



## Índice

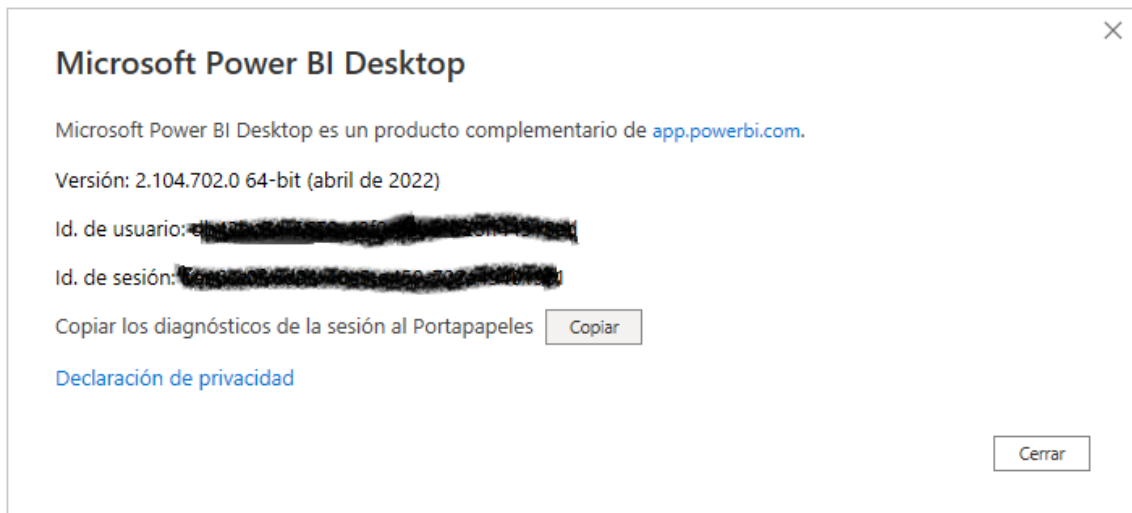
1. Introducción .....	2
Presentación.....	2
Análisis de la situación .....	2
Contexto .....	2
Alcance .....	3
Destinatarios .....	3
2. Modelo Entidad – Relación .....	4
3. Transformación de las Tablas.....	5
Transformación de las tablas con Power Query .....	5
Columnas y definición de tipo de datos .....	11
4. Tabla de Medidas Calculadas .....	12
5. Análisis Funcional .....	14
Portada .....	14
Análisis General .....	15
Análisis Específico.....	18

## 1. Introducción

### Presentación

Nosotros somos una empresa dedicada al servicio de la consultoría tecnológica que genera valor para la toma de decisiones del negocio. Ayudamos a purificar los datos de la organización para que las áreas del negocio puedan analizarlos para tomar mejores decisiones. Transformamos tus datos en información ágil, relevante, e inteligente.

Para solucionar los problemas del negocio utilizamos como software de aplicación Microsoft Power BI Desktop



### Análisis de la situación

Se nos ha encomendado que analicemos los datos de tres plataformas de video streaming muy populares como Disney +, Amazon Prime Video, y Netflix, con el fin de realizar un análisis general, y otro específico sobre cada plataforma, haciendo énfasis en la cantidad de títulos (películas y series) que ofrecen cada una de las plataformas a analizar, como también ver cuál es la plataforma de streaming más popular, o saber cuáles son las películas de mayor duración.

### Contexto

Descripción de cada uno de los servicios a analizar

**Netflix:** es un servicio de streaming que ofrece una amplia variedad de series, películas, títulos de anime, documentales y otros contenidos premiados en miles de dispositivos conectados a internet. Puede ver todo el contenido que quieras, cuando quieras y sin un solo anuncio por una tarifa mensual reducida.

**Disney+:** Disney no solo ofrece contenido clásico de sus canales Disney Channel o Espn, sino que también ofrece documentales de National Geographic, series y películas de Marvel, series y películas de Star Wars, y contenido y cortos de Pixar.

**Amazon Prime Video:** Prime Video es un servicio de streaming de películas y series creado y gestionado por Amazon. La plataforma ofrece miles de títulos a cambio de suscripción mensual o anual.

### Alcance

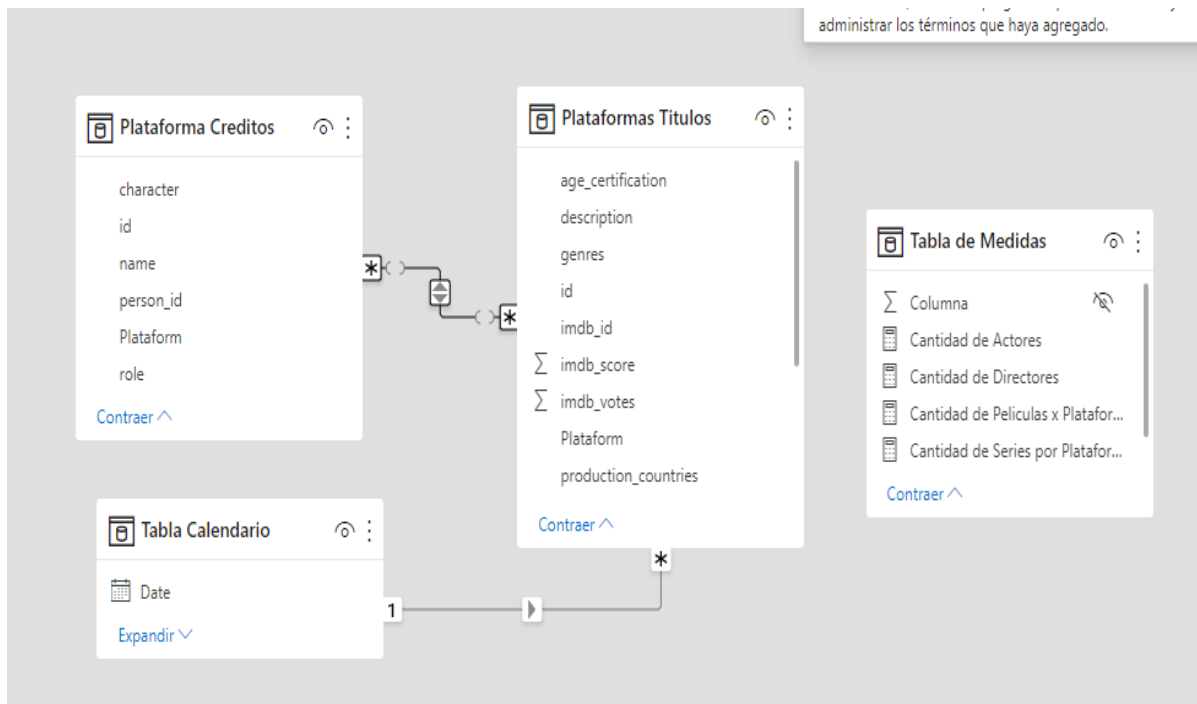
Se analizan todo el contenido de ofertas propuestas por cada una de plataformas en cuestión que abarcan títulos desde 1912 hasta la actualidad.

### Destinatarios

El presente análisis está orientado al público en general, que pueda ver y decidir que plataforma se adapta mejor a sus gustos y preferencias. Como así también ver como se desglosa cada plataforma por tipo de películas o series, o por género, o ver cuáles son las películas de mayor duración por ej.

Conocer la gran variedad de actores y directores que integran este universo de fantasía que se lo conoce como “el séptimo arte”.

## 2. Modelo Entidad – Relación



### 3. Transformación de las Tablas

Se contaba con una base de datos de cada una de las plataformas a analizar las cuales fueron extraídas de Kaggle;

Data set de Amazon

<https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows>

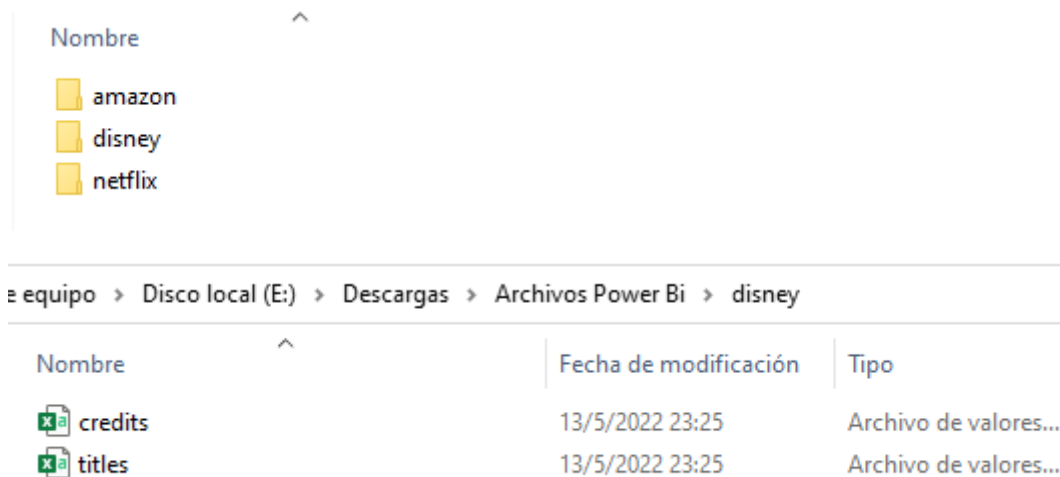
Data set de Disney

<https://www.kaggle.com/datasets/shivamb/disney-movies-and-tv-shows>

Data set de Netflix

<https://www.kaggle.com/datasets/victorsoeiro/netflix-tv-shows-and-movies>

Cada data base constaba de dos archivos a su vez, credits.csv y titles.csv



#### Transformación de las tablas con Power Query

El archivo titles.csv, contiene los id (clave única de identificación) de cada título, el nombre de cada una de las propuestas que ofrece, una breve sinopsis del contenido, el tipo, el género, el año de lanzamiento, la clasificación por edades, el tiempo de duración expresado en minutos, el país de producción, el índice de popularidad, de votos, y la cantidad de veces que fue vista.

En cambio, el archivo credits.csv, contiene los datos de los actores y actrices que participaron en cada uno de los shows, como del director de la misma (Id, y el nombre). Así también, el personaje que interpretaron.

Antes de anexas cada una de las tablas de acuerdo a su tipo (titles y credits), se procedió a realizar algunas modificaciones en particular sobre cada una.

### Tabla Amazon Titulos

En esta tabla se efectuaron las siguientes modificaciones:

- Se quitaron los errores de la columna “release\_year”

*Table.RemoveRowsWithErrors("#Filas filtradas", {"release\_year"})*

- Se cambio el formato número entero de la columna “release\_year” a fecha
- Se agrego columna “Plataform”

*Table.AddColumn("#Errores quitados", "Plataform", each "Amazon", type text)*



### Tabla Disney Titulos

- Se agrego columna “Plataform”

*Table.AddColumn("#Tipo cambiado", "Plataform", each "Disney", type text)*

- Se cambio el formato número entero de la columna “release\_year” a fecha



### Tabla Netflix Titulos

- Se agrego columna "Plataform"

*Table.AddColumn("#Tipo cambiado", "Plataform", each "Netflix", type text)*

- Se cambio el formato número entero de la columna "release\_year" a fecha



### Tabla Creditos Netflix

- Se agrego columna "Plataform"

*Table.AddColumn("#Tipo cambiado", "Plataform", each "Netflix", type text)*



### Tabla Creditos Amazon

- Se agrego columna "Plataform"

*Table.AddColumn("#Tipo cambiado", "Plataform", each "Amazon", type text)*





### Tabla Credits Disney

- Se agrego columna "Plataform"

*Table.AddColumn(#"Tipo cambiado", "Plataform", each "Disney", type text)*



*Se procede a anexar las tablas....*

### Tabla Plataforma Titulos

- Se cambio el nombre de la tabla anexada por Plataforma Titulos
- Se procedió a deshabitarla carga de las tablas titles.csv de cada una de las plataformas
- Se agrego una nueva columna "genres", dado que varias películas y series tenían más de un género, se decidió tomar como género principal el primer género de cada evento.

```
#"Texto insertado entre delimitadores" = Table.AddColumn(Origen, "genres_", each
Text.BetweenDelimiters([genres], "", ""), type text),
```

- Se procedió a quitar la columna original “genres”

```
#"Columnas quitadas" = Table.RemoveColumns("#Columnas reordenadas",{"genres"}),
```

- Se agrego una nueva columna “production\_countries\_”, dado que varias películas y series tenían más de un país de producción, se decidió tomar como país principal el primero de cada evento.

```
#"Texto insertado entre delimitadores1" = Table.AddColumn("#Filas filtradas",
"production_countries_", each Text.BetweenDelimiters([production_countries], "", ""), type text),
```

- Se procedió a quitar la columna original “production\_countries”

```
#"Columnas quitadas1" = Table.RemoveColumns("#Columnas
reordenadas1",{"production_countries"}),
```

- Se cambio el nombre de las columnas “genres\_” por “genres” y “production\_countries\_” por “production\_countries”

```
#"Columnas con nombre cambiado" = Table.RenameColumns("#Columnas quitadas1",{{"genres_",
"genres"}, {"production_countries_", "production_countries"}}),
```

- Se borro la columna “seasons”, por considerarla de poca utilidad dado la gran cantidad de filas en blanco que posee

```
#"Columnas quitadas2" = Table.RemoveColumns("#Columnas con nombre cambiado",{"seasons"}),
```

- Para subsanar el tema de los errores y demás espacios en blanco se decidió filtrar las columnas “genres”, “type”, “production\_countries”, considerando que de esta manera se obtenía la menor pérdida de datos posibles

```
#"Filas filtradas1" = Table.SelectRows("#Columnas quitadas2", each ([genres] <> "" and [genres] <>
"AU" and [genres] <> "CA" and [genres] <> "CN" and [genres] <> "GB" and [genres] <> "JP" and
[genres] <> "US") and ([production_countries] <> "") and ([type] = "MOVIE" or [type] = "SHOW"))
```

- Se quito de la columna “id” las filas en blanco

```
#"Filas filtradas" = Table.SelectRows("#Columnas quitadas", each true),
```

```

let
    Origen = Table.Combine({#"Amazon Titulos", #"Disney Titulos", #"Netflix Titulos"}),
    #"Texto insertado entre delimitadores" = Table.AddColumn(Origen, "genres_", each Text.BetweenDelimiters([genres], "", ""), type text),
    #"Columnas reordenadas" = Table.ReorderColumns(#"Texto insertado entre delimitadores",{"id", "title", "type", "description", "release_year", "genres_"}),
    #"Columnas quitadas" = Table.RemoveColumns(#"Columnas reordenadas",{"genres_"}),
    #"Filas filtradas" = Table.SelectRows(#"Columnas quitadas", each true),
    #"Texto insertado entre delimitadores1" = Table.AddColumn(#"Filas filtradas", "production_countries_", each Text.BetweenDelimiters([production_countries], "", ""), type text),
    #"Columnas reordenadas1" = Table.ReorderColumns(#"Texto insertado entre delimitadores1",{"id", "title", "type", "description", "release_year", "production_countries_"}),
    #"Columnas quitadas1" = Table.RemoveColumns(#"Columnas reordenadas1",{"production_countries_"}),
    #"Columnas con nombre cambiado" = Table.RenameColumns(#"Columnas quitadas1",{"genres_", "genres"}, {"production_countries_", "production_countries"}),
    #"Columnas quitadas2" = Table.RemoveColumns(#"Columnas con nombre cambiado",{"seasons"}),
    #"Filas filtradas1" = Table.SelectRows(#"Columnas quitadas2", each ([genres] <> "" and [genres] <> "AU" and [genres] <> "CA" and [genres] <> "US")),
in
    #"Filas filtradas1"
  
```

✓ No se han detectado errores de sintaxis.

Listo Cancelar

### Tabla Plataforma Creditos

No se efectuaron modificaciones

```

let
    Origen = Table.Combine({#"Netflix Creditos", #"Disney Creditos", #"Amazon Creditos"})
in
    Origen
  
```

✓ No se han detectado errores de sintaxis.

Listo Cancelar

### *Se procedió a aplicar los cambios*

En el menú Datos, en el panel Herramientas de tablas, a la columna “release\_year” se le cambio el formato de fecha a (Short Date)

Se saco de la Tabla Plataforma Titulos, en Herramientas de columna/Propiedades/Resumen a la columna “runtime” la suma y se optó por “No resumir”, al igual que para la columna “tmdb\_popularity”.

Lo mismo se hizo con la columna “person\_id” de la Tabla Plataforma Creditos.

**Tabla Calendario**

Se procedió a realizar una tabla calendario, para que cuando trabajemos con las fechas arroje resultados mucho más apropiados a la realidad. Se considero la mínima fecha de estreno de una película o serie, hasta nuestros días.

Tabla Calendario =

//Tabla Calendario, que muestra el día por día desde la fecha min, hasta la actualidad

```
CALENDAR(
  Min('Plataformas Titulos'[release_year]),
  MAX('Plataformas Titulos'[release_year])
)
```

Columnas y definición de tipo de datos

Al anexar las tablas, entonces obtuvimos dos nuevas tablas, las cuales pasamos a continuación a detallar como quedaron conformadas

**Tabla Plataforma Titulos**

Columna	Tipo de Dato
Id	Texto
title	Texto
type	Texto
description	Texto
release_year	Fecha
age_certification	Texto
runtime	Número entero
Plataform	Texto
imdb_id	Texto
imdb_score	Número entero
imdb_votes	Número entero
tmdb_popularity	Número entero
tmdb_score	Número entero
genres	Texto
production_countries	Texto

**Tabla Plataforma Creditos**

Columna	Tipo de Dato
person_id	Número entero
id	Texto
name	Texto
character	Texto
role	Texto
Plataform	Texto

#### 4. Tabla de Medidas Calculadas

1. **Cantidad de Actores;** cuenta la cantidad de actores, no repetidos, que participan en las tres plataformas

```
Cantidad de Actores =
//Contar valores únicos
CALCULATE(
    COUNT('Plataforma Creditos'[person_id]),
    'Plataforma Creditos'[role] = "ACTOR"
)
```

2. **Cantidad de Directores;** cuenta la cantidad de directores, no repetidos, que participan en las tres plataformas

```
Cantidad de Directores =
//Contar valores únicos
CALCULATE(
    COUNT('Plataforma Creditos'[person_id]),
    'Plataforma Creditos'[role] = "Director"
)
```

3. **Cantidad de películas;** Devuelve la cantidad de películas

```
Cantidad de Peliculas x Plataforma =
//Cuenta la cantidad de peliculas de las 3 plataformas
CALCULATE(
    COUNT('Plataformas Titulos'[type]),
    'Plataformas Titulos'[type] ="Movie"
)
```

4. **Cantidad de Series;** Devuelve la cantidad de series dentro de las tres plataformas

```
Cantidad de Series por Plataforma =
//La cantidad de series dentro de la columna type
CALCULATE(
    COUNT('Plataformas Titulos'[type]),
    'Plataformas Titulos'[type] ="Show"
)
```

5. **Cantidad de títulos;** Cuenta los id, para saber cuantos títulos ofrecen en tres las plataformas

Cantidad de Titulos =

```
//La cantidad de titulos que hay en las tres plataformas  
COUNT(  
    'Plataformas Titulos'[id]  
)
```

## 5. Análisis Funcional

### Tipografía

Se utilizó como tipografía para todo el trabajo la fuente SegoeUI



[https://es-academic.com/pictures/eswiki/50/255px-SegoeUI\\_SP.svg.png](https://es-academic.com/pictures/eswiki/50/255px-SegoeUI_SP.svg.png)

### Portada



Para la portada se usó la imagen

<https://www.latercera.com/practico/noticia/guia-para-elegir-la-mejor-plataforma-de-streaming/37QKBRSTJBG3PE25HGVRML6I4/>

**Título:** Se usó como tipo de letra SegoeUI, tamaño 80, formato negrita, Alineación Centrado, todo en Mayúsculas

Para el resto del texto se usó un tamaño 20, negrita, alineación a la izquierda.

**Botones:** Para los botones usamos un tamaño 12, color #FFFFFF, alineación horizontal Centrado, y alineación vertical Medio.

**Aclaraciones:** Todos los gráficos tendrán una transparencia de fondo del 70%, para los títulos de los gráficos se usarán Título tipo 3, tamaño 12, negrita, color de la letra #000000, alineación Centrado. Pudiendo variar el color de fondo del título

### Análisis General



Como fondo de lienzo se usó como imagen

<https://www.marketeroslatam.com/5-mejores-plataformas-de-streaming-2021/>, con un nivel de transparencia del 50%.

**Título:** Se usó un tamaño 20, negrita, alineación Centrado. Dicho cuadro de texto cuenta con un tamaño de 41 de alto y 1280 de ancho. Para el fondo se usó el color #208B20.

Se agregó un botón de retroceso.

**KPI:** Tenemos tres KPIs, uno con la cantidad de títulos, cantidad de actores, y otro con la cantidad de directores.



Se uso un tamaño para la fuente de 45, negrita, color #000000.

Para la etiqueta de categoría se uso un tamaño 12, negrita, color #000000.

**Gráfico de áreas;** Muestra la cantidad de títulos por año, mostrando el aumento de la producción a través del tiempo, desde sus inicios hasta nuestros días, y a su vez se desagrega por tipo (película o serie).

Eje X; Tenemos la columna “reléase\_year”.

Valores; Usa un tamaño 10, negrita, color #000000.

Eje y; Tenemos la medida cantidad de títulos.

Valores; Usa un tamaño 10, negrita, color #000000.

Rango; min de 10, y máx. 1200

Leyenda;

Texto; Fuente 10, negrita, color #000000.

Líneas;

Estilo; Puntos

Colores; Movie; #FFB310

Show; #A7ACD3

Título;

Fondo; color#D0D3A7

**Gráfico embudo:** Muestra las películas de máxima duración, contenidas en las tres plataformas, mostrando en el eje Y el nombre de la película y en el eje X, la duración en minutos. Este grafico no interactúa con el segmentador “type” cuando se selecciona la opción “show”.

Color de las barras; #414FB1

Etiqueta de datos; tamaño de la fuente 10, negrita, color #FFFFFF, muestra los valores en el interior de la barra.

Etiqueta de categoría; tamaño de la fuente 10, negrita, color #000000

Título; Color de fondo del título #F5C4AF

Filtro; En la pestaña filtro se aplico

Tipo de filtro; Top N

Cantidad a mostrar; 4

Por valor; Máximo de la columna “runtime”

**Gráfico de columnas agrupados:** Muestra la cantidad de títulos producidos por país. A este gráfico se le aplicó un tooltip.

Eje X: columna “production\_countries”, con una fuente tamaño 10, negrita, color #000000 y un ancho máximo de las columnas de un 25%.

Eje Y: la medida cantidad de títulos

Color de las barras; #118DFF

Etiqueta de datos; tamaño de la fuente 10, negrita, color #000000, muestra los valores por encima de la barra, los cuales tienen un fondo de color #000000 con una transparencia del 88%

Color de fondo del título es #E66C37

Filtro: En la pestaña filtro se aplicó

Tipo de filtro: Top N

Cantidad a mostrar: 7

Por valor: Recuento de la columna “production\_countries”

### **Tooltips**

**Gráfico circular:** Muestra la cantidad de títulos de todas las plataformas, segregadas por tipo (películas o series). Muestra el valor de los datos y el porcentaje total. Este gráfico se muestra de manera conjunta con el de barras agrupadas, al pasar el cursor por encima de alguna de las barras, mostrando la cantidad de películas y series que produce ese país en particular.

Título: Color de fondo #8CB1DE

Legenda: Tamaño 10, negrita, color #000000

Color: Movie #357F69

Show #C32626

**Treemap:** Muestra la gran cantidad de géneros que componen los diferentes títulos de las plataformas en cuestión, esta visualización está dividida en rectángulos a los que se les asigna un tamaño y orden en función de la cantidad veces que se repite ese género.

Etiqueta de datos: tamaño de la fuente 10, negrita, color #000000.

Etiqueta de categoría: tamaño de la fuente 10, negrita, color #FFFFFF

Color de fondo del título: #5E864B

Este gráfico contiene un filtro,

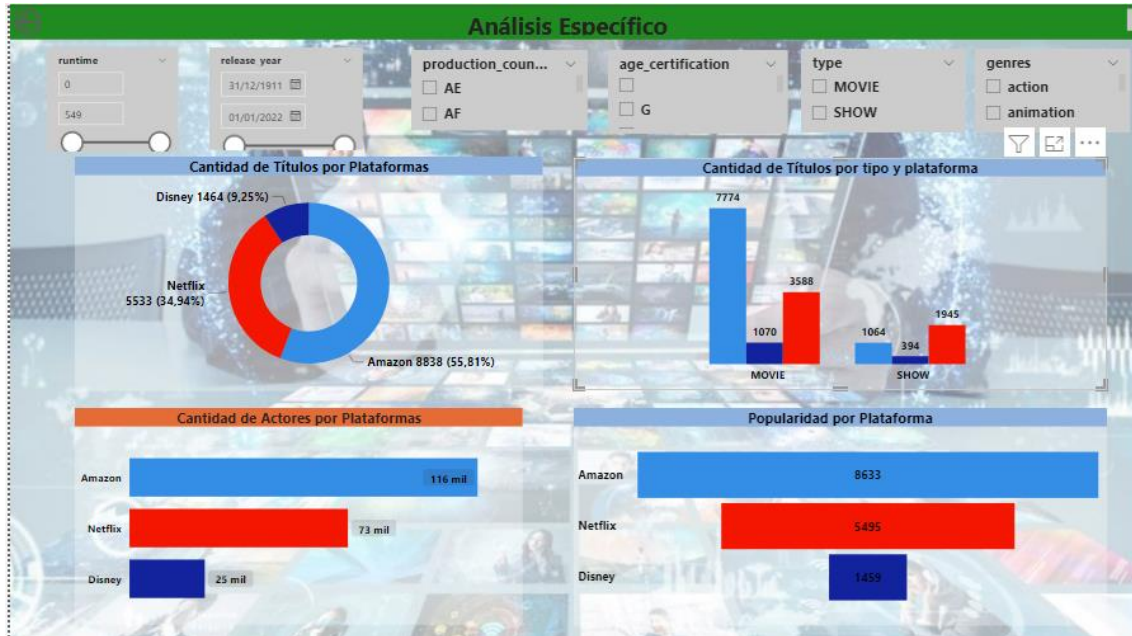
Tipo de filtro: Top N

Cantidad a mostrar: 7

Por valor: Medida Cantidad de títulos.

**Segmentadores:** Encontramos segmentadores de datos por género, tipo, clasificación por edades, país de producción, por plataforma, año de lanzamiento y duración

### Análisis Específico



### Colores

Para el análisis específico, en los gráficos, se tuvo en cuenta los colores predominantes de cada una de las plataformas.



Plataformas	Color
Netflix	#F41600
Disney	#12239E
Amazon Prime	#358EE6

Todos los títulos tienen el mismo color de fondo #8CB1DE, aparte de lo ya mencionado que anteriormente que tienen el mismo formato.

**Gráfico de anillos:** Muestra la cantidad de títulos discriminados por plataforma

*Etiqueta de datos;* tamaño de la fuente 10, negrita, color #000000.

**Gráfico de columnas agrupadas:** Muestra la cantidad de títulos, diferenciándolo por tipo y plataforma.

Eje X: columna “type”, con una fuente tamaño 10, negrita, color #000000y un ancho máximo de las columnas de un 25%.

Eje Y: La medida Cantidad de títulos

Después tenemos otro gráfico de columnas agrupadas, el que muestra la cantidad de actores por plataforma, este grafico tiene un color de título distinto al resto por tratarse de una tabla distinta de donde se extrajeron los datos. (Color #E66C37)

**Gráfico de embudo:** Muestra que plataforma es más popular. Haciendo un recuento de los valores de la columna. “tmdb\_popularity”