



Team: **Brisna Huerta, Bernardo Bartelucci & Arturo Suárez**

Comisión 29655

Data Analytics

Agosto 2022

Contenido

1	INTRODUCCIÓN	3
2	DESCRIPCIÓN DE LA TEMÁTICA DE LOS DATOS	4
3	OBJETIVOS	4
4	DATASET	4
5	USUARIO FINAL Y NIVEL DE APLICACIÓN DE ANÁLISIS	7
6	DIAGRAMA ENTIDAD-RELACIÓN	7
7	LISTADO DE TABLAS	8
7.1	Netflix Titles.....	8
7.2	Type.....	8
7.3	Director and Actor	8
7.4	Country.....	8
7.5	Rate	8
7.6	Listed_in	9
8	LISTADO DE COLUMNAS POR TABLAS.....	9
9	TRANSFORMACIÓN DE DATOS.....	10
10	MODELO RELACIONAL	11
11	MEDIDAS CALCULADAS	11
11.1	Medida con una variable y una función de agregación	11
11.2	Medida con una variable y una función de conteo:.....	12
11.3	Medida con dos variables (una función de agregación y una función de inteligencia de tiempo).....	12
11.4	Medida que contenga un parámetro (con una función de agregación)	12
12	SEGMENTACIONES ELEGIDAS:	13
13	ANÁLISIS FUNCIONAL DEL TABLERO	14
13.1	Menu	14
13.2	GENERAL:.....	15
13.3	Contents	17
13.4	CLASSIFICATION.....	18
14	HERRAMIENTAS TECNOLÓGICAS UTILIZADAS.....	19
15	CONCLUSIÓN	20

1 INTRODUCCIÓN

NETFLIX es mundialmente conocido como el pionero de la televisión por internet, también conocido este servicio como *streaming video*.

Hasta hace unos años, lideraba el mercado por lo cuál su catalogo fue enriquecido con una gran cantidad de títulos e incluso, empezó a generar sus propios formatos.

Sin embargo, con la llegada de una terrible pandemia, muchas otras empresas comenzaron a buscar posibilidades de negocios y, sobre todo, apostaron por innovar sus servicios, siendo el internet el mejor medio para no morir en el intento.

Es por ello que la plataforma de Netflix comenzó a buscar nuevos títulos con temas diversos, convenios de exclusividad, etc.; con la finalidad de atraer y mantener la cantidad de suscriptores.

En los últimos años, el panorama no ha sido nada fácil para la compañía. Es por ello, que surgió la idea de tomar información de su catálogo completo (desde sus inicios hasta finales de 2021) para así, junto con la información de las calificaciones de IMBD y Rotten Tomatoes, poder observar la situación actual de la empresa en cuanto a contenido y su aceptación entre los consumidores.

2 DESCRIPCIÓN DE LA TEMÁTICA DE LOS DATOS

Para el presente proyecto, se extrajo información del catálogo de la plataforma de *streaming* “Netflix” y la recepción de la sociedad al mismo.

Se obtuvieron varios datos sobre los títulos de la plataforma. Por ejemplo:

- Director
- Actor
- Fecha del título
- Fecha en la que fue añadida al servicio

A su vez, mediante otro dataset se adquirieron datos sobre las puntuaciones de los títulos en el catálogo. Se tuvieron en cuenta dos de las páginas más populares: IMDB y Rotten Tomatoes.

3 OBJETIVOS

Este proyecto tiene como finalidad:

- Analizar la influencia cultural de distintos países en el catálogo como podrían ser EEUU e India, los cuales poseen una gran parte del directorio y, con ello, la predominancia de títulos pertenecientes a un idioma o continente.
- Estudiar las películas más exitosas y populares de la plataforma con base en las puntuaciones de la crítica.
- Analizar actores y directores que participaron en éstas y determinar si hay alguna relación en la influencia cultural.

4 DATASET

El dataset que se utilizó para el estudio de sus datos se muestra a continuación:

[Dataset NETFLIX](#)

Dicha recopilación de datos tiene en su mayor parte información cualitativa correspondiente al tipo de títulos vigentes en la plataforma NETFLIX hasta diciembre de 2021. Aunado a esto, cuenta con dos calificaciones de cada uno de los títulos hechas a través de los portales IMDB Y Rotten Tomatoes, los cuales tienen como puntuación máxima 5 y mínima 0.

Este dataset está compuesto por seis tablas, las cuales se muestran a continuación:

TABLA 1. NETFLIX TITLES

K22 fx FLI03

	A	B	C	D	E	F	G	H	I	J	K
1	show_id	ID type	title	ID country	date_added	release_year	ID_Rate	rating_IMDB	rating_rotten	duration	ID_1stlistedIn
2	s1	T01	Dick Johnson Is Dead	C80	septiembre 25, 2021	2020	PG-13	3.20	5.00	90 min	FLI01
3	s2	T02	Blood & Water	C67	septiembre 24, 2021	2021	TV-MA	5	1.6	2 Seasons	FLI02
4	s3	T02	Ganglands		septiembre 24, 2021	2021	TV-MA	1	3.9	1 Season	FLI03
5	s4	T02	Jailbirds New Orleans		septiembre 24, 2021	2021	TV-MA	4.9	1.5	1 Season	FLI04
6	s5	T02	Kota Factory	C30	septiembre 24, 2021	2021	TV-MA	4.2	5	2 Seasons	FLI02
7	s6	T02	Midnight Mass		septiembre 24, 2021	2021	TV-MA	4.8	3.8	1 Season	FLI05
8	s7	T01	My Little Pony: A New Generation		septiembre 24, 2021	2021	PG	1.4	3.7	91 min	FLI06
9	s8	T01	Sankofa	C80	septiembre 24, 2021	1993	TV-MA	0.9	1.1	125 min	FLI07
10	s9	T02	The Great British Baking Show	C79	septiembre 24, 2021	2021	TV-14	4.7	5	9 Seasons	FLI08
11	s10	T01	The Stairing	C80	septiembre 24, 2021	2021	PG-13	0.7	4.3	104 min	FLI09
12	s11	T02	Vendetta: Truth Lies and The Mafia		septiembre 24, 2021	2021	TV-MA	2.8	5	1 Season	FLI03
13	s12	T02	Bangkok Breaking		septiembre 23, 2021	2021	TV-MA	3.5	4.2	1 Season	FLI03
14	s13	T01	Je Suis Karl	C23	septiembre 23, 2021	2021	TV-MA	5	3.4	127 min	FLI07
15	s14	T01	Confessions of an Invisible Girl		septiembre 22, 2021	2021	TV-PG	2	1.4	91 min	FLI06
16	s15	T02	Crime Stories: India Detectives		septiembre 22, 2021	2021	TV-MA	5	4.7	1 Season	FLI08
17	s16	T02	Dear White People	C80	septiembre 22, 2021	2021	TV-MA	3.3	4.5	4 Seasons	FLI10
18	s17	T01	Europe's Most Dangerous Man: Otto Skorzeny in Spain		septiembre 22, 2021	2020	TV-MA	1.3	5	67 min	FLI01
19	s18	T02	Falsa identidad	C45	septiembre 22, 2021	2020	TV-MA	4.6	5	2 Seasons	FLI03
20	s19	T01	Intrusion		septiembre 22, 2021	2021	TV-14	3.8	0.9	94 min	FLI11
21	s20	T02	Jaguar		septiembre 22, 2021	2021	TV-MA	4.9	2.7	1 Season	FLI02
22	s21	T02	Monsters Inside: The 24 Faces of Billy Milligan		septiembre 22, 2021	2021	TV-14	0.8	1.4	1 Season	FLI03
23	s22	T02	Resurrections: Ertugrul	C76	septiembre 22, 2021	2018	TV-14	2.8	0.2	5 Seasons	FLI02
24	s23	T01	Avvai Shanmughi		septiembre 21, 2021	1996	TV-PG	4.4	1.5	161 min	FLI09
25	s24	T01	Go! Go! Cory Carson: Chrissy Takes the Wheel		septiembre 21, 2021	2021	TV-Y	4.4	0.8	61 min	FLI06

TABLA 2. DIRECTOR AND ACTOR

A1 fx show_id

	A	B	C
1	show_id	1st Director	1st Actor
2	s1	Kirsten Johnson	
3	s2		Ama Qamata
4	s3	Julien Leclercq	Sami Bouajila
5	s4		
6	s5		Mayur More
7	s6	Mike Flanagan	Kate Siegel
8	s7	Robert Cullen	Vanessa Hudgens
9	s8	Haile Gerima	Kofi Ghanaba
10	s9	Andy Devonshire	Mel Giedroyc
11	s10	Theodore Melfi	Melissa McCarthy
12	s11		
13	s12	Kongkiat Komesiri	Sukollawat Kanarot
14	s13	Christian Schwab	Luna Wedler
15	s14	Bruno Garotti	Klara Castanho
16	s15		
17	s16		Logan Browning
18	s17	Pedro de Echave García	
19	s18		Luis Ernesto Franco
20	s19	Adam Salky	Freida Pinto
21	s20		Blanca Suárez
22	s21	Olivier Megaton	
23	s22		Engin Altan Düzyatan
24	s23	K.S. Ravikumar	Kamal Hassan
25	s24	Alex Woo	Maisie Benson
26	s25	S. Shankar	Prashanth

TABLA 3. TYPE

B3 fx TV Sho

	A	B
1	ID type	type
2	T01	Movie
3	T02	TV Show

TABLA 4. TYPE

A1 ▾ fx ID_Country

	A	B
1	ID_Country	Country
2	C01	Argentina
3	C02	Australia
4	C03	Austria
5	C04	Bangladesh
6	C05	Belarus
7	C06	Belgium
8	C07	Brazil
9	C08	Bulgaria
10	C09	Cambodia
11	C10	Cameroon
12	C11	Canada
13	C12	Chile
14	C13	China
15	C14	Colombia
16	C15	Croatia

TABLA 5. COUNTRY

A1 ▾ fx ID_Country

	A	B
1	ID_Country	Country
2	C01	Argentina
3	C02	Australia
4	C03	Austria
5	C04	Bangladesh
6	C05	Belarus
7	C06	Belgium
8	C07	Brazil
9	C08	Bulgaria
10	C09	Cambodia
11	C10	Cameroon
12	C11	Canada
13	C12	Chile
14	C13	China

TABLA 6. LISTED_IN

B1 ▾ fx 1stlisted_in

	A	B
1	ID 1stListedIn	1stlisted_in
2	FLI01	Documentaries
3	FLI02	International TV Shows
4	FLI03	Crime TV Shows
5	FLI04	Docuseries
6	FLI05	TV Dramas
7	FLI06	Children & Family Movies
8	FLI07	Dramas
9	FLI08	British TV Shows
10	FLI09	Comedies
11	FLI10	TV Comedies
12	FLI11	Thrillers
13	FLI12	Horror Movies
14	FLI13	Kids' TV
15	FLI14	Action & Adventure
16	FLI15	Reality TV
17	FLI16	Anime Series

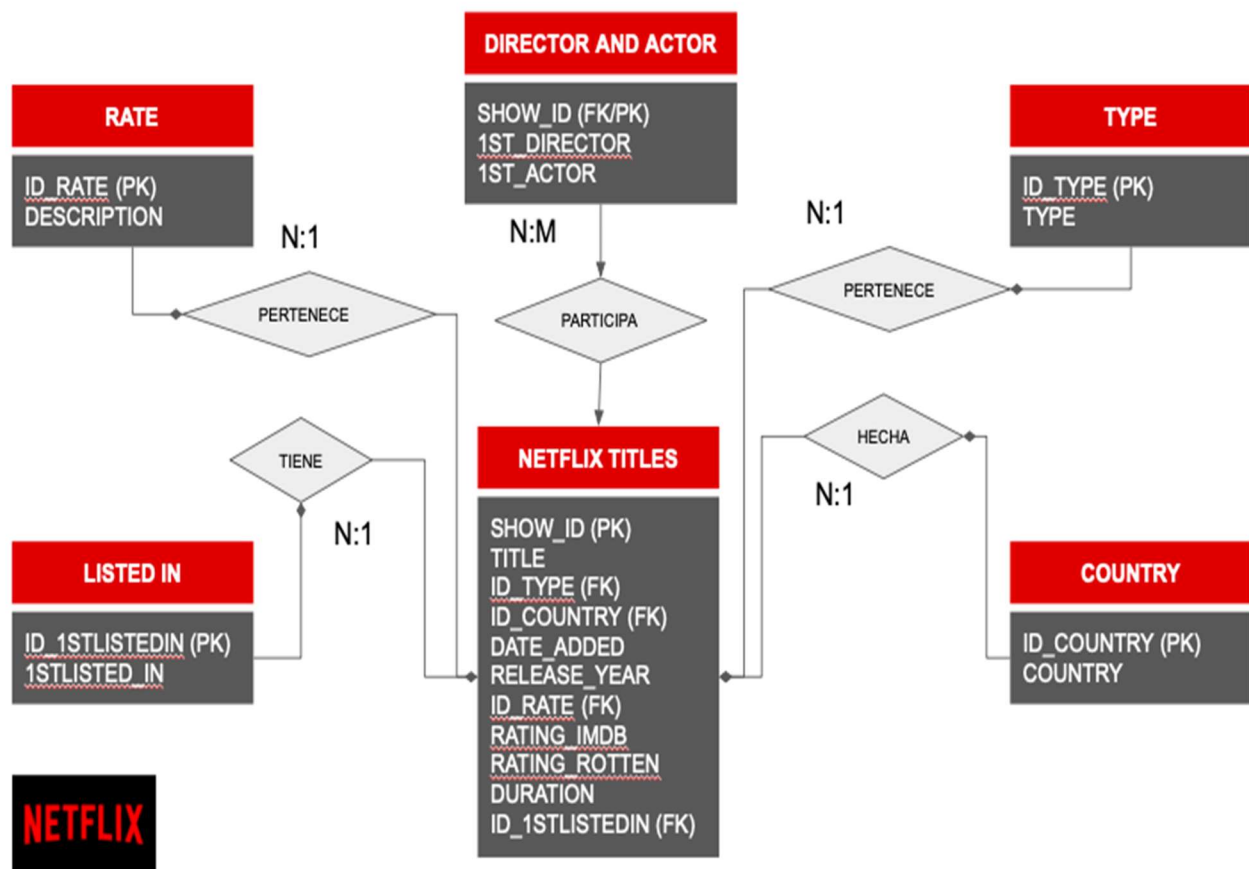
5 USUARIO FINAL Y NIVEL DE APLICACIÓN DE ANÁLISIS

El análisis de los datos, así como su visualización, está diseñado para **niveles tácticos**. Es decir, la información que se presenta es ideal para la toma de decisiones de los líderes del área encargada de la actualización del catálogo de manera anual.

Dicha información está estrechamente relacionada con el contenido del catálogo disponible en la plataforma, el tipo de título, su categoría y clasificación. También contiene el nivel de aceptación de los usuarios mediante dos evaluaciones.

6 DIAGRAMA ENTIDAD-RELACIÓN

Se elaboró un diagrama DER, el cuál se presenta a continuación:



7 LISTADO DE TABLAS

En este apartado, se hará mención de cada una de las tablas junto a una breve descripción de las mismas y la definición de la clave primaria y foránea:

7.1 Netflix Titles

Contiene el título de la película o serie, el ID de la misma, el rating IMDB, el rating Rotten, la fecha en la que fue producida y la fecha en que fue añadida al catálogo.

PK:SHOW_ID

FK:ID_COUNTRY

FK:ID_TYPE

FK:ID_RATE

FK:ID_1STLISTEDIN

7.2 Type

Contiene información acerca del título, es decir si es una película o una serie. Se relaciona con la tabla principal a través del ID_TYPE.

PK:ID_TYPE

7.3 Director and Actor

Tiene el nombre del director y actor relacionado con el showID.

PK:SHOW_ID

7.4 Country

Posee el país de producción de la obra y su respectivo ID.

PK:ID_COUNTRY

7.5 Rate

Esta tabla contiene una descripción sobre la clasificación de los títulos y quién es apto para verla.

PK:ID_RATE

7.6 Listed_in

Contiene el género de los títulos y el ID correspondiente.

PK:ID_1STLISTEDIN

8 LISTADO DE COLUMNAS POR TABLAS

Tabla: Netflix Titles

Tipo de Clave	Campo	Tipo de Campo
PK	show_id	varchar (5)
	title	varchar (255)
FK	id_type	varchar (3)
FK	id_country	varchar (3)
	date_added	date
	release_year	int
FK	id_rate	varchar (255)
	rating_IMDB	decimal
	rating_rotten	decimal
	duration	varchar (255)
FK	id_1stListedin	varchar (5)

Tabla: Type

Tipo de Clave	Campo	Tipo de Campo
PK	id_type	varchar (3)
	type	varchar (255)

Tabla: Director and Actor

Tipo de Clave	Campo	Tipo de Campo
PK/FK	show_id	varchar (5)
	1st_director	varchar (255)
	1st_actor	varchar (255)

Tabla: Country

Tipo de Clave	Campo	Tipo de Campo
PK	id_country	varchar (3)
	country	varchar (255)

Tabla: Rate

Tipo de Clave	Campo	Tipo de Campo
PK	id_rate	varchar (10)
	description	varchar (255)

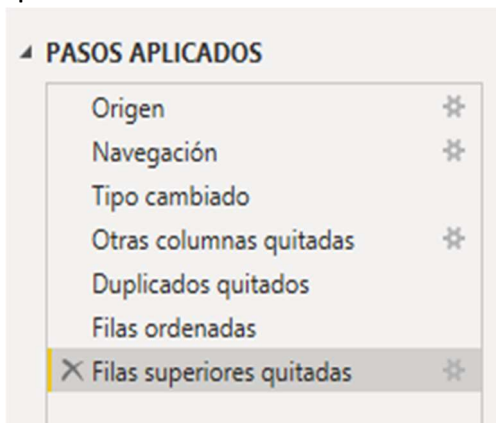
Tabla: Listed_in

Tipo de Clave	Campo	Tipo de Campo
PK	id_1stListedin	varchar (5)
	1stlisted_in	varchar (255)

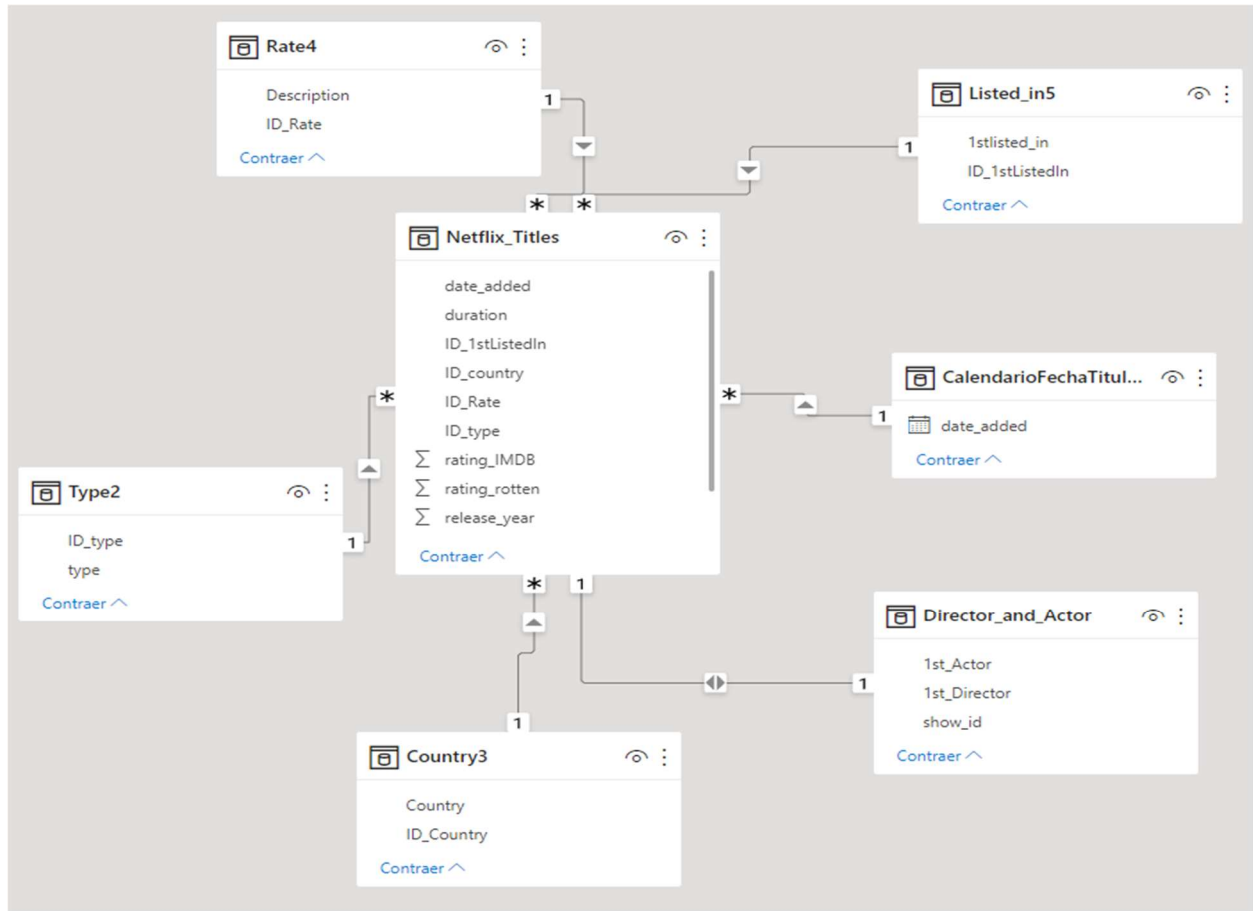
9 TRANSFORMACIÓN DE DATOS

Se duplicó la tabla de Títulos, para luego eliminar todas las columnas a excepción de “date_added” la cual tiene información acerca de cuándo fueron añadidos los títulos a la plataforma de Netflix.

Posteriormente se quitaron los duplicados de la columna; además fue necesario ordenar la columna de manera ascendente y remover la primera fila superior ya que tenía un dato “null”:



10 MODELO RELACIONAL



11 MEDIDAS CALCULADAS

Se realizaron una serie de medidas, en su mayoría relacionadas con las columnas Rating_IMDB Y Rating_Rotten. Los hallazgos fueron los siguientes:

11.1 Medida con una variable y una función de agregación

Busca el promedio del rating Rotten en el año 2021 y obtiene la diferencia de dicho rating con respecto al año pasado. El resultado final es de 0.03

AVG ROTTEN 2021 VS 2020 =

```
VAR                                AVG2021                                =
CALCULATE (AVERAGE (Show1[rating_rotten]), Show1[release_year]=2021)

RETURN                             AVG2021                                -
CALCULATE (AVERAGE (Show1[rating_rotten]), Show1[release_year]=2020)
```

11.2 Medida con una variable y una función de conteo:

Busca el promedio mensual de los títulos realizados durante el 2020. El resultado que se obtiene es de 79 títulos por mes durante dicho año.

AVG TítulosRealizadosPorMes2020 =

```
VAR TítulosRealizados2020=CALCULATE (COUNT (Show1[show_id]), FILTER (ALL (Show1[release_year]), Show1[release_year]=2020))

RETURN TítulosRealizados2020/12
```

11.3 Medida con dos variables (una función de agregación y una función de inteligencia de tiempo)

Esta medida es otra opción para buscar la diferencia del promedio mensual de los títulos realizados durante el periodo 2020 - 2021 (tal como se calculó en el primer ejemplo) pero esta vez utilizando dos variables y una función de inteligencia de tiempo. La diferencia obtenida es de 0.03

AVG2 ROTTEN 2021 VS 2020 =

```
VAR                                AVG2021                                =
CALCULATE (AVERAGE (Show1[rating_rotten]), Show1[release_year]=2021)

VAR AVG2020=CALCULATE (AVERAGE (Show1[rating_rotten]), DATEADD ('Calendar'[date_added], -1, YEAR))

RETURN AVG2021-AVG2020
```

11.4 Medida que contenga un parámetro (con una función de agregación)

En este caso, se está buscando una proyección para el 2022. Con un porcentaje que va del 1 al 100. En esta medida se muestra el promedio de ambos ratings (IMBD y Rotten) de todos los títulos del catálogo. El resultado obtenido es 2.85.

AVG RATING IMDB & ROTTEN =

```
VAR AVGIMDB = AVERAGEA (Show1[rating_IMDB])

VAR AVGROTTEN = AVERAGEA (Show1[rating_rotten])

VAR AVGRATING = (AVGIMDB + AVGROTTEN) / 2

RETURN AVGRATING
```

Si utilizamos un parámetro con esta medida podríamos observar cómo aumentaría la calificación el próximo año con base en un porcentaje, por ejemplo:

10% = El resultado obtenido es 3.13

25% = El resultado obtenido es 3.56

50% = El resultado obtenido es 4.27

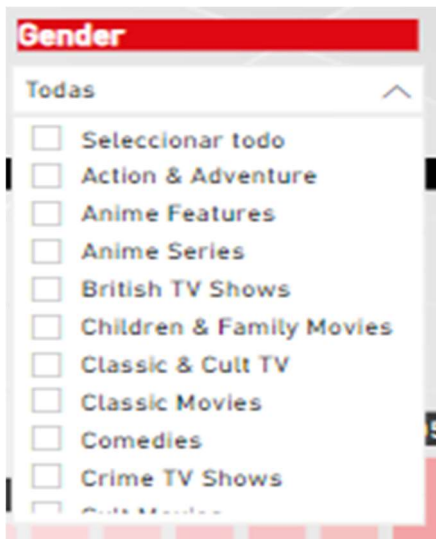
12 SEGMENTACIONES ELEGIDAS:

Para este proyecto se hizo uso de los siguientes recursos:

Listas desplegables

- Type: Filtra la información por tipo de título (TV shows/TV movies)
- Country: Muestra un listado de los países que conforman el catálogo de la plataforma
- Gender: Se puede seleccionar el género para filtrar información específica.
- Classification: Muestra las diferentes clasificaciones que se le dan a los títulos.

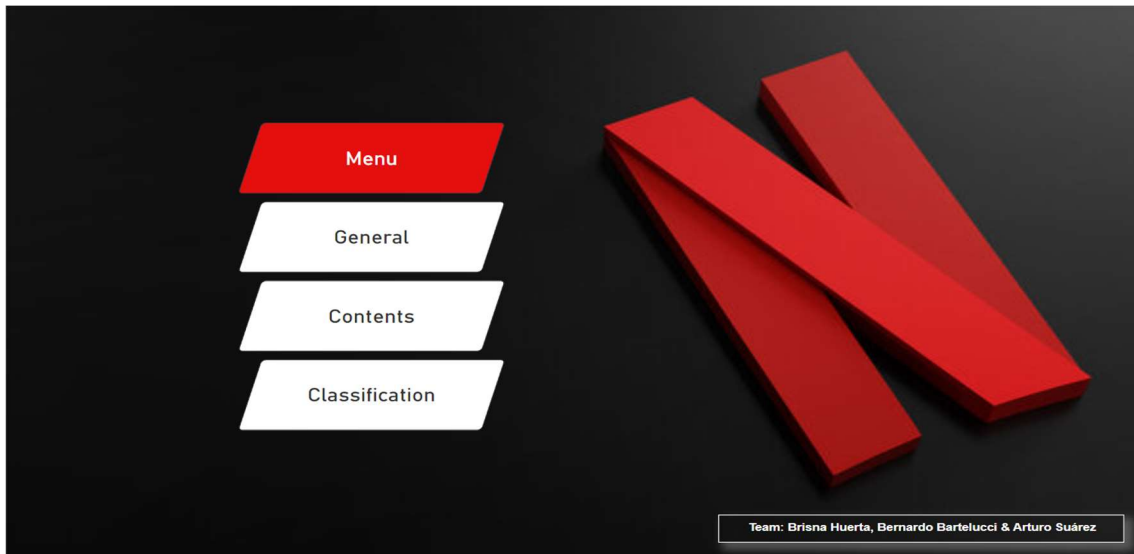
Este es un ejemplo de una lista desplegable:



13 ANÁLISIS FUNCIONAL DEL TABLERO

El diseño del tablero fue realizado pensando en que la información brindada será de utilidad y uso para líderes del área a nivel táctico. Este dashboard está compuesto de 4 solapas, las cuales se describen a continuación:

13.1 Menu

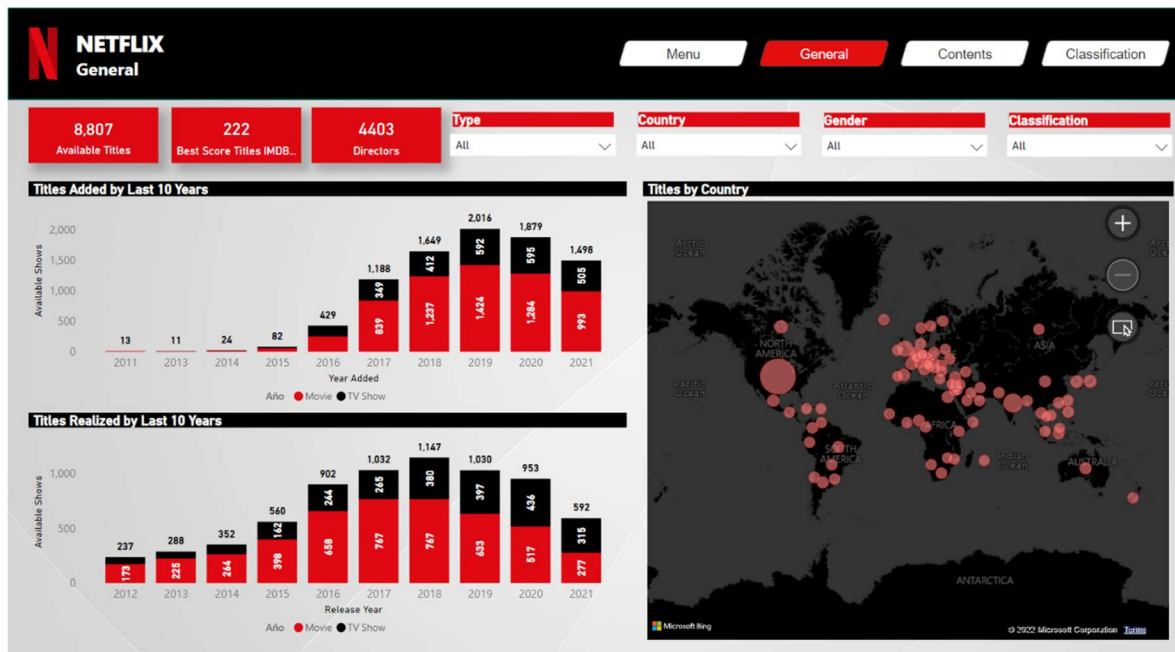


Esta solapa tiene como objetivo mostrar de manera general el contenido del dashboard. Los botones son útiles si se desea ir a alguna de las solapas sin orden; es decir, se puede ir directamente a la solapa “Clasificación” dando click en el botón con el mismo nombre.

En nuestro Mockup, la página de inicio fue diseñada de la siguiente manera:



13.2 GENERAL:



Esta solapa contiene gráficos con información general. Está distribuido de la siguiente manera:

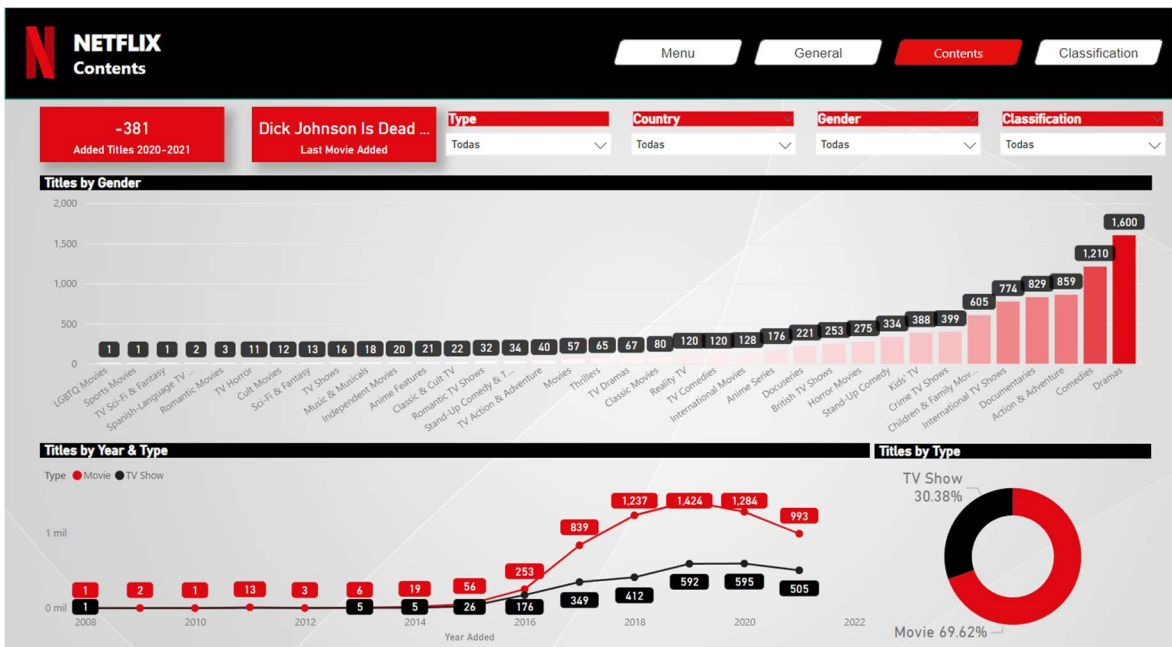
- En la parte superior Izquierda, se puede observar un KPI con el total de títulos disponibles en la plataforma.
- Siguiendo por la derecha, se encuentra un KPI que muestra el total de títulos con la máxima calificación tanto en Rotten Tomatoes como en IMDB.
- A la derecha, se encuentra un KPI con el conteo total de directores que participaron en los títulos del catálogo de NETFLIX.
- Posteriormente, a la derecha se observan cuatro filtros. Uno de ellos sirve para poder obtener datos por Tipo de Título (ya sea TV movies o TV shows). El siguiente es útil para conocer información del tablero de acuerdo con el país seleccionado. Además, hay otra segmentación por género del título y finalmente un por clasificación o rating
- En la parte izquierda del informe se pueden observar dos gráficos de columnas. En el superior se muestra la cantidad de títulos añadidos en el respectivo año.
- En el gráfico inferior izquierdo se muestra en qué año fueron producidos los títulos del catálogo/plataforma de Netflix

- Finalmente, en el sector derecho del informe se encuentra un mapa interactivo con un lazo para seleccionar regiones del mundo y cantidad de títulos producidos en la misma

En el Mockup esta solapa se diseñó de la siguiente manera:



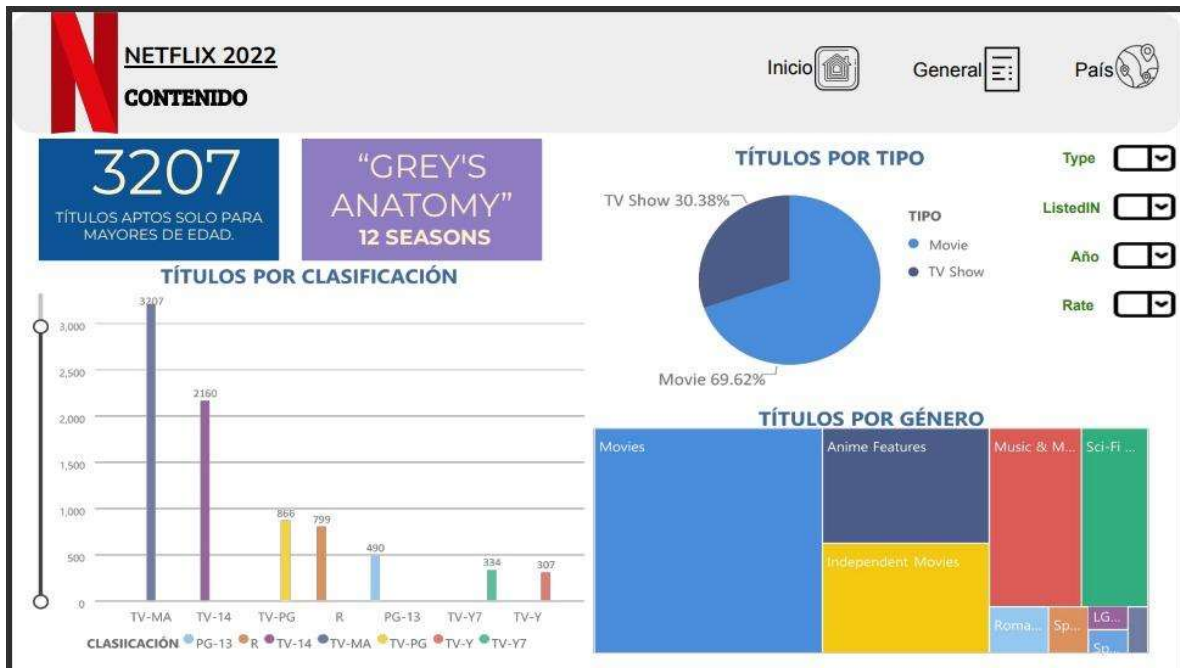
13.3 Contents



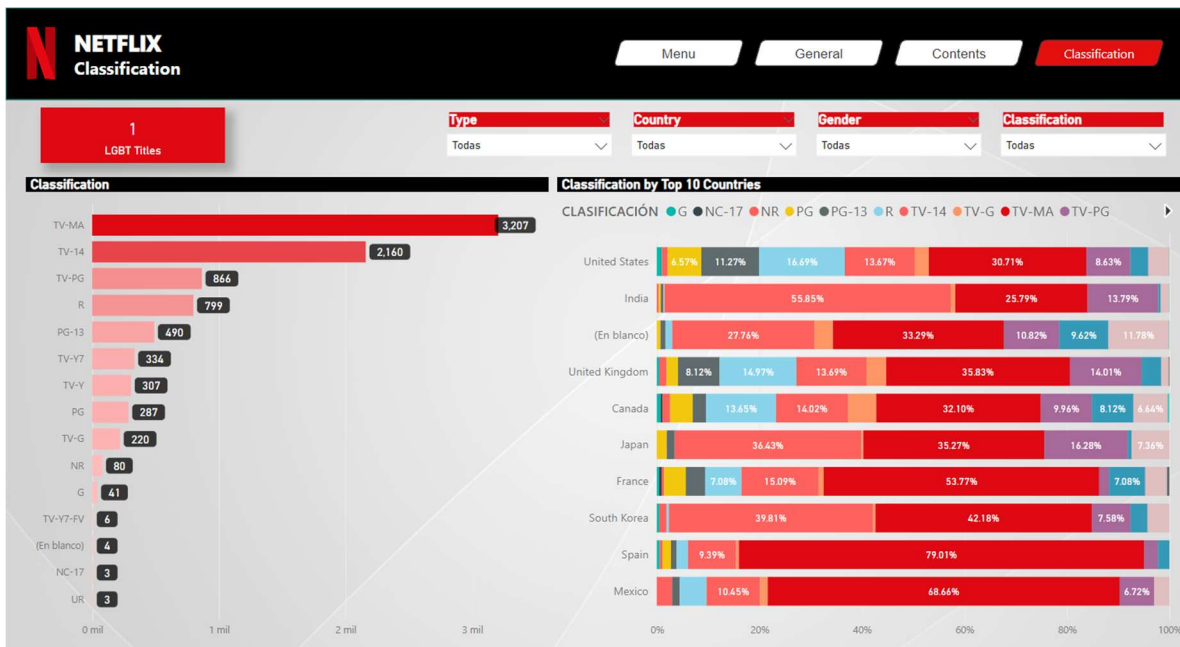
Esta solapa contiene información relevante del contenido del catálogo de NETFLIX. Dicha información está dividida en 3 gráficos ordenados de la siguiente manera:

- En la parte superior se puede observar un gráfico de barras que muestra la cantidad de títulos por géneros en la plataforma
- Debajo del gráfico de barras por género se encuentra un gráfico de líneas que muestra el crecimiento del catálogo de Netflix a través de los años tanto para películas como para series
- El ultimo grafico de esta solapa se encuentra en el extremo inferior derecho, el mismo es un gráfico circular que muestra el porcentaje de películas y TV shows presente en el servicio de streaming
- Al igual que en la solapa general esta tiene los mismos 4 filtros
- En el caso de las tarjetas, en esta página se encuentran dos, uno muestra el último título añadido a Netflix y el otro muestra la diferencia de obras audiovisuales añadidas en el 2021 respecto del 2020, en esta se observa que se añadieron 381 títulos menos en 2021 que en 2020

Esta solapa en el Mockup se veía de la siguiente manera:



13.4 CLASSIFICATION



La tercera solapa tiene como nombre "CLASSIFICATION". Inicialmente se había diseñado con el nombre de "PAÍS" pero se llegó a la conclusión de que, debido a su contenido, era más certero cambiarle el nombre.

La información de esta solapa se divide en:

- En el extremo superior derecho se encuentran los mismos 4 filtros que presentaban las solapas anteriores.
- En el otro extremo encontramos una tarjeta que muestra la cantidad de títulos LGBT en el catálogo.
- En la parte inferior derecha se puede ver un gráfico de barras 100% apiladas con el país y la cantidad de títulos que tiene por clasificación. Si te sitúas en alguno de los colores de cada barra también se podrá observar cómo se despliega un cuadro informativo que te dirá la descripción de dicha clasificación.
- Para concluir, el gráfico de la izquierda es un gráfico de barras lateral el cual indica la cantidad de títulos disponibles por clasificación

Inicialmente, esta solapa en el Mockup se veía así:



14 HERRAMIENTAS TECNOLÓGICAS UTILIZADAS

Para el ejercicio de este proyecto, se hizo uso de las siguientes tecnologías:



15 CONCLUSIÓN

Este dashboard deja ver a simple vista una tendencia de declive por parte de Netflix desde 2019, esta situación se puede atribuir principalmente a la pandemia mundial que provoco el Covid-19, la cual afectó al sector de entretenimiento limitando la generación de nuevo contenido no solo para Netflix si no para todos los generadores de contenido.

Adicional a esta situación, otras empresas han decidido generar su propia plataforma de streaming lo cual sin duda genera una mayor repartición de mercado considerando que cada una de ellas tiene a su favor contenido original creado durante años y con presupuesto para generar nuevo contenido, dando a Netflix la posibilidad de utilizar su catálogo solo a través de licenciamientos o bien negociaciones por exclusividad de franquicias ya posicionadas, siendo las películas el principal formato que ofrece Netflix en su catálogo.

Actualmente el contenido que muestra Netflix es muy variado (abarcando géneros desde dramas, comedias, acción y aventura como documentales, familiares e infantil, así como otros géneros más especializados como lo son animes, sci-fi, stand up, etc.) que en los últimos años ha tenido un auge cultural importante. Sin duda, esto es una gran ventaja que debe explotar ya que al ser pionera en streaming cuenta con información más precisa en cuanto a la segmentación de públicos y sus gustos.

Otro tema importante que debe considerarse en este dashboard es el uso de las clasificaciones. Mientras que en países como México o España se tiene mayor apertura y libertad de publicación de diferencias clasificaciones, países como Japón o Corea del Sur muestran poca permisibilidad a mostrar contenido considerado para un público adulto.

En conclusión, Netflix debe reposicionar su plataforma a través de un análisis más crítico a su portafolio, invertir en generar mayor contenido original que le permita competir y tener las exclusividades como lo hacen su similares y continuar ofreciendo un catálogo amplio como lo ha venido realizando en los últimos años.