

Tercera entrega del proyecto final

Análisis de los programas emitidos

CURSO: DATA ANALITYCS FLEX
COMISION 53470
ALUMNO: Pedro Jofre
PROFESOR: Agustin Escofet
TUTOR: Enzo Pavez

Descripción Temática:

Los programas de televisión han sido una parte integral de la cultura popular durante décadas, y su popularidad sigue siendo un área de interés constante tanto para la audiencia como para los profesionales de la industria del entretenimiento. Con el advenimiento de plataformas de streaming y la proliferación de contenido en línea, la competencia por la atención del espectador ha aumentado significativamente, desde el año 2018 según lo reportado por Aaron Smith y Monica Anderson en estudio de [pewresearch](#).

El consumo de servicios de streaming y contenido en línea ha experimentado un notable aumento en los últimos años, particularmente durante períodos de pandemia, donde se ha observado un incremento significativo en el tiempo dedicado al consumo de medios digitales y contenido en streaming durante los periodos de confinamiento según reportado por [Nielsen](#). Este cambio en el comportamiento del consumidor ha generado un entorno altamente competitivo para los creadores y distribuidores de contenido.

En este contexto, comprender los factores que contribuyen a la popularidad de los programas de televisión se ha vuelto más importante que nunca. Los análisis tradicionales basados en datos de audiencia y ratings han sido complementados por el uso de datos digitales, como los recopilados por TMDb, que proporcionan una visión más completa y detallada de las preferencias del público y los patrones de consumo de contenido televisivo.

Objetivo

Analizar los datos para identificar patrones y tendencias de los distintos programas frente a una encuesta de popularidad, para entender las características en común que comparten el top 100 de estas y los periodos de emisión.

Alcance

Generar una visualización interactiva que permita entregar de manera comprensible y fácil los resultados obtenidos y que permita generar nuevas conclusiones con respecto a la popularidad de los show

Hipótesis

La popularidad de los shows está relacionada al género, tipo de show y época de emisión

Después del estudio se podrá responder las siguientes preguntas

1. Cuales son los show mas populares de los últimos años?
2. Como a variado la preferencias de los usuarios a través del tiempo?
3. Que combinación de genero es efectiva?

Usuario final y nivel de aplicación:

La información recolectada y procesada puede ser de gran valor para distintas áreas y grupos, desde consumidores de streaming quienes pueden analizar para encontrar nuevos show o saber

cuales son los show del momento. Empresas de streaming les puede ser deseable el insights del análisis para optimizar su catálogo de contenido, identificar tendencias de visualización y mejorar la recomendación de contenido para sus usuarios. El área de marketing puede utilizarlo identificar programas populares y audiencias específicas a las que dirigirse con sus campañas publicitarias.

Por esto es que el nivel de aplicación puede darse en diversos niveles:

Estratégico: El análisis de popularidad puede dar una visión estratégica sobre expansión a nuevos mercados a agregar show que no se tenían presente con anterioridad. Mientras que los paneles informativos servirían para ver el rendimiento de los géneros y tipos de show que tienen actualmente.

Táctico: los insights pueden ayudar a generar campañas de marketing y/o promocionar distintos programas o servicios.

Operativo: a nivel operativo este análisis puede ayudar a un gerente de programación de televisión para determinar la duración óptima de los programas en función de la popularidad

Listado de tablas

La base de datos utilizada se obtuvo de:

<https://www.kaggle.com/datasets/denizbilginn/tv-shows/data?select=status.csv>

A continuación se detalla las tablas del dataset utilizado, incluyendo columnas, descripción y tipo de datos, además de incluir la llaves PK (Primary Key) y FK (Foreign Key)

1. Air_date: Contiene la fecha en que se emitió por primera vez cada show y cuando se emitió por última vez.

Nombre columna	descripción	Tipo de dato	PK	FK
Is_first	Indica si la fecha es de inicio o fin	INT	No	no
Show_id	Indica el id del show	INT	No	Si
date	Indica la fecha	date	No	No
Indice	Indice	INT	Si	No

2. Calendario:

3. Genre_types: Indica los tipos de genero que puede tener un show

Nombre columna	descripción	Tipo de dato	PK	FK
Is_genre_type_id	Indica id del genero	INT	Si	no
Genre_name	Indica el nombre del genero	VARCHAR	No	no

4. Genres: tabla relacional entre los show y la tabla de genero

Nombre columna	descripción	Tipo de dato	PK	FK
Show_id	Indica id del show	INT	No	Si
Genre_type_id	Indica el id del genero	INT	No	Si
Indice	Indice	INT	Si	No

5. Language_types: Tabla de idiomas

Nombre columna	descripción	Tipo de dato	PK	FK
Language_type_id	Indica id de lenguaje	INT	Si	No
Language_name	Nombre del lenguaje	VARCHAR	No	No

6. Languages: tabla relacional entre los show y la tabla de idiomas

Nombre columna	descripción	Tipo de dato	PK	FK
Show_id	Indica id del show	INT	No	Si
Language_type_id	Indica id de lenguaje	INT	No	Si
Index	Indice agregado	INT	Si	No

7. Origin_country_types: tabla de país de origen

Nombre columna	descripción	Tipo de dato	PK	FK
Origin_country_type_id	Indica id del pais	INT	Si	No
Origin_country_name	Indica código de pais	VARCHAR	No	No

8. Production_countries: tabla relacional entre show con países de producción y país de origen

Nombre columna	descripción	Tipo de dato	PK	FK
Show_id	Indica id del show	INT	No	Si
Production_country_type_id	Indica código de país de producción	INT	No	Si
Origin_country_type_id	Indica id del país de origen	INT	No	Si
Index	Índice agregado	INT	Si	No

9. Production_country_types: tabla de país de producción

Nombre columna	descripción	Tipo de dato	PK	FK
Production_country_type_id	Indica id de país de producción	INT	Si	No
Production_country_name	Indica nombre de país de producción	VARCHAR	No	No

10. Show: Tabla de los show donde se incluye diversa información, nombre, numero de temporadas y episodios, popularidad, entre otros

Nombre columna	descripción	Tipo de dato	PK	FK
Show_id	Indica el id del show	INT	Si	No
Name	Indica el nombre del show	VARCHAR	No	No
Number_of_seasons	Numero de temporadas	INT	No	No
Number_of_episodes	Numero de episodios	INT	No	No
overview	Sinopsis del show	VARCHAR	No	No
adult	Indica si el show es para adultos	BOOLEAN	No	No
Original_name	Nombre original del show	VARCHAR	No	No
Popularity	Indica la popularidad del show según tmdb	INT	No	No
Type_id	Indica el id del tipo de show	INT	No	Si
Status_id	Indica el id del status del show	INT	No	Si
popularity_places	Indica el lugar según popularidad	INT	No	No

11. Show_votes: esta tabla indica la información relacionada a los votos obtenidos por los show

Nombre columna	descripción	Tipo de dato	PK	FK
vote_count	Cantidad de votos obtenido por el show	INT	No	no
Showvote_average	Promedio de los votos del show	FLOAT	No	No
Show_id	Indica el id del show	INT	No	Si
Indice	Indice Agregado	INT	Si	No

12. Status: tabla que indica los distintos status que puede tener un show

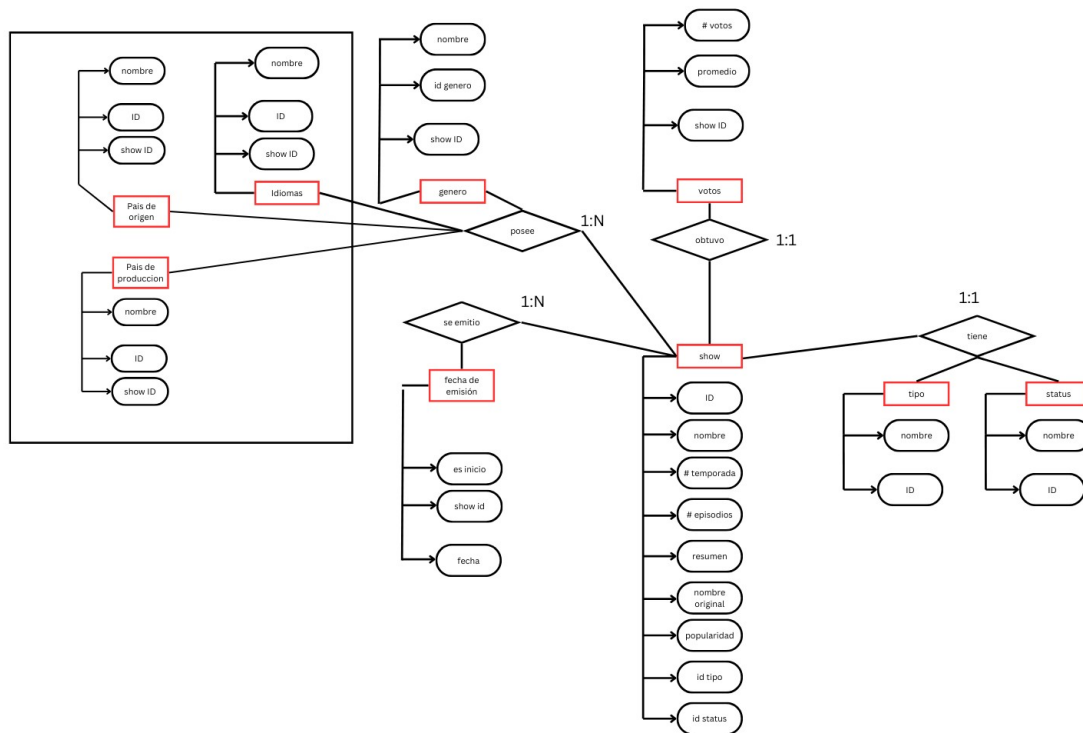
Nombre columna	descripción	Tipo de dato	PK	FK
Status_id	Id del status	INT	Si	No
status_name	Nombre del status	VARCHAR	No	No

13. Types: indica el tipo de show

Nombre columna	descripción	Tipo de dato	PK	FK
type_id	Id del tipo	INT	Si	No
Type_name	Nombre del tipo	VARCHAR	No	No

Diagrama Entidad-relación

A continuación, se detalla el diagrama entidad relación del proyecto, el recuadro en negro muestra nueva información que se agrega al modelo:



Transformación Dataset

A Continuación se detalla la transformaciones realizada a la base de datos para poder ser trabajadas en power bi.

Tabla "Calendario"

A partir de la tabla "air_dates" se crea la tabla auxiliar de calendario a través de los siguientes pasos:

1. Se busca la fecha mas antigua en la tabla "air_dates" en el campo "date".
2. Se crea una fila por cada día entre la fecha y el día de hoy.
3. Se agregan las siguientes columnas, numero de día, numero de mes, numero de la semana, trimestre, nombre del día y nombre del mes

Date	Años	Mes	Día	Semana	Mes Nombre	día semana	Día nombre	Trimestre
09-01-1917	1917	1	9	2	Enero	3	Martes	1
10-01-1917	1917	1	10	2	Enero	4	Miercoles	1
11-01-1917	1917	1	11	2	Enero	5	Jueves	1
12-01-1917	1917	1	12	2	Enero	6	Viernes	1
13-01-1917	1917	1	13	2	Enero	7	Sabado	1
14-01-1917	1917	1	14	3	Enero	1	Domingo	1
15-01-1917	1917	1	15	3	Enero	2	Lunes	1

Tabla “Show”

Se debe hacer una limpieza de la tabla de show:

1. Se eliminan las filas que tienen alguna columna con error o null
2. Se ordena las filas por nombre y popularidad.
3. Se crea una nueva columna índice (popularity_places) en la cual se le asigna un lugar basado en la popularidad que posee, en el caso de tener un empate se asigna un mejor lugar por orden alfabético del nombre del show.
4. Se agrega la columna “Cantidad de generos”, la cual entrega cuanta veces el show id se repite en la tabla “genres”, un valor null indica que la información de genero no se ha completado por los usuarios.

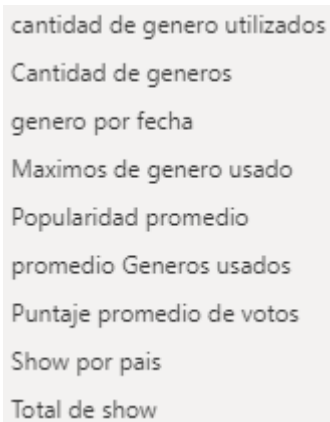
Cambios generales

A continuación, se detallan los cambios realizados en varias tablas:

1. Se compara los “show_id” de las diversas tablas con el de “show” y se elimina todo aquellos que no estén en la tabla show.
2. Se agrega una columna de índice en las tablas: air_dates, Genres, Languages, Production_countries, Show_votes.

Medidas

Para un mejor análisis de la tabla se crea la tabla MEDIDAS donde se guardan las medidas creadas, en su mayoría corresponden a conteos, promedios y conteos distintos de valores de las diferentes tablas.



A vertical list of nine measures to be stored in the MEDIDAS table. The list is contained within a light gray rectangular box with rounded corners. The text is in a dark gray, sans-serif font.

- cantidad de genero utilizados
- Cantidad de generos
- genero por fecha
- Maximos de genero usado
- Popularidad promedio
- promedio Generos usados
- Puntaje promedio de votos
- Show por pais
- Total de show

Visualización

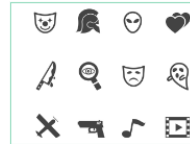
El reporte se ordena en 4 solapas, 1 de portada y 3 de información, a continuación se detalla brevemente cada una de estas paginas.

Portada

Se crea la portada con la siguiente información, datos del curso, logo de la pagina donde se obtuvo la base y los botones de navegación a las distintas paginas



Resumen



Genero

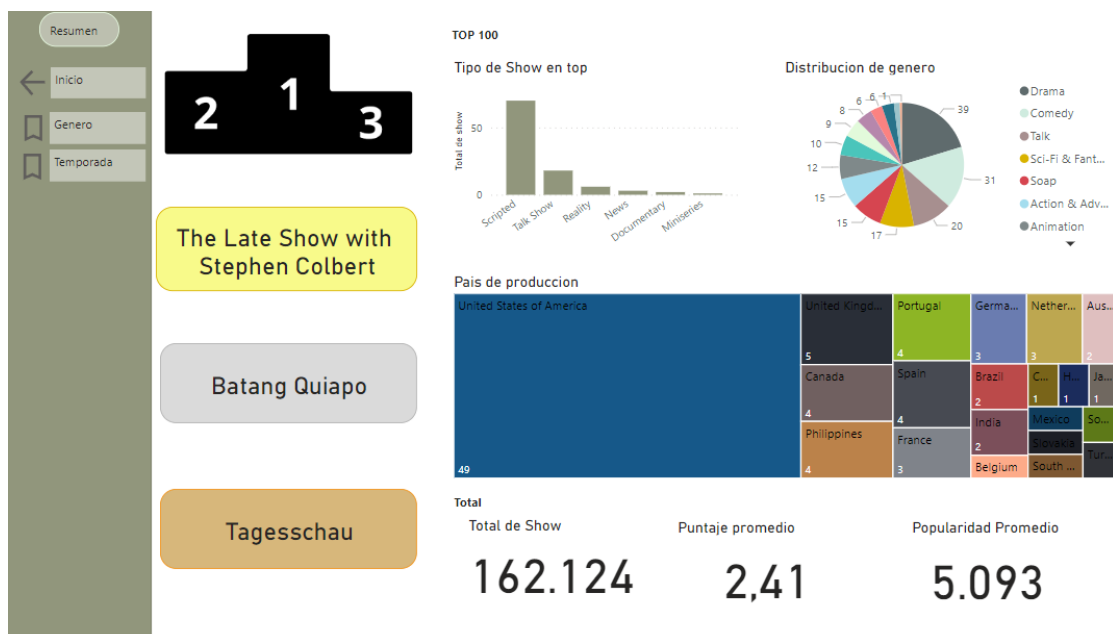


Temporada

CURSO: DATA ANALYTICS FLEX
COMISION 53470
ALUMNO: Pedro Jofre
PROFESOR: Agustin Escofet
TUTOR: Enzo Pavez

Resumen

La primera portada sirve para ver información del top 3 histórico, resumen de los top 100 show y por ultimo un resumen de la cantidad promedio en popularidad y puntaje de todos los show creados. También se agrega botones de navegación para moverse entre cualquiera de las pestañas, dado que es una pestaña de resumen general de la tabla, no se agregan filtros.



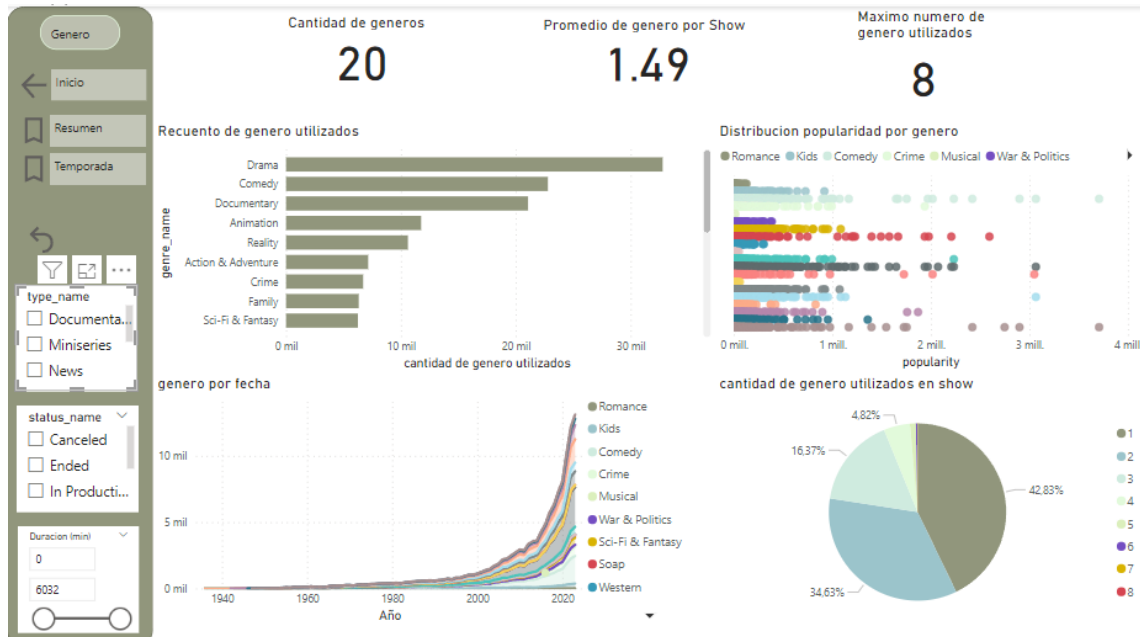
Se hicieron los siguientes ajustes:

- Se agrega el nombre del primer, segundo y tercer lugar según el atributo popularity_places de la tabla shows.
- Se agrega un grafico de barra resumiendo os tipos de show utilizados en el top 100.

- Se agrega un grafico circular indicando el recuento de los géneros utilizados en el top 100.
- Se agrega un grafico de Treemap mostrando los países de distribución del top 100.
- Se agrega una tarjeta con el conteo de los show existente en la tabla de shows.
- Se agrega una tarjeta con el puntaje promedio obtenido en votos.
- Se agrega una tarjeta con la popularidad promedio de los show.

Genero

En esta pestaña se muestra información relacionada a los géneros utilizados en los show, así como botones de navegación y se agregaron segmentaciones pertinentes al estudio.

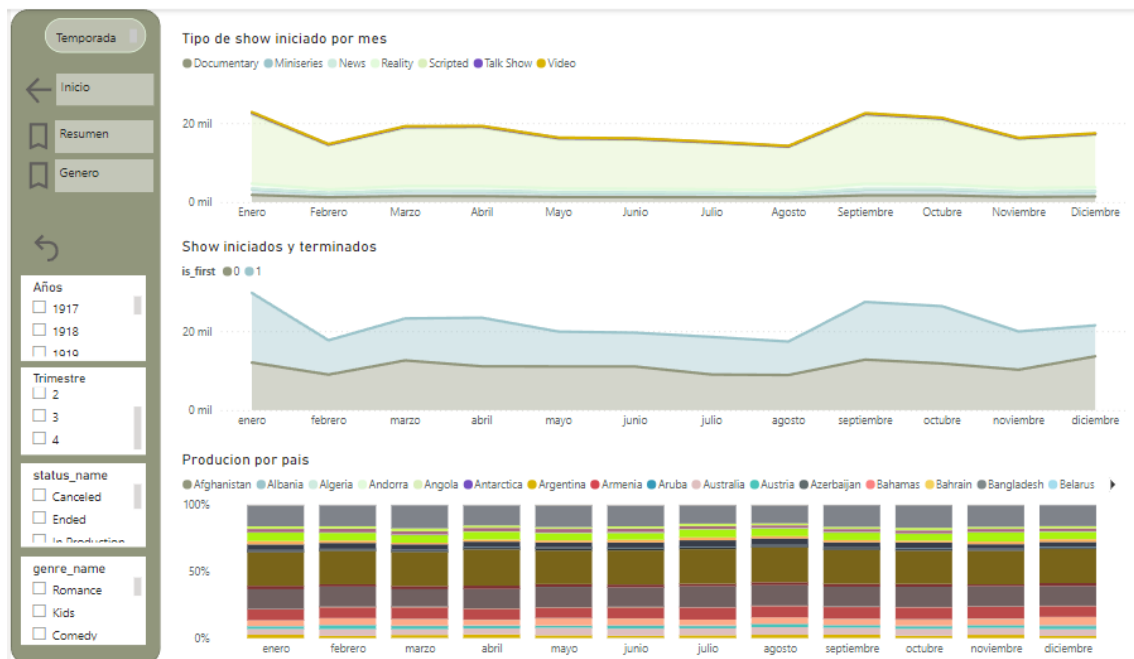


Se hicieron los siguientes ajustes:

- Se agrega una tarjeta con el numero distintivo de géneros.
- Se agrega una tarjeta con el promedio de géneros utilizado por show.
- Se agrega una tarjeta con el máximo de géneros utilizado por un show.
- Se agrega un grafico de barra mostrando la cantidad de veces que se usa cada género.
- Se agrega un gráfico de dispersión de popularidad y género.
- Se agrega un grafico de línea mostrando cuantos show se creaban cada año
- Se agrega un grafico circular que muestra la distribución de cuantos genero se utilizaba por show.
- Se agrega los filtros de tipo show.
- Se agrega los filtros de tipo status.
- Se agrega un filtro de intervalo para duración del show.
- Se agrega botón para borrar los filtros.
- Se agrega los botones de navegación

Temporada

Por ultimo se agrega una pestaña para mostrar los cambios durante el año, así como botones de navegación.



Se hicieron los siguientes ajustes:

- Se agrega un grafico de línea mostrando cuantos show y tipos de show se creaban cada mes.
- Se agrega un grafico de línea mostrando cuantos show iniciaron y terminaron por mes.
- Se agrega un grafico de barra porcentual que muestra la distribución de los países productores cada mes.
- Se agrega los filtros de año.
- Se agrega un filtro trimestre.
- Se agrega un filtro genero.
- Se agrega un filtro status.
- Se agrega botón para borrar los filtros.
- Se agrega los botones de navegación

Futuras líneas

Basado en el análisis de datos presentado en este informe sobre los shows, se identificaron varias áreas de enfoque y posibles acciones para mejorar la popularidad de los show a estrenar:

- 1- **Genero y estilo populares:** dentro del análisis se tiene que dentro de los 100 show mas populares históricamente, la mitad corresponde a tipo scripted, el 40% poseen el genero drama y 31% usan el genero de comedia, lo cual indica que estas combinaciones generan un impacto duradero en los espectadores.
- 2- **Estilo de producción:** se tiene que la mitad de los 100 show mas populares fueron producidos en estados unidos, lo cual hace que sea un lugar de estudio para entender que hacen distinto al resto de los países.
- 3- **Tendencia de producción:** se encuentra que en los últimos ha habido un incremento acelerado de los show creados, específicamente en los genero de drama y comedia,

por lo que si la idea es producir algo que impacte a la gente y diferenciarse de la mayoría de show es escogiendo dentro de los géneros menos usados como historia y romance.

- 4- Producción anual: se descubre que la mayoría de los inicios y finales de los show corresponde con el cambio de estación.