

PROYECTO FINAL



SILVINA HERRERA
SIMÓN ESCOBAR GALENDE
SEBASTIÁN RIZZUTO

COMISIÓN 18550

Profesora: Rebeca Figueroa Soriano

Tutor: Julian David Fernández

Año:2021

Prólogo

En el presente proyecto, desarrollado para el curso de Data Analytics dictado por Coder House, se podrá observar el desarrollo y puesta en escena de un Dashboard que proveerá métricas relacionadas a la industria cinematográfica, otorgando una mejor perspectiva y comprensión de las demandas en el sector.

Para la realización del proyecto se utilizó un dataset; una base de datos para el mismo, con sus respectivas tablas y diagrama, que permitan observar el modelo relacional, siendo SQL la tecnología seleccionada para el desarrollo del mismo; y por último, Power BI como servicio de análisis de datos, con el objeto de proporcionar visualizaciones interactivas y capacidades de inteligencia empresarial con una interfaz lo suficientemente simple como para que los usuarios finales puedan crear por sí mismos sus propios informes y paneles.

Índice

Página 1: Carátula introductoria

Página 2: Prólogo e índice

Página 3: Descripción de la temática de datos

Página 4: Listado de Tablas

Página 6: Diagrama entidad-relación

Página 7: Glosario

Paso a Power BI

Página 9: Medidas calculadas

Página 10: Análisis funcional del tablero

Página 22: Medidas calculadas avanzadas

Página 23 : Futuros pasos

Página 24: Correcciones y ajustes

DESCRIPCIÓN DE LA TEMÁTICA DE DATOS

El contenido del dataset detalla registros de popularidad sobre más de 4800 películas y su correspondiente metadata; creado con la intención de predecir el éxito de futuras producciones. El período comprendido será el año 2017.

Se buscará informar sobre:

- Relación popularidad y género: ¿Hay alguna correlación entre el género de una película y su rating?
- Relación popularidad y productora: ¿Hay alguna correlación entre la compañía que produce una película y su rating?
- Relación presupuesto y ganancias: ¿Llevara un mayor presupuesto a un aumento de ganancias?
- Palabras claves más buscadas: ¿Tienen las compañías ciertos temas en los que se enfoquen?
- Puntuación: ¿Qué aspectos afectan mayormente a esta medida?

De esta manera se otorgará la información necesaria para poder distinguir apropiadamente los pros y contras que provee el desarrollo de cada género, así como la popularidad, presupuesto y ganancias que promedian los mismos.

A continuación se observan las tablas utilizadas, datatypes y su correspondiente diagrama de entidad-relación para una comprensión más profunda del proyecto. También se provee un glosario de los campos utilizados.

Listado de tablas

Tabla 1: Películas populares

Tipo de clave	Campo Tipo de campo
---------------	---------------------

PK - INDEX movie_id int

- title varchar

-	budget int
-	revenue int
-	release_date datetime
-	vote_average decimal
-	vote_count int
-	original_language varchar

Tabla 2: Género

Tipo de clave	Campo Tipo de campo
PK - INDEX	genre_id int
-	genre text

Tabla 3: Compañía de produccion

Tipo de clave	Campo Tipo de campo
PK - INDEX	production_company_id int
-	production_company varchar

Tabla 4: Keywords

Tipo de clave	Campo Tipo de campo
---------------	---------------------

PK - INDEX id_keyword int
- keyword varchar

Tabla 5: Relación id_pelicula y géneros

Tipo de clave	Campo Tipo de campo
PK - INDEX	relaciongen int
FK	movie_id int
FK	genre_id int

Tabla 6: Relación id_pelicula y keywords

Tipo de clave	Campo Tipo de campo
PK - INDEX	relacionkey int
FK	movie_id int
FK	keyword_id int

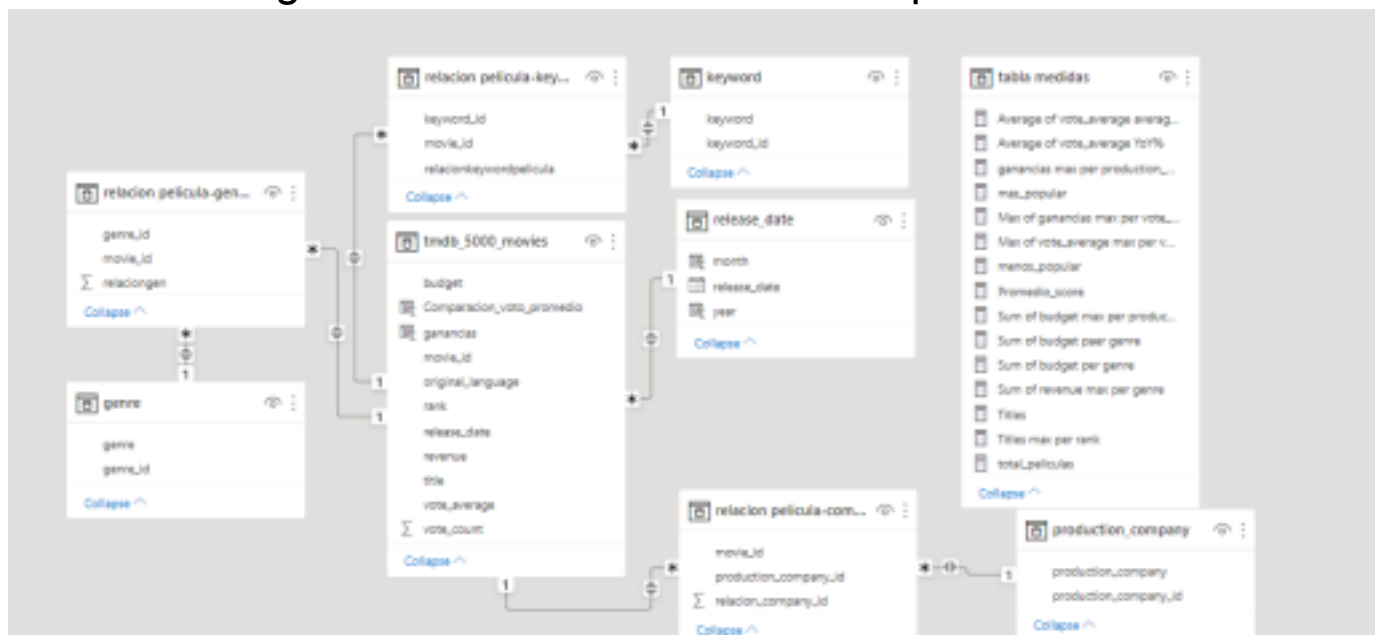
Tabla 7: Relación id_pelicula y compañía

Tipo de clave	Campo Tipo de campo
PK - INDEX	relacioncom int
FK	movie_id int
FK	company_id int

Diagrama entidad relación



Diagrama final de entidad-relación por Power BI



Glosario

Tabla 1:

movie_id: id de cada película
title: Título de la película
budget: Presupuesto de la película
revenue: Ingresos de la película
release_date: Fecha de lanzamiento
vote_average: Rating promedio de la película
vote_count: Cantidad de votos
original_language: Idioma original de la película

Tabla 2:

genre_id: id de cada género
genre: Género

Tabla 3:

production_company_id: id de cada compañía de producción
production_company: Compañía de producción

Tabla 4:

id_keyword: id de palabras clave
keyword: Palabras clave

Tabla 5:

relaciongen: id para la relación entre película y género
movie_id: id de cada película
genre_id: id de cada género

Tabla 6:

relacionkey: id para la relación entre película y palabras clave
movie_id: id de cada película
keyword_id: id de palabras clave

Tabla 7:

relacioncom: id para la relación entre película y compañía

movie_id: id de cada película

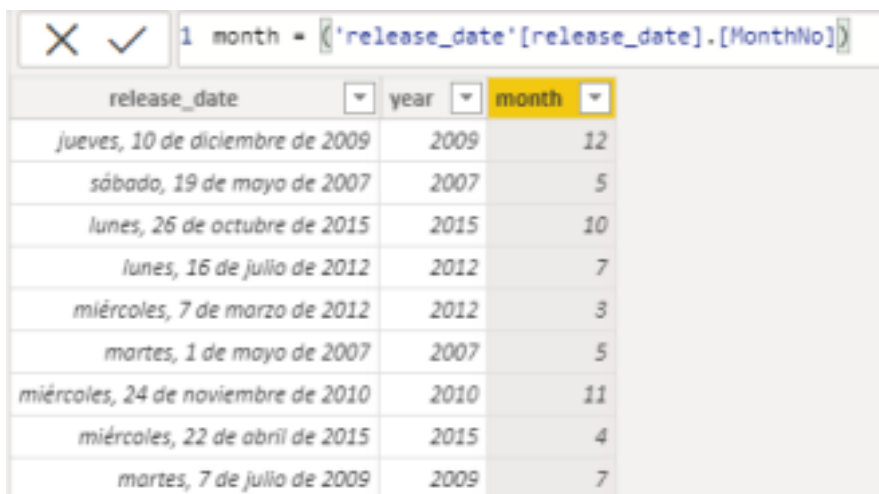
company_id: id de cada compañía de producción

TRANSFORMACIONES DE DATOS

Se creó la relación entre la tabla keyword y la tabla relación película-keyword.

COLUMNAS Y MEDIDAS CALCULADAS

1) Creación de la Tabla Calendario



The screenshot shows a data transformation interface. At the top, a formula bar displays: `1 month = {{'release_date'[release_date].[MonthNo]}}`. Below this is a table with three columns: `release_date`, `year`, and `month`. The `month` column is highlighted in yellow. The table contains the following data rows:

release_date	year	month
jueves, 10 de diciembre de 2009	2009	12
sábado, 19 de mayo de 2007	2007	5
lunes, 26 de octubre de 2015	2015	10
lunes, 16 de julio de 2012	2012	7
miércoles, 7 de marzo de 2012	2012	3
martes, 1 de mayo de 2007	2007	5
miércoles, 24 de noviembre de 2010	2010	11
miércoles, 22 de abril de 2015	2015	4
martes, 7 de julio de 2009	2009	7

2) Creación de las Medidas Calculadas

```
1 mas_popular = MAX{{tmdb_5000_movies[vote_average]}}
```

Cálculo del máximo del score de votación (más popular).

```
1 menos_popular = MIN{{tmdb_5000_movies[vote_average]}}
```

Cálculo del mínimo de score de votación(menos popular).

```
1 total_peliculas = COUNT{{'tmdb_5000_movies'[movie_id]}}
```

Conteo del total de películas evaluadas

```
1 Promedio_score = AVERAGE(tmdb_5000_movies[vote_average])
```

Voto promedio del total de las películas.

3) Creación de las Columnas Calculadas

```
1 ganancias = ('tmdb_5000_movies'[revenue])-( 'tmdb_5000_movies'[budget])
```

Cálculo de ganancias por película(revenue - budget)

```
1 Comparacion_voto_promedio = ('tmdb_5000_movies'[vote_average])-AVERAGE(tmdb_5000_movies[vote_average])
```

Cálculo del rendimiento por sobre o debajo de la media (según voto promedio)

ANÁLISIS FUNCIONAL DEL TABLERO

General

En todas las solapas se encuentran los botones de navegación y los filtros por Año, Género y Compañía.

Solapa 1

Homepage del proyecto donde se encuentran el título del proyecto y los integrantes del mismo.

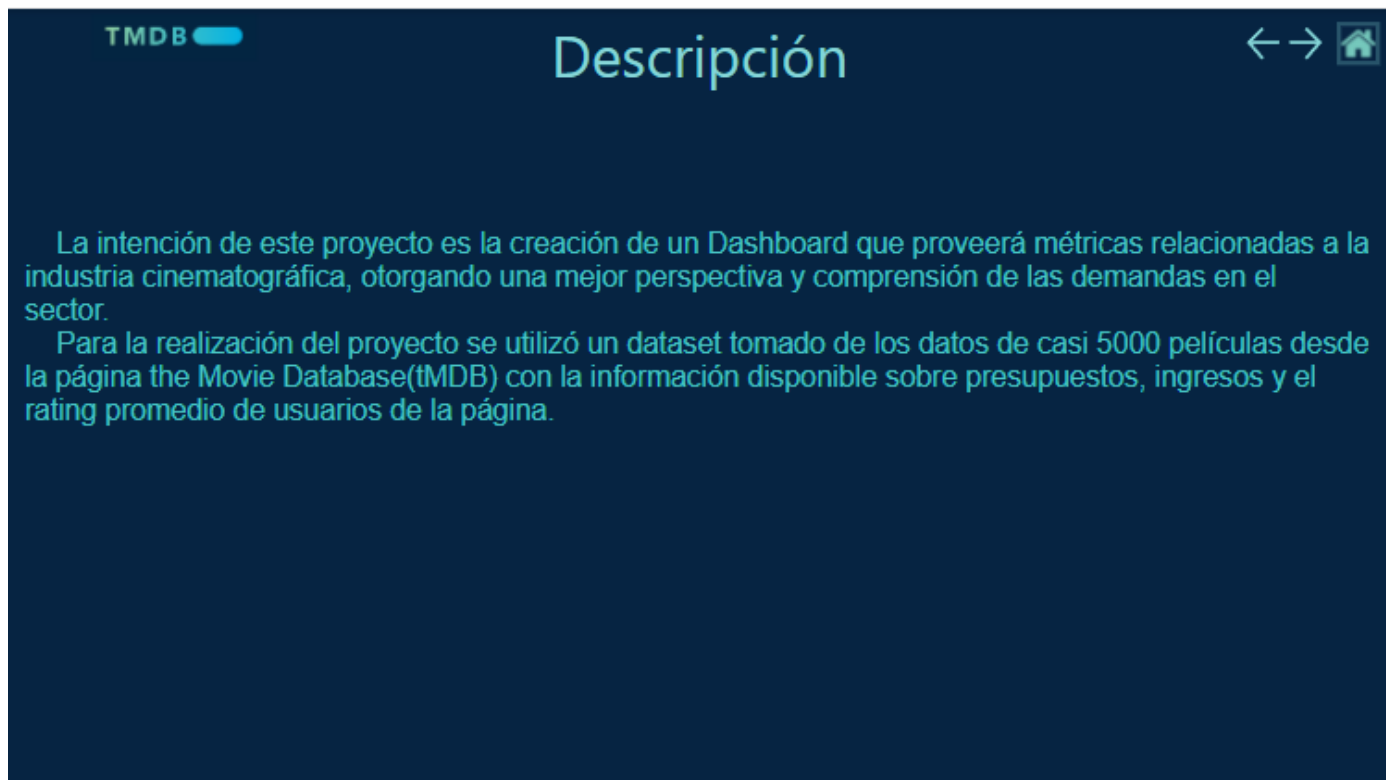
Homepage:



Solapa 2

Breve descripción del proyecto y algunos datos particulares.

Descripción:



Solapa 3

Overview:



Esta solapa proporciona un paneo general de las solapas que le adyacen. Proporciona un top10 de rating de películas, presupuestos e ingresos.

Filtro:

- Año

Filtro según año de lanzamiento.

- Género

Filtro según género.

- Compañía



Filtro con barra de búsqueda según compañía.

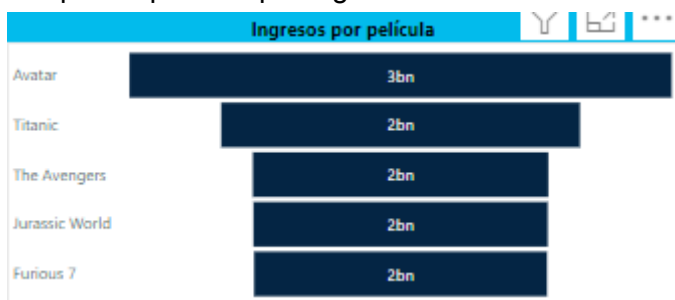
Tablas:

- Top 5 de película por presupuesto



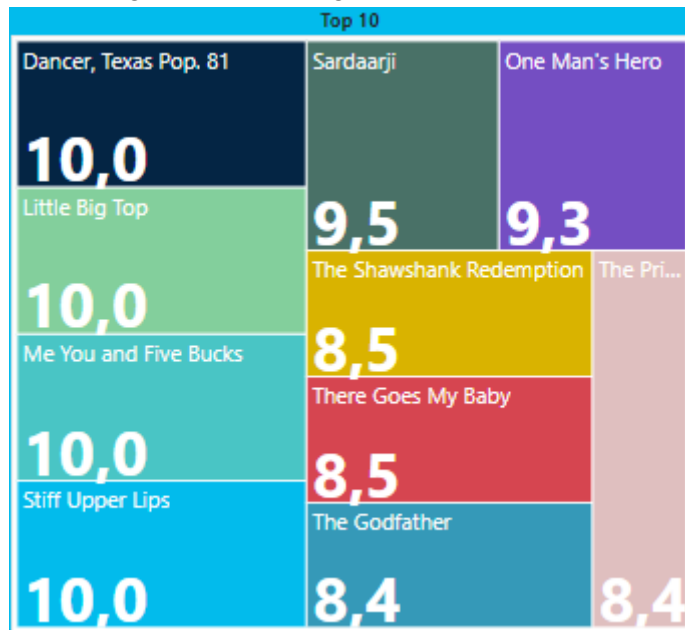
Utilizamos el gráfico Funnel con el propósito de otorgar una comparación clara en el apartado de Presupuesto por Película e Ingresos por Película entre las distintas filmaciones.

- Top 5 de película por ingreso



Utilizamos el gráfico Funnel con el propósito de otorgar una comparación clara en el apartado de Presupuesto por Película e Ingresos por Película entre las distintas filmaciones.

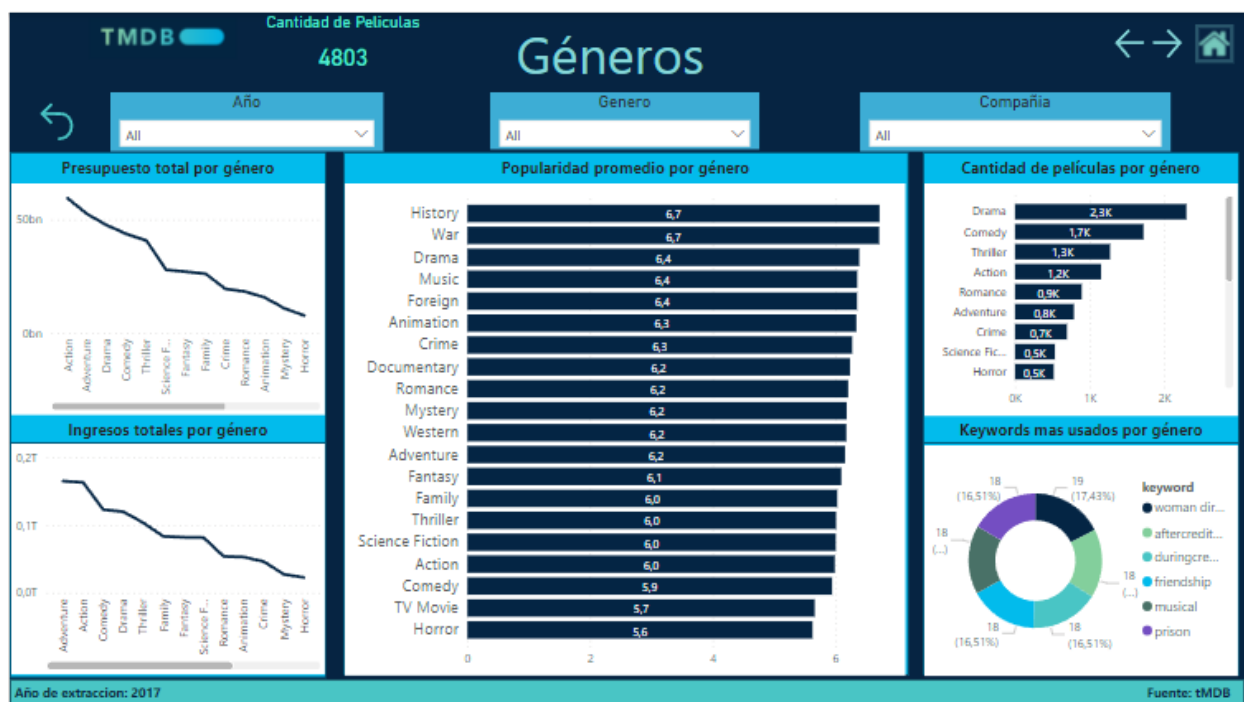
- Top 10 general de ratings



Utilizamos un Trim Up que provee una visualización directa de los puntajes de las mejores películas.

Solapa 4

Género:



Esta solapa se proporciona toda la información relacionada al género de la película: presupuesto total, ingresos totales, popularidad promedio, cantidad de películas producidas y las palabras claves más utilizadas para referenciarse a dicho género.

Filtro:

- Año

Año

All

- Género

Género

All

- Compañía

Compañía

All

Search

☐ Select all

☐ (Blank)

☐ \DIA\ Productions GmbH & Co. KG

☐ 1.B5 Films

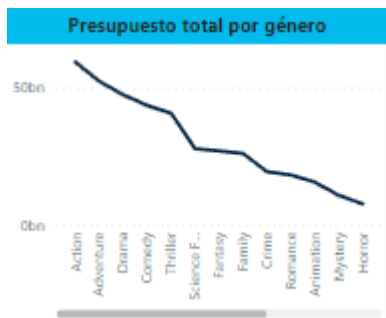
☐ 10 West Studios

☐ 100 Bares

☐ 1019 Entertainment

Tablas:

- Presupuesto total por género



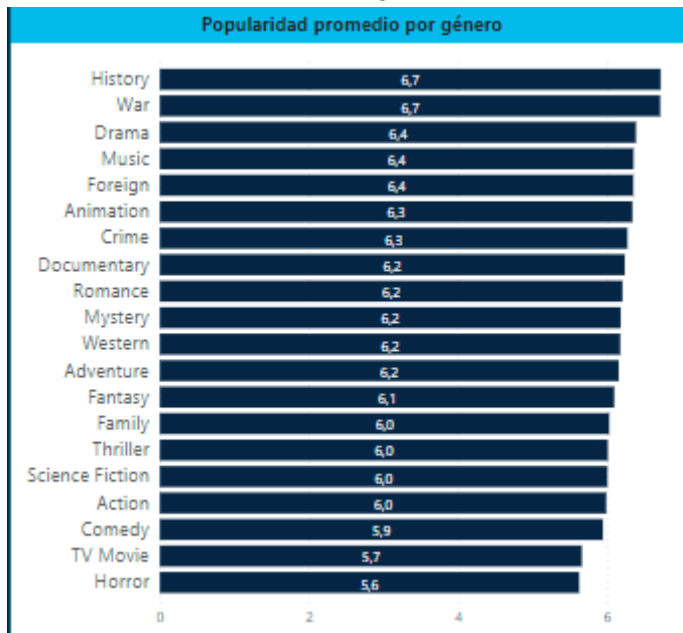
Utilizamos gráficos Line Chart. Considerando que nos permitía visualizar a buena escala qué tan abrupta es la disminución de presupuesto al recorrer los distintos géneros.

- Ingreso total por género



Utilizamos gráficos Line Chart. Considerando que nos permitía visualizar a buena escala qué tan abrupta es la disminución de ingresos al recorrer los distintos géneros.

- Popularidad promedio por género



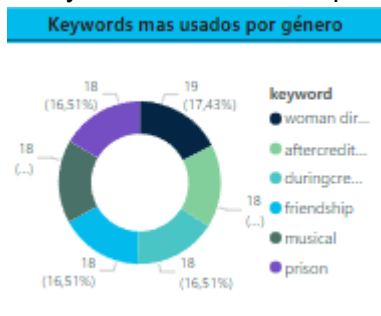
Se recurrieron a Funnel para observar la Popularidad Promedio, de esta manera podríamos incluir una mayor cantidad de géneros y que la lectura de dichos datos sea amena.

- Cantidad de películas por género



Se recurrieron a Funnel para observar la Cantidad de Películas por Género, de esta manera podríamos incluir una mayor cantidad de géneros y que la lectura de dichos datos sea amena.

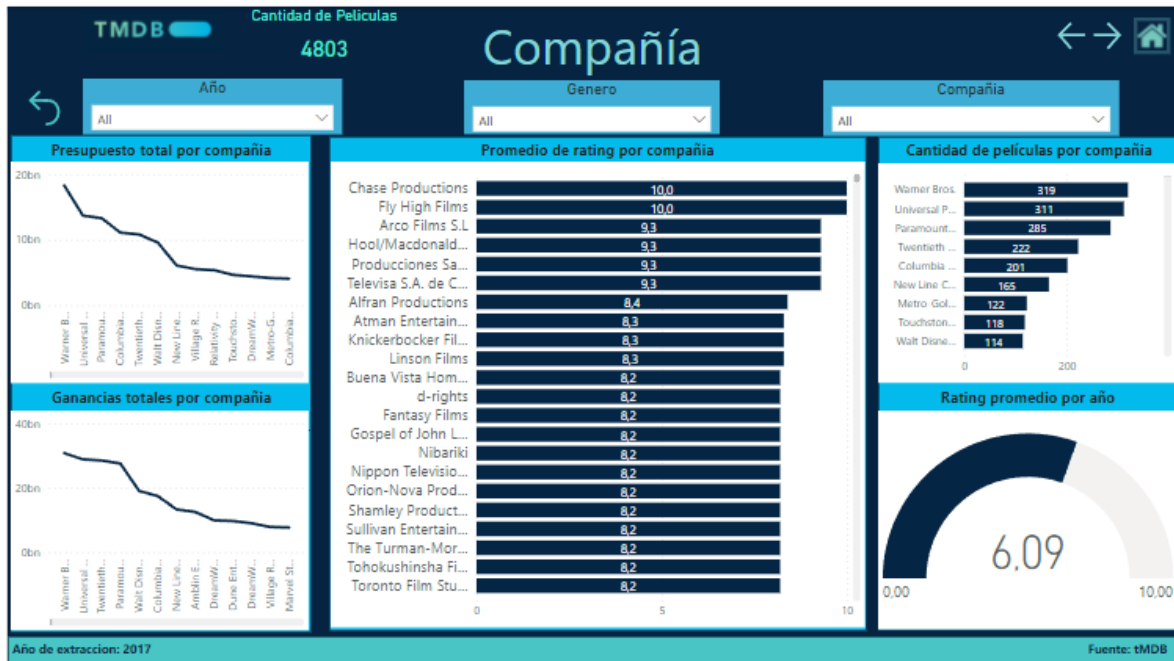
- Keywords más usados por género



Donut chart demostrando los keywords más usados limitándose a 5 (extendiéndose si hay más con la misma cantidad de usos).

Solapa 5

Compañía:



Proporciona información relacionada a la/s compañía/s que produjeron la película: Presupuesto total, Ganancia total, promedio de rating, promedio de rating anual y cantidad de películas.

Filtro:

- Año

Año

All

- Género

Género

All

- Compañía

Compañía

All

Search

- ☐ Select all
- ☐ (Blank)
- ☐ \DIA\ Productions GmbH & Co. KG
- ☐ 1.85 Films
- ☐ 10 West Studios
- ☐ 100 Bares
- ☐ 1019 Entertainment

Tablas:

- Presupuesto total por compañía



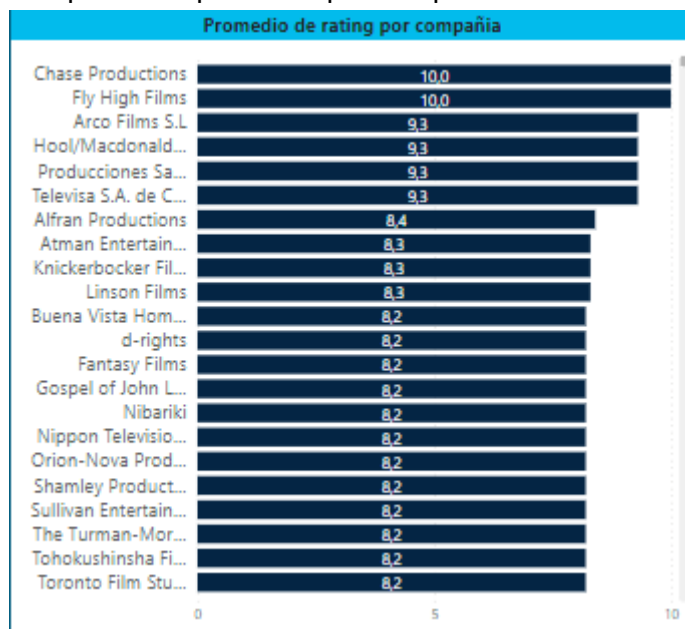
Line chart comparando el presupuesto total invertido por cada compañía.

- Ganancia total por compañía



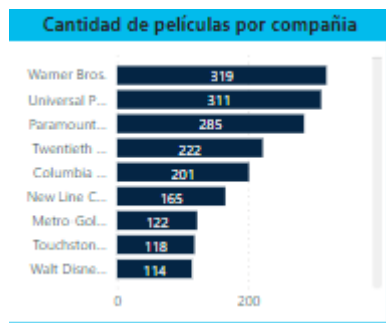
Line chart comparando el total de ganancias por cada compañía.

- Popularidad promedio por compañía



Clustered bar chart con la popularidad promedio por compañía de las películas a las que aplican los varios filtros, permite comparar la popularidad con el presupuesto y ganancias de cada compañía.

- Cantidad de películas por compañía



Clustered bar chart mostrando la cantidad total de películas que creó cada compañía para dar contexto a las visualizaciones anteriores.

- Rating promedio según filtros aplicados



Una tarjeta de tipo Gauge mostrando el rating promedio de todas las películas a las cuales se aplican los filtros elegidos.

Solapa 6

Película:



Proporciona información específica de la película con mayor rating según los filtros aplicados: Título, Ganancias(respecto de la película más taquillera), Presupuesto (respecto de la película más cara en cuanto a producción), Rating y diferencia con el promedio de rating total.

Información por película individual

Filtro:

- Año

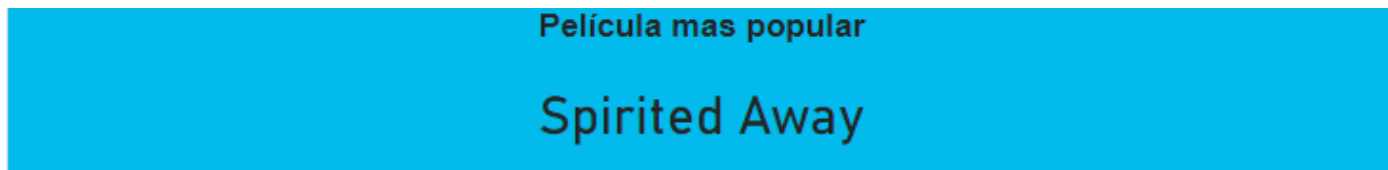
- Género

- Compañía

- ☐ Select all
- ☐ (Blank)
- ☐ \DIA\ Productions GmbH & Co. KG
- ☐ 1.85 Films
- ☐ 10 West Studios
- ☐ 100 Bares
- ☐ 1019 Entertainment

Visualizaciones:

- Nombre de película con mayor ranking según filtros



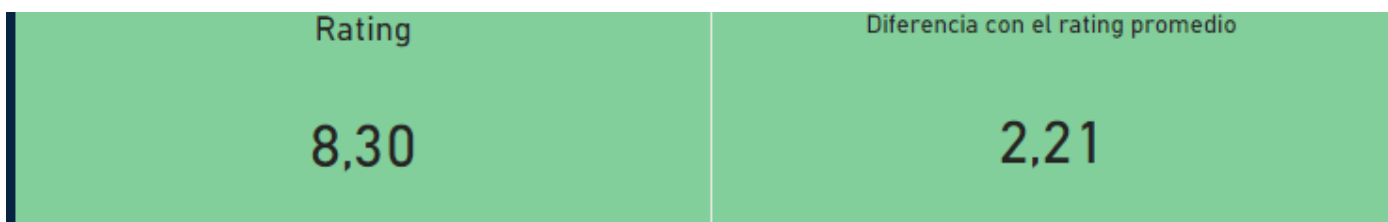
Tarjeta simple mostrando el nombre de la película con mayor ranking.

- Ganancias y presupuesto de la película mostrada



Se utilizaron gráficos tipo Gauge para poder comparar con las varias otras películas ya que tienen como máximo los valores mayores de todos los datos extraídos.

- Rating y diferencia con el rating promedio de la película mostrada



Se utilizaron tarjetas simples mostrando los datos para facilitar la comparación visual al cambiar los filtros.

Correcciones 21/12/21

Ajustamos la dirección de CrossFilter entre las varias tablas.

En la pestaña género:

En la tabla Presupuesto Total, se cambió el filtro max of budget y ahora es sum of budget per genre

En la tabla Ingresos Totales, se cambió el filtro max revenue y ahora es sum of revenue

per genre

En la tabla Compañía:

En la tabla Presupuesto Total, Se cambió el filtro max of budget y ahora es sum of budget per production_company

En la tabla Ganancias Totales, se cambió el filtro max revenue y ahora es sum of ganancias per production_company

Medidas calculadas avanzadas

Nuevas medidas para filtración de las tablas:

```
1 Average of vote_average YoY% =
2 IF(
3     ISFILTERED('release_date'[release_date]),
4     ERROR("Time Intelligence quick measures can only be grouped or filtered by the Power BI-provided date hierarchy or primary date column."),
5     VAR __PREV_YEAR =
6         CALCULATE(
7             AVERAGE('tmdb_5000_movies'[vote_average]),
8             DATEADD('release_date'[release_date].[Date], -1, YEAR)
9         )
10    RETURN
11    DIVIDE(
12        AVERAGE('tmdb_5000_movies'[vote_average]) - __PREV_YEAR,
13        __PREV_YEAR
14    )
15 )
```

Fórmula para el kpi de cambios en el promedio de rating año a año

```
1 Average of vote_average average per genre =
2 AVERAGEX(
3     KEEPFILTERS(VALUE('genre'[genre])),
4     CALCULATE(AVERAGE('tmdb_5000_movies'[vote_average]))
5 )
```

Fórmula de Rating promedio por género

```
1 Sum of revenue max per genre =
2 MAXX(
3     KEEPFILTERS(VALUE('genre'[genre])),
4     CALCULATE(SUM('tmdb_5000_movies'[revenue]))
5 )
```

Fórmula para calcular la suma de ingresos por género

```

1 Sum of budget per genre =
2 SUMX(
3     KEEPFILTERS(VALUES('genre'[genre])),
4     CALCULATE(SUM('tmdb_5000_movies'[budget]))
5 )

```

Fórmula para calcular la suma total de presupuesto por género

```

1 Sum of budget max per production_company =
2 MAXX(
3     KEEPFILTERS(VALUES('production_company'[production_company])),
4     CALCULATE(SUM('tmdb_5000_movies'[budget]))
5 )

```

Fórmula para calcular el presupuesto total por compañía

```

1 ganancias max per production_company =
2 MAXX(
3     KEEPFILTERS(VALUES('production_company'[production_company])),
4     CALCULATE(SUM('tmdb_5000_movies'[ganancias]))
5 )

```

Fórmula para calcular las ganancias totales por compañía

Futuros Pasos

Las intenciones a futuro son incorporar nuevos Bookmarks y tooltips que provean información, guía y fluidez a la hora de navegar por y entre las solapas.

Respecto a los gráficos, se realizará una revisión de los mismos para procurar otorgar una mejor visualización de los datos, así como la opción de incorporar nuevos a los ya existentes en pos de agudizar la calidad de éstos.

Por último, se incorporarán imágenes que refieran a la producción realizada en pos de una más rápida identificación y estética.

Correcciones 23/12/21:

Limpieza de códigos de unicode.

Ajuste de color de fondo y funcionalidad de filtros(se agrego opcion de buscar texto en filtro de compañía).

Ajustes de títulos de visualizaciones y nombres de ejes.

Ajuste de color de contador de películas.

Agregado del logo de tMDB a cada pestaña.

Creación de columna "rank" en la tabla principal para poder utilizarla en la pestaña

"película" Se completó la pestaña "película" con los siguientes elementos:

Tarjeta de película más popular: muestra el título de la película más popular según los filtros actuales usando esta función

```
1 Titles max per rank =  
2 MAXX(  
3     KEEPFILTERS(VALUES('tmdb_5000_movies'[rank])),  
4     CALCULATE([Titles])  
5 )
```

Tarjeta de ganancias de la película más popular: Muestra las ganancias de la película más popular según los filtros aplicados.

Tarjeta de presupuesto de la película más popular: Muestra el presupuesto de la película más popular según los filtros aplicados.

Tarjeta de rating de la película más popular: Muestra el rating de la película más popular según los filtros aplicados.

Tarjeta de comparación de rating de la película más popular con el rating promedio

total: Muestra la diferencia entre el rating de la película más popular y el rating promedio de todas las películas analizadas según los filtros aplicados.

Ajustes 29/12/21

Se agregó una descripción del proyecto en la pestaña "Descripción".

Cambio de lugar de los botones de navegación en la Homepage.

Ajustes 02/01/22

Se agregó pie de página a las pestañas "Overview", "Compañía", "Género" y "Película" con la fecha de extracción de datos y fuente.

Se cambió el diseño de las visualizaciones para hacer que el título y los datos sean más prominentes.

Se cambió la visualización de kpi por una visualización que muestra el rating promedio según los filtros seleccionados.

Se cambió el diseño de la pestaña "Película" para generar más atracción a los datos. Se

cambiaron algunas visualizaciones a otros tipos más apropiados y para generar más variedad.

Expansión sobre los varios puntos en la sección "Descripción de la temática de los datos".

10/01/2022

Se agregó un botón para resetear los filtros en cada pestaña.

13/01/2022

Screenshots finales de las pestañas del dashboard.

30/01/2022

Se extendió la sección de análisis funcional del tablero.

Se incorporaron los motivos para el uso de los gráficos.

Se incorporó la sección "Pasos futuros".