
Análisis de Datos: Películas y Shows en Disney+

Proyecto Personal
Torres Leandro Joel
Gmail: leandrojtorres95@gmail.com
Marzo-2022

Introducción

Este proyecto es realizado con el fin analizar de forma descriptiva un dataset de Kaggle:

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

El fin es como portfolio, está destinado a un aspecto técnico en el ámbito de la extracción y limpieza de datos, pero no técnico para el informe final. Para ello utilizaremos de herramientas MySQL y Power BI

Primer vistazo

Si abrimos el archivo csv con un bloc de notas

```
disney_plus_titles: Bloc de notas
Archivo Edición Formato Ver Ayuda
show_id,type,title,director,cast,country,date_added,release_year,rating,duration,listed_in,description
s1,Movie,Duck the Halls: A Mickey Mouse Christmas Special,"Alonso Ramirez Ramos, Dave Nasson","Chris Diamantopoulos, Tony Anselmo, Tress MacNeille, Bill Farmer, Russi Taylor, Corey Burton
s2,Movie,Ernest Saves Christmas,John Cherry,"Jim Varney, Noelle Parker, Douglas Seale","November 26, 2021",1988,PG,91 min,Comedy,Santa Claus passes his magic bag to a new St. Nic.
s3,Movie,Ice Age: A Mammoth Christmas,Karen Disher,"Raymond Albert Romano, John Leguizamo, Denis Leary, Queen Latifah",United States,"November 26, 2021",2011,TV-G,23 min,"Animation, Comec
s4,Movie,The Queen Family Singalong,Hanish Hamilton,"Darren Criss, Adam Lambert, Derek Hough, Alexander Jean, Fall Out Boy, Jimmie Allen","November 26, 2021",2021,TV-PG,41 min,Musical,"I
s5,TV Show,The Beatles: Get Back,"John Lennon, Paul McCartney, George Harrison, Ringo Starr","November 25, 2021",2021,1 Season,"Docuseries, Historical, Music",A three-part documentary
s6,Movie,Becoming Cousteau,Liz Garbus,"Jacques Yves Cousteau, Vincent Cassel","November 24, 2021",2021,PG-13,94 min,"Biographical, Documentary",An inside look at the legends
s7,TV Show,Hawkeye,"Jeremy Renner, Hailee Steinfeld, Vera Farmiga, Fra Fee, Tony Dalton, Zahn McClarnon","November 24, 2021",2021,TV-14,1 Season,"Action-Adventure, Superhero",Clint Bart
s8,TV Show,Port Protection Alaska,"Gary Huellbarger, Mary Miller, Curly Leach, Sam Carlson, Stuart Andrews, David Squibb",United States,"November 24, 2021",2015,TV-14,2 Seasons,"Docuseri
s9,TV Show,Secrets of the Zoo:Tampa,"Dr. Ray Ball, Dr. Lauren Smith, Chris Massaro, Tiffany Burns, Mike Burns, Helinda Mendolusky",United States,"November 24, 2021",2019,TV-PG,2 Seasons
s10,Movie,A Muppets Christmas: Letters to Santa,Kirk R. Thatcher,"Steve Whitmire, Dave Goetz, Bill Barretta, Eric Jacobson",United States,"November 19, 2021",2008,G,45 min,"Comedy, Fami
s11,Movie,Adventure Thru the Walt Disney Archives,John Gleim,"Don Hahn, Kathryn Beaumont, Pete Docter, Kevin Feige, Mark Hamill, Bob Iger","November 19, 2021",2020,TV-PG,59 min,Documenta
s12,Movie,Puppy for Hanukkah,"November 19, 2021",2020,TV-G,4 min,"Comedy, Family, Music","Check out Daveed Dig's official music video ""Puppy for Hanukkah""""
s13,Movie,The Pixar Story,Lesslie Iwerks,"Stacy Keach, John Lasseter, Brad Bird, John Musker, Ron Clements, Frank Thomas",United States,"November 19, 2021",2007,G,91 min,"Documentary, Fami
s14,TV Show,"Dr. Oakley, Yukon Vet","Dr. Michelle Oakley, Zachary Fine",United States,"November 17, 2021",2013,TV-PG,10 Seasons,"Action-Adventure, Animals & Nature, Docuseries","Meet Dr.
s15,Movie,America the Beautiful,Michael B. Jordan,"November 12, 2021",2021,TV-PG,2 min,Animals & Nature,"Epic, grand portrait of the land across four regions on the most extreme contine
s16,Movie,Baymax!,"November 12, 2021",2021,TV-G,1 min,Animation,"Baymax!" premieres on Disney+ summer 2022.
s17,Movie,Ciao Alberto,McKenna Harris,"Jack Dylan Grazer, Marco Barricelli","November 12, 2021",2021,TV-PG,8 min,"Animation, Comedy, Family",Fun-loving sea monster Alberto whole-heartedl
s18,TV Show,Disney Fancy Nancy,"Mia Jenness, Alyson Hannigan, Rob Riggle, Spencer Moss, Dana Heath",United States, Canada,"November 12, 2021",2018,TV-PG,3 Seasons,"Animation, Kids",Nar
s19,TV Show,Disney Intertwined,"Carolina Domenech, José Jiménez Zapola, Kevsho, Elena Roger, Clara Alonso, Emilia Mernes","November 12, 2021",2021,TV-PG,1 Season,"Comedy, Coming of Age
s20,Movie,Enchanted,Kevin Lima,"Amy Adams, Patrick Dempsey, James Marsden, Timothy Spall, Idina Menzel, Susan Sarandon",United States,"November 12, 2021",2007,PG,110 min,"Comedy, Family,
s21,Movie,Faust,Patrick Osborne,"Stephen Apostolina, Kirk Baily, Ben Bledsoe, David Cogswill, Terri Douglas, Fuschia",United States,"November 12, 2021",2014,G,8 min,"Animation, Comedy, Dr
s22,Movie,Frozen Fever,"Chris Buck, Jennifer Lee","Kristen Bell, Idina Menzel, Jonathan Groff, Josh Gad, Santino Fontana, Chris Williams",United States,"November 12, 2021",2015,G,11 min,"
s23,Movie,Get a Horse!,Lauren MacMillan,"Walt Disney, Marcellite Garner, Russi Taylor, Billy Blitcher, Will Ryan, Bob Bergen",United States,"November 12, 2021",2013,G,7 min,"Action-Advent
s24,Movie,Home Sweet Home Alone,Dan Mazer,"Ellie Kemper, Rob Delaney, Archie Yates, Aisling Bea, Kenan Thompson, Tim Simons",United States,"November 12, 2021",2021,PG,95 min,"Comedy, Fami
s25,Movie,Jungle Cruise,Jaume Collet-Serra,"Dwayne Johnson, Emily Blunt, Edgar Ramirez, Jack Whitehall, Jesse Plemons, Paul Giamatti",United States,"November 12, 2021",2021,PG-13,123 min,
s26,Movie,Limitless with Chris Hemsworth,"Chris Hemsworth","November 12, 2021",2021,TV-14,2 min,Lifestyle,"""Limitless with Chris Hemsworth"" is coming to Disney+ in 2022."
s27,Movie,Marvel Studios' 2021 Disney+ Day Special,"November 12, 2021",2021,TV-PG,14 min,Movies,"Marvel Studios' Disney+ Day Special explores the Disney+ series of the MCU - past, pres
s28,Movie,Obi-wan Kenobi,"November 12, 2021",2021,TV-PG,1 min,"Action-Adventure, Science Fiction","A behind the scenes look of the ""Obi-wan Kenobi"" special event series. Coming to DI
s29,TV Show,Olaf Presents,"Josh Gad","November 12, 2021",2021,TV-PG,1 Season,"Animation, Family",Olaf goes from snowman to showman for his unique retelling of five favorite Disney Animati
s30,Movie,Paperman,John Kahrs,"John Kahrs, Karl Wahlgren",United States,"November 12, 2021",2012,G,7 min,"Animation, Comedy, Family",A young man strives to find girl of his dreams.
s31,Movie,Pixar 2021 Disney+ Day Special,"Pete Docter, Larry the Cable Guy, Jack Dylan Grazer, Domee Shi, Angus MacLane","November 12, 2021",2021,G,5 min,Movies,Join Pete Docter for a t
s32,Movie,Shang-Chi and the Legend of the Ten Rings,Destin Daniel Cretton,"Simu Liu, Awkwafina, Tony Leung, Michelle Yeoh, Peng'er Zhang, Ben Kingsley",United States, Australia,"Novembe
s33,Movie,Spin,Hanjari Makijany,"Avantika, Michael Bishop, Anna Cathcart, Jahbrill Cook, Kerri Medders, Aryan Simhadri","November 12, 2021",2021,TV-G,94 min,"Coming of Age, Music",Rhea di
s34,Movie,Tangled Ever After,"Nathan Greno, Byron Howard","Mandy Moore, Zachary Levi, Alan Dale, Paul F. Tompkins, Karl Wahlgren, Mark Allen Stewart",United States,"November 12, 2021",201
s35,Movie,The Ballad of Nessie,"Stevie Nermers-Skelton, Kevin Deters",Billy Connolly","November 12, 2021",2011,G,8 min,"Animation, Family, Fantasy",A colorful tale about the friendly Loch
s36,Movie,The Little Hatchgirl,Roger Allers,"United States,"November 12, 2021",2006,G,7 min,"Animation, Drama",A poor matchgirl sees happy images in the flames of her matches.
s37,Movie,The Making of Happier Than Ever: A Love Letter to Los Angeles,David Clair,"Billie Eilish, FINNEAS, Gustavo Dudamel, Patrick Osborne, Robert Rodriguez","November 12, 2021",2021,
s38,Movie,The Proud Family: Louder and Prouder,"Kvla Pratt, Tommy Davidson, Paula Jai Parker, JoMarie Payton, Cedric the Entertainer"...November 12, 2021",2021,TV-G,1 min,"Animation. Com
```

En la primera fila podemos ver los títulos de cada columna que usaremos y también podemos ver que están separados por comas y un “enter” (“/n”) entre filas, es recomendable en lo posible dar un pequeño vistazo a nuestros datasets para ver su estructura y guiarnos en la carga de los datos

En caso de MySQL la palabra “title” y “description” son palabras reservadas para funciones por lo que dentro del bloc de notas cambiaremos estos títulos con
title=ttitle
description=ddescription

MySQL

Carga y transformación de datos

Lo próximo a hacer es crear un tabla en MySQL ya que lo utilizaremos para limpiar la data y realizar transformaciones, ya que si lo cargamos directamente en Power BI y realizamos estás acciones allí no sería lo mejor para un rendimiento óptimo.

Lo próximo a hacer para cargar el dataset, necesitamos crear una nueva tabla en nuestra base de datos, para ello vamos a crear una nueva base de datos (“no hace falta”), y dentro la tabla. Utilizaremos una tabla.

Entonces en el siguiente script creamos la base de datos “disneyt” y la tabla temporal “dataset” con las variables declaradas con el titulo que notamos en el bloc de notas.

```

CREATE DATABASE disneyt;
USE disney;

CREATE TEMPORARY TABLE dataset (
show_id VARCHAR(20)
, ttype VARCHAR (20)
, title VARCHAR(200)
, director VARCHAR(200)
, cast VARCHAR(200)
, country VARCHAR(100)
, date_added VARCHAR(100)
, release_year VARCHAR(20)
, rating VARCHAR(50)
, duration VARCHAR(50)
, listed_in VARCHAR(100)
, ddescription VARCHAR(200)
);

```

El tipo de variable utilizamos todos en VARCHAR porque luego lo cambiaremos a entero o float según lo requiramos, además si por ejemplo declarando un entero puede ser que entre uno de los miles de registros no posea un entero o esté vacío por lo que ocurrirá un error.

Luego cargamos los datos, en este paso puede ocurrir algún error de acuerdo alguna medida de seguridad de la versión de MySQL del usuario, por lo que quedará para el lector resolver, la siguiente página puede servir de ayuda, también hay que tener cuidado con usar "/" o "\"

<https://computingforgeeks.com/how-to-solve-mysql-server-is-running-with-the-secure-file-priv-error/>

```

#CARGA DE DATOS
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/dis-
ney_plus_titles.csv'
INTO TABLE dataset
FIELDS TERMINATED BY ',' ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 LINES;

#OBSERVAMOS LOS DATOS
SELECT * FROM dataset;

```

Por otro lado creamos nuestras tablas dimensiones para luego cargarlas directamente a Power BI para ser más eficientes así no usar Power Query, ellas son la principal, la fecha de agregado, la dirección y la descripción de las películas o shows

```

#---TABLA DIMENSION FECHA_DE_INGRESO ( PODEMOS AGREGAR MAS FILAS COMO
CUATRIMESTRE O SEMESTRE QUEDA A CRITERIO)----

# (pd: iniciar el contador para crear la id de la tabla)
SET @contador=0;
CREATE TABLE tb_fecha_de_agregado AS
WITH CALENDARIO AS(SELECT show_id ,date_added,SUBSTRING_INDEX(date_ad-
ded," ",1) AS MES
, SUBSTRING_INDEX(SUBSTRING_INDEX(date_added,"",1),"",-1) AS DIA
, SUBSTRING_INDEX(date_added,"",-1) AS ANIO FROM dataset
)
SELECT DISTINCT (SELECT @contador:=@contador+1) AS id_fecha_de_agre-
gado,
date_added AS fecha_completa, CAL.dia, CAL.mes, SUBSTRING(CAL.MES, 1,
3) AS mes_corto
, CASE
WHEN MES="January" THEN 1
WHEN MES="February" THEN 2
WHEN MES="March" THEN 3
WHEN MES="April" THEN 4
WHEN MES="May" THEN 5
WHEN MES="June" THEN 6
WHEN MES="July" THEN 7
WHEN MES="August" THEN 8
WHEN MES="September" THEN 9
WHEN MES="October" THEN 10
WHEN MES="November" THEN 11
WHEN MES="December" THEN 12
END num_mes, CAL.anio
FROM CALENDARIO AS CAL;

# este id contiene el valor nulo por eso lo borramos
DELETE FROM tb_fecha_de_agregado WHERE id_fecha_De_agregado=168;

#VISUALIZAMOS LA TABLA CREADA
SELECT * FROM tb_fecha_de_agregado;

#--- TABLA DIMENSION DIRECCION---
SET @contador=0;
CREATE TABLE tb_direccion AS
SELECT DISTINCT (SELECT @contador:=@contador+1) AS id_direccion ,di-
rector,cast FROM dataset
WHERE (director <> "NULL" AND cast <> "NULL");

```

```
SELECT * FROM tb_direccion;
```

```
#---TABLA DIMENSION DESCRIPCION---
```

```
SET @contador=0;
```

```
CREATE TABLE tb_descripcion AS
```

```
SELECT DISTINCT (SELECT @contador:=@contador+1) AS id_descripcion, ddescription AS descripcion FROM dataset
```

```
WHERE ddescription <> "NULL";
```

```
SELECT * FROM tb_descripcion;
```

```
#---TABLA PRINCIPAL---
```

```
CREATE TABLE tb_disney_principal AS
```

```
SELECT SUBSTRING_INDEX(A.show_id,"s",-1) AS id, A.ttype AS tipo, A.title AS titulo, B.id_direccion AS id_direccion,
```

```
A.country AS pais, C.id_fecha_de_agregado AS id_fecha_de_agregado,
```

```
A.release_year AS fecha_de_lanzamiento, A.rating AS edad_clasificacion,
```

```
A.duration AS duracion, A.listed_in AS genero,
```

```
D.id_descripcion AS id_descripcion
```

```
FROM dataset A
```

```
LEFT JOIN tb_direccion B ON (A.cast=B.cast AND A.director=B.director)
```

```
LEFT JOIN tb_fecha_de_agregado C ON A.date_added=C.fecha_completa
```

```
LEFT JOIN tb_descripcion D ON A.ddescription=D.descripcion;
```

Notamos que tenemos valores nulos, los dejamos en blanco ya que Power BI los detecta de esa manera, entonces transformamos, también alteramos el tipo de dato para lograrlo en caso de que sean enteros.

```
ALTER TABLE tb_disney_principal MODIFY id_direccion VARCHAR(20);
```

```
UPDATE tb_disney_principal
```

```
SET id_direccion=""
```

```
WHERE id_direccion IS NULL;
```

```
UPDATE tb_disney_principal
```

```
SET pais=""
```

```
WHERE pais IS NULL;
```

```
ALTER TABLE tb_disney_principal MODIFY id_fecha_de_agregado VARCHAR(20);
```

```
UPDATE tb_disney_principal
```

```
SET id_fecha_de_agregado=""
```

```
WHERE id_fecha_de_agregado IS NULL;
```

```
ALTER TABLE tb_disney_principal MODIFY fecha_de_lanzamiento VARCHAR(20);
```

```
UPDATE tb_disney_principal
```

```
SET fecha_de_lanzamiento=""
```

```
WHERE fecha_de_lanzamiento IS NULL;
```

```
UPDATE tb_disney_principal
```

```
SET edad_clasificacion=""
```

```
WHERE edad_clasificacion IS NULL;
```

```
UPDATE tb_disney_principal
```

```
SET duracion=""
```

```

WHERE duracion IS NULL;

UPDATE tb_disney_principal
SET genero=""
WHERE genero IS NULL;

ALTER TABLE tb_disney_principal MODIFY id_descripcion VARCHAR(20);
UPDATE tb_disney_principal
SET id_descripcion=""
WHERE id_descripcion IS NULL;

UPDATE tb_direccion
SET director=""
WHERE director IS NULL;

UPDATE tb_direccion
SET cast=""
WHERE cast IS NULL;

SELECT * FROM tb_disney_principal;
SELECT * FROM tb_direccion;

```

Terminando, creamos las vistas y las vemos

```

#CREAMOS LAS VISTAS
CREATE VIEW v_disney_principal AS
SELECT * FROM tb_disney_principal;

CREATE VIEW v_disney_tb_descripcion AS
SELECT * FROM tb_descripcion;

CREATE VIEW v_disney_tb_direccion AS
SELECT * FROM tb_direccion;

CREATE VIEW v_disney_fecha_de_agregado AS
SELECT * FROM tb_fecha_de_agregado;

# vemos los views que creamos
SELECT * FROM v_disney_principal;
SELECT * FROM v_disney_tb_descripcion;
SELECT * FROM v_disney_tb_direccion;
SELECT * FROM v_disney_fecha_de_agregado;

```

Por último paso en MySQL vamos a exportar los datos, con la opción que nos brinda luego de realizar las consultas de los datos de nuestras visualizaciones. Entonces exportamos con el mismo nombre de las vistas en formato CSV con separación de “;”, ya que con “,” puede confundir con datos de un mismo registro, por ejemplo “November 26, 2021” podría separar en “November 26” y “2021”

```

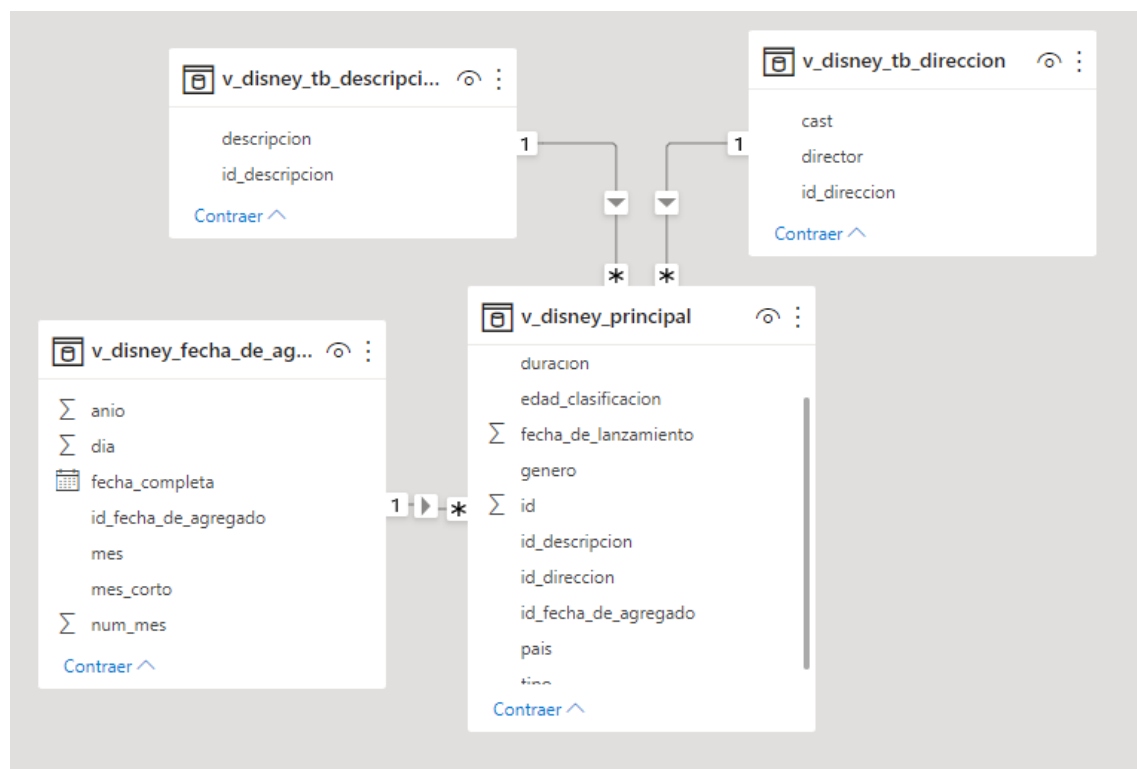
# vemos los views que creamos
SELECT * FROM v_disney_principal;
SELECT * FROM v_disney_tb_descripcion;
SELECT * FROM v_disney_tb_direccion;
SELECT * FROM v_disney_fecha_de_agregado;

```

id_fecha_de_agregado	fecha_completa	mes_corto	num_mes	anio
1	November 26, 2021	26	November	Nov
2	November 25, 2021	25	November	Nov
3	November 24, 2021	24	November	Nov
4	November 19, 2021	19	November	Nov
5	November 17, 2021	17	November	Nov

Power BI

Cargamos los CSV en el programa y luego realizamos el modelo relacional, entre “ids” de las tablas hacia la principal (puede que PBI ya haya realizado las relaciones automáticamente), quedaría de esta forma



Las relaciones deberán tener la siguiente configuración, con cardinalidad de “varios a uno” y dirección de filtro cruzado “en ambas”

Editar relación

Permite seleccionar tablas y columnas relacionadas.

v_disney_principal

id_agregado	fecha_de_lanzamiento	edad_clasificacion	duracion	genero	id_descripcion
166	1998	G	98 min	Animation, Comedy, Family	713
166	1946	TV-G	7 min	Animation, Comedy, Family	716
166	2018	TV-G	7 min	Animation, Comedy, Family	734

v_disney_tb_descripcion

id_descripcion	descripcion
1	Join Mickey and the gang as they duck the halls!
2	Santa Claus passes his magic bag to a new St. Nic.
3	Sid the Sloth is on Santa's naughty list.

Cardinalidad

Varios a uno (*:1)

☒ Activar esta relación

☐ Asumir integridad referencial

Dirección del filtro cruzado

Ambas

Única

Ambas

Aceptar

Cancelar

Para el informe utilice un fondo azul desde google imágenes, no es necesario usar el mismo o tener un fondo, es más un aspecto estético. En el diseño de la solapa podemos configurar ello.

Por otro lado, creamos una tabla nueva para almacenar todas las medidas que calculemos.

Solapas

Portada

Es la primera solapa, por ende la presentación del dashboard, colocaremos el título del mismo, el autor ("nosotros") y el mes de elaboración, también un botón para dirigir al usuario a la siguiente solapa.

Análisis de Películas y Shows

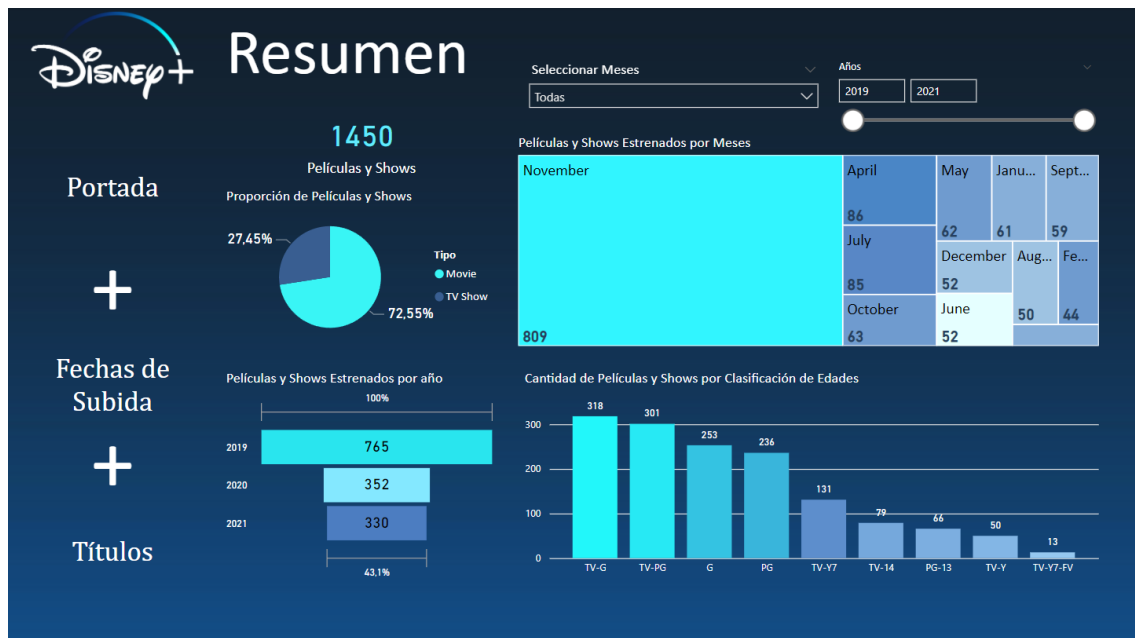


Resumen + Fechas de Subida + Títulos



Resumen

Un pequeño resumen de los datos que tenemos, como la cantidad de películas y shows, la proporción, la distinta distribución de fechas de subidas en plataforma. La misma tiene 2 tooltips, uno de la cantidad de shows y películas, y otro con un glosario en la clasificación de edades.



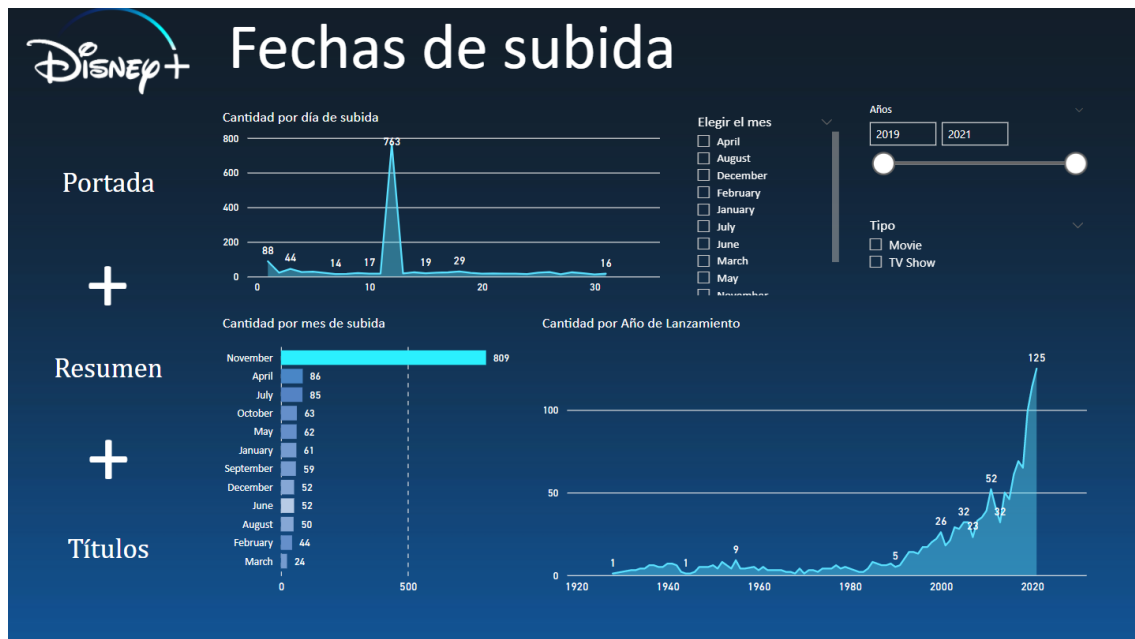
El glosario se obtuvo con datos de las siguientes páginas webs:

[https://es.wikipedia.org/wiki/Clasificaci%C3%B3n_por_edades_\(TV_Parental_Guidelines\)](https://es.wikipedia.org/wiki/Clasificaci%C3%B3n_por_edades_(TV_Parental_Guidelines))

https://es.wikipedia.org/wiki/Sistema_de_calificaciones_de_la_Motion_Picture_Association

Fechas de subida

En esta solapa está diseñada con la intención de observar cuando la película o shows fue subido en plataforma y cuando fue su fecha de lanzamiento



Títulos

Por último mostramos todos los títulos que tenemos, con sus características como director, cast, una breve descripción entre otros datos. Esta solapa también tiene un tooltip para un glosario en la clasificación por edad

The dashboard for 'Títulos' (Titles) on Disney+ features a dark blue background with the Disney+ logo in the top left. On the left, a vertical sidebar contains navigation links: 'Portada' (Home), '+', 'Resumen' (Summary), '+', and 'Fechas de Subida' (Upload Dates). The main content area is divided into several sections. The top section, 'Título' (Title), lists titles with checkboxes: '10 Things I Hate About You', '101 Dalmatian Street', '101 Dalmatians', '101 Dalmatians (Series)', '101 Dalmatians II: Patch's London Adventure', '102 Dalmatians', and '12 Dates of Christmas'. The middle section, 'Tipo' (Type), has a checkbox for 'Movie'. The bottom section, 'Cast', lists the cast members: Barry Bostwick, Jason Alexander, Martin Short, Bobby Lockwood, Susanne Blakeslee, and Sam... The right section, 'Descripción' (Description), shows a description for 'Pongo and Perdita's pup Patch gets the chance to meet his TV hero.' Below this are the 'Fecha de Lanzamiento' (Release Date) as 2003 and 'Fecha de Subida en Disney+' (Upload Date on Disney+) as November 2019. The bottom of the dashboard features a row of information: 'Jim Kammerud, Brian Smith' as the Director, 'Action-Adventure, Animation' as the Genre, 'G' as the Age Rating, and '77 min' as the Duration.

Conclusiones

Para finalizar cabe aclarar que los procesos de ETL (“extracción, transformación y carga”) de datos es de vital importancia para poder crear con más facilidad nuestro dashboard, también lo importante de usar SQL para lograrlo y una vez terminado llegar así al proceso final de elaborar el informe en PoweBI. También del producto final, podemos notar que en 2019 se subieron por mucho más la cantidad de contenido, que la plataforma sube mas contenido en noviembre, se suben más contenidos para un público adulto (“TV-G y TV-PG”) , que el día 12 es cuando mayormente se sube contenido y que el contenido procede de un lanzamiento en el presente siglo, entre más conclusiones.