

Amazon Prime Video: Global Catalog Strategic Analytics



**End-to-End Data Lakehouse &
Business Intelligence Solution**

Name: Beatriz Gomes

Date: January 2026

Conteúdo

Sumário	3
Business Context and Scope	4
Analytical questions	4
Key Performance Indicators (KPI).....	5
Análise da Qualidade dos Dados.....	Erro! Marcador não definido.
Arquitetura Data Lakehouse	Erro! Marcador não definido.
Desenvolvimento Data Lakehouse	Erro! Marcador não definido.
Fase Bronze	Erro! Marcador não definido.
Fase Silver.....	Erro! Marcador não definido.
Fase Gold.....	Erro! Marcador não definido.
Dashboards	Erro! Marcador não definido.
Conclusão	Erro! Marcador não definido.
Distribuição de Notas.....	Erro! Marcador não definido.
Link vídeo	Erro! Marcador não definido.
Referências	Erro! Marcador não definido.
Anexos.....	Erro! Marcador não definido.
Anexo 1 – Relatório de <i>Data Profiling</i>	Erro! Marcador não definido.

Sumário

Introdução do relatório.

Business Context and Scope

The project, titled "Amazon Prime Video: Global Catalog Strategic Analytics," is driven by the need to transform vast amounts of streaming metadata into competitive advantages. In an era where content acquisition and production costs are soaring, the motivation lies in moving beyond simple volume-based metrics. The goal is to identify the underlying patterns of "content health"—balancing critical quality with commercial popularity—to ensure that every title in the Amazon Prime Video catalog serves a specific strategic purpose, whether it be audience retention, global expansion, or brand prestige.

This project consists of an end-to-end analytical solution that integrates title metadata and talent credits into a unified business intelligence framework. The mission is to provide a "North Star" for content decision-makers, moving from intuitive-based selections to evidence-based investments. The project operates within the complex ecosystem of global digital entertainment, where production hubs are shifting and audience fatigue is a constant risk. Key business processes addressed include Strategic Portfolio Balancing, Talent ROI Assessment, and Global Market Penetration, all managed through a centralized Power BI interface that translates complex correlations into visual insights.

The overarching objective of this project is to develop a Big Data Intelligence System focused on the multi-dimensional optimization of the Amazon Prime Video catalog. To support high-level decision-making, the project aims to:

- Establish a "Content Health" Monitoring System: Develop mechanisms to analyze the life-cycle of genres and production formats, identifying areas of diminishing returns and high-growth potential to optimize budget reallocation and prevent catalog stagnation.
- Create a Talent & Production Performance Framework: Develop analytical models to measure the impact of specific talent (cast/crew) and technical attributes—such as runtime and format—on audience satisfaction and long-term engagement.
- Optimize Global Content Acquisition & Marketing Synergy: Support international expansion by identifying high-quality production hubs and analyzing the correlation between marketing popularity and actual user satisfaction to ensure efficient promotional spending.
- Provide a Scalable Interactive Visualization Component: Deliver a Power BI dashboard that enables stakeholders to explore the catalog's performance through dynamic geographical and categorical lenses, facilitating rapid response to global market shifts.

Analytical questions

The analytical framework of this project is designed to address the core challenges of global content management. Each question targets a specific business pain point, ranging from quality control to geographic expansion. By integrating these questions into a centralized Data Lakehouse, we enable a holistic view of the catalog where financial performance, audience trends, and talent influence are interconnected.

Question 1:

- Question: How has the distribution of genres evolved over the last decade in terms of production volume, and which genres are showing signs of "Content Rot" (High volume but declining IMDb scores)?
- Target Stakeholder: Chief Content Officer and Acquisition Strategy.
- Objective: To identify the life-cycle of each genre, allowing the platform to phase out underperforming categories and reallocate budget to high-growth areas.

Question 2:

- Question: Who are the top 10 most frequent actors in the catalog, and does their presence guarantee higher audience ratings and engagement?
- Target Stakeholder: Talent Relations and Casting Directors.
- Objective: To rank talent based on frequency and performance, providing data-driven evidence for future casting decisions and high-value talent contracts.

Question 3 :

- Question: What is the relationship between content length (Runtime) and user satisfaction? Is there an "ideal duration" that maximizes scores for Movies vs. Shows?
- Target Stakeholder: Creative Production and Editing.
- Objective: To guide the technical format of new "Originals" by identifying the runtimes that historically prevent audience fatigue and maximize ratings.

Question 4:

- Question: Is there a strong correlation between marketing success (Popularity) and user satisfaction (Score), or is the platform relying on "overhyped" low-quality content?
- Stakeholder: Marketing and Acquisition Strategy.
- Goal: Distinguish between "Blockbusters" and "Niche Gems" to optimize promotional spend.

Question 5:

- Question: Which international production hubs are delivering the highest "Quality-per-Title" ratio, and where are the geographical gaps for content expansion?
- Target Stakeholder: International Expansion & Localization Team.
- Objective: To identify untapped geographical markets where high-quality content is being produced but is underrepresented in the catalog, enabling a data-driven global acquisition strategy.

Key Performance Indicators (KPI)

To ensure the strategic alignment of the Amazon Prime Video catalog, five key metrics were developed. These KPIs were defined following an Exploratory Data Analysis, which established the current performance baselines of the platform.

KPI	Business Value	Calculation Logic	Target	Business Users/Stakeholders
Genre Decay Index	Detects "Content Rot" by identifying genres where production is growing but quality is falling.	Difference between the annual growth rate of title count and the annual change in average IMDb scores per genre.	Green: < 0 Amber: 0 Red: > 10	Department of Content Strategy
Commercial Star Power (Reach)	Measures the actor's ability to attract a large audience. High average votes per title indicate "Bankable Stars" who guarantee visibility.	The total sum of IMDb votes divided by the total count of titles associated with a specific actor.	Green: > 26,000 Amber: 10,000 to 26,000 Red: < 10,000	Department of Talent Relations
Format Satisfaction	Compares the average rating and engagement	Average IMDb score segmented by content type	Green: > 7.1 for Shows	Department of Creative Production

Index	across content types (Movies vs. Shows) to guide production investment.	(Movie>Show) and duration tiers: Short-form (<60 min), Feature/Standard (60-120 min), and Epic/Long-form (>120 min)	and > 6.0 for Movies Amber: 6.0 - 7.0 for Shows and 5.0 - 6.0 for Movies Red: < 6.0 for Shows and < 5.0 Movies	
Hype Efficiency Ratio	Checks if marketing popularity translates into real user satisfaction. Identifies "overhyped" content.	Total TMDB popularity points divided by the total sum of IMDb scores.	Green: < 1.20 (Efficient) Amber: 1.20 to 2.00 Red: > 2.00 (Overhyped)	Department of Marketing and Acquisition
Regional Quality Hubs	Ranks production countries to identify premium content markets for future localized investment.	Intersection of Title Count (Saturation) and Mean IMDb Score (Intensity).	Priority 1 (Green): Score > 6.5 & Vol > 10 Priority 2 (Blue): Score > 7.0 & Vol < 10 Low Priority (Red): Score < 6.0	Department of International Expansion