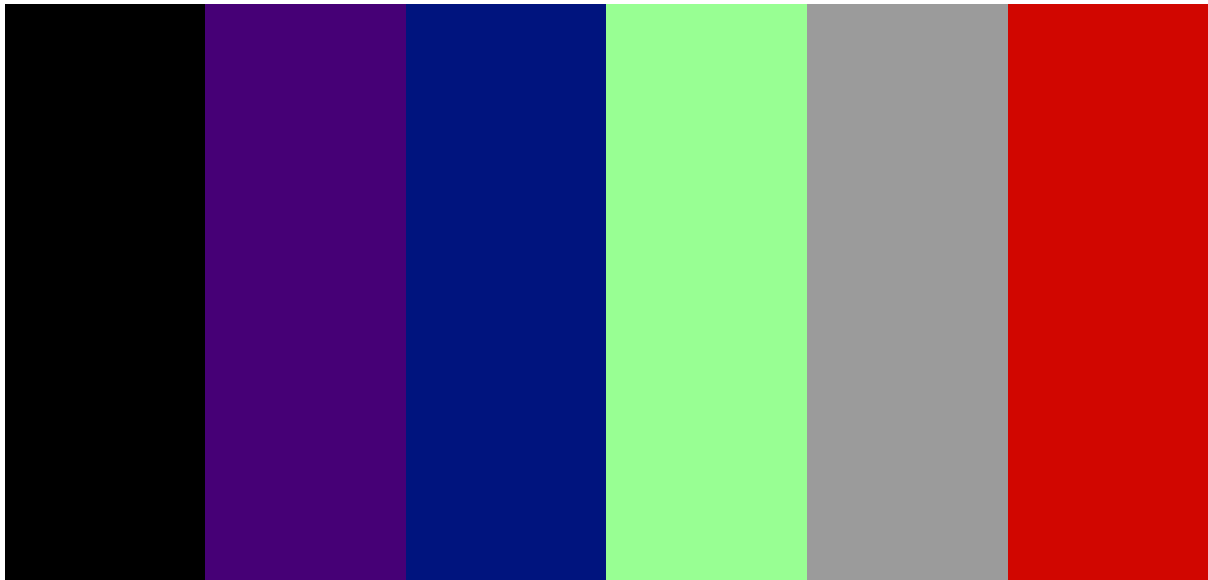


Plataformas de Streaming



Análisis Exploratorio de Datos del Contenido de las Principales Plataformas de Streaming en EE.UU.

Laura Ledo, Junio 2022

Las plataformas SVOD (Streaming Video On-Demand) empezaron a popularizarse hace poco más de 10 años, convirtiéndose a día de hoy en la opción principal de una parte mayoritaria de la población.

Desde entonces, y a la vez a la que las principales plataformas aumentaban significativamente el contenido de su catálogo, han ido apareciendo nuevas, incluyendo medios 'tradicionales', como HBO o Disney+, dejando al consumidor con un gran abanico de opciones entre las que elegir pero, ¿muchas opciones pueden ser demasiadas?

El panorama cada vez más amplio hace más difícil la elección, siendo muy común que estemos suscritos a más de una. Según un informe de Deloitte (2021), a pesar de que la aparición de nuevas plataformas está acelerando la rotación, la cantidad promedio de plataformas a las que estará suscrita una persona aumentará.

Objetivo

Si bien muchas variables entran en juego a la hora de elegir la o las plataformas de video on streaming, en este análisis (debido a la falta de datos gratuitos y que no hemos realizado web scrapping) nos centraremos en analizar el contenido de las principales plataformas con el objetivo de ver si existen, o no, diferencias que respalden nuestra tendencia a estar suscritos a varias.

¿Qué plataformas tienen más variedad de contenido?, ¿cuánto contenido comparten entre ellas?, ¿qué les diferencia a cada una?, ¿en qué destacan?

Hipótesis

Es necesario estar suscrito a más de una plataforma de streaming para poder tener acceso al contenido que queremos.

Análisis

Para poder confirmar o anular nuestra hipótesis, analizaremos el contenido de cada plataforma en función de:

Tipo de contenido: película o serie. Por falta de datos, no podremos saber si en serie engloban también contenido en directo, o retransmitido como Talk Shows o espectáculos deportivos.

Década: clasificamos en función de la década de estreno. También habría sido interesante tener las fechas en las que se ha subido a la plataforma cada título.

Edad recomendada: con esta variable intentaremos ver si las plataformas están dirigidas a un público específico o, si por el contrario, son generalistas.

Duración del contenido.

Género: clasificando en función de los géneros podríamos ver si las plataformas, que aunque están dirigidas al público masivo, se caracterizan por algún género concreto.

Países de producción: en caso de que hubiese diferencias o variedad entre las plataformas, nos podría indicar también el público determinado al que están dirigidas.

Temporadas: esta variable habría sido muy útil si tuviésemos la información del número de capítulos en cada serie. Ya que aunque en nuestros datos hay muchos menos títulos de series que de películas, una serie engloba mucho más contenido que una película, ya que esta se multiplica por sus temporadas y episodios.

Puntuación Imdb: con esta variable intentaremos ver la popularidad o 'calidad' del contenido de cada plataforma.

Puntuación Tmdb: la utilizaremos para comprobar si los datos de Imdb están sesgados, ya que pertenece a Amazon.

También, a partir de las variables de referencia ID y Título, veremos el porcentaje de contenido exclusivo en cada plataforma. Anotar en este punto que no podemos determinar si este contenido exclusivo ha sido producido por la propia plataforma o no, ya que no tenemos ese dato.

Los datos de los que partiremos para realizar este análisis son datasets de Kaggle de 6 (en negrita) de las principales plataformas de streaming en EE.UU.

Según el medio 'US News', las principales plataformas de video en streaming son:

1. **Netflix**
2. **Amazon Prime Video**
3. **Disney+**
4. Funimation
5. **Hulu**
6. **HBO Max**
7. Peacock
8. Crunchyroll
9. **Paramount+**
10. Apple+

En cada dataset eliminamos las columnas de descripción de los títulos, votaciones de Imdb, porque ya teníamos la puntuación de cada título, y la columna de popularidad de Tmdb.

Las variables con las que tratamos son:

VARIABLES DATASETS			
id	title	type	release_year
Limpieza	Comprobar nulos y duplicados	Clasificar entre película y serie	Representa el año de estreno del contenido
ID	Título	Tipo	Decada
Estas dos columnas las usamos como base para el análisis, cada fila debe tener un valor único que representa un contenido en la plataforma		Para poder separar el análisis entre películas y series. De cada dataset, creamos dos.	Agrupamos en décadas
age_certification	runtime	genres	production_countries

VARIABLES DATASETS			
Indica la edad recomendada de cada contenido		Varios géneros por contenido	En algunos casos, varios géneros por contenido
Edad_Recom	Duración	Géneros	País Producción
Reagrupamos en edad mínima recomendada: 0, 7, 13 o 17		Limpiamos para poder acceder a cada elemento de la lista, reasignamos el que aparece primero como género principal	Lo mismo que con géneros
seasons	imdb_score	tmdb_score	
	0 -10	0 -10	
Temporadas	Imdb Puntos	Tmdb Puntos	
	Para ver el éxito del contenido	Para comparar con los puntos de Imdb y asegurarnos que no están sesgados	

Debido a la diferencia en la cantidad de contenido entre series y películas, dividimos cada dataset en dos: uno de películas y otro de series. A su vez, al haber mucha diferencia entre la cantidad de contenido por plataforma, tratamos el análisis ponderando las cantidades sobre los totales.

Conclusiones

Tras el análisis comprobamos que 5 de las 6 plataformas a analizar ofrecen, principalmente, contenido exclusivo.

Podría resultar extraño que solo 1 de ellas tenga un peso de casi un 72% de contenido no exclusivo, ya que las estamos comparando entre ellas, pero esto se debe a que no todas tienen la misma cantidad de títulos, por lo que podemos determinar que Paramount tiene en su mayoría películas populares no producidas por ellos, los 'clásicos' de cada época.

Amazon y Netflix, como podíamos imaginar, son las principales en cantidad de contenido, y en los dos géneros principales: drama y comedia. Destacando Amazon para películas y Netflix para series -es la que tiene una mejor media de voto por título en serie-, seguida por Hulu, que por los datos del análisis es principalmente una plataforma de contenido tipo serie: contenido por episodios y temporadas, con un peso relevante de contenido de ciencia ficción y reality en series y documentales y thriller para películas. HBO, sigue un poco por detrás a Netflix, intercambiándose posiciones con Hulu según la clasificación, destacando que está en la segunda

posición por los puntos de su catálogo de películas, catálogo en su mayoría compuesto por los géneros drama, comedia y documentales

Y, por último Disney, la mejor valorada por su catálogo de películas. Es la que más se diferencia del resto. Prácticamente todo su catálogo de películas es exclusivo, y tiene el porcentaje más pequeño de series también en otras plataformas. No es sorpresa para nadie que destaque con diferencia en animación y fantasía (también destaca en documentales), y es la única que tiene un peso muy bajo en la categoría de drama (siendo drama el género con mejores puntuaciones).

A pesar de que pueden parecer plataformas muy similares, podemos concluir que cada una de ellas se va diferenciando poco a poco en alguna de las categorías más ¿Qué tienen en común? Todas ellas tienen en su catálogo contenido de los últimos 20 años.

Herramientas utilizadas

Python

Numpy

Pandas

Plotly

Fuentes

Obtención de datos: Kaggle.

Amazon: <https://www.kaggle.com/datasets/victorsoeiro/amazon-prime-tv-shows-and-movies?select=titles.csv>

Netflix: <https://www.kaggle.com/datasets/victorsoeiro/netflix-tv-shows-and-movies?select=titles.csv>

HBO: <https://www.kaggle.com/datasets/victorsoeiro/hbo-max-tv-shows-and-movies?select=titles.csv>

Hulu: <https://www.kaggle.com/datasets/victorsoeiro/hulu-tv-shows-and-movies?select=titles.csv>

Disney +: <https://www.kaggle.com/datasets/victorsoeiro/disney-tv-shows-and-movies?select=titles.csv>

Paramount +: <https://www.kaggle.com/datasets/victorsoeiro/paramount-tv-shows-and-movies?select=titles.csv>

Información sobre las plataformas de streaming:

Informe 'State of Play', Nielsen 2022: <https://www.nielsen.com/us/en/insights/report/2022/state-of-play/>

Artículo 'As the world churns: The streaming wars go global', Deloitte 2021: <https://www2.deloitte.com/xe/en/insights/industry/technology/technology-media-and-telecom-predictions/2022/streaming-video-churn-svod.html>

Resultados encuesta 'After a boom year in video streaming, what comes next?', PWC 2021: <https://www.pwc.com/us/en/services/consulting/library/consumer-intelligence-series/consumer-video-streaming-behavior.html>

Artículo 'Best Streaming Services of 2022', US News, 2022: <https://www.usnews.com/360-reviews/technology/streaming-services>