# Streaming Platforms Content Analysis and Sentiment Insights

## 1. Problem Statement

With the growing number of content streaming platforms, such as Netflix, Hulu, Amazon Prime, and Disney+, competition is fierce, and understanding their content landscape has become vital for both business intelligence and user experience research.
This project aims to:
- Understand the distribution of content types (Movies vs TV Shows) across platforms.
- Identify the most popular genres and ratings.
- Analyze trends in content duration by platform and type.
- Conduct time-series analysis of monthly content additions.
- Perform sentiment analysis on titles and descriptions to evaluate the emotional tone across platforms.

This analysis supports strategic decision-making for content acquisition, user engagement strategies, and platform positioning.

## 2. Documentation Sections

### A. Introduction

#### Background

Over-the-top (OTT) platforms have revolutionized media consumption. With original content, global availability, and fierce competition, platforms like Netflix, Hulu, Amazon Prime, and Disney+ continue to compete for audience attention. Understanding their content offerings, user sentiment, and strategic additions over time can provide insights into market trends and user preferences.

#### Objective

The objective of this project is to perform a comparative analysis of four major streaming platforms, focusing on content characteristics and sentiment, using Python-based data science tools.

### Scope

- Analyze data from Netflix, Hulu, Amazon Prime, and Disney+.
- Perform EDA, time-series, and sentiment analysis.
- Visualize insights using Python libraries like Seaborn and Matplotlib.

## B. Dataset Overview

Datasets Used:
- netflix_titles.csv
- hulu_titles.csv
- amazon_prime_titles.csv
- disney_plus_titles.csv

Source: Publicly available datasets from Kaggle or similar repositories containing metadata about the titles available on each platform.

Dataset Features:
- show_id, type, title, director, cast, country, rating, date_added, release_year, duration, listed_in, description

Handling Missing Data:
- Missing textual fields (e.g., director, cast) filled with 'Unknown'.
- Null date_added values were parsed with error coercion.
- Duration cleaned and converted into integer minutes where possible.

## C. Data Preprocessing

Column Standardization:
All datasets were standardized to the same format, with a new platform column added.

Handling Missing Values:
Text fields → 'Unknown'
description → 'No Description'
date_added → parsed with error coercion

Feature Engineering:
- added_year, added_month
- duration_int (converted duration)
- title_sentiment, desc_sentiment (TextBlob)

## D. Exploratory Data Analysis (EDA)

1. Content Type by Platform – Bar chart
2. Top 10 Genres Overall – Horizontal bar chart
3. Rating Distribution – Count plot of content ratings
4. Duration Distribution – Boxplots by platform and type
5. Monthly Trends – Time-series analysis of content addition

## E. Sentiment Analysis

Method Used: TextBlob for polarity scores.

Fields Analyzed: title_sentiment, desc_sentiment

Visualizations:
1. Boxplot of sentiment polarity
2. Violin Plot of distribution shape and spread
3. Barplot of average sentiment by platform
4. Histogram + KDE per platform
5. KDE Overlay for all platforms

## F. Results & Insights

- Netflix and Amazon Prime have more Movies than TV Shows.
- Drama, Comedy, and Action dominate across platforms.
- TV-MA, PG-13, and R are the most common ratings.
- Netflix and Amazon Prime show longer content durations.
- Sentiment varies: Netflix has slightly more positive descriptions, and Disney+ has more neutral tones.

## G. Conclusion

This multi-platform analysis highlights how content libraries differ in terms of type, genre, and tone. Each platform seems to adopt unique strategies to engage viewers.

- Netflix leads in volume and genre diversity.
- Disney+ maintains a brand-safe, family-oriented tone.
- Hulu balances between modern and traditional formats.
- Amazon Prime has international and indie variety.

Sentiment analysis provides further insight into content tone and audience targeting.

## H. Recommendations (Optional)

- Optimize libraries based on genre gaps and user sentiment trends.
- Guide marketing or trailer messaging using sentiment analysis.
- Future work: include user reviews, content clustering, or popularity prediction using NLP.

## I. Technical Stack

Languages: Python

Libraries:
- Pandas, NumPy (data handling)
- Seaborn, Matplotlib (visualization)
- TextBlob (sentiment analysis)
- AdjustText (plot label adjustment)

Tools: Jupyter Notebook or VS Code

## J. File Output

Final cleaned data saved as: allstreamingplatforms_cleaned.csv