

# **Catálogo Netflix**

## **Análisis de resultados IMDb**

Melisa Caputo  
Coder House  
25/05/2022

# Tabla de contenidos

- 1) Introducción y objetivo del análisis
- 2) Descripción de la temática y alcance
- 3) Herramientas tecnológicas utilizadas
- 4) Diagrama entidad relación
- 5) Detalle de las tablas
- 6) Transformación de los datos
- 7) Medidas calculadas
- 8) Detalle de las solapas
- 9) Diagrama de relaciones en Power BI

# Introducción

La temática seleccionada para el análisis es el catálogo histórico de Netflix.

En los últimos años, el consumo de contenido de entretenimiento en línea creció exponencialmente, y uno de los principales actores de este mercado es Netflix, la plataforma de streaming más popular del mundo. Netflix ofrece una amplia gama de contenido que incluye series de televisión, películas y documentales, y se convirtió en la opción de entretenimiento preferida para millones de personas en todo el mundo.

El objetivo de este análisis es explicar las causas que hacen que una película tenga mayor cantidad de votos y mejor puntaje IMDb, considerando diferentes variables como por ejemplo su género, clasificación, director, y año de lanzamiento. Se evaluarán las posibles relaciones entre las variables buscando responder a la incógnita planteada.

# Dataset

Los datos utilizados para el análisis fueron contruidos por un estudiante originario de India, Syed Mubarak, quien publicó en Kaggle una base de datos que contiene 9425 registros de películas y series disponibles en la plataforma Netflix. La base fue publicada en agosto de 2021 por lo que es una de las más recientes. [Puede consultarse en el sitio de Kaggle](#).

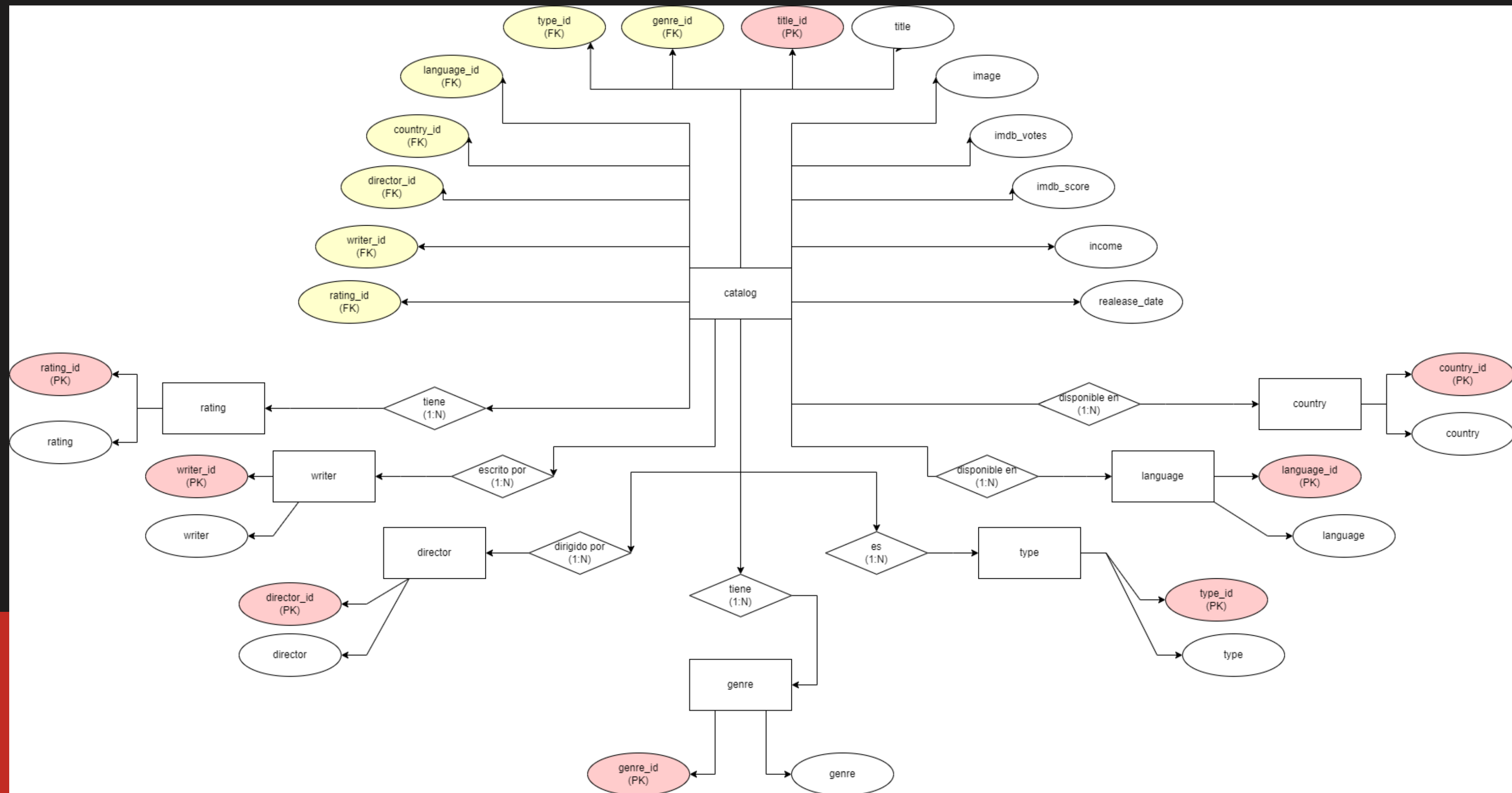
El dataset parte de un conjunto de datos que ofrece 29 variables de las que se eliminaron los siguientes campos porque no se consideraron relevantes para el análisis: tags, hidden gem score, runtime, actors, rotten tomatoes score, metacritic score, awards received, awards nominated for, Netflix release date, production house, netflix link, IMDb link, summary, poster, TMDb trailer, trailer site.

El dataset final se compone de las siguientes variables: title, genre, language, type (series or movie), country, director, writer, view rating, IMDb score, IMDb votes, income, release date, image.

# Herramientas utilizadas

- Hojas de calculo de google para la importación del dataset, limpieza de datos y normalización de los datos en tablas.
- Canva para la confección de la presente documentación.
- Power BI para la creación del tablero, Power Query para la transformación de datos.
- [app.diagrams.net](https://app.diagrams.net) para la creación del diagrama entidad relación.

# Diagrama entidad relación



# Detalle de las tablas

CAMPO	TIPO DE DATO	CLAVE
title	VARCHAR(MAX)	
title_id	INT	PK
genre_id	INT	FK
type_id	INT	FK
language_id	INT	FK
country_id	INT	FK
director_id	INT	FK
writer_id	INT	FK
rating_id	INT	FK
imdb_votes	INT	
imdb_score	DECIMAL(3,1)	
income	INT	
release_date	DATE	

## Tabla catalog

Es la tabla de hecho, en esta tabla se muestra la información general del título (serie o película) y todas las claves foraneas que permiten relacionarla con las otras tablas.

# Detalle de las tablas

CAMPO	TIPO DE DATO	CLAVE
genre_id	INT	PK
genre	VARCHAR(MAX)	

## Tabla genre

En esta tabla se muestran los géneros de los títulos junto con sus ids.

CAMPO	TIPO DE DATO	CLAVE
type_id	INT	PK
type	VARCHAR(6)	

## Tabla type

En esta tabla se muestran los posibles tipos de títulos (serie o película) junto con sus ids.



# Detalle de las tablas

CAMPO	TIPO DE DATO	CLAVE
language_od	INT	PK
language	VARCHAR(MAX)	

## Tabla language

En esta tabla se muestran los idiomas posible que puede tener un título junto con sus ids.

CAMPO	TIPO DE DATO	CLAVE
country_id	INT	PK
country	VARCHAR(MAX)	

## Tabla country

En esta tabla se muestran los países en los que se encuentran disponibles los títulos junto con sus ids.

# Detalle de las tablas

CAMPO	TIPO DE DATO	CLAVE
director_id	INT	PK
director	VARCHAR(MAX)	

## Tabla director

En esta tabla se muestran los directores junto con sus ids.

CAMPO	TIPO DE DATO	CLAVE
escritor_id	INT	PK
escritor	VARCHAR(MAX)	

## Tabla writer

En esta tabla se muestran los escritores junto con sus ids.

# Detalle de las tablas

CAMPO	TIPO DE DATO	CLAVE
rating_id	INT	PK
rating	VARCHAR(20)	

## Tabla rating

En esta tabla se muestran las posibles clasificaciones de audiencia con sus ids.

# Transformación de datos

- Se agregó la columna personalizada `release_year` a partir de los datos de la columna `release_date` y se cambió el tipo de dato a número entero para poder armar un filtro dinámico por año.
- Se eliminó el registro con título “Fiction.” que hacía referencia a la película “Pulp Fiction” porque estaba repetido. También se quitaron los duplicados en base a la columna `title` para evitar otros errores.
- Se corrigieron los links de las imágenes que figuran en el mosaico de películas, algunos estaban rotos.
- Como las series no contaban con datos de recaudación, use la función rellenar para completar los registros de forma ficticia.
- Se creó la tabla calendario duplicando la tabla `catalog` y dejando únicamente la columna `release_date`, luego se quitaron valores nulos y se relaciono con la tabla principal.
- Se creó la tabla "Ultima actualización" que usa la función `DateTime.LocalNow()`.
- Transforme los registros que tenían muchos valores delimitados por comas de forma que únicamente quede visible el primer valor para facilitar el análisis. Por ejemplo, de “United States, Canada, Japan, Australia, South Korea” pasa a “United states”. Los pasos aplicados fueron los siguientes: dividir columna por delimitador, eliminar todas las columnas menos el primer valor y el `id`, quitar duplicados. Tuve que dejar los valores repetidos para no volver a empezar de cero con todo el tablero.

# Medidas calculadas

- Cantidad de títulos = COUNTA('catalog'[country\_id]) > contabiliza la cantidad de títulos por país para poder aplicar el número en el mapa. Se usa en la solapa "global".
- Cantidad votos IMDb = SUM('catalog'[imdb\_votes]) > suma los votos IMDb, permite mostrar una tarjeta con el indicador al seleccionar alguno de los objetos. Se usa en las solapas "películas" y "series".
- Promedio votos IMDb = AVERAGE('catalog'[imdb\_votes]) > promedia la cantidad de votos IMDb, permite mostrar una tarjeta con el indicador al seleccionar alguno de los objetos. Se usa en la solapa "género".
- Promedio puntaje IMDb = AVERAGE('catalog'[imdb\_score]) > promedia el puntaje de votos IMDb, permite mostrar una tarjeta con el indicador al seleccionar alguno de los objetos. Se usa en la solapa "género".
- Total Recaudación = SUM('catalog'[income]) > muestra la recaudación total según la variable que se seleccione, por ejemplo género. Se usa en las solapas "películas" y "géneros".
- Última actualización = LASTDATE('calendar'[release\_date]) > muestra la fecha del último registro ingresado en el dataset. Se usa en la portada.

# Detalle de las solapas

## Películas

Se pueden visualizar las 10 películas más votadas así como las 10 mejor puntuadas junto con su cantidad de votos y puntaje IMDb. Además, se muestran las 100 películas más taquilleras y los 100 directores más populares, pudiendo filtrar las películas por director. Toda la solapa se puede filtrar por año (2012 - 2021).

## Series

Se pueden visualizar las 10 series más votadas así como las 10 mejor puntuadas junto con su cantidad de votos y puntaje IMDb. Además, se muestran la cantidad de estrenos por mes y la división por rating de audiencia, pudiendo filtrar los meses por rating. Toda la solapa se puede filtrar por año (2012 - 2021).

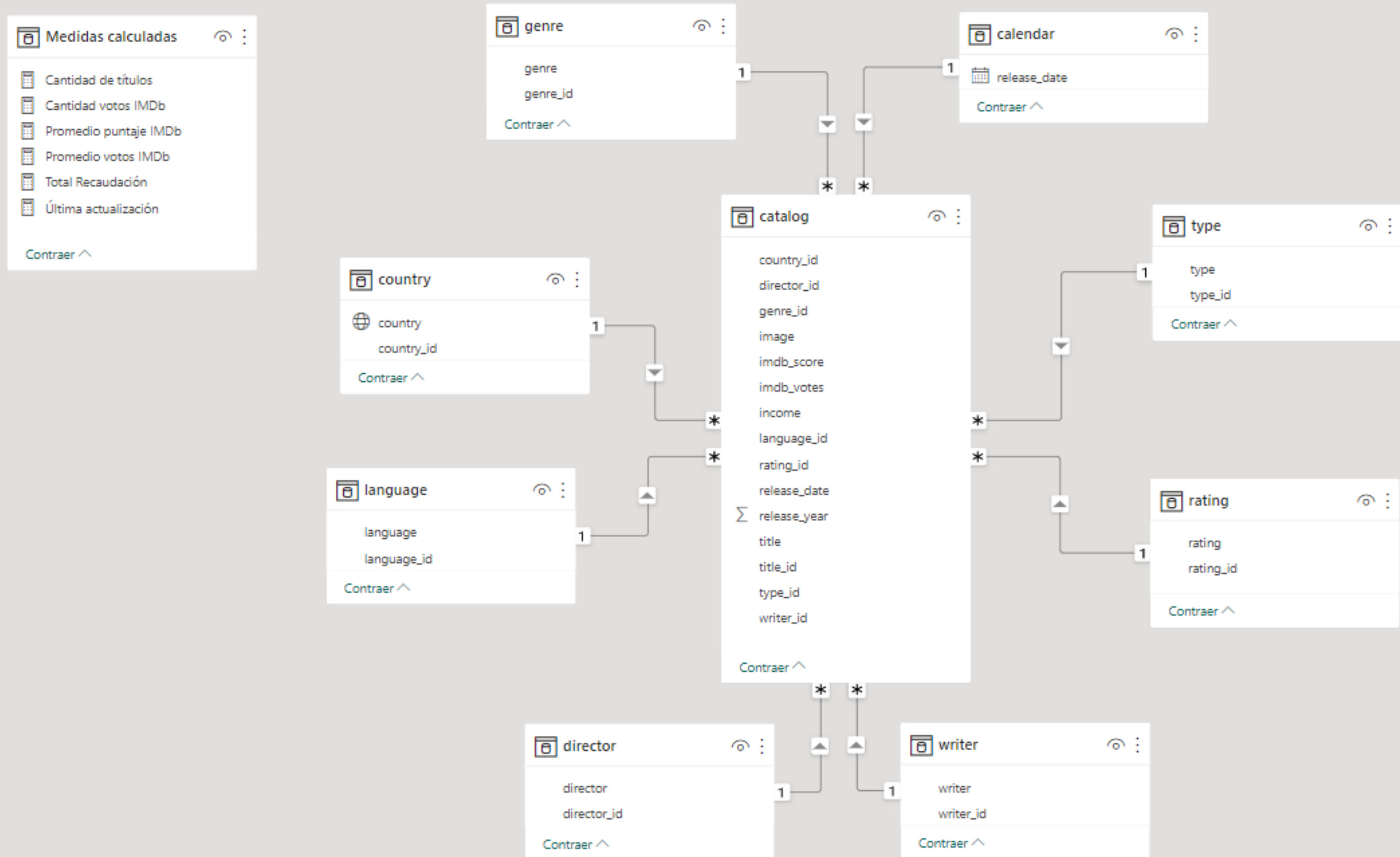
## Género

El foco de la solapa está en la cantidad de títulos por género, pudiendo seleccionar cada uno de ellos a modo de filtro. También se cuenta con otros filtros principales: categoría, rating, director, escritor y año de lanzamiento. En la parte inferior se visualizan el promedio de puntaje IMDb (mostrando el menor y mayor puntaje a partir de los filtros aplicados), el promedio de votos IMDb y la recaudación por año.

## Global

El foco de la solapa está en la cantidad de títulos por país, pudiendo seleccionar cada uno de ellos a modo de filtro. También se cuenta con otros filtros principales: categoría, rating, director, escritor y año de lanzamiento. Además, se visualiza una tabla con la división por idioma a partir de los filtros aplicados.

# Diagrama entidad relación - Power BI



**¡Muchas gracias!**