

Segundo informe PowerBI:
Comparación de contenido entre plataformas digitales

Jaime García Lozano y Bernat Amengual Mesquida

10 de abril de 2023

1. Introducción

En este informe se han analizado datos sobre el contenido de las plataformas *Netflix*, *Disney +* y *Amazon Prime*. Es posible acceder a él a través del siguiente enlace: [1]. Los datos han sido extraídos de la página *Kaggle* [2].

Se plantean las siguientes preguntas:

1. ¿Las compañías apuestan por contenido internacional o se centran en producciones estadounidenses?
2. ¿Qué características principales predominan en el contenido de cada plataforma?
3. ¿Cómo ha sido la evolución temporal del número de publicaciones ?
4. ¿Cómo es la distribución de la fecha de estreno de cada publicación en cada plataforma?

2. Limpieza de los datos

Disponemos de tres *csv*'s cada uno con los datos de una plataforma. Los tres archivos tienen la misma estructura y se componen de las siguientes columnas:

- **show__id**: id de la película o serie.
- **type**: si es una película o serie.
- **title**: título.
- **cast**: actores principales.
- **country**: país donde se ha producido.
- **date__added**: fecha en la que se publicó en la plataforma.
- **release__year**: fecha de estreno.
- **rating**: clasificación por edad.
- **duration**: duración.
- **listed__in**: género.
- **description**: descripción breve del argumento.

La limpieza se ha llevado a cabo en *Python*, se adjunta un *notebook* con el código utilizado.

Se han hecho los siguientes pasos:

1. Concatenación de los tres datasets en uno nuevo con una columna adicional que indica la plataforma a la que corresponde cada dato.
2. En la columna *rating*, hay diferentes formatos refiriéndose a una misma clasificación por edad, debido a que cambia en cada país. Por ejemplo, algunas celdas tienen el formato '18+', mientras que otras está escrito como 'AGE_18_'. Por ello, se ha establecido un mismo formato en las clasificaciones equivalentes. En bastantes casos no ha sido posible por tener un criterio completamente distinto al resto de países.
3. En la columna *listed__in*, sucede algo similar. Los géneros son escritos de manera diferente dependiendo de la plataforma. Por ejemplo, tenemos en *Netflix* el género *documentary*, mientras que en *Amazon Prime* está escrito en plural.

Para solucionar parcialmente el problema, se ha utilizado una librería de *Python* con el fin de convertir los plurales en singulares, mientras que con otros géneros no ha habido más remedio que cambiarlos a mano.

4. Tanto en la columna *listed_in* como en la de *country* puede haber celdas que pertenezcan a varias categorías. Asumiendo que la primera de ellas es la de mayor importancia, se ha decidido prescindir del resto.
5. Por último, se guarda este dataset llamado *all_data* en formato *excel* para posteriormente ser abierto en *Power Bi*.

3. Países productores

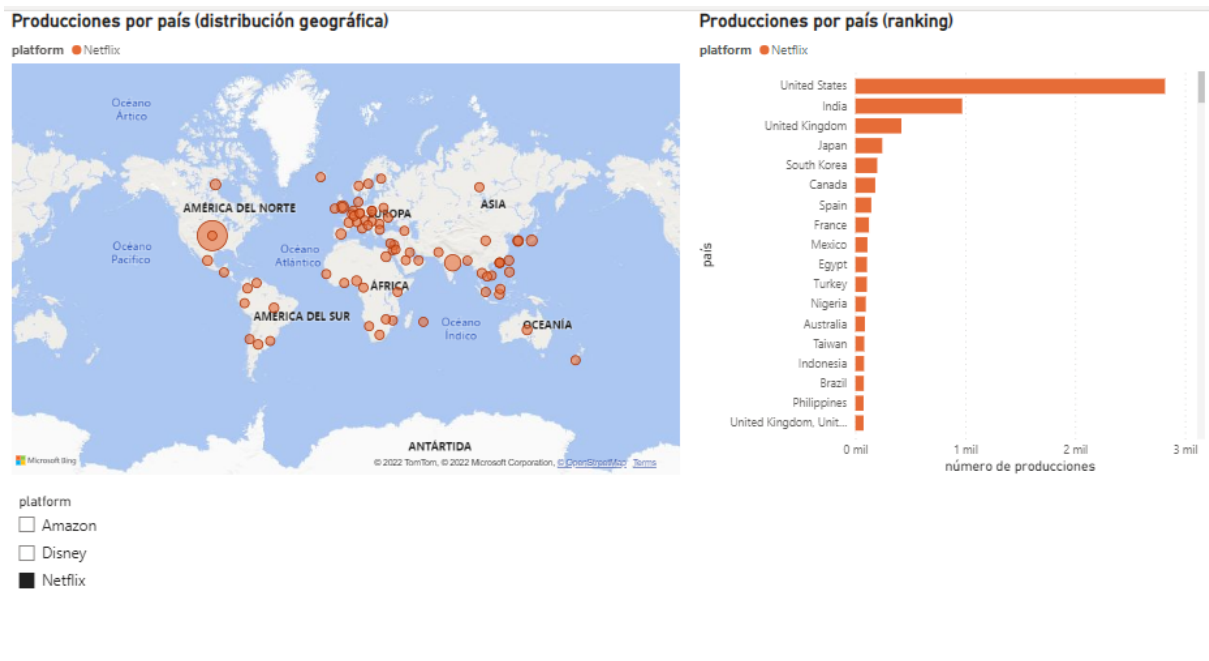


Figura 1: A la izquierda, distribución geográfica de la cantidad de contenido producido. A la derecha, ranking de los países productores.

Lo que podemos ver claramente, es que el origen del contenido es mayoritariamente estadounidense en las tres plataformas. Sin embargo, *Netflix* es la que más apuesta por contenido internacional, seguida de *Amazon Prime* (nota: usar el filtro para cambiar de plataforma). Resulta curioso que el segundo país con más producciones en ambas plataformas sea la India.

4. Tipo de contenido

En este apartado se analizan los siguientes puntos:

- Proporción de películas y series.
- Proporción de contenido por clasificación de edad.
- Proporción de contenido por género.

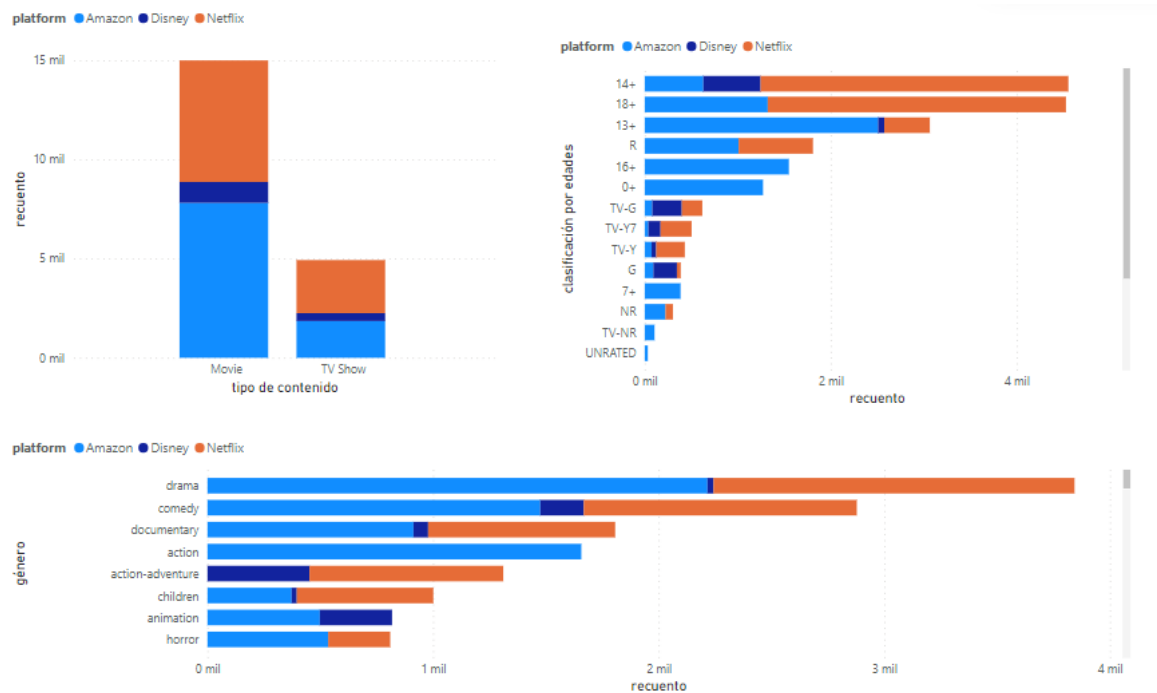


Figura 2: Arriba a la izquierda, cantidad de películas y series. Arriba a la derecha contenido clasificado por clasificación de edad. Abajo, contenido clasificado por género.

Vemos que en las tres plataformas hay muchas más películas que series. Además, en general, la mayor parte del contenido es para mayores de 13 o 18 años, lo que indica que el catálogo está enfocado en general a adolescentes o adultos. Por último, drama es el género más habitual en *Netflix* y *Amazon Prime*, seguido de comedia, mientras que en *Disney Plus* predomina el cine de acción y aventuras.

5. Publicación de contenido

En la tercera hoja del informe, tenemos la evolución temporal de la cantidad de contenido publicado en cada plataforma:

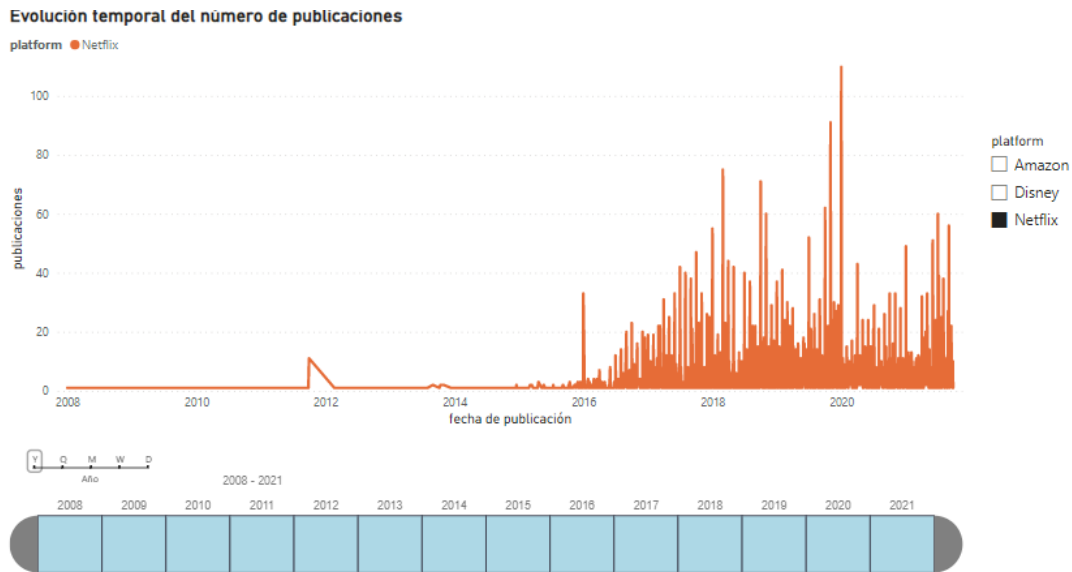


Figura 3: Evolución temporal de la cantidad de contenido publicado en cada plataforma con diversos filtros: el de abajo permite centrarse en un periodo concreto, mientras que el otro permite seleccionar la plataforma que nos interese.

En el caso *Amazon Prime* y de *Disney Plus*, la evolución no nos dice demasiado. Su apertura fue mucho después que la de *Netflix* y no se dispone de tantos datos. Por lo que podemos ver, incluyeron bastante contenido en el momento en el que iniciaron la plataforma y después la frecuencia se mantuvo más o menos constante.

Netflix, en cambio, incrementa la cantidad de contenido enormemente a partir del 2016 , con un pico destacable en el 2022.

6. Antigüedad del contenido

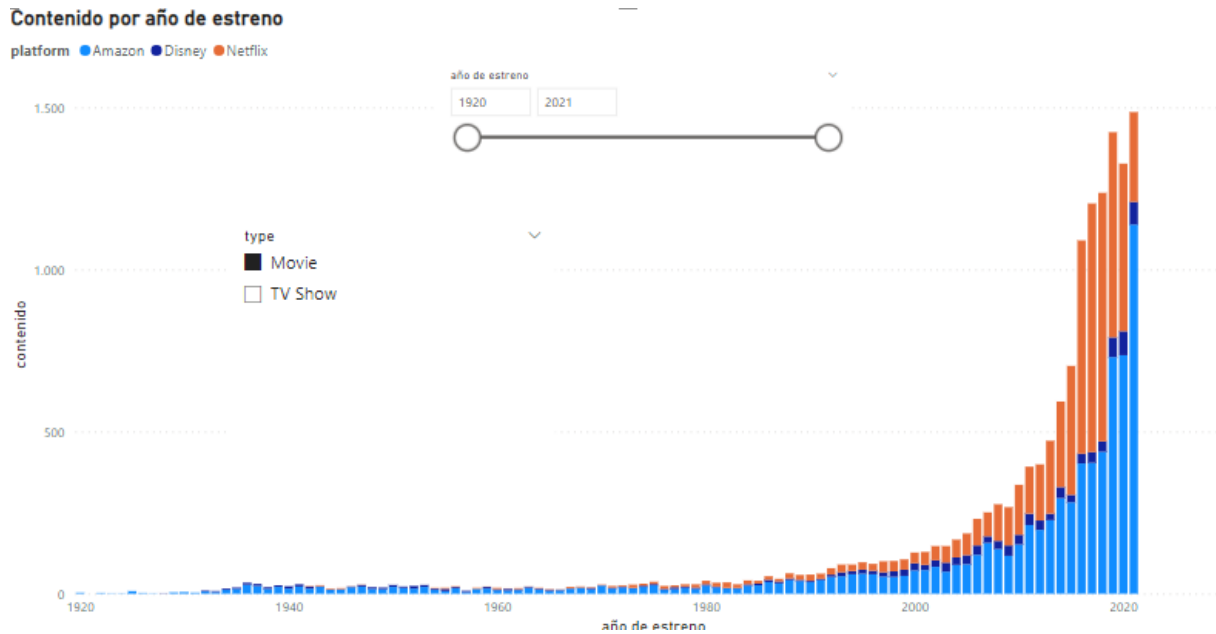


Figura 4: Cantidad de contenido por año de estreno. Se ha puesto un filtro para poder visualizar por separado las películas y series, además de otro para poder seleccionar distintos intervalos de tiempo.

Las tres empresas apuestan por contenido de los últimos años y poco por contenido clásico.

Referencias

- [1] *Informe*, https://app.powerbi.com/links/V0UXnm__Pp?ctid=71bb3458-d75f-4716-a5cb-647cdacbc0af&pbi_source=linkShare.
- [2] *Datasets*, <https://www.kaggle.com/datasets/shivamb/netflix-shows>.