**Netflix Content Analysis Project**

**Project overview**

A data analysis project on this Netflix dataset would provide valuable insights into content strategy, audience demographics, and production trends. This project analyses the Netflix content Catalog (netflix1.csv) to uncover key trends in content type, production, and distribution over time. The goal is to provide **strategic recommendations** for content investment and geographical expansion.

**Tools and Techniques**

* **Power BI /Excel –** Data cleaning, modelling, and Dashboard creation
* **SQL** – Querying and aggregating data
* **Python** ( pandas)- For Deeper Exploratory Data Analysis (EDA) and forecasting.

**Process**

**1. Data Cleaning and Pre-processing**

The initial steps of loading and inspecting the data (8,790 rows, 10 columns) revealed a clean structure but highlighted specific columns requiring transformation for effective analysis.

| Column | Data Type | Cleaning Action |
| --- | --- | --- |
| **director** | Object | Replace 'Not Given' with **'Missing'** (as an analysis category). |
| **country** | Object | Replace 'Not Given' with **'Unknown'** (as an analysis category). Split multi-country entries (e.g., 'United States, Canada') to count each country individually. |
| **Date\_added** | Object | Convert to **datetime** objects. Extract add\_year and add month for time-series analysis. |
| **duration** | Object | **Split** into two columns: Duration\_Value (numeric) and Duration\_Unit ('min' or 'Season'). This is crucial for calculating average movie length and season count for TV Shows. |
| **Listed\_in** | Object | **Split** the comma-separated genres into individual rows for genre-level analysis (one-hot encoding or melting for 'genre-based' metrics). |
| **rating** | Object | **Standardize** similar rating categories (e.g., group different forms of 'PG' if needed, or handle cases like 'TV-Y7-FV'). |

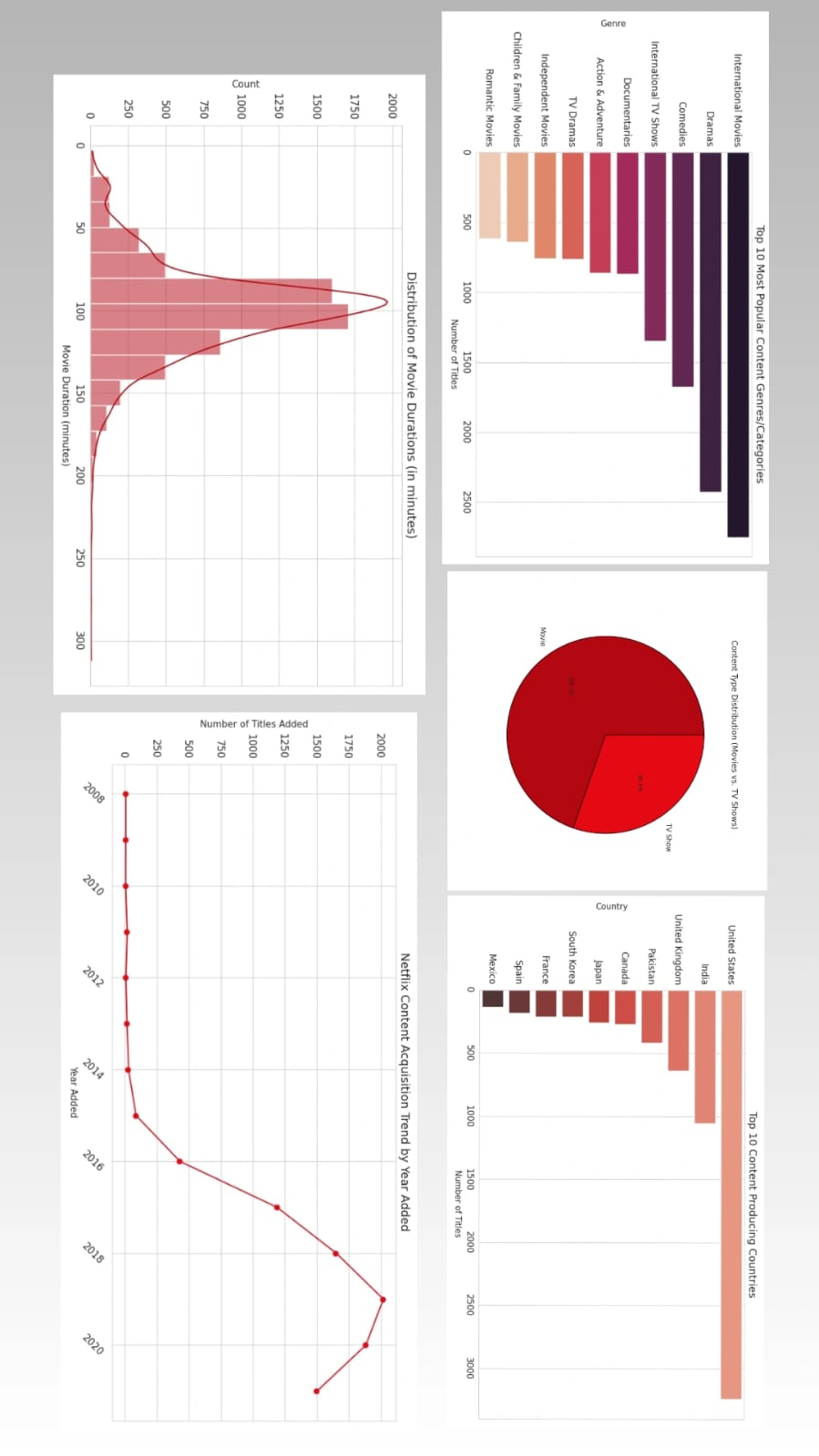


**2. Data Analysis and Visualization**

The core analysis will focus on five key areas to generate impactful visualizations and findings.

**Key Visualizations**

| Focus Area | Metric/Analysis | Visualization | Insight Goal |
| --- | --- | --- | --- |
| **Content Mix** | Ratio of **Movies vs. TV Shows** and their trend over time. | **Pie Chart & Stacked Bar Chart** | Understand the platform's primary content focus and recent changes. |
| **Geographic Focus** | **Top 10 Content Producing Countries** (excluding 'Unknown'). | **Horizontal Bar Chart** | Identify key markets for production and assess geographical diversity. |
| **Content Growth** | Number of Titles **Added to Netflix by Year** (using add\_year). | **Line Chart** | Visualize the platform's expansion and growth rate. |
| **Genre Popularity** | **Top 10 Most Popular Genres** (based on individual counts from listed\_in). | **Horizontal Bar Chart** | Guide future investment in high-demand genres. |
| **Movie Duration** | Distribution of **Movie Durations** (in minutes). | **Histogram** | Find the optimal/most common movie runtime. |

****

**3. Project Report Structure**

The final output will be a comprehensive report structured to deliver maximum strategic value to stakeholders.

**I. Executive Summary**

* **Objective:** State the project's goal (e.g., analyze content catalog to inform content strategy).
* **Key Finding Snapshot:** Summarize 2-3 most critical findings (e.g., "Netflix heavily favors movies but is aggressively increasing its TV Show catalog" and "The United States and India dominate content production.").
* **Actionable Recommendation:** Provide the top 1-2 strategic recommendations (e.g., "Increase investment in international TV drama co-productions from emerging markets.").

**II. Data Overview**

* **Dataset Description:** Explain the source, size (8,790 rows, 10 columns), and time frame (oldest/newest release\_year).
* **Data Cleaning Summary:** Briefly detail the process: handling of 'Not Given' values, and the creation of new columns like add\_year and Duration\_Unit.
* **Initial Observations:** Note the prevalence of the 'object' data type and the necessity of splitting the listed\_in column.

**III. Key Visualizations**

* *Presentation of the 5 key charts (Content Mix, Top Countries, Content Growth, Top Genres, Movie Duration).* Each visualization will be accompanied by a brief caption explaining what it shows.

**IV. Analytical Findings**

* **Content Strategy:** Detail the precise ratio of movies to TV shows. Discuss which type of content is being added more aggressively recently.
* **Geographical Concentration:** Quantify the production dominance of the top 3 countries and highlight content sources that are currently underrepresented.
* **Audience Segmentation (via Ratings):** Analyze the distribution of ratings to understand which audience segments (e.g., 'TV-MA' for mature, 'TV-G' for general) are most catered to.
* **Genre Deep Dive:** Report on the specific genres that are most common across the entire catalog and note the genres most prevalent in **newly added** content.

**V. Strategic Recommendations**

These recommendations translate the analytical findings into business actions:

1. **Content Focus:** Recommend a target ratio for future investment (e.g., shift from 70/30 Movie/TV Show to 60/40) to capitalize on the success of TV series.
2. **Market Expansion:** Suggest specific countries (e.g., Brazil, Turkey, Mexico) as high-potential targets for original content co-production, based on their existing content output and local market demand.
3. **Genre Investment:** Advise on prioritizing investment in high-performing genres (e.g., 'International TV Shows' or 'Documentaries') while exploring a potential gap in a currently underrepresented category.
4. **Content Length Optimization:** Based on the Movie Duration Histogram, recommend an ideal runtime for future movie acquisitions or productions (e.g., target minutes).

**VI. Conclusion**

This comprehensive analysis successfully achieved its objective of dissecting the Netflix content catalog, providing a clear map of the platform's current content strategy and historical growth.

The findings confirm that Netflix remains primarily a movie platform, with films constituting the majority of the total catalog. However, the aggressive increase in titles added yearly, particularly in recent years, signals a decisive expansion phase. Geographically, content sourcing is highly concentrated, with the **United States** and **India** serving as the indisputable production leaders, emphasizing a reliance on these key markets. Furthermore, the high prevalence of **International TV Shows**, **Dramas**, and **Comedies** highlights the core genres driving current audience engagement.