

## Netflix Data: Cleaning, Analysis & Visualization

Internship Project – Unified Mentor Pvt. Ltd.

Submitted by: B. Archana

---

### 1. Project Overview

This project focuses on analyzing Netflix's dataset to understand content trends across movies and TV shows. It involves:

Data cleaning

Exploratory data analysis (EDA)

Visualizations

Building an interactive Streamlit dashboard

The goal is to derive insights such as top genres, content distribution, yearly trends, and country-wise analysis.

---

### 2. Dataset Information

Source: Netflix titles dataset

Rows: 8,790

Columns: 12

Includes:

show\_id

type (Movie / TV Show)

title

director

country

date\_added

release\_year

rating

duration

listed\_in (Genres)

---

### 3. Project Folder Structure

```
NETFLIX-PROJECT/ | └── data/ | └── raw/ | | └── netflix1.csv | └── clean/ | └── netflix_cleaned.csv |  
| └── notebooks/ | └── analysis.ipynb | └── dashboards/ | └── app.py | └── reports/ | └──  
| └── Netflix_Report.md | └── requirements.txt └── venv/
```

---

## ❖ 4. Data Cleaning Process

### ✓ Missing Values Handling

Replaced empty director and country fields with "Not Given"

Filled missing dates with most frequent value

Cleaned inconsistent formats

### ✓ Converted Data Types

date\_added → datetime

duration → numeric + unit separated

Cleaned genre strings

### ✓ Removed duplicates

Duplicate rows were identified and removed

### ✓ Saved cleaned dataset

data/clean/netflix\_cleaned.csv

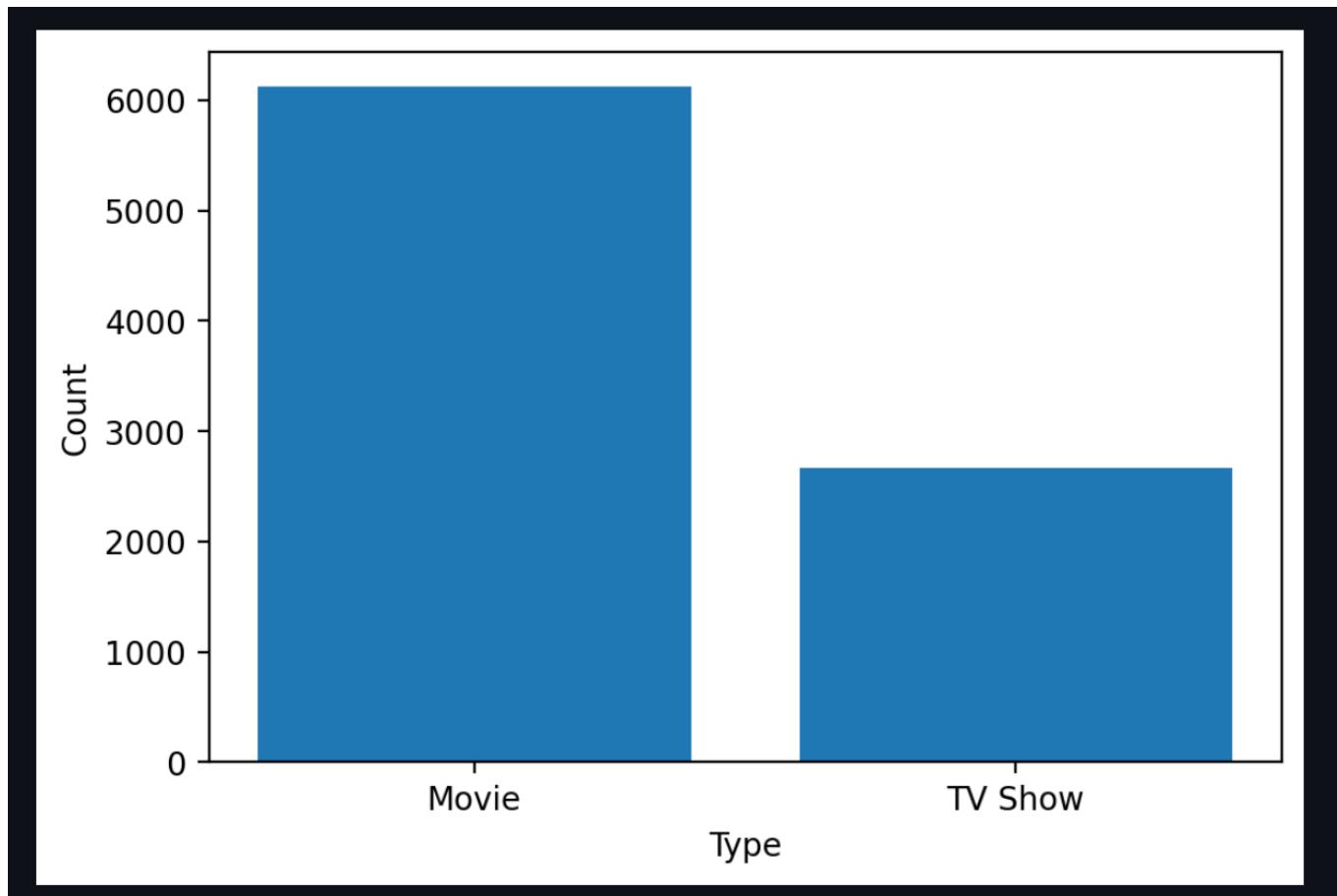
---

## ❖ 5. Exploratory Data Analysis (EDA)

Below visuals were generated from analysis.ipynb.

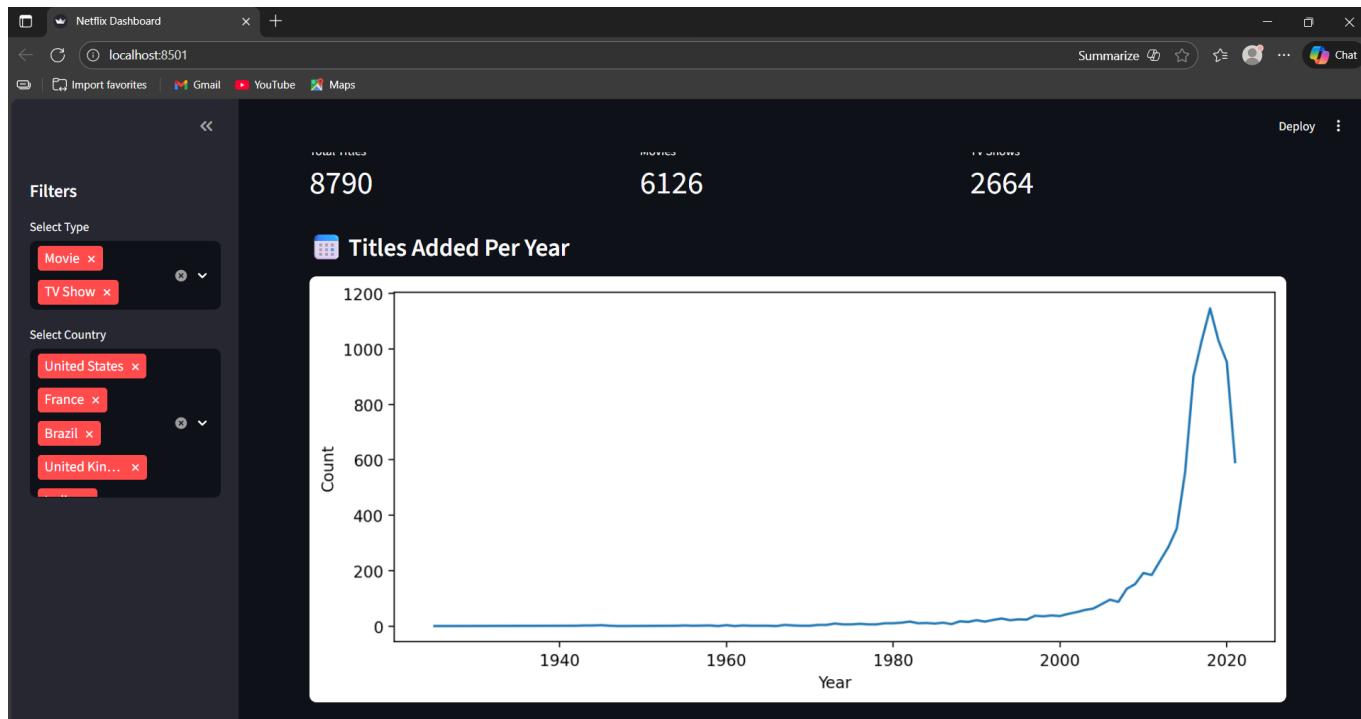
---

## ❖ 5.1 Distribution of Movies vs TV Shows



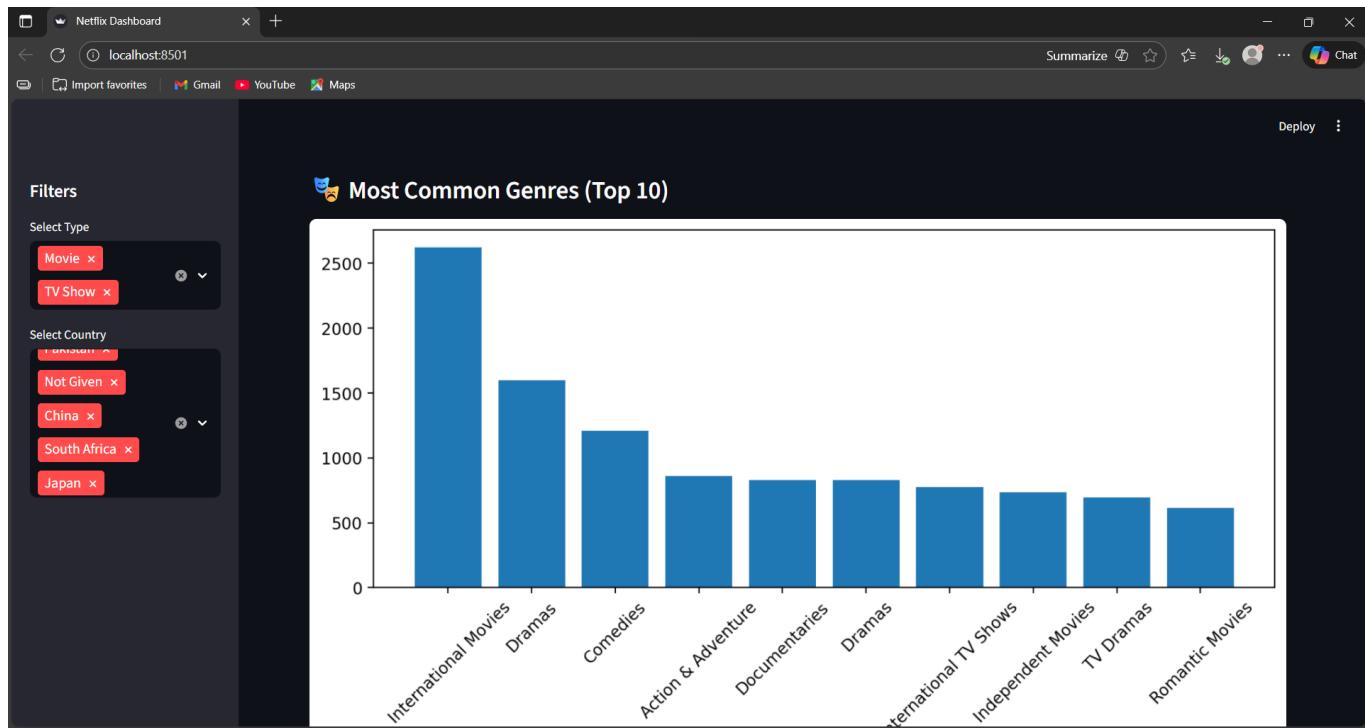
### ❖ 5.2 Titles Added Per Year

Netflix drastically increased content after 2015, peaking around 2020.



### ❖ 5.3 Top 10 Most Common Genres

International Movies & Dramas dominate the platform.



## 5.4 Word Cloud of All Genres

A word cloud was generated to visualize frequency of genres.



## 6. Interactive Dashboard

A fully interactive dashboard was created using Streamlit.

## Features:

### Filter by Type (Movie / TV Show)

Filter by Country

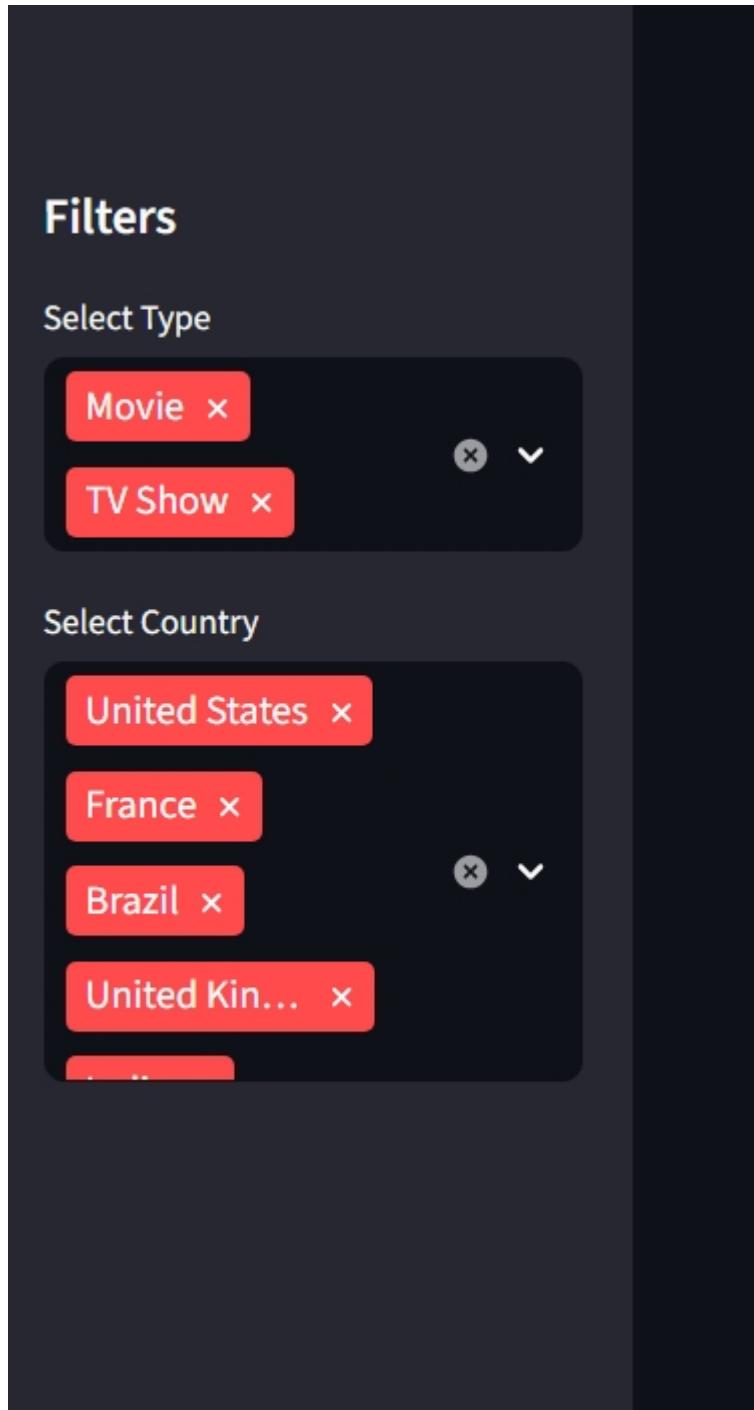
Yearly trend visualization

Genre distribution

Dynamic table view

---

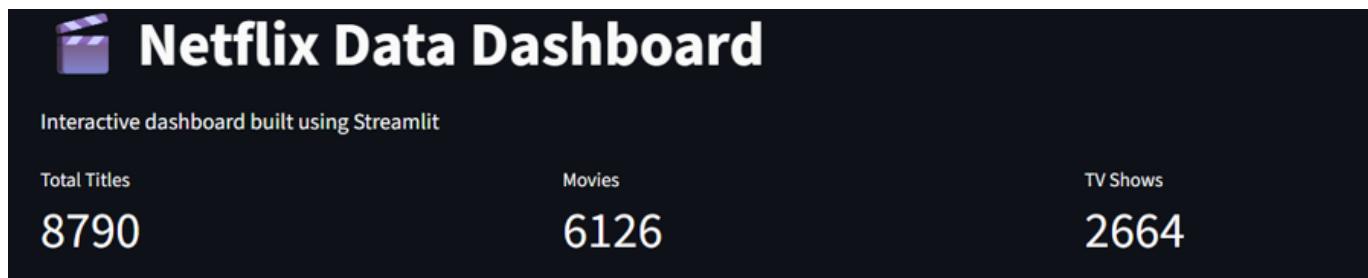
#### ❖ 6.1 Dashboard Filters Section



---

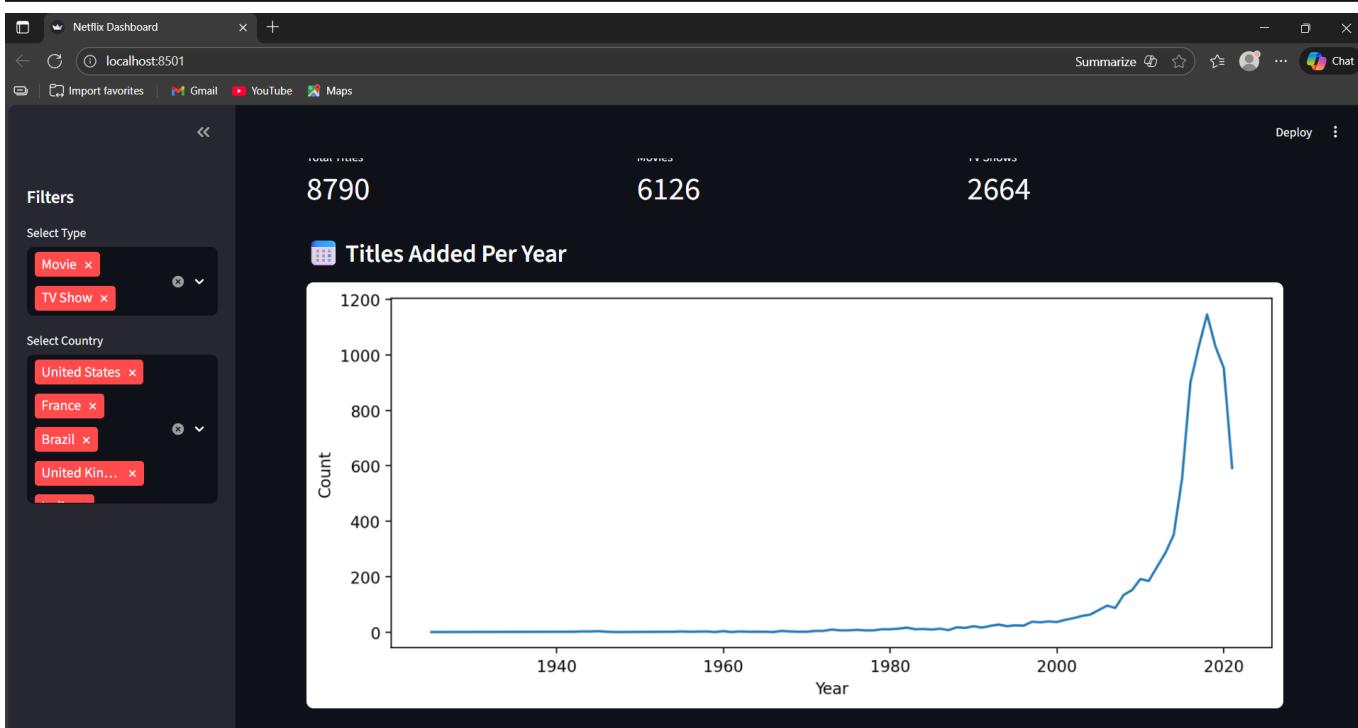
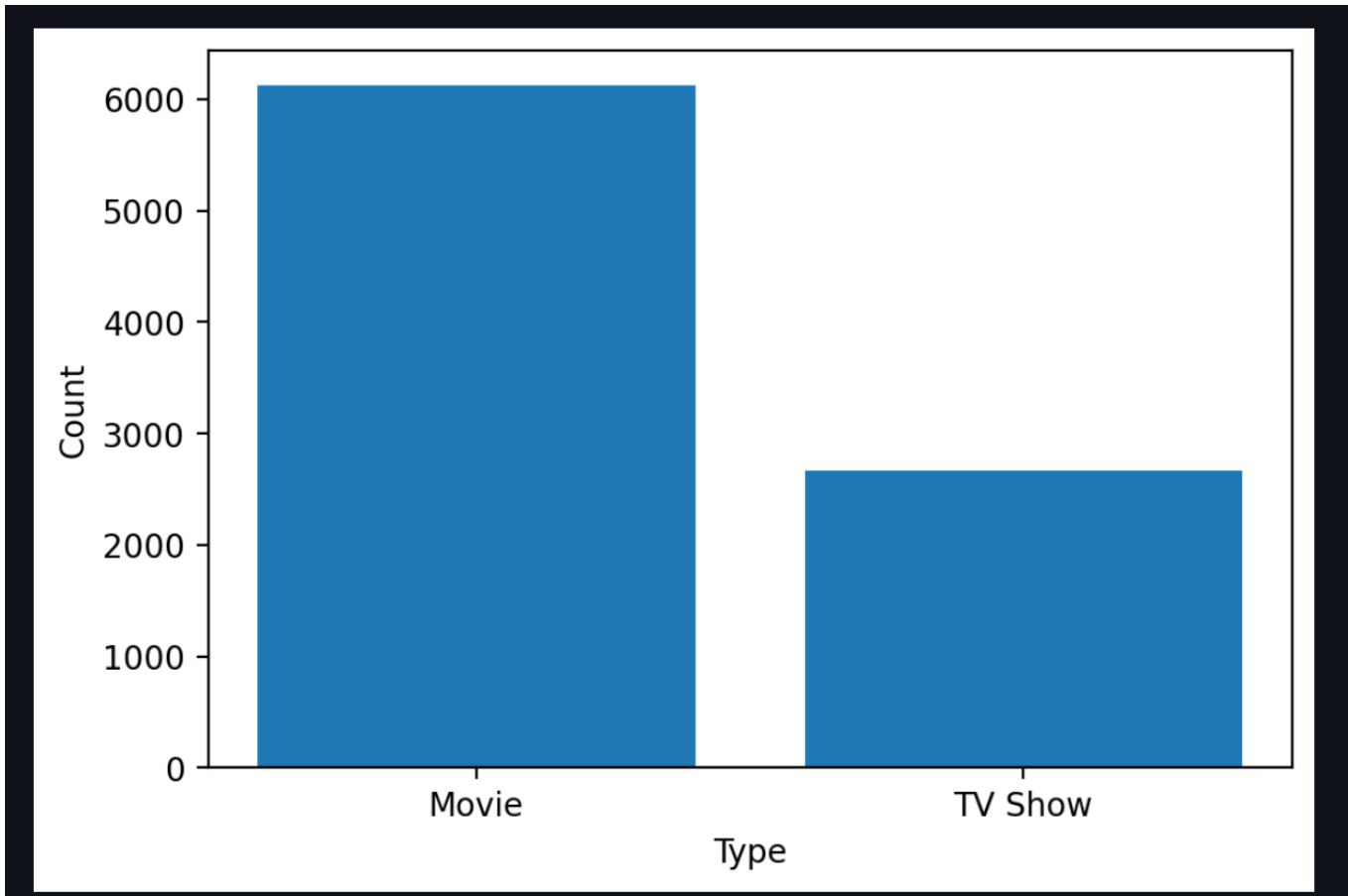
#### ❖ 6.2 Dashboard Metrics Section

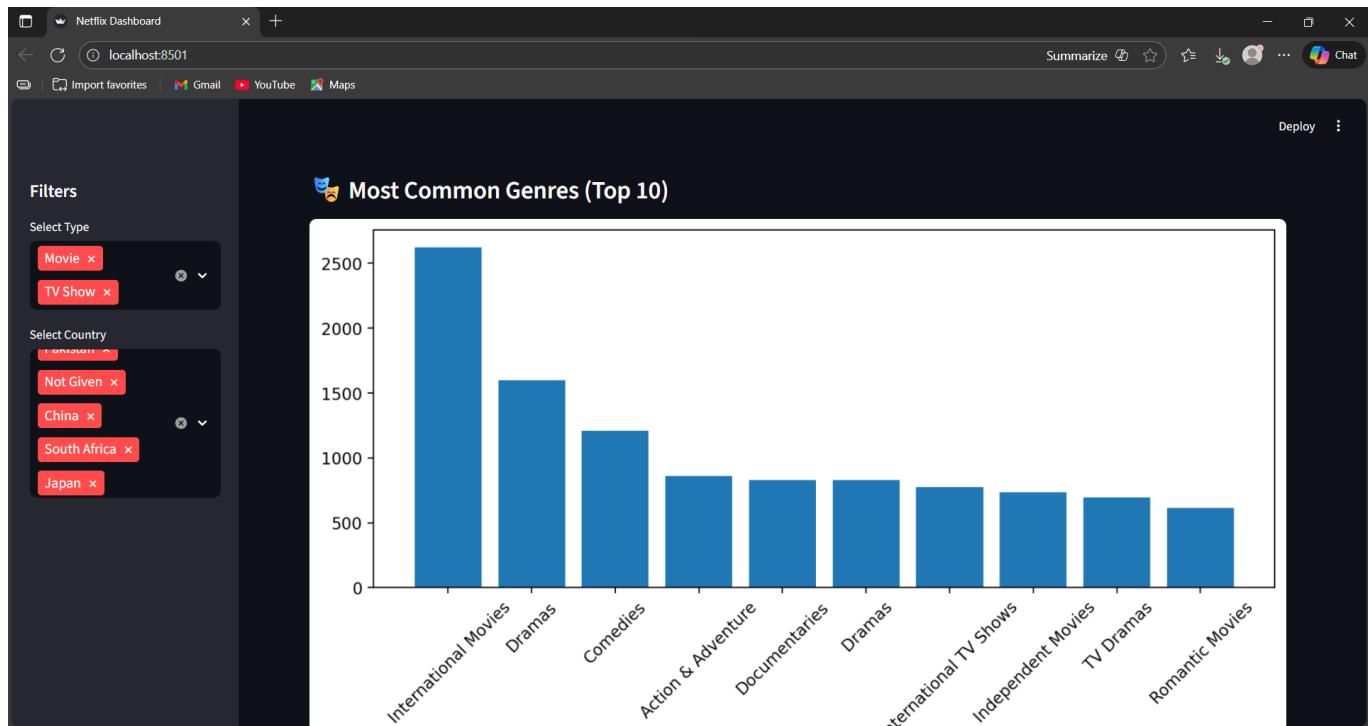
Shows counts of Movies, TV Shows & Total Titles.



### ❖ 6.3 Dashboard Visualizations

Includes line charts & bar charts for trends.





## ❖ 6.4 Dashboard Table View

Displays filtered Netflix dataset.

The figure shows a table titled "Filtered Dataset". The table has 14 rows and 11 columns. The columns are: show\_id, type, title, director, country, date\_added, release\_year, rating, duration, and listed\_in. The table displays a list of movies and TV shows with their respective details.

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	2021-09-25	2020	PG-13	90 min	Documentaries
1	s3	TV Show	Ganglands	Julien Leclercq	France	2021-09-24	2021	TV-MA	1 Season	Crime TV Shows, Internation
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	2021-09-24	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	2021-09-22	2021	TV-PG	91 min	Children & Family Movies
4	s8	Movie	Sankofa	Halle Gerima	United States	2021-09-24	1993	TV-MA	125 min	Dramas, Independent Mo
5	s9	TV Show	The Great British Baking Show	Andy Devonshire	United Kingdom	2021-09-24	2021	TV-14	9 Seasons	British TV Shows, Reality
6	s10	Movie	The Starling	Theodore Melfi	United States	2021-09-24	2021	PG-13	104 min	Comedies, Dramas
7	s939	Movie	Motu Patlu in the Game of Zones	Suhas Kadav	India	2021-05-01	2019	TV-Y7	87 min	Children & Family Movies
8	s13	Movie	Je Suis Karl	Christian Schwochow	Germany	2021-09-23	2021	TV-MA	127 min	Dramas, International Mo
9	s940	Movie	Motu Patlu in Wonderland	Suhas Kadav	India	2021-05-01	2013	TV-Y7	76 min	Children & Family Movies
10	s941	Movie	Motu Patlu: Deep Sea Adventure	Suhas Kadav	India	2021-05-01	2014	TV-Y7	76 min	Children & Family Movies
11	s942	Movie	Motu Patlu: Mission Moon	Suhas Kadav	India	2021-05-01	2013	TV-Y7	71 min	Children & Family Movies
12	s852	Movie	99 Songs (Tamil)	Not Given	Pakistan	2021-05-21	2021	TV-14	131 min	Dramas, International Mo
13	s471	Movie	Bridgerton - The Afterparty	Krysia Plonka, Kristian	United States	2021-07-13	2021	TV-14	39 min	Movies
14	s730	Movie	Bling Empire - The Afterparty	Krysia Plonka, Kristian	United States	2021-06-12	2021	TV-MA	36 min	Movies

## ❖ 7. Key Insights

- ⌚ 1. Netflix content increased massively after 2015
- ⌚ 2. Movies dominate the platform (6126 movies vs 2664 shows)
- ⌚ 3. "International Movies" is the most common genre
- ⌚ 4. USA, India, UK produce the most content
- ⌚ 5. Many entries have missing director names

## ❖ 8. Conclusion

This project successfully cleaned, analyzed, visualized, and built an interactive dashboard for the Netflix dataset. The results offer strong insights for understanding content trends and can be extended into:

Recommendation systems

Country-based insights

Rating-based studies

---

## ❖ 9. Tools Used

Python

Pandas, NumPy

Matplotlib, Seaborn

WordCloud

Streamlit

Jupyter Notebook

---

## ✿ Report Completed Successfully