

**UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA**

**FACULTAD DE CONTADURÍA Y ADMINISTRACIÓN**



Preprocesamiento de Datos

**Programación para la Extracción de Datos**

PRESENTADO POR

**Gonzales Preciado Angel Javier**

**Gonzales Ramirez Biron Jahdiel**

**Mascareño Zendejas Aylin Yael**

**Ramirez Velazquez Sherlyn**

**Tapia Carro Angelica Vanesa**

NOMBRE DEL DOCENTE

**Josue Miguel Flores Parra**

NOMBRE DE LA CARRERA

**Inteligencia de Negocios**

FECHA

**de noviembre de 2024**

# Índice

<b>Descripción del Problema.....</b>	<b>3</b>
<b>Objetivos.....</b>	<b>4</b>
<b>Recolección de Datos.....</b>	<b>5</b>
<b>Transformación.....</b>	<b>6</b>
Diagrama de Relación.....	6
Diccionario de Datos.....	6
Películas.....	6
Series.....	7
<b>Visualización.....</b>	<b>8</b>
<b>Conclusión.....</b>	<b>15</b>
<b>Referencias.....</b>	<b>16</b>

# Descripción del Problema

En la actualidad, la cantidad de información disponible sobre películas y series ha crecido enormemente, lo que genera una saturación de datos que hace difícil para aquellos usuarios que quieren acceder a la información de manera rápida y eficiente. Muchos usuarios se ven obligados a realizar búsquedas manuales en diversas plataformas para encontrar tendencias actuales, calificaciones, géneros o detalles específicos sobre películas y series de su interés. Este proceso no es solo tedioso, sino que también puede llevar a resultados incompletos o erróneos debido a la gran cantidad de información dispersa en diversas fuentes.

Esta situación se complica aún más por la falta de herramientas centralizadas y estructuradas que organicen toda esta información de manera accesible. Los usuarios necesitan poder buscar fácilmente cuál es la película o serie con la mejor calificación, qué otros trabajos ha realizado un director específico o qué contenido está disponible dentro de un género determinado. La ausencia de un sistema que facilite estas consultas de forma automatizada y eficiente limita la capacidad de los usuarios para obtener lo que buscan sin tener que navegar entre múltiples fuentes y perder tiempo en procesos manuales.

# Objetivos

El objetivo de este proyecto es ofrecer una solución que permita a los usuarios explorar información relevante sobre películas y series de manera sencilla y visual.

Se busca responder preguntas clave como: ¿Qué películas y series tienen las mejores y peores calificaciones? ¿Qué géneros son más populares y mejor valorados? ¿Qué directores o creadores se destacan por la cantidad y calidad de sus obras? Además, se analizará cómo se distribuyen las calificaciones en diferentes géneros, la duración promedio de las películas por categoría, y los picos de estrenos a lo largo del tiempo.

Este proyecto busca ayudar a los usuarios a tomar decisiones informadas sobre qué contenido consumir sin necesidad de buscar manualmente y perder tanto tiempo, contando con datos claros y accesibles para todo aquel que busque entretenimiento.

# Recolección de Datos

La página utilizada para nuestro proyecto es SensaCine, una plataforma española dedicada a brindar información detallada sobre cine y series. Entre las características principales de la página destaca sus fichas completas de películas y series, incluyendo información sobre actores, directores, trailers y noticias relevantes con relación en el mundo cinematográfico. Además, cuenta con un apartado de críticas donde SensaCine asigna su propia calificación a las películas, complementando la información con reseñas tanto de expertos como de los usuarios.

Para nuestro proyecto, utilizamos este apartado para la recolección de datos de esta página. El enfoque principal de la recolección se centra en extraer las calificaciones otorgadas por SensaCine a diferentes películas y series, junto con información relevante de las fichas, como el título, género y fechas de lanzamiento si es que cuentan con esta información. Estos datos son extraídos y limpiados para su posterior análisis.

Liga utilizada para la recolección de datos de películas:

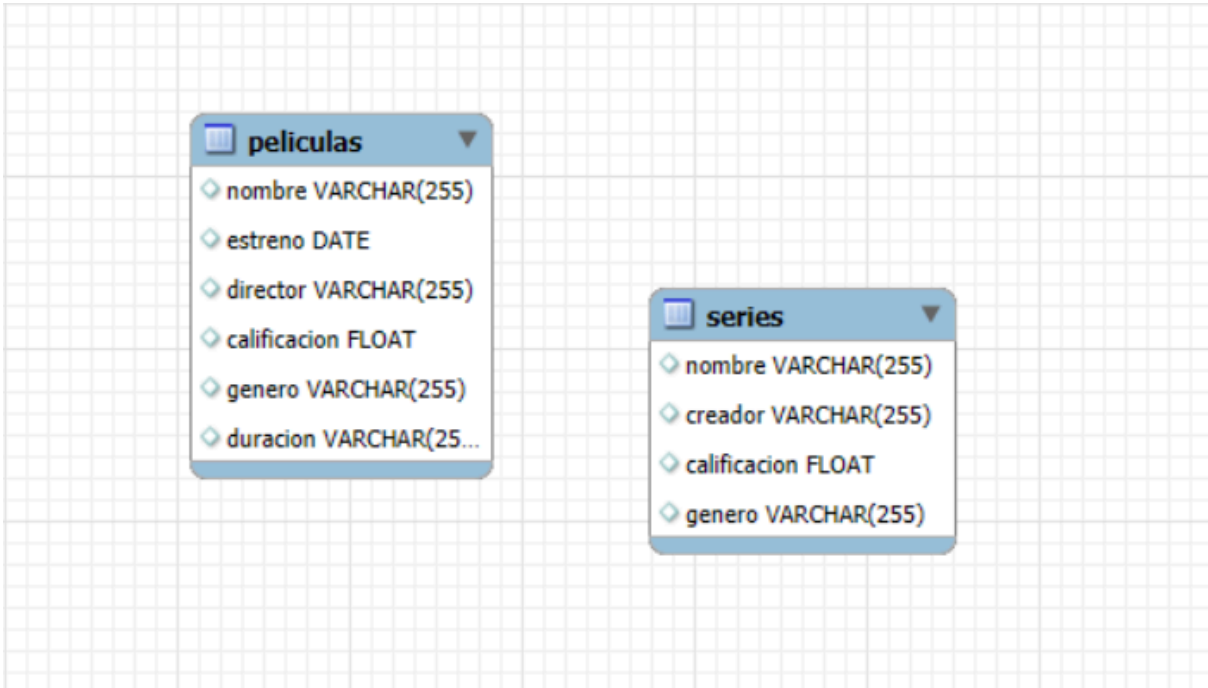
<https://www.sensacine.com/peliculas/criticas-sensacine/>

Liga utilizada para la recolección de datos de series:

<https://www.sensacine.com/series-tv/mejores/>

# Transformación

## Diagrama de Relación



## Diccionario de Datos

### Películas

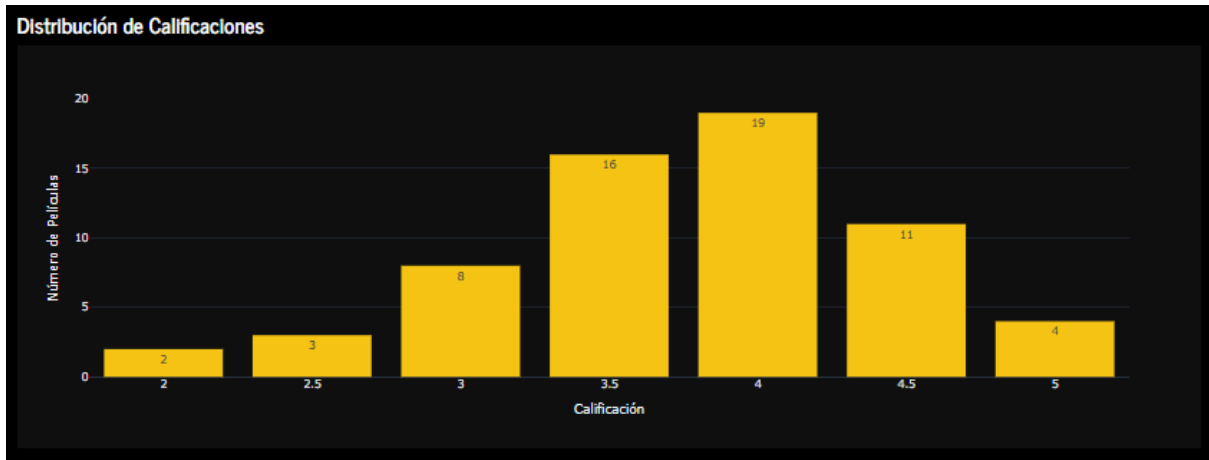
Nombre del Campo	Clave	Tipo de Datos	Descripción del Campo
nombre	NN	VARCHAR(255)	El título de la película.
estreno	NN	DATE	La fecha de estreno de la película. Se utiliza el formato de fecha estándar (YYYY-MM-DD).
director	NN	VARCHAR(255)	El nombre del director de la película.
calificación	NN	FLOAT	La puntuación asignada a la película, expresada como un número flotante.

género	N	VARCHAR(255)	El género o categoría principal de la película (por ejemplo, "Drama", "Comedia").
duración	NN	VARCHAR(255)	La duración de la película, normalmente expresada en horas y minutos.

## Series

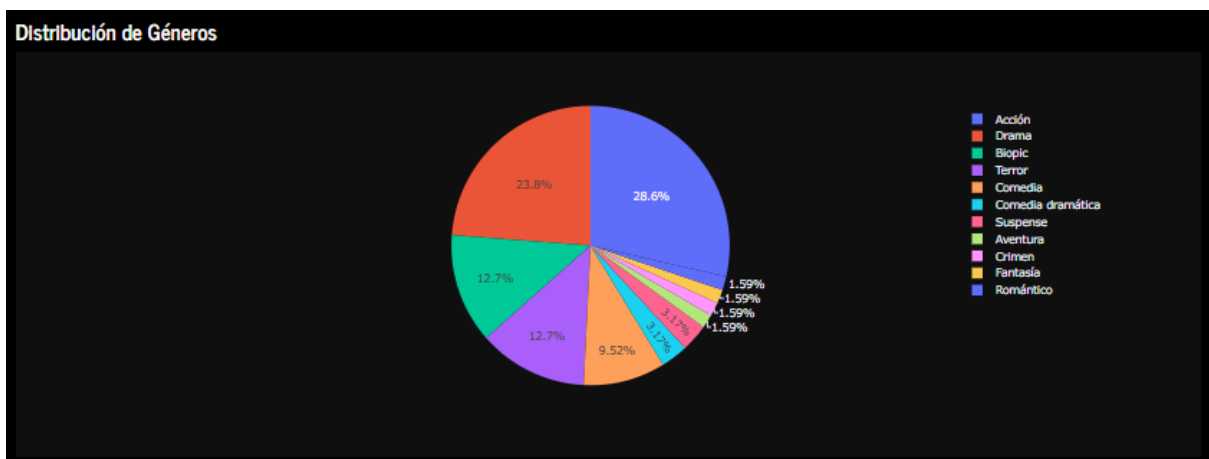
Nombre del Campo	Clave	Tipo de Datos	Descripción del Campo
nombre	NN	VARCHAR(255)	El título de la serie.
creador	NN	VARCHAR(255)	El nombre del creador de la serie.
calificación	NN	FLOAT	La puntuación asignada a la serie, expresada como un número flotante.
género	NN	VARCHAR(255)	El género o categoría principal de la serie (por ejemplo, "Drama", "Ciencia ficción").

# Visualización



Este gráfico es un histograma que muestra la frecuencia de las calificaciones que tienen las películas estrenadas en el año seleccionado.

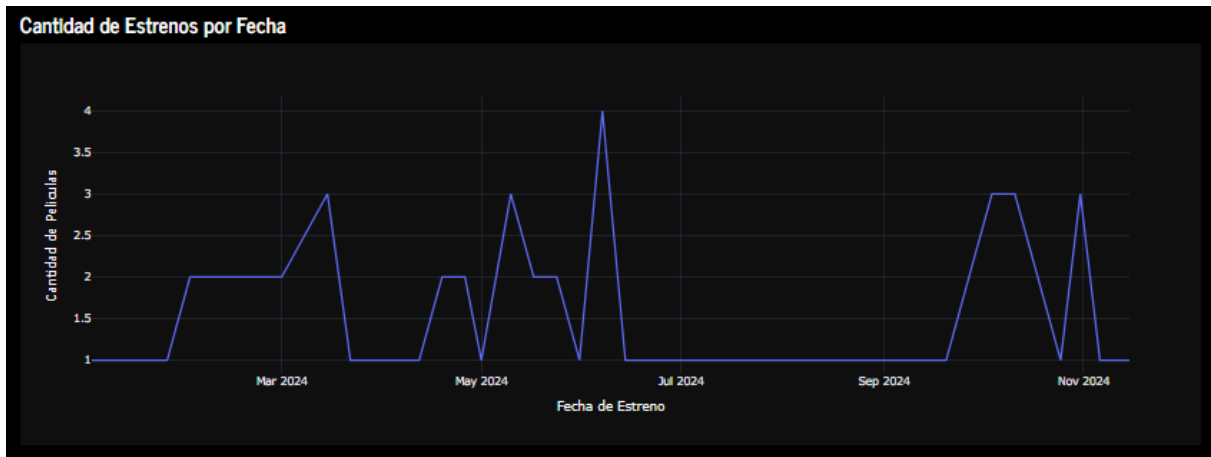
Propósito: Ayuda a identificar si la mayoría de las películas tienen altas, medias o bajas calificaciones y observar la variabilidad en la calidad percibida de las películas en ese año.



Este es un gráfico de pastel que muestra la proporción de películas por género en el año seleccionado.

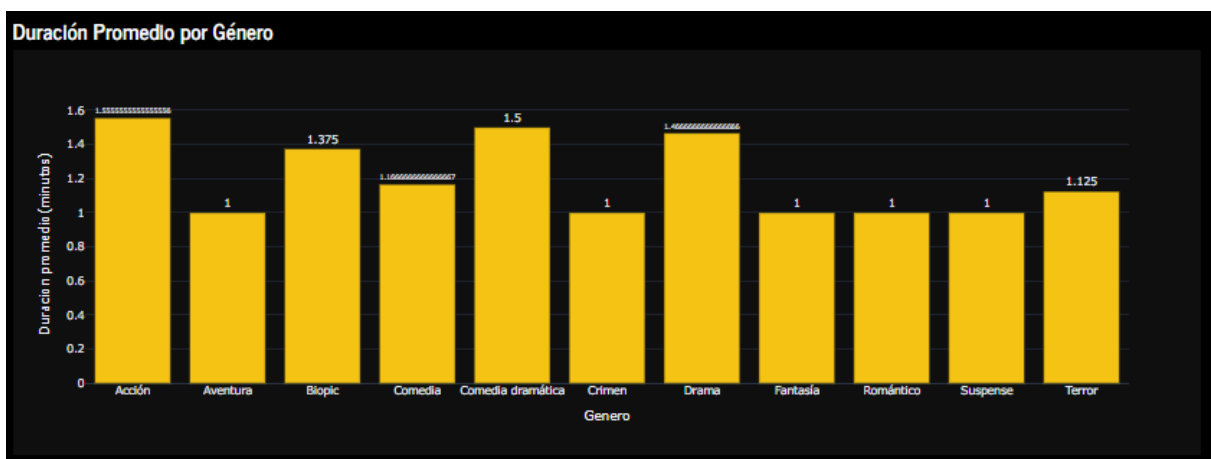
Propósito: Ayuda a visualizar qué géneros fueron más populares en términos de cantidad de películas estrenadas durante el año seleccionado.





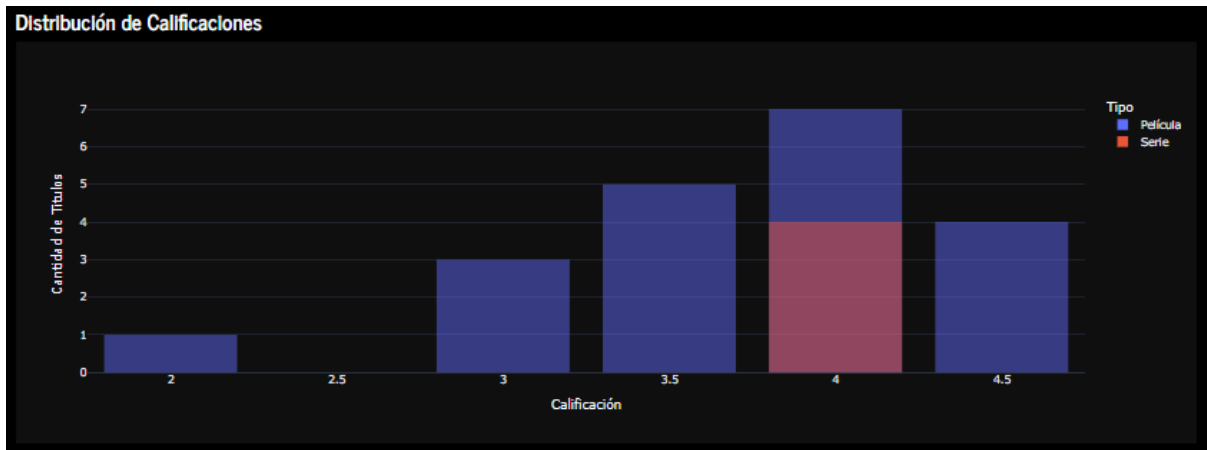
Este gráfico es una línea temporal que muestra la cantidad de películas estrenadas en diferentes fechas a lo largo del año seleccionado.

Propósito: Permite identificar picos de estrenos en fechas específicas, como temporadas altas y patrones de lanzamientos a lo largo del año.



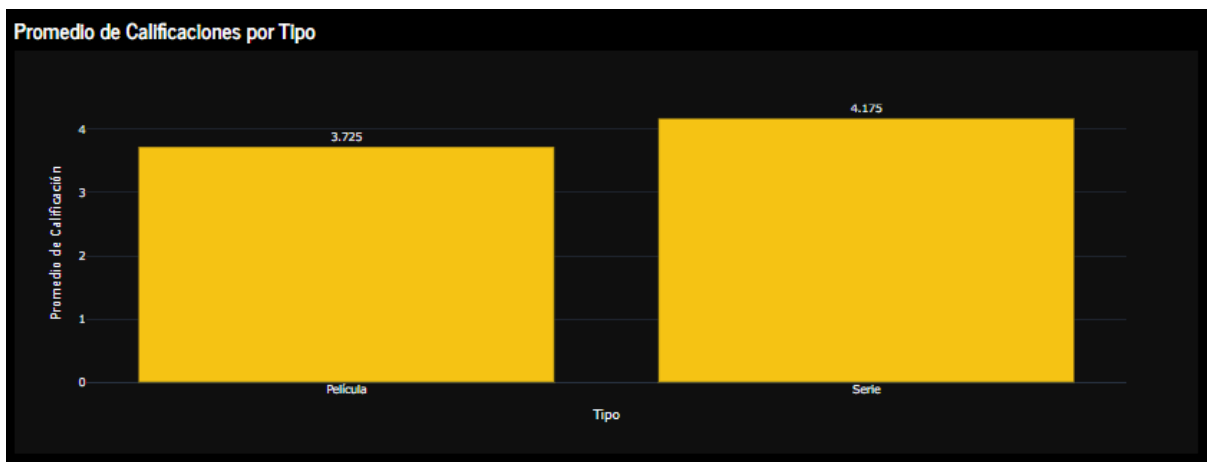
Este gráfico de barras muestra la duración promedio (en minutos) de las películas agrupadas por género en el año seleccionado.

Propósito: Proporciona información sobre cuánto tienden a durar las películas de cada género, lo cual es útil para identificar tendencias en la industria cinematográfica.



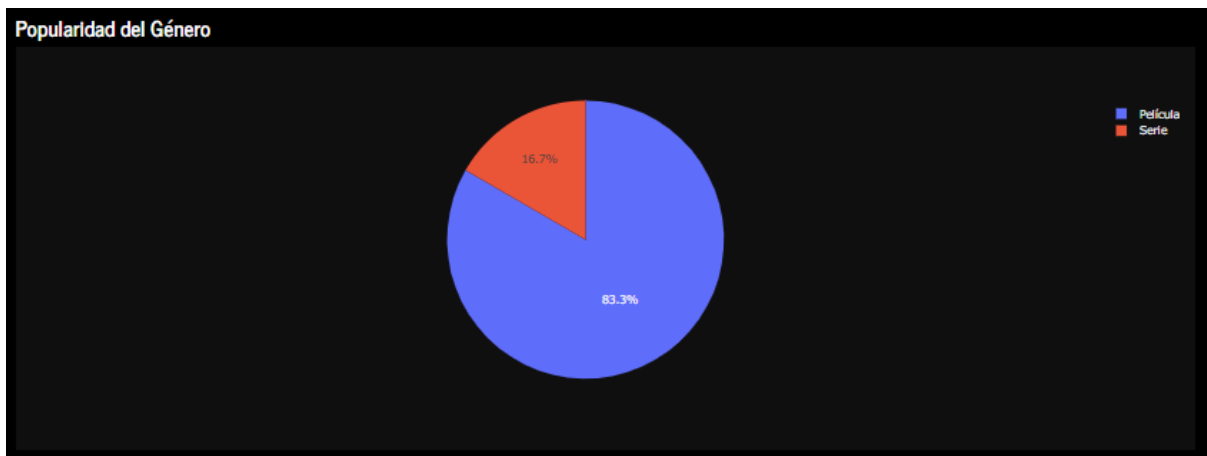
Este histograma muestra la distribución de calificaciones otorgadas a películas y series dentro del género seleccionado.

Propósito: Ayuda a visualizar qué tan dispersas o concentradas están las calificaciones, identificando si el género tiene más títulos con calificaciones altas, medias o bajas. También permite comparar el comportamiento entre películas y series en términos de calidad percibida.



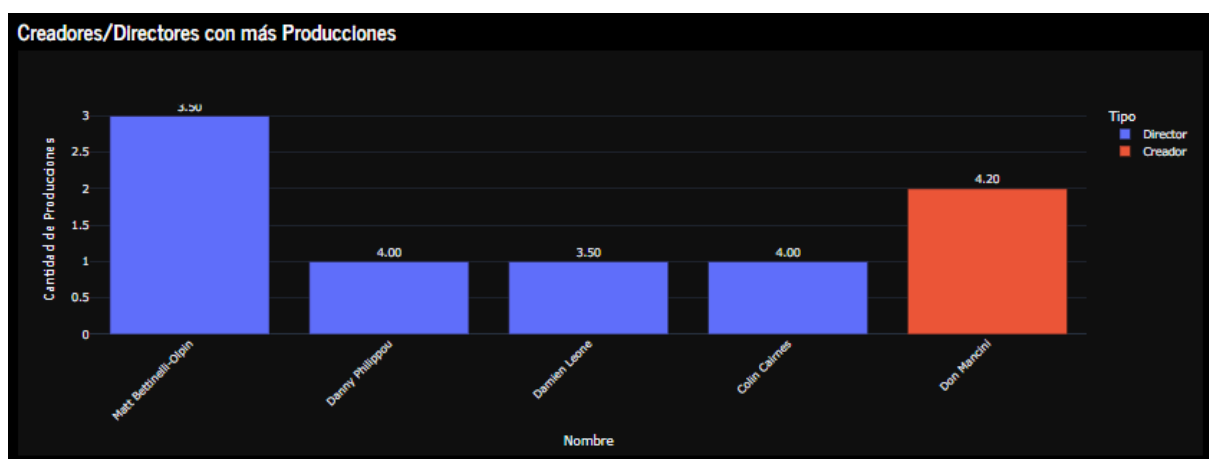
Este gráfico de barras compara el promedio de calificaciones de películas y series dentro del género seleccionado. Las barras indican el valor promedio de calificación, y el texto sobre ellas resalta la cifra exacta.

Propósito: Proporciona una comparación directa entre la calidad promedio de películas y series en un género, permitiendo identificar cuál tipo tiende a tener mejor recepción por parte de los espectadores.



Este gráfico de pastel muestra la proporción de títulos (películas y series) que pertenecen al género seleccionado. Cada segmento del pastel representa la cantidad relativa de producciones de cada tipo.

Propósito: Facilita la identificación de si el género está dominado por películas o series, ayudando a entender la distribución de producción en la industria.

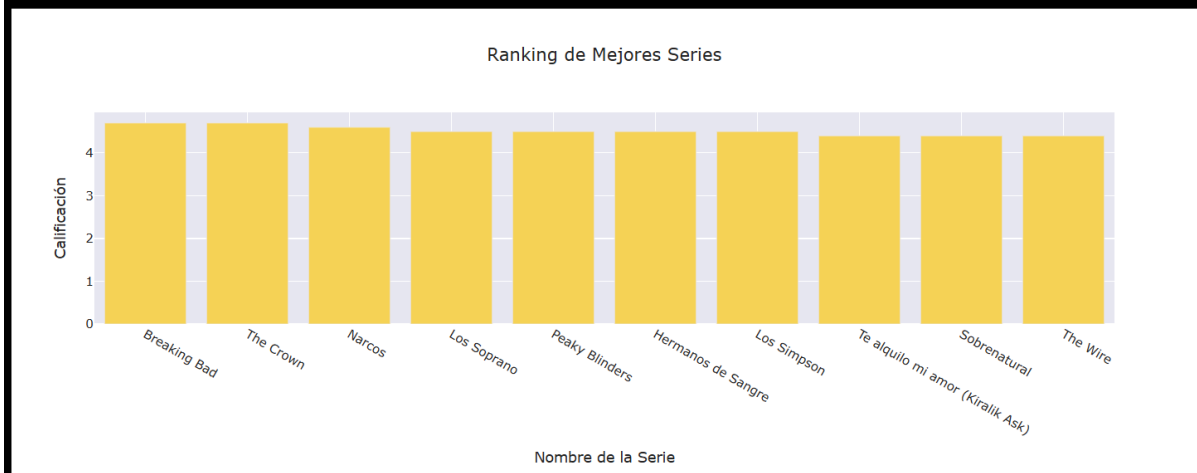


Este gráfico de barras destaca a los cinco creadores o directores con mayor cantidad de producciones en el género seleccionado. También incluye el promedio de calificación de sus obras, indicado como texto sobre las barras.

Propósito: Identifica a los principales responsables creativos en el género, permitiendo evaluar no solo su productividad, sino también la calidad promedio de sus producciones, ya sea como creadores (para series) o directores (para películas).

### Selecciona una gráfica:

Mejores Series

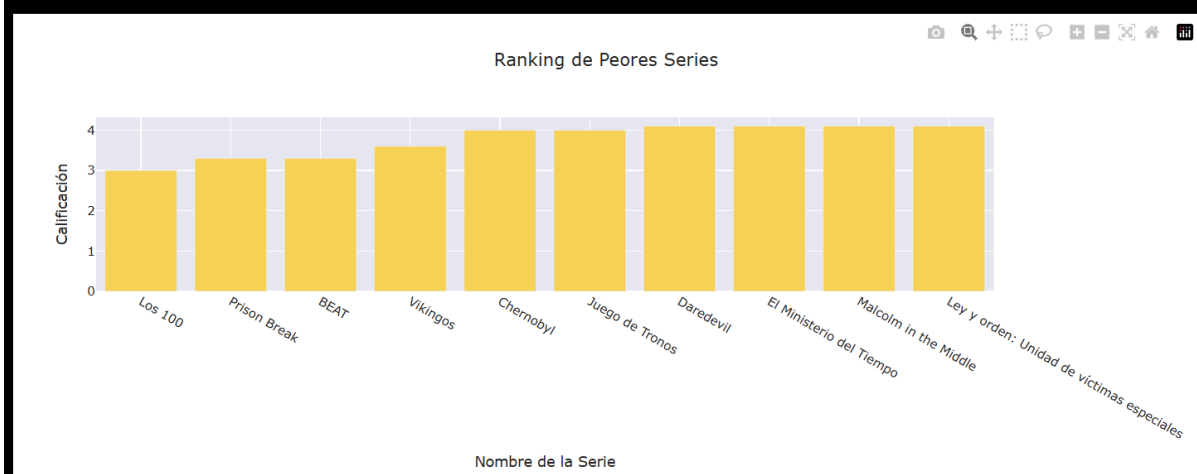


Este gráfico de barras muestra el ranking de las mejores series según su calificación. Se muestra en una escala del 0 al 5, ordenadas del top 1 al top 5. Cada fila cuenta con su calificación.

Propósito: Destacar las series con las calificaciones más altas, proporcionando una visión clara de la preferencia de los espectadores.

### Selecciona una gráfica:

Peores Series



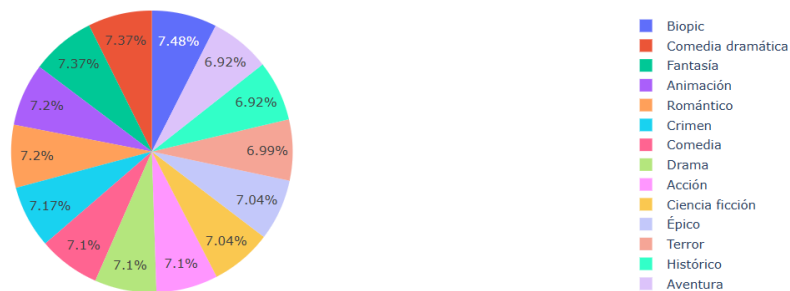
Este gráfico de barras muestra el ranking de las peores series según su calificación. Se muestra en una escala del 0 al 5, ordenadas del top 1 al top 10. Cada fila cuenta con su calificación.

Propósito: Destacar las series con las calificaciones más bajas, proporcionando una visión clara de los espectadores.

### Selecciona una gráfica:

Promedio de Calificación por Género

Ranking Promedio por Género



Este gráfico muestra el promedio de calificaciones por género de series, y cada segmento del pastel está representado por un género en específico, su tamaño muestra el porcentaje del promedio de calificación que está ligado a ese género.

Propósito: Analiza cómo se distribuyen las calificaciones promedio entre los diferentes géneros, ayudando a identificar cuales son los mejores valorados.

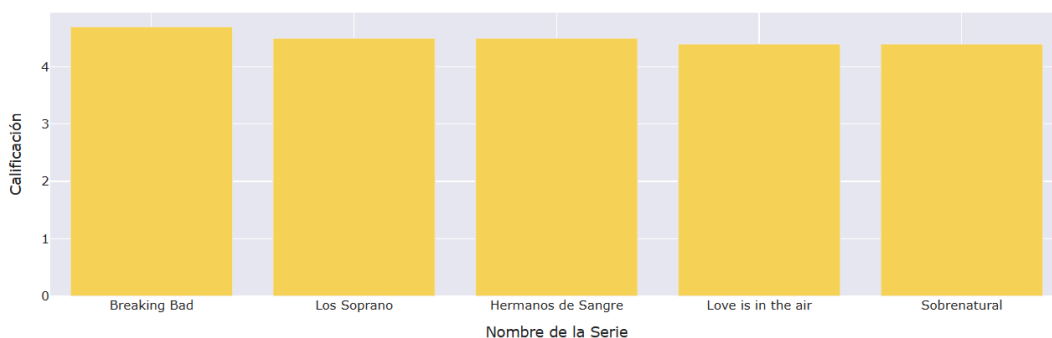
### Selecciona una gráfica:

Top 5 por Género

### Selecciona un Género:

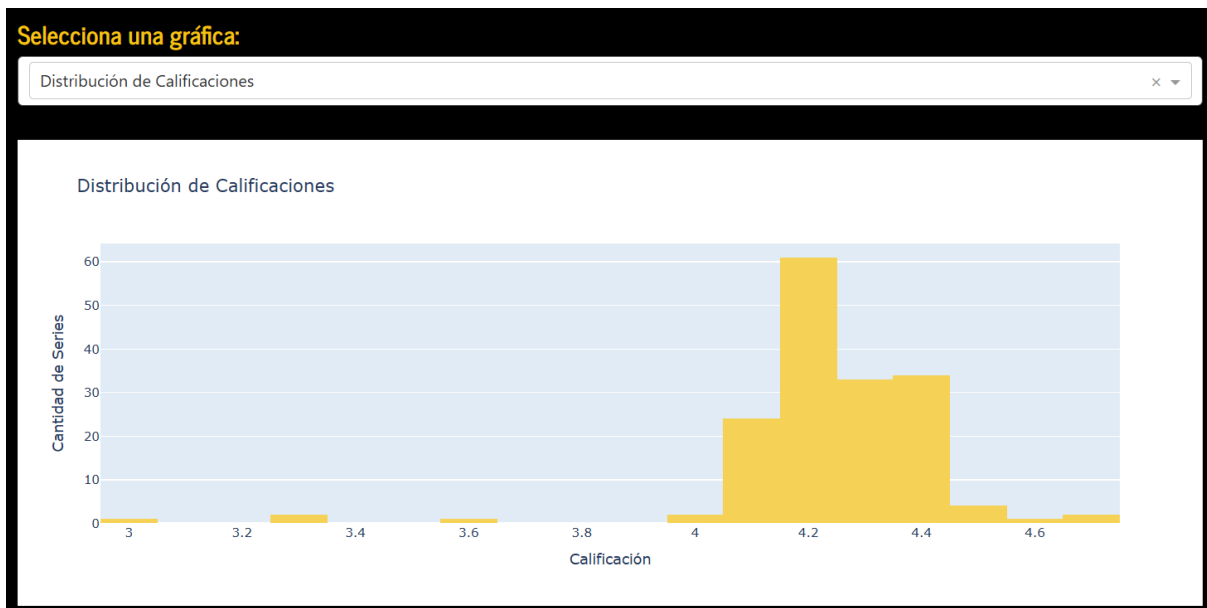
Drama

Top 5 Series en Género: Drama



Este gráfico de barras muestra el ranking de las mejores series según su género. Se muestra en una escala del 0 al 5, ordenadas del top 1 al top 5. Cada fila cuenta con su calificación.

Propósito: Destacar las series con las calificaciones más altas según su género, proporcionando una visión clara de las preferencias de los espectadores.



Esta gráfica de barras muestra la distribución de calificaciones de series en una escala de 3 a 4.7 y con la cantidad de series que obtuvieron calificaciones dentro de cada rango.

Propósito: El objetivo de esta gráfica es mostrar cómo se distribuyen las calificaciones de las series e identificar los rangos de calificaciones más comunes.

# Conclusión

En este proyecto logramos cumplir con los objetivos planteados al diseñar un sistema que permita recolectar, procesar y visualizar datos sobre películas y series de manera eficiente. La creación de un tablero interactivo facilitó el análisis de tendencias, calificaciones y géneros, haciendo más accesible la información para los usuarios.

Durante el desarrollo, enfrentamos retos como la limpieza y estructuración de los datos, así como la optimización del proceso de extracción, lo que nos permitió identificar oportunidades de mejora en la automatización y validación de la información. Estos desafíos fortalecieron nuestras habilidades técnicas y nos brindaron una mejor comprensión de la importancia de los procesos iterativos en proyectos tecnológicos.

# Referencias

SensaCine. (s. f.). *Las críticas de SensaCine.com*. SensaCine.com.

<https://www.sensacine.com/peliculas/criticas-sensacine/>

SensaCine. (s. f.-b). *Las mejores series*. SensaCine.com.

<https://www.sensacine.com/series-tv/mejores/>

Github. (2024, 3 diciembre). Proyecto Final.

<https://github.com/AylinSenju/ProyectoFinal>