

Project Title: Netflix Content Analysis Project

Executive Summary

This project analyzes a cleaned dataset of Netflix titles to uncover patterns in content release trends, genre distribution, country-level production, and platform growth. Using SQL for exploration and Power BI for visualization, I created an interactive dashboard that offers insights into how Netflix curates and expands its global content library. These insights can help content strategists identify underrepresented regions and genre saturation.

Project Objective

To analyze Netflix's global content library to understand content trends over time, dominant genres, country-level contributions, and shifts in production focus, with a goal of revealing how Netflix caters to diverse audiences.

Business Questions

- What's the growth trend of content over the years?
- Which countries produce the most content on Netflix?
- What are the most common genres across TV shows and movies?
- What type of content (Movie vs TV Show) dominates by country?
- How has content availability changed over time?

Tools & Technologies

- Excel: Initial data cleaning (dates, genre formatting)
- SQL (MySQL): Data exploration, filtering, aggregation
- Power BI: Visualizations, interactive dashboards

Dataset Information

- Source: Netflix Titles dataset from Kaggle
- Size: ~8,800 records, 12 columns
- Key columns:

title, type, director, cast, country, date_added, release_year, rating, duration, listed_in, description

Data Dictionary Snippet:

Column	Description
Type	Movie or TV Show
Country	Country of production
listed_in	Genres/categories
date_added	Date Netflix added the title

release_year

Year the title was released

Data Cleaning Process

- Removed null values in critical columns (type, date_added, country)
- Extracted month_added and year_added from date_added
- Split multi-valued listed_in genres for analysis
- Standardized country names and filtered unrealistic release years (<1940)

	A	B	C	D	E	F	G	H	I	J	K	L
1	show_id	type	Title	director	Country	month_ad	year_ad	release_y	rating	listed_in		
2		1 Movie	Dick Johns	Kirsten Jof	United Sta	Sep		2021	2020	PG 13	Documentaries	
3		2 TV Show	Blood & W	Unknown	South Afri	Sep		2021	2021	TV MA	International TV Shows	
4		3 TV Show	Ganglands	Julien Lecl	Unknown	Sep		2021	2021	TV MA	Crime TV Shows	
5		4 TV Show	Jailbirds N	Unknown	Unknown	Sep		2021	2021	TV MA	Docuseries	
6		5 TV Show	Kota Facts	Unknown	India	Sep		2021	2021	TV MA	International TV Shows	
7		6 TV Show	Midnight I	Mike Flan	Unknown	Sep		2021	2021	TV MA	TV Dramas	
8		7 Movie	My Little P	Robert Cu	Unknown	Sep		2021	2021	PG	Children and Family Movies	
9		8 Movie	Sankofa	Haile Geri	United Sta	Sep		2021	1993	TV MA	Dramas	
10		9 TV Show	The Great Andy Dev	United Kin	Sep			2021	2021	TV 14	British TV Shows	
11		10 Movie	The Starlin	Theodore	United Sta	Sep		2021	2021	PG 13	Comedies	
12		11 TV Show	Vendetta: Unknown	Unknown	Unknown	Sep		2021	2021	TV MA	Crime TV Shows	
13		12 TV Show	Bangkok B	Kongkiat K	Unknown	Sep		2021	2021	TV MA	Crime TV Shows	
14		13 Movie	Je Suis Kar	Christian S	Germany	Sep		2021	2021	TV MA	Dramas	
15		14 Movie	Confessio	Bruno Gar	Unknown	Sep		2021	2021	TV PG	Children and Family Movies	
16		15 TV Show	Crime Sto	Unknown	Unknown	Sep		2021	2021	TV MA	British TV Shows	
17		16 TV Show	Dear Whit	Unknown	United Sta	Sep		2021	2021	TV MA	TV Comedies	
18		17 Movie	Europe s	Pedro de l	Unknown	Sep		2021	2020	TV MA	Documentaries	
19		18 TV Show	Falsa iden	Unknown	Mexico	Sep		2021	2020	TV MA	Crime TV Shows	
20		19 Movie	Intrusion	Adam Salk	Unknown	Sep		2021	2021	TV 14	Thrillers	

Exploratory Data Analysis (EDA)

SQL revealed several content trends:

- Movies make up 70% of total content, while TV Shows account for the rest
- United States, India, UK are the top 3 countries by content volume
- The most common genres include Dramas, International Movies, and Comedies
- Netflix experienced a content surge in 2018 and 2019
- A noticeable drop in new titles after 2020, likely due to COVID-19

■ Top 10 countries producing the most content

```
SELECT country, COUNT(*) AS total_titles
FROM netflix_titles_cleaned
WHERE country != 'Unknown'
GROUP BY country
ORDER BY total_titles DESC
LIMIT 10;
```

■ Monthly distribution of content added

```
SELECT month_added, COUNT(*) AS total
FROM netflix_titles_cleaned
WHERE month_added != 'Unknown'
GROUP BY month_added
ORDER BY FIELD(month_added,
'Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec');
```

■ Most frequent content ratings (e.g., TV-MA, PG, etc.)

```
SELECT rating, COUNT(*) AS total
FROM netflix_titles_cleaned
WHERE rating != 'Unknown'
GROUP BY rating
ORDER BY total DESC;
```

■ Top 10 most common genres or categories

```
SELECT listed_in, COUNT(*) AS total
FROM netflix_titles_cleaned
GROUP BY listed_in
ORDER BY total DESC
LIMIT 10;
```

Key Insights

- 2018 and 2019 were peak years for new content additions
- Movies dominate Netflix's content library (~70%)
- USA, India, and UK are the top producers, while countries like Nigeria and South Korea are rising
- Drama is the most consistent genre across all years and regions
- A drop in content growth post-2020 may reflect global production slowdowns

Challenges & How I Solved Them

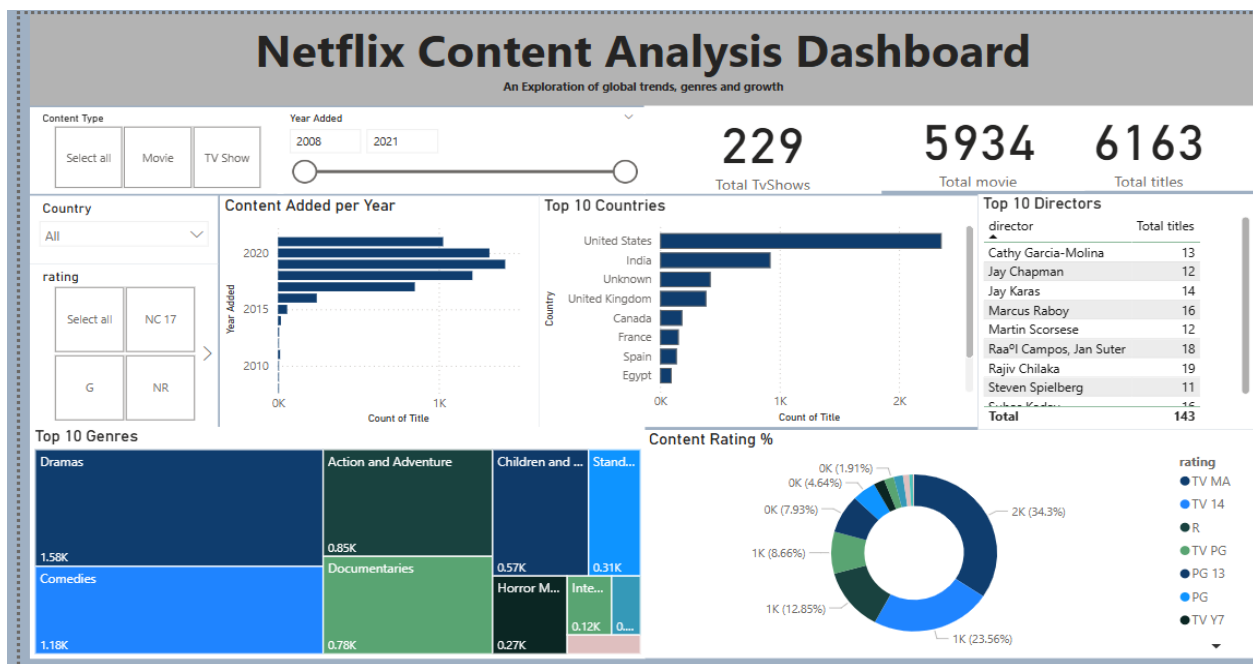
- Issue: Multi-valued listed_in genre field
- Solution: Used SQL string functions and Power BI slicers to visualize genre overlaps
- Issue: Missing country or date_added in some rows
- Solution: Filtered out those rows for cleaner trend analysis

- Issue: Time series inconsistency (release vs added year)
- Solution: Created separate views for release_year and year_added trends

Dashboard or Visual Output

Created an interactive Power BI dashboard with:

- Line chart showing content additions by year
- Country map showing content volume distribution
- Genre breakdown by type and year
- Movie vs TV show ratio over time
- Filters: Year, Type, Country, Genre



Business Recommendations

- Target emerging markets like South Korea, Nigeria, and Turkey with more content
- Diversify genres in overrepresented regions (e.g., more Sci-Fi or Documentaries in the USA)
- Reinvest in original content development post-2020 dip
- Leverage high-performing non-English content for international audiences

Project Links

- GitHub Repository:

https://github.com/Deejarh-ops/Deejerh.github.io/blob/main/Netflix_%20Exploratory_analysis.sql