**Netflix Data Analysis Report**

**Problem Statement and Background**

**Problem Statement:** The objective of this project is to analyse the Netflix dataset to uncover insights and patterns regarding the content available on the platform. This includes understanding the distribution of content types, ratings, country of origin, production trends over the years, and key contributors in terms of directors and actors. The analysis aims to provide actionable insights for improving content strategy and enhancing viewer engagement.

**Background:** Netflix, a leading streaming service provider, offers a vast library of TV shows, movies, documentaries, and more. With the increasing competition in the streaming market, understanding the dynamics of its content library is crucial for Netflix to maintain its competitive edge. Analysing the dataset helps in identifying popular genres, viewing trends, and production patterns, which can guide future content acquisition and production decisions.

**Solutions**

To address the problem statement, the following solutions were implemented:

1. **Data Analysis using Python and Power BI:** Leveraging Python for data cleaning and initial analysis, and Power BI for creating interactive and dynamic visualizations.
2. **KPIs and Metrics:** Developing key performance indicators to measure and track important aspects of the Netflix content library.
3. **Visualizations:** Creating various visualizations to provide a clear and concise understanding of the dataset.

**Project Scope and Methodology**

**Scope:** The project focuses on the analysis of the Netflix dataset with columns: 'show\_id', 'type', 'title', 'director', 'cast', 'country', 'date\_added', 'release\_year', 'rating', 'duration', 'listed\_in', and 'description'. The analysis covers content types, ratings, country contributions, yearly production trends, and key contributors.

**Methodology:**

1. **Data Collection:** The dataset was obtained from a reliable source and loaded into the analysis environment.
2. **Data Cleaning:** Using Python, the dataset was cleaned to handle missing values, duplicate entries, and inconsistent data formats.
3. **Exploratory Data Analysis (EDA):** Conducted initial data exploration to understand the structure and main characteristics of the dataset.
4. **Visualization in Power BI:** Created interactive visualizations in Power BI to present the findings and insights.
5. **Analysis and Interpretation:** Analysed the visualizations to draw meaningful conclusions and recommendations.

**Goals and KPIs**

**Goals:**

1. Identify the distribution of content types (Movies vs. TV Shows).
2. Determine the most common ratings for the content.
3. Analyse the geographical distribution of content production.
4. Examine the production trends over the years.
5. Identify top directors and actors contributing to the Netflix library.

**KPIs:**

1. **Total Content:** Total number of shows and movies available on Netflix.
2. **Total Number of Directors:** Unique number of directors contributing to the content.
3. **Total Number of Actors:** Distinct actors featured across various titles.
4. **Total Movies:** Number of movies in the Netflix collection.

**Recommended Analysis**

1. **Content by Type:** Visualize the distribution of Movies vs. TV Shows.
2. **Content by Country:** Analyse the content based on the country of origin.
3. **Yearly Content Production:** Chart the number of titles released each year.
4. **Rating Distribution:** Examine how content is rated.
5. **Content Added Over Time:** Cumulative view of the content library growth.
6. **Show ID by Duration:** Investigate the duration of content.

**Conclusion**

1. **The rating "TV-MA" holds the highest percentage at 36.4%.**
   * This indicates a significant portion of Netflix content is intended for mature audiences.
2. **Most of the content was produced during 2015-2022.**
   * The recent years have seen a substantial increase in content production, reflecting Netflix's aggressive content expansion strategy.
3. **2018 was the year when most of the movies were released.**
   * 2018 stands out as a peak year for movie releases on Netflix, suggesting a strategic push to enhance their movie library.
4. **Type-Movie was the most played.**
   * Movies are a major part of Netflix's offering, attracting a large viewer base.
5. **The United States has the highest contribution.**
   * The US is the leading contributor to Netflix's content library, indicating strong domestic production and acquisition efforts.