing value in Actors column.

Replace with 0 for continuous variables having null values.

Extra:

Review the dataset structure, including the number of rows and columns, data types, and any missing values.

Check for any inconsistencies or errors in the data.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What does ‘good’ look like?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Find the counts of each categorical variable both using graphical and non-graphical analysis.

For Non-graphical Analysis:

Hint: We want you to find the values counts of each category for the given column

For graphical analysis:

Hint: We can use a count plot to get the counts of each category

Extra:

Calculate summary statistics for numerical columns such as release year and duration to understand the distribution of values.

Analyze categorical columns like type, country, rating, and genre to identify the frequency of different categories.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Comparison of tv shows vs. movies.

Find the number of movies produced in each country and pick the top 10 countries.

Hint: We want you to apply group by each country and find the count of unique titles of movies

Find the number of Tv-Shows produced in each country and pick the top 10 countries.

Hint: We want you to apply group by each country and find the count of unique titles of Tv-shows

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What is the best time to launch a TV show?

Find which is the best week to release the Tv-show or the movie. Do the analysis separately for Tv-shows and Movies

Hint: We expect you to create a new column and group by each week and count the total number of movies/ tv shows.

Find which is the best month to release the Tv-show or the movie. Do the analysis separately for Tv-shows and Movies

Hint: We expect you to create a new column and group by each month and count the total number of movies/ tv shows.

Extra:

Analyze trends over time, such as the number of titles added to Platform each year.

Identify any seasonal patterns or trends in viewer preferences.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Analysis of actors/directors of different types of shows/movies.

Identify the top 10 directors who have appeared in most movies or TV shows.

Hint : We want you to group by each actor and find the count of unique titles of Tv-shows/movies

Identify the top 10 directors who have appeared in most movies or TV shows.

Hint : We want you to group by each director and find the count of unique titles of Tv-shows/movies

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Which genre movies are more popular or produced more

Hint : We want you to apply the word cloud on the genre columns to know which kind of genre is produced

Extra:

Identify trends in genre popularity over time and across different countries.

Determine whether certain genres are more successful in specific regions or among particular demographics.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Find After how many days the movie will be added to platform after the release of the movie (you can consider the recent past data)

Hint : We want you to get the difference between the columns having date added information and release year information and get the mode of difference. This will give an insight into what will be the better time to add in platform

**My Steps:**

1. **Country Analysis**:
   * Analyze the distribution of shows and movies by country of production.
   * Investigate whether certain countries have a higher audience engagement or viewership.
   * Determine if there are any correlations between the country of production and the popularity of specific genres or types of content.
2. **Content Duration**:
   * Examine the distribution of content duration (in minutes or number of seasons) for both movies and TV shows.
   * Identify any patterns in viewer preferences regarding the duration of content.
   * Determine if there are optimal duration ranges for different types of content or genres.