

Streaming Smarter: A Netflix SQL Case Study

Objective - The goals of this project were to utilize pure SQL to explore, clean, and analyze Netflix's content catalog. With over 8,800 titles in the dataset, this case study focused on uncovering patterns in global content production, identifying the most common genres and directors, and performing a dirty data audit, all without any BI tool or dashboarding software. The intent was to showcase strong backend analysis skills and the ability to derive actionable insights using only SQL.

Tools Used

- MySQL: Data cleaning, transformation & querying
- Kaggle Dataset: Raw Netflix metadata
- Excel: Light pre-cleaning (encoding + UTF-8 fix)

Methodology

The dataset was imported into MySQL after light UTF-8 cleaning in Excel. Each column was inspected for NULLs, blanks, and inconsistencies. From there, a series of 8 queries were written to explore trends like production year spikes, regional dominance, popular genres, and movie durations. One final query focused on data quality to stimulate a real-world data validation scenario.

Key Queries & Insights

1. **Most Content-Heavy Years-** Netflix's content production peaked in 2018, with 1144 new titles added that year, the highest across the dataset. This surge aligns with Netflix's aggressive shift toward original programming and international expansion shift toward original programming and international expansion between 2016 and 2020. After 2020, content additions slightly declined, potentially reflecting pandemic-related slowdowns and increased competition in the streaming industry.
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 United States leads Netflix content production by a wide margin, with 2800+ titles, nearly triple that of the next highest, India, with 972 titles.

This reflects Netflix's roots in the US and its dominance in Western media markets.

However, India's strong second place highlights Netflix's growing investment in regional and non-English content, especially in Asia.

The top 10 includes countries across North America, Europe, Asia, and even Africa, showing Netflix's strategy to diversify and localize its global catalog.

3. Content Type Distribution-

Type	Total Titles
Movies	6119
TV Show	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, International Movies	251

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

Netflix's top genres lean heavily toward Dramas, International content, and Documentaries, all of which reflect its strategy to cater to diverse, globally distributed audiences.

- Drama appears in over half of the top genre combos
- International Movies are frequently paired with other themes, showcasing Netflix's global-first programming model
- Stand Up Comedy ranks high, indicating strong performance for solo, low-budget formats.

This genre breakdown proves that while Netflix is expanding globally, it also relies on emotionally rich and culturally specific content to drive viewer retention.

5. **Most Featured Directors on Netflix-** Netflix's most frequently featured directors include a mix of international and genre-specialized creators.

Rajiv Chilaka leads the chart with 19 titles, followed closely by Raul Campos & Jan Suter, creators of Latin American documentaries and stand-ups. Iconic filmmakers like Martin Scorsese and Steven Spielberg also appear, but Netflix's focus on regional diversity is clear with creators from India, Egypt, and the Philippines among the top 10.

6. Average Movie Duration

Metric	Value
Average Duration	99.57 minutes

The average Netflix movie has a runtime of 99.57 minutes, just under 1 hour and 40 minutes. This suggests that the majority of films on the platform are designed for quick, single-sitting experiences, aligning with the viewing habits of modern audiences. It also reinforces Netflix's investment in shorter, binge-ready content that's easily consumable and ideal for casual viewers and global audiences.

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

A review of the Netflix dataset revealed that several fields contain missing or blank values, which could impact analysis reliability if not handled carefully.

The 'Director' field is missing in over 2600 records, and both 'Cast' and 'Country' are missing in 800+ entries, likely due to incomplete metadata or international imports.

While critical categorical fields like 'Listed In' (genre) and 'Description' are complete, this highlights the importance of performing data cleaning and validation before drawing conclusions or building dashboards.

What This Project Showcases

- SQL-only analysis with real-world business questions
- Data profiling + cleaning directly in SQL

- Clear storytelling from query to insight
- Understanding of global content trends and platform strategy

Final Reflection

This project was a reminder that SQL alone can be incredibly powerful when used with the right mindset. From string manipulation and date parsing to trend analysis and data validation, it taught me how to balance business logic with data structure and how to tell stories using just raw data and queries. Cause sometimes, the best dashboard is a clean query and a sharp brain.