**Netflix Data Analysis Case Study**

**Introduction**

Netflix, one of the world's leading streaming platforms, provides a vast collection of movies, TV shows, and documentaries. Analyzing its dataset can help uncover patterns in content distribution, genre popularity, and release trends. This case study explores the Netflix dataset to derive meaningful insights through exploratory data analysis (EDA) and statistical techniques.

**Data Overview**

The dataset contains various attributes related to movies and TV shows available on Netflix. The key features include:

* **show\_id**: Unique identifier for each title.
* **type**: Indicates whether the entry is a Movie or TV Show.
* **title**: Name of the content.
* **director**: Director of the content (if available).
* **cast**: List of main actors.
* **country**: Country where the content was produced.
* **date\_added**: Date when the content was added to Netflix.
* **release\_year**: Year the content was released.
* **rating**: Content rating (e.g., PG, R, TV-MA).
* **duration**: Duration of the movie or number of seasons for TV shows.
* **genre**: Category of content such as Action, Drama, Comedy, etc.

**Problem Statement**

The goal of this case study is to analyze Netflix's content library to answer the following questions:

* What is the distribution of content types (Movies vs. TV Shows)?
* What are the most common genres available on Netflix?
* How has content availability changed over time?
* What are the trends in movie durations and TV show seasons?
* Which countries contribute the most content to Netflix?
* Are there any patterns in content ratings?

**Techniques Applied**

To explore the dataset and extract insights, the following techniques were used:

* **Data Cleaning**: Handling missing values and inconsistent data entries.
* **Descriptive Statistics**: Summarizing key numerical and categorical data points.
* **Data Visualization**: Using Seaborn and Matplotlib to create plots such as bar charts, histograms, and trend lines.
* **Correlation Analysis**: Identifying relationships between different attributes.
* **Time Series Analysis**: Understanding trends in content release over time.

**Observations & Insights**

* The dataset consists of a higher number of movies compared to TV shows.
* Certain genres like Drama and Comedy dominate the platform, while others are less common.
* Most content additions have been increasing over recent years, indicating Netflix's focus on expanding its library.
* The United States contributes the highest number of titles, followed by India and other countries.
* Ratings like TV-MA and TV-14 are the most prevalent, reflecting the platform's diverse audience base.
* Movie durations vary widely, with most falling between 80-120 minutes, while TV shows generally have 1-2 seasons.

**Conclusion**

Through this analysis, we have identified key trends in Netflix's content distribution. The findings help in understanding how Netflix curates its library, what genres are most popular, and how content availability has evolved. This study provides valuable insights for content creators, marketers, and analysts looking to understand streaming trends.