STREAMLINE ANALYSIS: Uncovering Trends in Netflix's Library

Problem Statement:

Analyse the Netflix dataset to uncover trends in content production, genre popularity, and platform distribution. Use data cleaning, analysis, and visualization techniques to understand how Netflix's content offering has evolved over time, identify patterns in genres, and explore content ratings by country.

20 Questions for Exploration:

Data Inspection:

1. What are the basic details of the dataset? (Shape, column names, data types)
2. Are there any missing values in the dataset? If yes, how do you handle them?
3. What is the distribution of missing data across columns?

Data Cleaning:

1. Are there any duplicate rows in the dataset? If so, how will you address them?
2. Are there any inconsistent or invalid entries (e.g., incorrect data types, typos)? How can you clean them?

Subset Creation:

1. How can you create a subset of movies released after 2015?
2. How can you filter the dataset to show content of a specific genre (e.g., "Drama" or "Comedy")?

Descriptive Statistics:

1. What are the most common genres in the dataset?
2. What is the distribution of content release years in the dataset?
3. What are the top 5 countries producing the most content on Netflix?

Visualizations:

1. Can you create a bar chart of the top 10 genres by content count?
2. How can you visualize the distribution of release years as a histogram or line graph?
3. Create a pie chart showing the proportion of different content ratings (e.g., TV-MA, PG-13, etc.).

Country-wise Analysis:

1. Which countries have the highest representation of Netflix content?
2. What is the trend of content production across different countries over the years?
3. How does the genre distribution vary between two specific countries (e.g., USA vs India)?

Time Series Analysis:

1. How has the content production evolved over the years? Can you plot a trend line for this?
2. What is the trend in releasing TV shows versus movies over time?

Advanced Visualizations:

1. Can you create a heatmap to show the correlation between different numerical columns (if any)?
2. Use a boxplot to visualize the distribution of content durations (if available) for movies.

Analysis of Cast and Directors:

1. Which directors have the highest number of titles on Netflix?
2. Can you visualize the number of titles per cast member, focusing on the top 10?

Content Ratings:

1. What are the most frequent content ratings on Netflix?

Is there a relationship between content rating and genre (e.g., are dramas more likely to be rated TV-MA)?