**Project 1: Netflix Data Cleaning, Analysis, and Visualization**

**Title**: Netflix Content Library: An Analysis of Trends and Distribution

**Author**: *Nilay Saha*

**Date**: August 24, 2025

**1. Introduction**

The purpose of this report is to analyze the Netflix content dataset, which includes titles added up to 2021. The analysis focuses on cleaning the data and visualizing key trends to understand the composition of the Netflix library, including its content distribution, geographical sources, and popular genres.

**2. Key Findings & Visualizations**

The following section details the primary insights derived from the exploratory data analysis.

* A graph showing a distribution of content by type

  AI-generated content may be incorrect.**Finding 1: The Netflix library is predominantly composed of movies.**
  + The dataset shows a clear majority of movies over TV shows, with movies making up approximately 70% of the content.
* A graph of a graph showing the number of years

  AI-generated content may be incorrect.**Finding 2: There was a dramatic increase in content added between 2016 and 2019.**
  + The number of titles added to the platform grew exponentially during this period, indicating a major strategic push to expand the content library.
* A graph of different colored squares

  AI-generated content may be incorrect.**Finding 3: The United States and India are the largest producers of content.**
  + The United States is by far the largest single contributor of content to the Netflix library.
  + India is the second-largest contributor, highlighting its importance as a major content market.
* A graph showing a number of content ratings

  AI-generated content may be incorrect.**Finding 4: The most common content rating is TV-MA (Mature Audience).**
  + The TV-MA rating is the most frequent, suggesting that a significant portion of Netflix's content is targeted towards adult audiences.
  + This is followed by TV-14 (Teens - 14 and up) and TV-PG (Parental Guidance suggested).
* **Finding 5: 'International Movies' and 'Dramas' are the most prevalent genres.**

A graph of different colored bars

AI-generated content may be incorrect.

* + When genres are analyzed individually, "International Movies" appears most frequently, reflecting Netflix's global content strategy.
  + "Dramas" is the second most common genre, indicating its broad appeal and popularity on the platform.

A graph of different colored bars

AI-generated content may be incorrect.**3. Conclusion**

The analysis reveals several key characteristics of the Netflix content library. It is a movie-heavy platform that experienced rapid growth in the late 2010s. The content is predominantly sourced from the United States and India and is largely aimed at a mature audience, with "International Movies" and "Dramas" being the most common genre tags. These insights provide a clear snapshot of Netflix's content strategy over the period covered by the dataset.