COMSATS UNIVERSITY ISLAMABAD

DATA VISUALIZATION TECHNIQUES

LAB TERMINAL (REPORT)

SUBMITTED TO:

Dr. Sheneela Naz

SUBMITTED BY:

Talal Atiq (FA22-BDS-052)
Shahab Hassan (FA22-BDS-052)

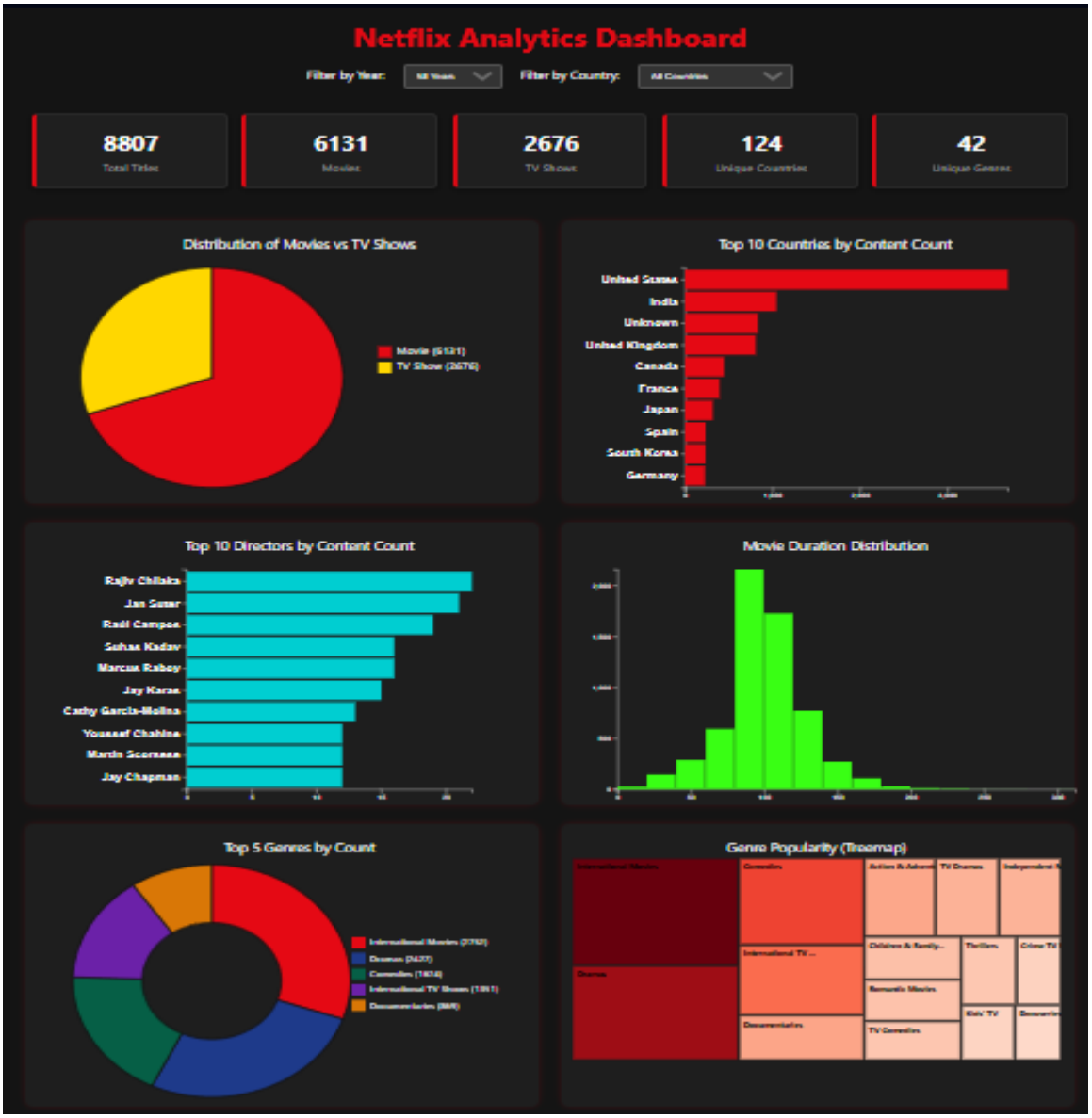
Project: **Netflix Analytics Dashboard**

Description: An Interactive Data Visualization Dashboard using D3.js

Introduction

The **Netflix Analytics Dashboard** is an interactive, browser-based tool built using D3.js, HTML, and CSS. It visualizes the Netflix global content catalog, offering users the ability to filter and analyze data by year and country. The dashboard helps surface key trends in production, genres, ratings, and content type through responsive KPIs and six insightful visualizations.

Dashboard Overview



The dashboard is laid out in a grid format with two charts per row across three rows, preceded by a section containing:

- Dashboard title
- Year filter (global)
- Country filter (affects selected charts)
- Five KPIs for quick insights

Color Justification:

- A dark theme was used, inspired by Netflix's UI aesthetic, to reduce visual strain and maintain brand consistency.
- Accent colors like Netflix red and variations of vibrant hues (orange, teal, purple, yellow, green) were used to distinguish chart elements.
- Tooltips and highlights use lighter tones to improve visibility on dark backgrounds.

Visualization Techniques Used:

- **Pie and Donut Charts:** For proportional comparison (e.g., content type and top genres)
- **Bar Charts (Horizontal):** To easily display top contributors like countries and directors
- **Histogram**: For distribution analysis of durations
- **Treemap:** For compact hierarchical genre views

The layout ensures readability, responsiveness, and interactivity, offering a cohesive analytical experience.

Dataset Overview

Dataset Name: *Netflix Movies and TV Shows*

Source: https://www.kaggle.com/datasets/shivamb/netflix-shows (kaggle)

Total Entries: ~8,800

Purpose: This dataset contains information about movies and TV shows available on Netflix as of 2021. It is used to explore content availability, genre trends, country-wise production, duration patterns, and more.

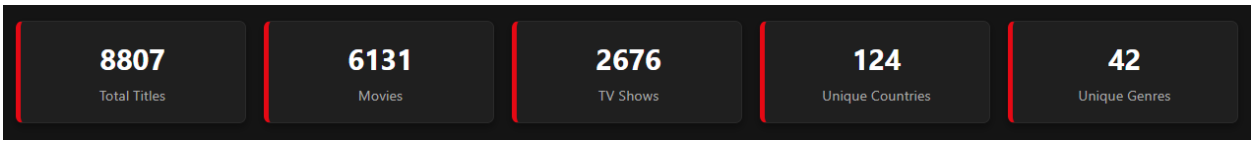
Attribute	Description
show_id	A unique identifier assigned to each title in the dataset.
type	Indicates whether the content is a *Movie* or a *TV Show*.
title	The name of the movie or TV show.
director	Name(s) of the director(s) responsible for the title. Can be blank for some entries.
cast	List of main actors/actresses in the title. May contain multiple names separated by commas.
country	Country or countries where the title was produced.
date_added	The date the title was added to Netflix. Helpful in analyzing content trends over time.
release_year	The year the title was originally released.
rating	Netflix content rating (e.g., *TV-MA*, *PG*, *R*, etc.), indicating age suitability.
duration	Duration of the title — in minutes for movies and number of seasons for TV shows.
listed_in	Genres or categories assigned to the title (e.g., *Dramas*, *Comedies*, etc.).
description	A short summary or synopsis of the title.

Data Preparation Techniques

To build an accurate and dynamic dashboard, we applied the following:

1. **Date and Time Conversion**:
    - release_year converted to integers
    - duration parsed:
        - Movies: "90 min" → 90
        - TV Shows: "2 Seasons" → 2
2. **Data Cleaning**:
    - Filled missing country and listed_in with "Unknown"
    - Standardized casing and whitespace
    - Dropped null title rows if present
3. **Feature Extraction**:
    - Extracted multiple values from comma-separated fields (cast, country, listed_in)
    - Counted appearances for top 10/5 visualizations
    - Ensured each record contributes accurately across visuals
4. **Filtering Mechanism**:
    - Year filter: affects **all** visuals and KPIs
    - Country filter: affects Pie Chart, Donut Chart, Treemap, and KPIs

KPIs

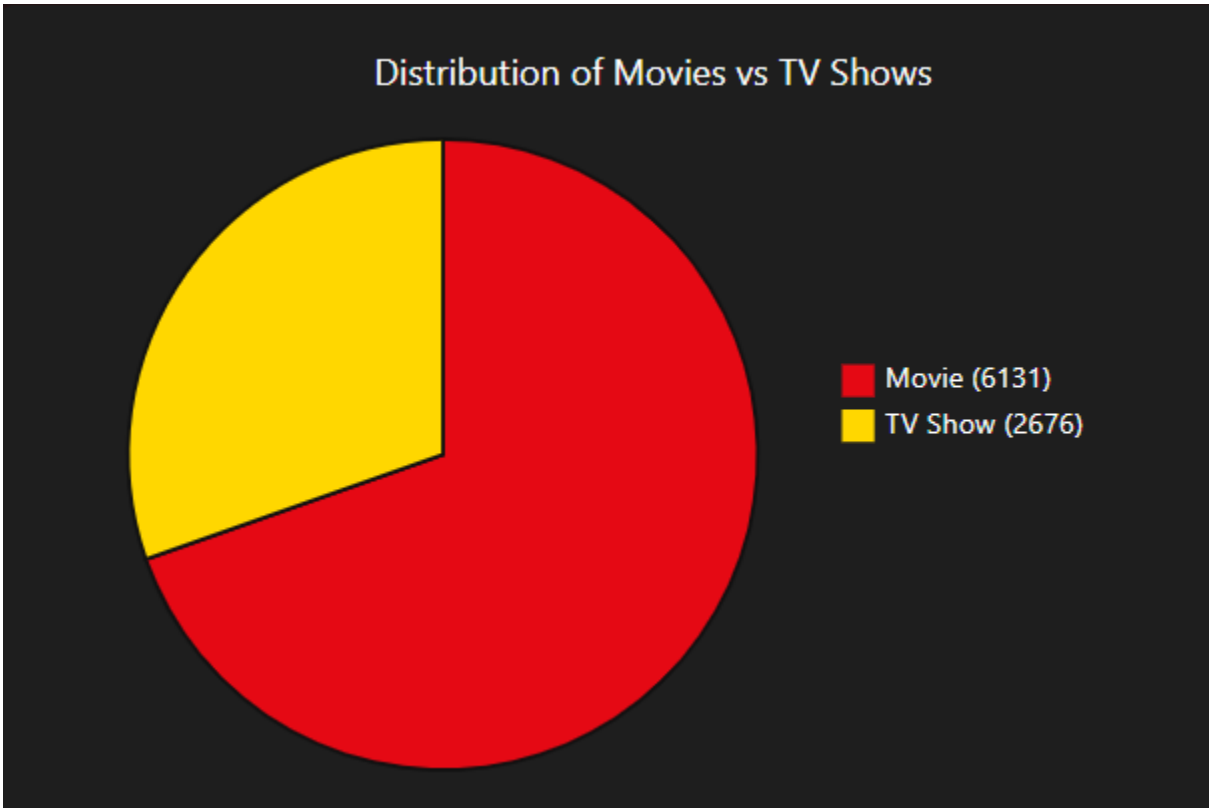


KPI	Justification
Total Titles	Indicates dataset size and helps monitor content growth. ![Insert Screenshot]
Movies Count	Shows standalone content availability, reflecting Netflix's film focus. ![Insert Screenshot]
TV Shows Count	Highlights serial content, key for binge-viewing trends. ![Insert Screenshot]
Unique Countries	Measures global content diversity — strategic for international markets. ![Insert Screenshot]
Unique Genres	Reflects genre richness and audience targeting. ![Insert Screenshot]

Dashboard Visualizations

Each visualization was carefully chosen based on the insights it could best provide from the dataset, interactivity potential, and alignment with user expectations. All visualizations dynamically respond to filters applied via year and country dropdowns (where applicable). Every chart features tooltips to enrich user interaction and make the insights more accessible at a glance.

1. *"What's the proportion of Movies vs TV Shows?"*



Purpose:

This chart gives users a quick and immediate understanding of the overall content split between Movies and TV Shows on Netflix.

Why a Pie Chart?

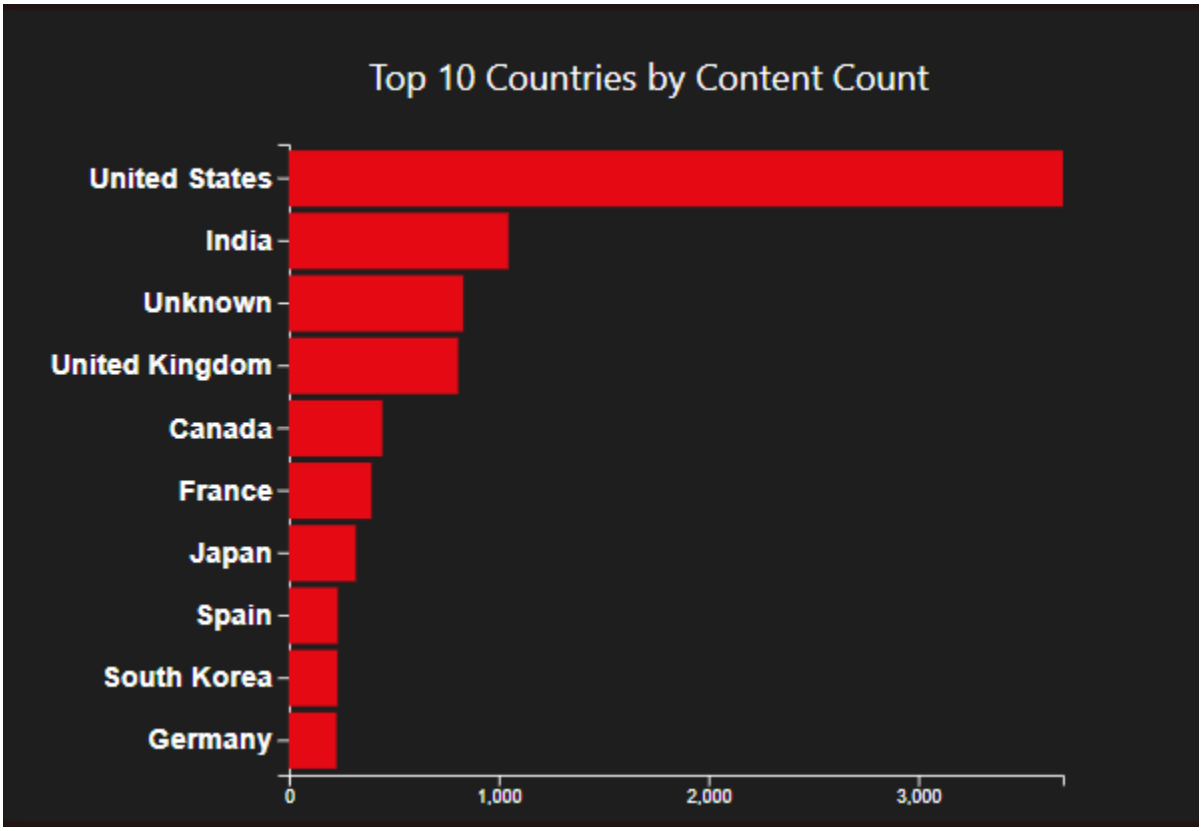
- Binary comparisons (Movies vs TV Shows) are naturally intuitive in a pie format.
- The circular form allows viewers to grasp proportion with minimal effort.
- It uses clear coloring to differentiate the two types.
- It's easy to scan and interpret.

Interactivity Features:

- Tooltip on hover shows exact count and percentage for each type.
- Dynamically updates based on both year and country filters.
- When country is filtered, it shows only the distribution of content released in that country for the selected year.

Visualization Analysis: We can conclude that Netflix has a greater share of TV Shows than movies in recent years from 2015 onwards, in years prior to that Netflix content type was mostly movies.

2. *"Which countries produce the most Netflix content?"*



Purpose:
This horizontal bar chart shows the top 10 countries based on the number of titles produced or released.

Interactivity Features:

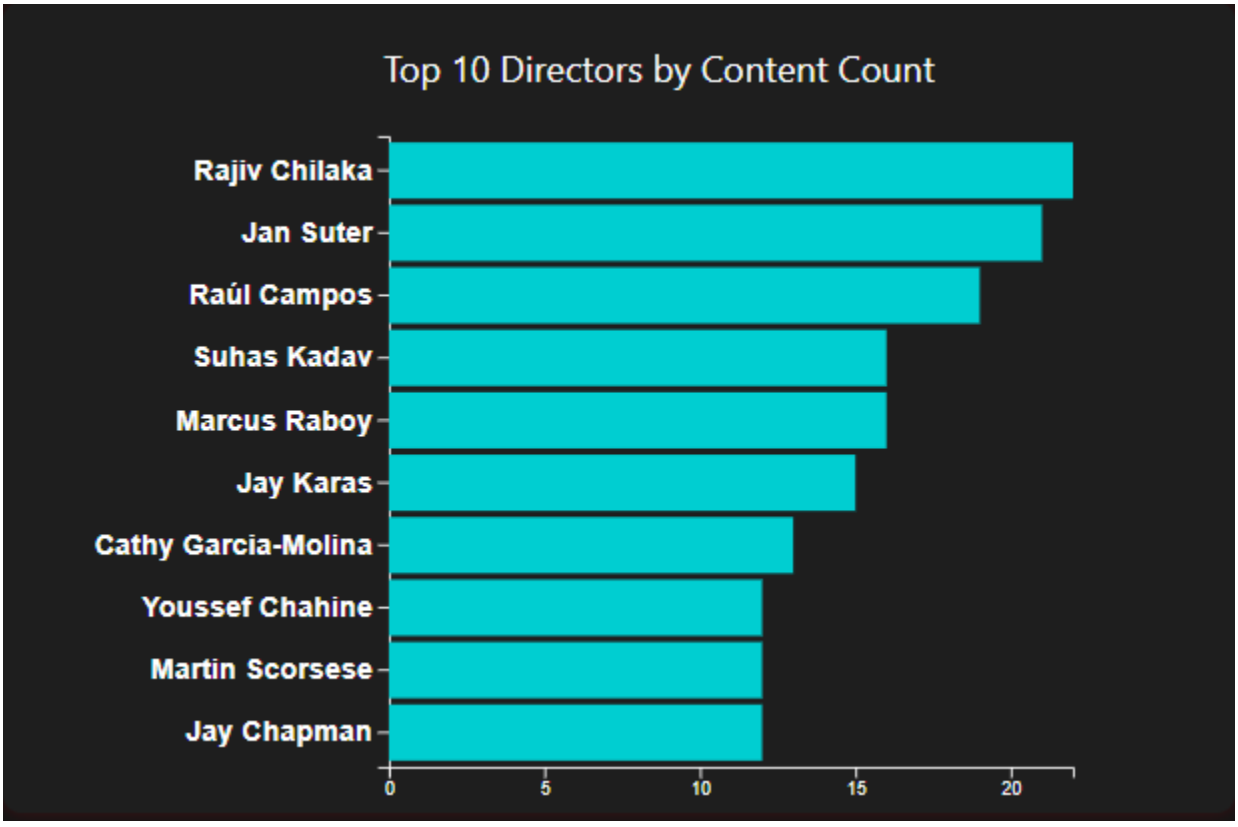
- Tooltip reveals the number of titles per country on hover.
- Dynamically responds to year filter (global count only; country filter not applied to avoid redundancy).

User Insight Gained:

- Users can quickly identify leading content-producing nations, and how that changes over time.

Visualization Analysis: By applying year filter for various years, we can draw a conclusion that most of the content uploaded on Netflix is from the United States. Countries like India, United Kingdom and France also contribute heavily.

3. *"Who are the most prolific directors on Netflix?"*



Purpose:
Provides insight into the top 10 directors with the most titles available on Netflix.

Interactivity Features:

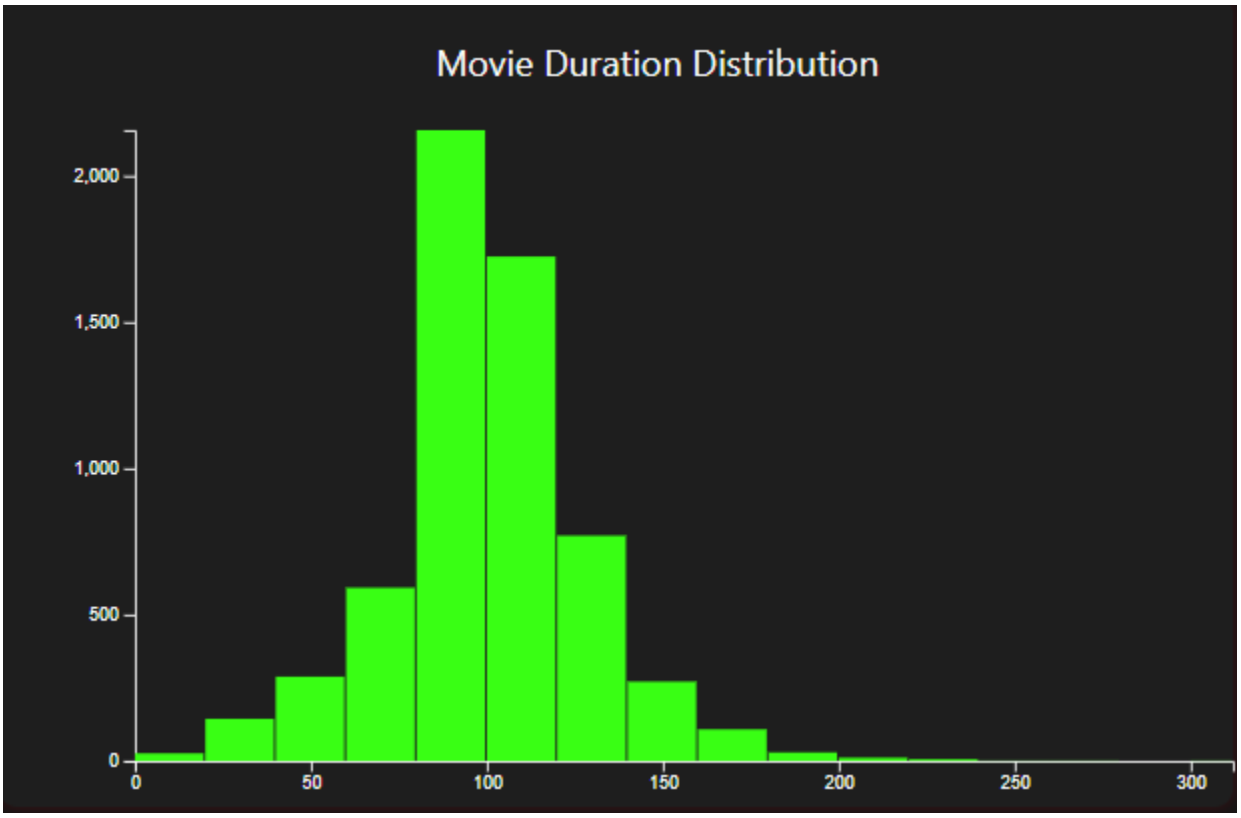
- Tooltip shows director name and number of titles.
- Updates dynamically based on year filter.
- Does not apply country filter since many titles span multiple countries or lack country metadata.

Value to Viewer:

- Reveals influential creators and trends over time.

Visualization Analysis: The visualization shows that Rajiv Chakla's content is the most common on Netflix over various years. Directors like Jan Suther and Yousseff Chahine also contribute significantly.

4. *"What is the distribution of movie durations?"*



Purpose:
Analyzes how long Netflix movies typically are, and whether there are clusters or outliers.

Data Preprocessing Note:

- The “duration” column, originally a string (e.g., “90 min”), was cleaned and converted to integer minutes for meaningful visualization.

Interactivity Features:

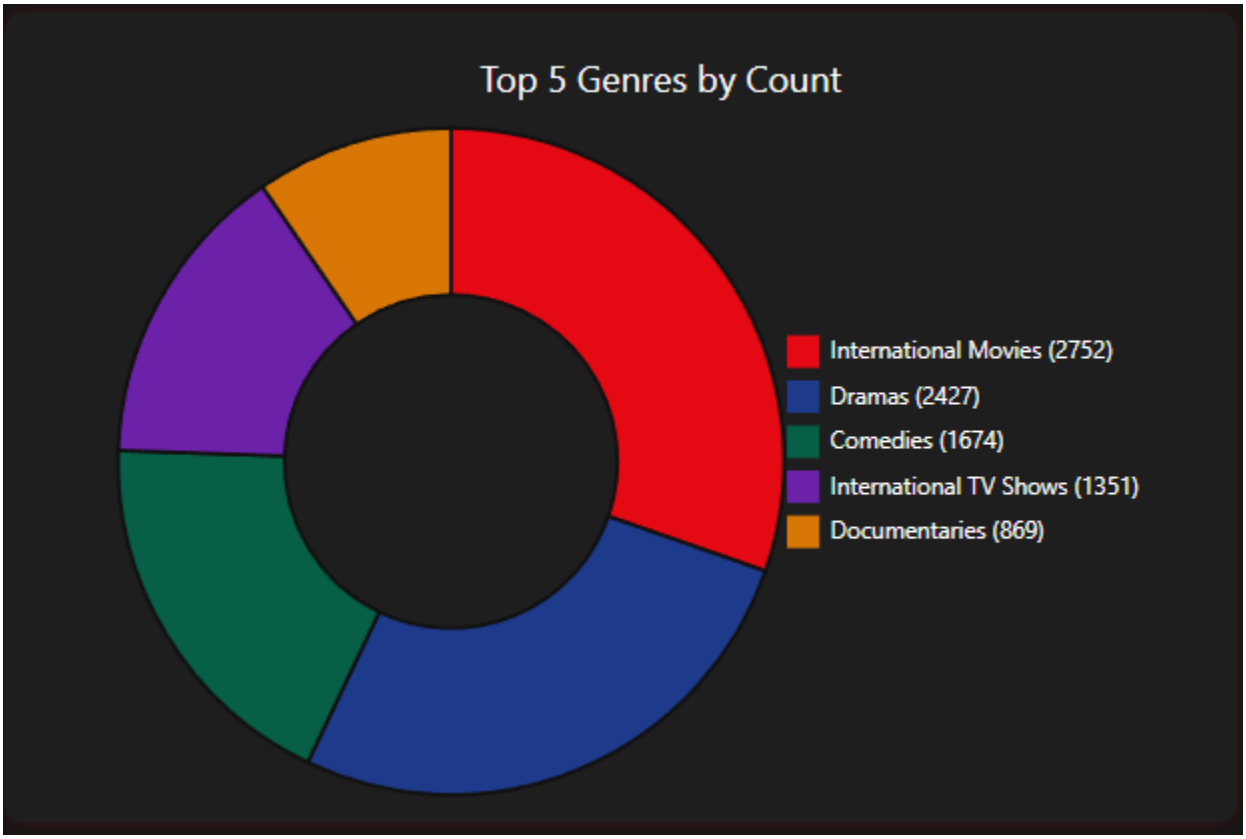
- Tooltip shows the number of movies in each time bin.
- Updates with the year filter (only Movies considered).
- Country filter is not applied here, focusing on global movie length trends.

User Value:

- Users can understand if content is skewed toward short or long-form content.

Visualization Analysis: The visualization shows that the duration of most movies uploaded on Netflix in recent years is between 90 to 120 minutes whereas movies uploaded in years prior to 2015 are often longer than the latest ones.

5. *“What are the most popular genres?”*



Purpose:
Shows the top 5 most frequent genres listed in Netflix titles.

Why a Donut Chart?

- A donut chart draws focus to proportional comparisons but with central space for additional emphasis (e.g., total genre count).
- It's more visually balanced than a pie chart when there are several categories.
- Uses distinct colors and legend for clarity.

Interactivity Features:

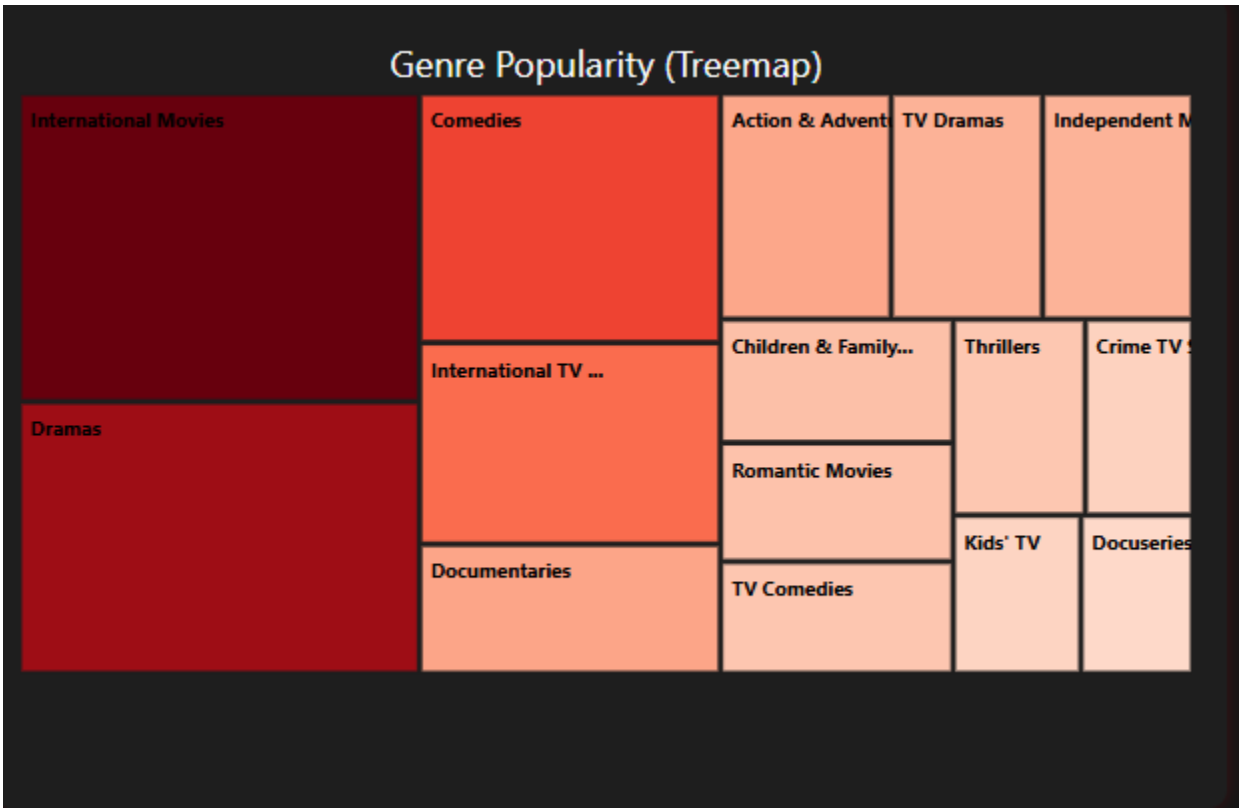
- Tooltip reveals genre name and number of titles.
- Fully responds to year and country filters.
- When a country is selected, only genres from that country's titles are shown.

Why It's Effective:

- Allows quick identification of dominant genres in a selected year/country context.
- Visually appealing and compact.

Visualization Analysis: The donut chart shows the top 5 genres by count, with *International Movies* (2,752) being the most common, followed by *Dramas* (2,427). *Comedies* and *International TV Shows* have moderate representation, while *Documentaries* (869) are the least frequent. This indicates a strong preference for international content and drama among viewers.

6. *"How is Netflix's content distributed across genres and content types?"*



Purpose:
Visually maps the hierarchy between content types (Movie/TV) and genres, highlighting which combinations dominate the catalog.

Why a Treemap?

- Treemaps are ideal for nested categorical data with quantity-based comparisons.
- They use block size to indicate volume—larger blocks represent genres with more titles.
- Color coding helps distinguish between parent categories (Movies vs TV Shows).

Interactivity Features:

- Tooltip displays genre and count.
- Updates with year and country filters.
- Hovering highlights the focus area for better readability.

Unique Insight:

- Offers a compressed, at-a-glance view of which genre-type combos dominate the platform.
- Easier to digest than long lists or bar charts when categories are many.

Visualization Analysis: The treemap visualization highlights genre popularity, with *International Movies* and *Dramas* occupying the largest areas, indicating their dominance. *Comedies*, *International TV Shows*, and *Documentaries* also show significant presence. Smaller blocks like *TV Comedies*, *Crime TV*, and *Kids' TV* reflect more niche viewer interest.

Difficulties Overcome

- **Incomplete Fields**: Fields like cast, country, director were often missing. We handled these gracefully with fallbacks like "Unknown".
- **Irregular Durations**: Parsing duration column into consistent numeric form required special logic for TV and Movies.
- **Tooltip and Responsiveness**: Ensured all charts had smooth interactive tooltips and dynamic resizing based on filters.
- **Data Normalization**: Properly accounted for multi-country and multi-genre entries to avoid skewed charts.

Tools & Technologies Used

Development & Programming Tools

- **HTML5** – Structured the web page and dashboard layout.
- **CSS3** – Designed the visual appearance with a custom dark theme inspired by Netflix, including responsive layout and red accent highlights.
- **JavaScript (ES6**) – Handled dynamic behavior, rendering, filtering, and interactivity
- **D3.js (v7**) – The core data visualization library used for creating all six interactive charts and KPIs. D3 was selected for:

Data Processing

- **D3.csv** – Used to load and parse the CSV dataset directly in the browser.

Version Control

- **Git** – Used for source control to track development and changes.
- **GitHub** – Hosted the project repository, enabling versioning, and public sharing.

Deployment

Hosting Platform: Vercel

Steps Taken:

1. Connected GitHub repository:
2. Uploaded all HTML, CSS, JS, and dataset files.
3. Deployed with automatic updates on GitHub push.
4. **Live Dashboard**: 🔗 https://netflix-analytics-dashboard.vercel.app

This project demonstrates the power of D3.js in building responsive, real-time data dashboards. By using advanced filtering and visualization techniques, we transformed raw data into an engaging tool for discovering insights from Netflix's content catalog. With modular JavaScript architecture, interactive visuals, and thoughtful design choices, the dashboard serves both as an analytical tool and a front-end data visualization showcase.

Conclusion: The Netflix Analytics Dashboard leverages key columns from the dataset such as type, country, listed_in, duration, release_year, and director to uncover meaningful patterns in Netflix's content catalog. The type column revealed a growing shift toward TV Shows in recent years, while country highlighted the United States as the dominant content producer, followed by countries like India and the UK. The listed_in (genre) field enabled a deep dive into user-preferred genres, showing a clear dominance of Dramas and International content. The duration column provided insight into content length trends, with most movies falling in the 90–120 minute range. Meanwhile, release_year allowed for time-based filtering to observe changes over the years, and director data helped spotlight the most prolific creators on the platform. These attributes, when visualized, provide a holistic view of Netflix's evolving content strategy.