Netflix Data Analysis Report

1. Project Overview

This project focuses on the cleaning, analysis, and visualization of Netflix content data from 2008 to 2021.

The aim is to uncover patterns in content type, release timing, geography, and genre preferences.

2. Dataset Description

- Total Records: 8,790 titles
- Key Columns: show_id, type, title, director, country, date_added, release_year, rating, duration, listed_in
- Data Source: Netflix titles dataset from Kaggle

3. Data Cleaning Process

- Removed duplicate entries
- Handled missing values in critical fields like director, cast, country
- Converted 'date_added' column to datetime format
- Extracted year, month, and day from 'date_added' for time analysis

4. Exploratory Data Analysis

- **Content Type**: 70% Movies, 30% TV Shows
- **Rating Analysis**: Most titles rated TV-MA and TV-14
- **Country Contribution**: US and India dominate content contribution
- **Monthly Trends**: Highest content uploads in July and August
- **Genre Insights**:
  - Movies: Dramas, Documentaries, Comedies are common
  - TV Shows: International and Kids programming are frequent
- **Top Directors**: Rajiv Chilaka, Steven Spielberg, and others have the most titles

5. Visual Insights

- Bar and pie charts for rating distribution
- Line plots for monthly and yearly releases
- Genre-wise content frequency bar charts
- Director popularity visualizations

6. Key Takeaways

- Netflix content has grown consistently from 2008 to 2021
- Majority of content is rated for mature audiences
- US-based content dominates, though international presence is growing
- Diverse genres allow Netflix to cater to global audiences

7. Applications and Future Scope

- Basis for developing recommendation engines
- Insights for regional content production strategies
- Can support machine learning models for predictive analytics