Netflix Insights

# Metrics and Insights

From the Netflix dataset, the following insights were generated:  
- Basic statistics (mean, median, count, etc.) of numeric columns were obtained using describe().  
- Distribution of content types (Movies vs. TV Shows).  
- Duration information was separated into minutes for Movies and seasons for TV Shows.  
- Temporal insights were extracted after converting 'date\_added' into datetime format.  
- Unique counts of directors, cast members, countries, and ratings provided an overview of diversity in content.

# Data Cleaning Steps

The following data cleaning steps were performed using Pandas:  
1. Removed duplicates from key columns such as director, cast, country, date\_added, and rating.  
2. Handled missing values:  
 - Filled missing 'director', 'cast', 'rating', 'duration', and 'date\_added' with 'Unknown'.  
3. Processed 'duration' column:  
 - Extracted numeric values for minutes and seasons.  
 - Differentiated between Movies (minutes) and TV Shows (seasons).  
4. Converted 'date\_added' column to datetime format for temporal analysis.  
5. Converted categorical/text columns (show\_id, type, title, director, cast, country, rating, listed\_in, description) into numeric codes for easier analysis.