Streaming Smarter: A Netflix SQL Case Study

TL;DR

Analyzed over 8,800 Netflix titles using pure SQL to uncover global content trends, dominant genres, top-performing countries, and dirty data risks, no dashboards, just sharp queries and sharper insights. The findings support Netflix's global-first, short-format strategy while highlighting key data health gaps.

Objective - To analyze Netflix's global content strategy using raw SQL, uncovering which genres perform best, which countries dominate production, and what viewing patterns define the platform.

Tools Used

- MySQL: Data wrangling, querying, profiling, and pattern extraction
- Microsoft Excel: Light preprocessing (UTF-8 encoding fixes)
- Dataset: Netflix Titles Dataset from Kaggle

Business Questions Answered

- What are Netflix's peak years for content releases?
- Which countries contribute the most to Netflix's catalog?
- Which genres dominate, and what does that say about their audience strategy?
- How does movie length align with modern viewer preferences?
- Are there red flags in Netflix's metadata that could hinder decision-making?

Key Queries & Insights

1. **2018 was Netflix's Peak Year**
    - 1,144 new titles were added, reflecting Netflix's international push and pivot to original content between 2016-2020.
    - ***Recommendation:*** Consider why 2018 peaked, align future analysis with marketing and global rollout milestones.

2. **Top Content Producing Countries**-

Rank	Country	Titles
1	USA	2806
2	India	972
3	UK	419

4	Japan	245
5	South Korea	199
6	Canada	181
7	Spain	145
8	France	123
9	Mexico	110
10	Egypt	106

- The USA and India lead content production
- ***Recommendation:*** Double down on localization efforts in regions with growing output but underrepresented viewership.

3. **Content Type Distribution**-

Type	**Total Titles**
Movies	6119
TV Shows	2674

The majority of content on Netflix consists of movies, which make up nearly 70% of the catalog.
While binge-worthy TV shows are a major part of the platform's strategy, this result shows that Netflix continues to invest heavily in single-session, high-turnover content like films likely to cater to global viewers seeking shorter watch experiences.
However, the presence of over 2,600 TV shows still reflects a strong commitment to serialized content.

4. **Most Popular Genres**

Genre Combination	**Count**
Dramas, International Movies	361
Documentaries	358
Stand-Up Comedy	334
Comedies, Dramas, International Movies	273
Dramas, Independent Movies,	251

International Movies	
Kids' TV	220
Children & Family Movies	215
Children & Family Movies, Comedies	201
Documentaries, International Movies	186
Dramas, International Movies, Romantic Movies	180

- **Top genre:** Dramas, International Movies (361 titles) dominate, revealing Netflix's 'global-first' curation strategy. Stand-up comedy, while lower budget, ranks 3rd, showing that relatability often beats production value.
- ***Recommendation:*** Continue leveraging global storytelling and low-budget formats like stand-up for high return.

5. **Most Featured Directors on Netflix**
    - Most frequent names include Rajiv Chilaka and Latin American creators. Spielberg & Scorsese appear, but less often.
    - ***Recommendation:*** Strategic partnerships should include niche regional talent, not just mainstream names.

6. **Average Movie Duration**

Metric	**Value**
Average Duration	99.57 minutes

- **Average movie duration** = 99.57 minutes, suggesting a single-sitting preference.
- ***Recommendation:*** Use this insight to guide acquisition and recommendation engines, favoring bite-sized formats.

7. **Longest Running Shows by Number of Seasons**

Show	**Seasons**
Grey's Anatomy	17
Supernatural	15

NCIS	15
Heartland	13
Criminal Minds	12

Netflix features several long-running series, with Grey's Anatomy leading at 17 seasons, followed by Supernatural and NCIS at 15 seasons each.
These legacy titles reflect Netflix's strategy of licensing high-engagement, evergreen shows that offer hundreds of hours of watch time.
Titles like Criminal Minds, Heartland, and Cheers highlight the platform's mix of drama, crime, and classic sitcom content designed to retain subscribers through long-form storytelling.

8. **Dirty Data Report**

Column	**Missing Entries**
Title	29
Director	2633
Cast	825
Country	830
Listed In	0
Description	0

- **Missing:** 2633 director fields, 825 cast entries, 830 country rows.
- ***Recommendation:*** Implement stronger metadata QA and validation pipelines before catalog expansion.

Methodology

The dataset was cleaned and imported using Excel for encoding fixes, then loaded into MySQL. I wrote 8 key queries covering content timeline trends, genre clustering, country contributions, and runtime averages. A ninth query flagged nulls to stimulate a real-life dirty data audit.

SQL Query Preview

Before running any visuals or exports. I wrote targeted queries to pull insights directly from raw data.

```sql
SELECT Listed_In, COUNT(*) AS Total_Titles
FROM netflix_clean
WHERE Listed_IN IS NOT NULL AND TRIM(Listed_In) != ''
GROUP BY Listed_In
ORDER BY Total_Titles DESC
LIMIT 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Listed_In	Total_Titles
Dramas, International Movies	361
Documentaries	358
Stand-Up Comedy	334
Comedies, Dramas, International Movies	273
Dramas, Independent Movies, International Mo...	251
Kids' TV	220
Children & Family Movies	215
Children & Family Movies, Comedies	201
Documentaries, International Movies	186
Dramas, International Movies, Romantic Movies	180

```sql
SELECT
SUM(CASE WHEN TRIM(Title) = '' OR Title IS NULL THEN 1 ELSE 0 END) AS Missing_Title,
SUM(CASE WHEN TRIM(Director) = '' OR Title IS NULL THEN 1 ELSE 0 END) AS Missing_Director,
SUM(CASE WHEN TRIM(`Cast`) = '' OR Title IS NULL THEN 1 ELSE 0 END) AS Missing_Cast,
SUM(CASE WHEN TRIM(Country) = '' OR Country IS NULL THEN 1 ELSE 0 END) AS Missing_Country,
SUM(CASE WHEN TRIM(Listed_In) = '' OR Listed_In IS NULL THEN 1 ELSE 0 END) AS Missing_Listed_In,
SUM(CASE WHEN TRIM(`Description`) = '' OR `Description` IS NULL THEN 1 ELSE 0 END) AS Missing_Description
FROM netflix_clean;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Missing_Title	Missing_Director	Missing_Cast	Missing_Country	Missing_Listed_In	Missing_Description
29	2633	825	830	0	0

Why This Project Matters

This project isn't just about Netflix; it shows what you can do with SQL alone. No dashboards. No front-end visuals. Just structured thinking, clean logic, and narrative-rich insights pulled

straight from raw data. In a real-world setting, this is the kind of backend thinking that powers the best dashboards before they're even built.

What I'd Explore Next

- Run genre trends over time to spot evolving viewer preferences
- Build a Python script to flag dirty rows automatically
- Segment by target audience (kids, teens, adults) using genre tags

Explore the Project

- GitHub
- Portfolio