

Del caos de la oferta al valor para el espectador

Estrategias de contenido y calidad percibida en el streaming

Laura Soto Alsina

Resumen:

Este proyecto disecciona el caos de los catálogos *streaming* para descubrir las estrategias reales que hay detrás de cada gigante. Mediante un proceso integral de Extracción, Transformación y Carga (ETL) en Python y MySQL, se unifica los catálogos de las plataformas con datos del 2021 (Bansal, S., 2021) y se vinculan con *ratings* de IMDb (IMDb.com, 2024) mediante un *Matching* Híbrido, con la finalidad de medir la calidad real a ojos de los espectadores. El análisis revela si existe una coherencia estratégica hacia públicos concretos o si se apuesta por la masividad. Los resultados demuestran que, más allá del volumen, el valor real lo genera la especialización y la calidad percibida por el usuario, confirmando que un catálogo curado conecta mejor con los espectadores que uno indiscriminado.

Introducción:

Hoy en día, el espectador no sufre por falta de opciones streaming, sino por una saturación de títulos en cada catálogo, que hace casi imposible elegir qué ver (Simon-Kucher & Partners, 2023). El objetivo de este proyecto ha sido poner orden a ese caos y descubrir la estrategia real tras cada oferta: ¿las plataformas buscan ser un almacén gigante o un espacio curado para nichos específicos? Al integrar valoraciones de los espectadores de la base de datos IMDb, se ha podido medir la calidad real más allá del marketing, validando si el volumen masivo de títulos se traduce en valor para el usuario o si la verdadera riqueza reside en la especialización.

Metodología:

El trabajo se ha estructurado como un proceso integral de ETL, orientado a transformar datos en bruto en información analítica sobre la estrategia de contenido de las plataformas de streaming y la calidad percibida por los espectadores.

Para la adquisición y arquitectura de datos, el punto de partida ha sido la selección de fuentes sólidas y contrastadas. Se ha trabajado con los catálogos de Netflix, Amazon Prime Video, Disney+ y Hulu, obtenidos de Kaggle. Debido a la opacidad de las plataformas, los datos más recientes disponibles corresponden al año 2021. De forma complementaria, se ha incorporado la base de datos oficial de IMDb, referencia internacional en contenidos audiovisuales, con información actualizada hasta 2024, con el objetivo de aportar una validación externa de calidad.

Para centralizar el análisis y garantizar su trazabilidad, se ha construido una base de datos relacional en MySQL denominada *streaming_analysis*. La carga no ha sido directa, ya que los asistentes de importación fallaban por el volumen de los archivos, por lo que se ha utilizado Python para asegurar una carga completa, limpia y sin pérdidas de información.

Una vez integrados los datos de catálogos en bruto, para la transformación, modelado y normalización de estos, se ha realizado una limpieza técnica general, ajustando la tipología de las variables y eliminando columnas irrelevantes. Posteriormente, se han unificado los registros de las cuatro plataformas en una única tabla maestra.

Una de las primeras decisiones de modelado ha sido la normalización de la duración, separando los minutos de las películas y el número de temporadas de las series en columnas independientes, con el fin de permitir comparaciones coherentes entre formatos.

El tratamiento de los géneros ha supuesto uno de los principales retos del proyecto. Dado que los títulos incluían múltiples géneros en una misma celda, se ha utilizado Python para crear una relación título-género mediante una tabla puente. Posteriormente, se ha abordado la normalización semántica, reduciendo un total de 121 géneros originales a 18 categorías coherentes, tras una revisión manual exhaustiva.

De forma análoga, se ha aplicado una lógica de segmentación para unificar los distintos sistemas de clasificación por edades, creando grupos de audiencia homogéneos (*Universal, Teens, Adults*, etc.) mediante reglas definidas en MySQL. Ante la detección de una brecha de información en títulos sin catalogar, se ha procedido a un enriquecimiento de datos mediante consultas de cruce. Para ello, se ha empleado la inteligencia colectiva del catálogo global, asignando clasificaciones basadas en la consistencia de títulos de otras plataformas, logrando recuperar un 3,18% de información adicional.

La fase final ha consistido en la integración de calidad en la base de datos mediante *matching* híbrido de los ratings de IMDb. Tras una limpieza y normalización previa de los archivos mediante Python, se ha diseñado una estrategia de *matching* híbrido: una primera unión exacta por título normalizado, año y tipo de contenido, y un segundo proceso de *fuzzy matching* mediante la librería RapidFuzz para resolver discrepancias de escritura en el título. Este enfoque ha permitido alcanzar una cobertura aproximada del 70–80% del catálogo.

De esta forma se ha consolidado una tabla de hechos centralizada bajo un modelo relacional, que integra los 23.0000 registros procesados. Esta arquitectura permite cruzar la estrategia de catálogo de cada plataforma con indicadores de audiencia y calidad.

Para las visualizaciones y presentación de resultados se ha utilizado Python (Pandas, Seaborn y Matplotlib), Power BI, Flourish y Canva.

Resultados:

Al mirar los datos de los catálogos, se ve una diferencia muy clara en cómo trabaja cada plataforma. Si nos fijamos en la cantidad (Figura 1), Netflix y Amazon Prime Video son los reyes del volumen; tienen muchísimos más títulos que Hulu o Disney+. Hay un detalle: Netflix equilibra algo sus películas y sus series pero Amazon apuesta muchísimo por las películas. Además, si miramos qué tan nuevo es lo que ofrecen (Figura 2), casi todo se ha lanzado en los últimos años, aunque hay diferencias interesantes. Disney+ destaca por tener una mediana de año más antigua, lo que refleja un catálogo más clásico apoyado en su archivo histórico. En cambio, las otras plataformas se perciben más modernas, con una concentración de títulos mucho más reciente.

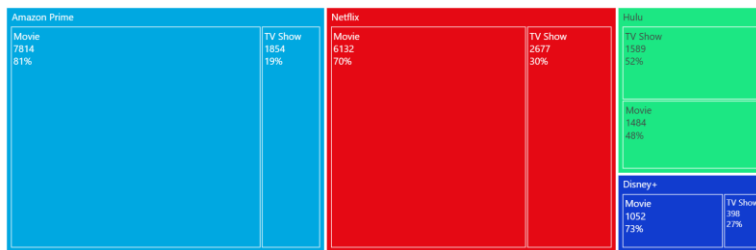


Figura 1.

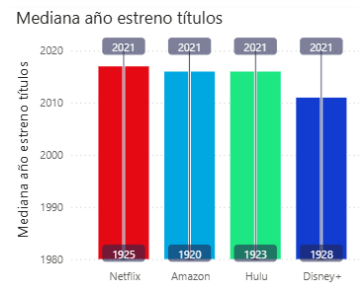


Figura 2.

Sobre a quién intentan atraer, los datos nos dicen que cada una tiene una identidad muy marcada (Figura 3.). Disney+ es la más enfocada de todas, con un catálogo muy pensado para niños y jóvenes; en cambio, Amazon no se focaliza en un público concreto. Esta identidad también se nota en sus géneros. Como vemos en el índice de concentración (Figura 4.), Netflix y Disney+ concentran más del 70% de sus títulos en solo tres géneros lo que demuestra una especialización muy fuerte. Aunque Amazon y Hulu se dispersan un poco más, siguen concentrando más de la mitad de su contenido en su “top 3” de géneros.

Porcentaje de audiencia por plataforma

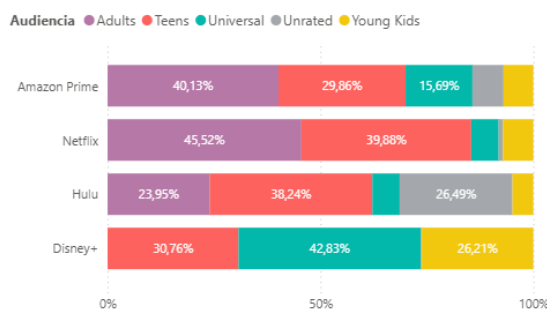


Figura 3.

¿Cuánto abarca en los catálogos el Top3 géneros?

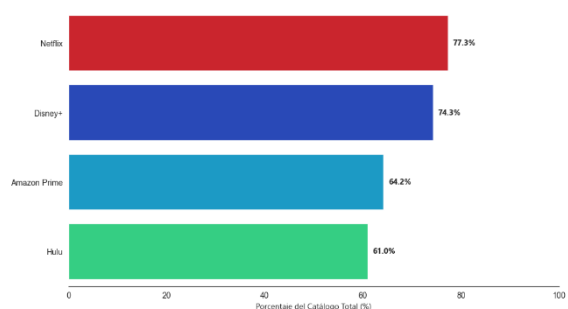


Figura 4.

Porcentaje de solapamiento respecto al total de cada plataforma

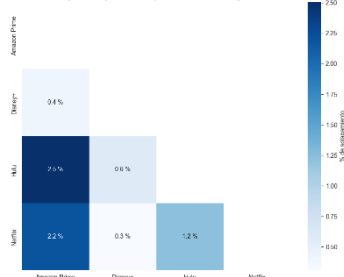


Figura 5.

Si nos centramos en el solapamiento de títulos entre plataformas, se observa como los porcentajes son bajos y que cada plataforma tiene bastante contenido exclusivo (Figura 5.).

Cuando metemos en la ecuación la opinión de la audiencia (Figura 6. Ver Anexo A.) observamos como Hulu es la que tiene las notas más estables y altas, con muchos títulos en el “top” de calidad.

Por el contrario, Netflix y Amazon son más irregulares: tienen grandes éxitos, pero también los fracasos son más sonados en el análisis. Algo muy curioso que aparece en todas las plataformas es que las series siempre están mejor valoradas que las películas (Figura 7.), y es Hulu quien mejor lo aprovecha.

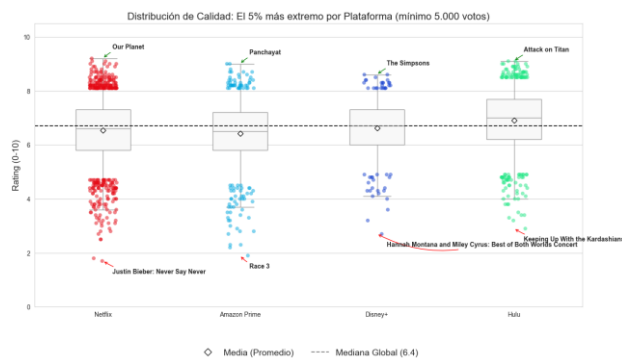


Figura 6.

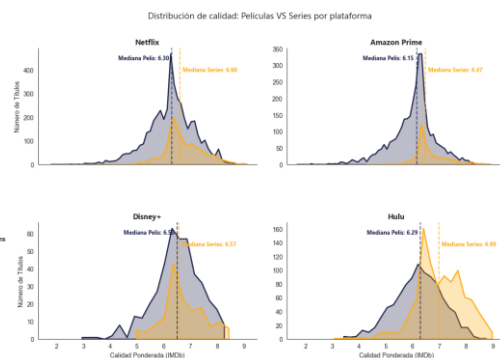


Figura 7.

Por último, si comparamos la nota media con el tamaño del catálogo (Figura 8. Ver Anexo B), vemos algo muy interesante: las plataformas con menos títulos (Hulu y Disney+) son las que consiguen mejores puntuaciones. Esto contrasta con uno de los gigantes del volumen (Amazon) que, aunque ofrece de todo, tiene la nota media más baja.

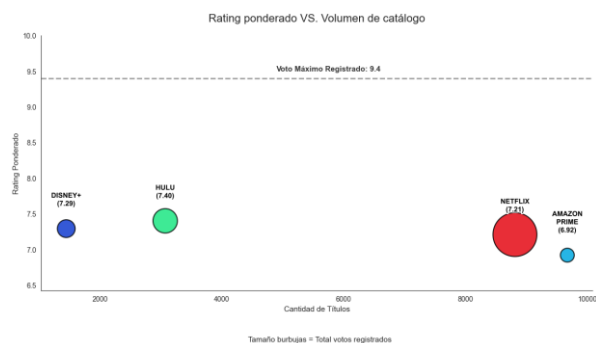


Figura 8.

Discusión:

Después de analizar los datos, queda claro que las plataformas no intentan parecerse entre ellas, sino que cada una ha buscado su propio rincón. Por ejemplo, siempre se dice que el mercado está saturado de contenido repetido, pero el solapamiento de títulos es tan bajo que confirma que cada plataforma funciona como una "isla" de contenido exclusivo. Esto obliga al usuario a pagar varias suscripciones si quiere toda la variedad existente, lo que explica la creciente "fatiga del suscriptor".

Es importante señalar que este análisis se basa en datos de 2021, y desde entonces el panorama ha evolucionado de forma agresiva. Un claro ejemplo es la fusión de Hulu dentro de la interfaz de Disney+ (The Walt Disney Company, 2024), lo que habrá disparado el volumen y la variedad del catálogo de Disney de forma masiva. Por otro lado, Netflix ha seguido reforzando su dominio con series globales de alto impacto que seguramente habrán movido sus métricas de fidelización. Como idea para el futuro, sería increíble poder cruzar estos datos actualizados con el precio actual de las cuotas o los tiempos de visualización reales, pero las plataformas mantienen esos datos muy "capados" y son muy opacas con sus métricas. Conseguir ese acceso permitiría ver cómo ha afectado la concentración de plataformas a la calidad real que recibe el usuario por su dinero.

Conclusión:

Este proyecto ha permitido transformar miles de filas de datos desordenados en una fotografía clara de un momento clave para el *streaming*: año 2021 (punto post-pandemia donde vivimos el verdadero auge del consumo digital y las plataformas se lanzaron a una carrera frenética por ganar suscriptores). La principal conclusión extraída es que el volumen no siempre es sinónimo de éxito. Mientras que algunas plataformas nacieron para ser un "archivo masivo", el análisis de los *ratings* demuestra que la audiencia valora mucho más los catálogos curados y especializados. Gracias a la opinión de los usuarios, se ha podido comprobar que la verdadera fuerza de una plataforma no está en cuántos títulos ofrece, sino en su capacidad para conectar con públicos muy concretos.

A lo largo del análisis, se ha confirmado que cuando un catálogo se vuelve demasiado masivo y disperso, como ocurre en el caso de Amazon, es mucho más difícil mantener el nivel de calidad que el espectador espera. Por el contrario, las plataformas con una identidad más definida y una mayor especialización (como ocurre con el claro predominio de las series en Hulu o el enfoque familiar de Disney+) tienden a obtener mejores valoraciones. Esto nos indica que, al final, las plataformas que mejor funcionan son aquellas que prefieren especializarse en lo que saben que les sale bien. En un mercado saturado de opciones, el usuario prefiere un catálogo coherente y alineado con sus expectativas por encima de una oferta gigante pero sin identidad.

Finalmente, un hallazgo muy revelador es la superioridad de la serie frente a la película en términos de crítica. En todas las plataformas analizadas, el formato episódico logra una conexión y una valoración más alta, lo que sugiere que el modelo de negocio del *streaming* favorece hoy más la fidelidad de una buena serie que el consumo rápido de un largometraje.

Referencias bibliográficas:

Bansal, S. (2021). *Amazon Prime movies and TV shows* [Dataset]. Kaggle. Recuperado en enero de 2026 de <https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows>

Bansal, S. (2021). *Disney+ movies and TV shows* [Dataset]. Kaggle. Recuperado en enero de 2026 de <https://www.kaggle.com/datasets/shivamb/disney-movies-and-tv-shows>

Bansal, S. (2021). *Hulu movies and TV shows* [Dataset]. Kaggle. Recuperado en enero de 2026 de <https://www.kaggle.com/datasets/shivamb/hulu-movies-and-tv-shows>

Bansal, S. (2021). *Netflix movies and TV shows* [Dataset]. Kaggle. Recuperado en enero de 2026 de <https://www.kaggle.com/datasets/shivamb/netflix-shows>

IMDb.com. (2024). *IMDb dataset subsets* [Base de datos en línea]. Recuperado en enero de 2026 de <https://developer.imdb.com/non-commercial-datasets/>

Simon-Kucher & Partners. (2023). *Global streaming study 2023*.

Recuperado en enero de 2026 de <https://www.simon-kucher.com/en/insights/global-streaming-study-2023>

The Walt Disney Company. (2024, March 27). *Disney+ officially launches Hulu on Disney+ for bundle subscribers in the U.S.* [Press release].

Recuperado en enero de 2026 de <https://thewaltdisneycompany.com/news/hulu-on-disney-launch-streaming/>

Anexo:

Anexo A. La figura 6 presenta la distribución de calidad de cada plataforma, destacando los "puntos extremos" del catálogo (el 5% mejor y peor valorado). Este gráfico de cajas con dispersión permite observar no solo la puntuación promedio, sino también los títulos específicos que se sitúan en los rangos de excelencia o que han tenido una recepción crítica negativa, siempre bajo un filtro de representatividad estadística (mínimo 5.000 votos).

Anexo B. La figura 8 de realizó mediante rating ponderado