Project Summary: Netflix Content Analysis

1. Introduction: Understanding the "Why"

The main goal of this project is to structure and analyze Netflix content data to extract key trends and insights regarding the evolution and diversity of its catalog. To achieve this, the analysis had to overcome two major challenges related to raw data preparation:

- **Data Normalization:** This step was crucial to avoid redundancy and efficiently organize the database by linking multiple tables (movies, actors, directors).

- **Data Cleaning:** This step was essential to correct inconsistencies, missing values, and formatting errors, ensuring the reliability of analytical results.

A rigorous, multi-step methodology was implemented to address these challenges.

2. Methodology: How the Data Was Prepared

Data preparation was carried out in three main phases, transforming a simple Excel file into a clean, relational database ready for analysis.

2.1 Import and Inspection

The initial Excel file (netflix_titles.xlsx) was imported into SAS. A preliminary inspection verified the overall structure and quality of the raw data, identifying potential issues for correction.

2.2 Normalization and Structuring

To organize the data logically and efficiently, a relational data warehouse was created. Information initially contained in a single table was distributed across multiple specialized, interconnected tables:

Table Type	Role and Example
Main Table	Contains core information about each title (e.g., titles table with title_id, type, release_year).
Dimension Tables	Store unique information to avoid redundancy (e.g., directors, actors, genres, countries).
Link Tables	Manage many-to-many relationships (e.g., title_directors linking a movie to its directors, title_genres linking a title to its genres).

2.3 Cleaning and Insertion

Key cleaning steps included trimming unnecessary spaces in text fields and converting dates to a standard numeric format. The cleaned data was then inserted into the normalized database structure, making it ready for analysis.

This clean, structured database enabled revealing analyses of the platform's catalog.

3. Key Analyses and Findings

3.1 Netflix Catalog: Volume, Duration, and New Releases

The analysis of catalog growth shows a steady increase in movies and series added over the years, with a marked acceleration recently, reflecting a massive investment strategy to continuously expand the offering and retain subscribers.

Regarding movie durations:

- **The Norm:** The average movie duration on the platform is approximately 99.32 minutes, aligning with industry standards.

- **The Exception:** The longest title is *Black Mirror: Bandersnatch* (312 minutes). Its exceptional length is explained by its interactive nature, representing a key differentiator in a saturated market and designed to enhance user engagement through unique viewing experiences.

Analysis of recent additions, such as the eight new titles scheduled for May 1, 2025, confirms a continuous diversification strategy, including films, documentaries, and stand-up shows to appeal to a wide audience.

3.2 Key Contributors: Directors, Actors, and Countries

The analysis highlights the creators and countries shaping Netflix's catalog.

Director(s)	Number of Productions
Raúl Campos, Jan Suter	324
Marcus Raboy	256
Martin Scorsese	144
Steven Spielberg	100

The origin of productions shows U.S. dominance, followed by India and the U.K., highlighting the prevalence of English-language content while recognizing the strategic importance of the Indian market.

Actor analysis emphasizes the prominence of Bollywood stars like Shah Rukh Khan (729 titles) alongside popular Hollywood figures such as Adam Sandler (361 titles), reflecting a dual-casting strategy targeting both the large Indian market and global audiences familiar with American stars.

This suggests Netflix's content acquisition strategy strongly favors the U.S., leveraging popular genres such as stand-up comedy to generate high volumes of content efficiently.

3.3 Diversity at the Core of the Offering

Genre analysis confirms Netflix offers a highly varied catalog to appeal to multiple audiences. The three most popular genres are Dramas, Comedies, and Documentaries. However, the catalog is enriched with specialized categories like Anime Series, Stand-Up Comedy, and LGBTQ Movies, capturing both broad audiences and niche communities.

Some creators demonstrate notable versatility. Director Justin G. Dyck has worked across five genres, while actors like Nicolas Cage have appeared in six distinct genres. Collaborating with such flexible talent allows Netflix to continuously produce a wide variety of content.

Overall, Netflix combines standardized formats and prolific creators to scale volume while strategically diversifying genres and talent to serve both mass markets and niche audiences.

4. Conclusion: Summary and Next Steps

4.1 Key Takeaways

1. The analysis confirms continuous expansion and enrichment of Netflix's catalog over the years.

2. While standard formats exist (e.g., average movie duration), Netflix stands out through innovative formats and diverse genres to satisfy all audiences.

3. Content is driven by highly prolific creators and actors, with strong contributions from the U.S. and India, and specialized directors in popular genres such as stand-up comedy.

4.2 Limitations and Future Directions

Study Limitations	Future Improvements
Missing data (e.g., episode duration)	Expand the database with new dimensions (language, number of episodes).
Purely quantitative analysis	Integrate qualitative data such as user ratings and reviews.
Results presented in raw tables	Create visualizations (charts, maps) for more accessible, interactive insights.

This analysis provides a foundation for deeper studies, combining quantitative rigor with insights for strategic content planning and audience engagement.