



Universidad  
Internacional  
de Andalucía

## TÍTULO

DETERMINANTES DE LA RECAUDACIÓN EN TAQUILLA COMO  
INDICADOR DE ÉXITO CINEMATOGRAFICO

## AUTORA

Aimara Caridad Viera Alonso

	<b>Esta edición electrónica ha sido realizada en 2025</b>
Tutora	Dra. Mónica Carmona Arango
Instituciones	Universidad Internacional de Andalucía; Universidad de Huelva
Curso	<i>Máster Universitario en Economía, Finanzas y Computación (2023/24)</i>
©	Aimara Caridad Viera Alonso
©	De esta edición: Universidad Internacional de Andalucía
Fecha documento	2024



Universidad  
Internacional  
de Andalucía



**Atribución-NoComercial-SinDerivadas  
4.0 Internacional (CC BY-NC-ND 4.0)**

Para más información:

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>

# «Determinantes de la recaudación en taquilla como indicador de éxito cinematográfico»

by

«Aimara Caridad Viera Alonso»

A thesis submitted in conformity with the requirements  
for the MSc in Economics, Finance and Computer Science

University of Huelva & International University of Andalusia

**uhu**.es

**un**  
i Universidad  
Internacional  
de Andalucía  
**A**

# «Determinantes de la recaudación en taquilla como indicador de éxito cinematográfico»

«Aimara Caridad Viera Alonso»

Máster en Economía, Finanzas y Computación

«Supervisor/s»

Universidad de Huelva y Universidad Internacional de Andalucía

2024

## Abstract

The present work explores the factors that determine the box office success of movies, considering revenue as a key indicator. A database was constructed comprising 497 movies and their most relevant attributes (genres, award nominations, release dates, cast, directors, rating) to perform an analysis using a multiple linear regression model. This approach was used to identify and compare, with previous studies, the variables that most influence revenue.

The results show that variables such as genre, the involvement of well-known actors and actresses, and effective distribution strategies have a positive and significant impact on revenue, while other variables do not contribute significantly as they depend on external factors.

**JEL classification:** L82, C51, D22, M31

**Keywords:** box office success, genres, nominations, release dates, cast, directors, rating, linear regression, independent variables, strategies.

## Resumen

El presente trabajo explora los factores que determinan el éxito en taquilla de las películas, considerando la recaudación como un indicador clave. Se construyó una base de datos compuesta por 497 películas y sus atributos más relevantes (géneros, nominaciones a premios, fechas de estreno, reparto, directores, rating) para realizar un análisis mediante un modelo de regresión lineal múltiple para de esta forma identificar y contrastar con estudios anteriores, las variables que más influyen en la recaudación.

Los resultados muestran que variables como el género, la participación de actores y actrices conocidos y estrategias de distribución efectivas tienen un impacto positivo y significativo en la recaudación, mientras que otras variables no aportan significativamente ya que dependen de factores externos.

**Clasificación JEL:** L82, C51, D22, M31

**Palabras clave:** éxito en taquilla, géneros, nominaciones, fechas de estreno, reparto, directores, rating, regresión lineal, variables independientes, estrategias.

## Tabla de contenido

1.- PROPUESTA DE TRABAJO DE FIN DE MÁSTER.....	1
1.1 Introducción.....	1
1.2 Objetivos específicos.....	2
2.- ESTADO DEL ARTE .....	4
3.- METODOLOGÍA.....	8
3.1.- La base de datos. ....	8
3.2.- Los datos. ....	8
3.3.- Las variables.....	10
3.4.- Estadísticos descriptivos. Uso de STATA .....	12
3.5.- El modelo. ....	14
4.- RESULTADOS .....	16
4.1 Interpretación de los coeficientes de regresión.....	17
4.2 Diagnóstico del modelo. Evaluación de los supuestos de la regresión lineal.....	18
4.2.1 Normalidad de los residuos. ....	18
4.2.2 Homocedasticidad. ....	18
4.2.3 Multicolinealidad entre variables independientes. ....	20
5.- CONCLUSIONES.....	22
REFERENCIAS BIBLIOGRÁFICAS.....	24
ANEXOS .....	26
Anexo 1. Carga de películas y procesamiento de cada fila. ....	26
Anexo 2. Método principal.....	26
Anexo 3. Búsqueda de películas del listado en Filmaffinity. ....	27
Anexo 4. Búsqueda de la información de cada película.....	27
Anexo 5. Definición de variables cualitativas.....	28
Anexo 6. Estructuración y limpieza de la Base de Datos.....	28
Anexo 7. Ejemplos de métodos para convertir las variables cualitativas a binarias. ....	29

## Índice de tablas

Tabla 1. Número de espectadores y recaudación en taquilla de películas españolas en los últimos 5 años.	Página 1
Tabla 2. Descripción de las variables.	Página 10
Tabla 3. Estadísticos descriptivos variable log_recaudacion	Página 12
Tabla 4. Estadísticos descriptivos variables discretas	Página 13
Tabla 5. Estadísticas del modelo de regresión	Página 16
Tabla 6. Modelo de regresión lineal múltiple en STATA	Página 16
Tabla 7. Test de Shapiro-Wilk	Página 18
Tabla 8. Factor de inflación de la varianza (VIF)	Página 20

# 1. PROPUESTA DE TRABAJO DE FIN DE MÁSTER

## 1.1 Introducción.

La industria cinematográfica en España ha ido en constante evolución, adaptándose a los cambios sociales, económicos y tecnológicos a lo largo de los años. Desde sus inicios, ha experimentado transformaciones significativas que han influido en su producción, distribución y consumo. En las últimas décadas la digitalización ha abierto nuevas oportunidades y desafíos para la industria, permitiendo una mayor accesibilidad a las películas en mercados internacionales y facilitando la producción independiente. Asimismo, la diversificación de géneros y la colaboración con talentos internacionales han contribuido a enriquecer la oferta cinematográfica del país y a consolidar su posición en la escena mundial del cine.

Teniendo en cuenta que la industria del cine en España también enfrenta desafíos como la piratería, la competencia en las plataformas de streaming, y la búsqueda de un equilibrio entre la preservación de la identidad cultural española y la búsqueda de audiencias globales, es relevante destacar que, a pesar de las fluctuaciones en los ingresos, el cine continúa siendo la principal fuente de ganancias en la explotación de una película, mostrando que la rentabilidad en taquilla sigue siendo un factor de suma importancia en el panorama de la industria cinematográfica española, incluso en tiempos de incertidumbre como los provocados por la crisis sanitaria global; cambio de panorama significativo que se muestra en la siguiente tabla:

**Tabla 1.** Número de espectadores y recaudación en taquilla de películas españolas en los últimos 5 años.

<b>AÑOS</b>	<b>ESPECTADORES</b>	<b>RECAUDACIÓN</b>
<b>2019</b>	16 050 136	94 115 228 €
<b>2020</b>	7 379 125	43 052 242 €
<b>2021</b>	7 107 930	42 309 484 €
<b>2022</b>	13 806 764	82 715 451 €
<b>2023</b>	13 197 207	81 528 425 €

Según un estudio de Carballo Sánchez, A. (2022), el impacto de la pandemia de COVID-19 en la rentabilidad en taquilla en España fue considerable, afectando a toda la cadena de valor de la industria. La intermitencia en los cierres de las salas de cine, las restricciones de aforo y las medidas de distanciamiento social implementadas para contener la propagación del virus tuvieron

un efecto directo en la asistencia de espectadores a las salas, lo que resultó en una disminución drástica de los ingresos por taquilla.

Además, la estrategia adoptada por las grandes distribuidoras de retrasar estrenos o lanzar películas directamente en plataformas de streaming afectó directamente la rentabilidad de las salas de cine, ya que se vieron obligadas a competir con nuevas formas de consumo audiovisual.

Para mitigar esta situación, el Gobierno español implementó medidas de apoyo, como los Expedientes Temporales de Regulación de Empleo (ERTEs) para aliviar los gastos de personal de las empresas cinematográficas, líneas de crédito subsidiadas por el ICO (Instituto de Crédito Oficial) para cubrir necesidades de tesorería y ayudas directas a las salas de cine respaldadas por la Unión Europea. A pesar de estos esfuerzos, la rentabilidad en taquilla en España se vio significativamente afectada, con una estimación de pérdidas de alrededor de 250 millones de euros.

## 1.2 Objetivos específicos.

El **objetivo general** de esta investigación es identificar y analizar los factores que influyen en la recaudación en taquilla de las películas, utilizando la métrica como un indicador principal de éxito cinematográfico.

Para lograr esto se plantean los siguientes objetivos específicos:

- Recopilar, organizar y limpiar la información de 497 películas asegurando la calidad y precisión de los datos recopilados.
- Determinar cuáles son las variables de producción y comercialización que tienen un impacto positivo y significativo en la recaudación en taquilla.
- Estimar el efecto de cada una de las variables identificadas en la recaudación en taquilla mediante la aplicación de un modelo estadístico.
- Verificar la significancia estadística de las relaciones identificadas entre las variables y la recaudación en taquilla, asegurando que los resultados obtenidos sean fiables en comparación con estudios anteriores.

- Ofrecer recomendaciones estratégicas basadas en los resultados obtenidos que puedan ser utilizadas por profesionales de la industria cinematográfica para mejorar la planificación y ejecución de proyectos con el fin de maximizar la recaudación en taquilla.
- Contribuir al conocimiento académico, proporcionando evidencia empírica que respalde o cuestione los hallazgos previos y promueva un debate enriquecedor sobre el uso de la taquilla como indicador de éxito.

## 2.- ESTADO DEL ARTE

En la industria cinematográfica, evaluar el éxito de una película es una tarea de suma importancia que involucra diversos aspectos, desde la calidad artística hasta su desempeño comercial; siendo la recaudación en taquilla, uno de los indicadores más prominentes. Los ingresos generados por una película en las salas de cine no solo reflejan su popularidad y aceptación entre el público, sino que también son un indicador clave de su rentabilidad financiera y su impacto en la industria. Este trabajo se centra en explorar el análisis predictivo de la rentabilidad en taquilla del cine, utilizando modelos para optimizar los lanzamientos y las estrategias de marketing, se pretende examinar cómo este indicador puede ser utilizado como una medida efectiva para predecir el éxito de una película y guiar decisiones estratégicas en la producción y distribución cinematográfica.

Numerosos trabajos avalan la idea de utilizar los ingresos por taquilla (box-office, su denominación en inglés) como indicador del éxito de una producción cinematográfica.

Entre estos trabajos destaca el de De Vany y Walls (1997) que exploran la dinámica del mercado de películas cinematográficas desde una perspectiva económica, estableciendo el ingreso por taquilla como indicador de éxito en la industria cinematográfica por diversas razones:

- i) Proporciona una medida directa de los ingresos generados por una película en las salas de cine, representando una parte significativa de sus ganancias totales.
- ii) La cantidad de personas que acuden a ver una película en los cines es un reflejo directo de su aceptación y atracción. La película que alcanza un rendimiento alto en taquilla generalmente indica que ha capturado el interés y la atención del público
- iii) La taquilla proporciona una medida objetiva que permite evaluar el éxito relativo de una película en comparación con otras en el mismo período de tiempo.
- iv) Los resultados en taquilla pueden influir en la percepción de una película entre los inversores, productores y distribuidores, así como en la asignación de recursos y presupuestos para futuros proyectos. Además, puede aumentar la visibilidad y el prestigio de los cineastas y actores involucrados en la película, lo que puede tener un impacto positivo en sus futuras carreras y oportunidades en la industria.

En Hennig-Thurau et al. (2006), la taquilla se establece como un indicador crucial de éxito, explorando cómo este se entrelaza con la crítica, el reconocimiento público, y los premios. El prestigio puede estar intrínsecamente ligado a la obtención de nominaciones y galardones en festivales de renombre, lo que eleva el atractivo tanto para el público como para los inversores. Además, el prestigio puede moldear la percepción de la calidad de una película y la reputación de los talentos involucrados en su realización.

Además, en este trabajo, se explora la interacción entre el prestigio de una película y su éxito en taquilla, y viceversa. Por ejemplo, una película elogiada por la crítica y laureada con premios prestigiosos puede atraer a un público más amplio, lo que potencialmente incrementa su recaudación en taquilla. Del mismo modo, una película con un gran éxito en taquilla puede ganar reconocimiento y prestigio en la industria, abriendo puertas a nuevas oportunidades para los cineastas y actores involucrados.

Otros autores como Pangarker y Smit (2013), analizan la existencia de factores, que, a pesar de la evolución de la industria cinematográfica y los cambios en los gustos del público a lo largo de las décadas, siguen siendo consistentes en su influencia en el éxito medido a través de los ingresos en taquilla.

Pero ¿cuáles son los factores que hacen que una película se convierta en un éxito en taquilla?

De Vany y Walls (2006) logra darle respuesta a esta interrogante controlando dos aspectos de notable importancia:

- i) Características y reputación de la película: Considerando dentro de las características una combinación única de elementos como la trama, el elenco, la dirección, la cinematografía y el estilo, la reputación, la publicidad, las críticas y la recepción del público.
- ii) Factores temporales, dado que diferentes aspectos del tiempo o el momento influyen en la acogida, en el sentido de que una película estrenada con anterioridad en el mercado americano puede los ingresos en otros mercados.

Por otro lado, Pangarker y Smit (2013), destacan el papel relevante que el coste de producción tiene en la taquilla de las películas, obteniendo mejores resultados aquellas con presupuestos más elevados. Además, demuestran que las películas respaldadas por grandes estudios

cinematográficos gozan de mayores probabilidades de éxito, lo que subraya la importancia de la distribución y promoción respaldadas por grandes empresas. Por otro lado, observan que las películas con nominaciones a premios importantes, como aquellos que otorga la Academia de Artes y Ciencias Cinematográfica (premios Oscars), tienden a atraer más atención y a generar mayores ingresos en taquilla. Además, apuntan a la presencia de secuelas de películas exitosas previas como un factor clave para el éxito en taquilla.

Otro concepto que tiende a estar muy relacionado con el éxito en taquilla es el de "star power" (Selvaretnam y Yang, 2015), que identifica la capacidad única que poseen ciertas estrellas de cine para atraer al público hacia una película, debido a su reputación, carisma, talento y popularidad. Es esa fuerza magnética que despierta el interés del espectador y puede hacer que una película destaque en la taquilla, convirtiéndose en un fenómeno cultural.

También se debe tener en cuenta que el éxito de una película en taquilla es el resultado de una intrincada red de variables que influyen en su rendimiento financiero y su popularidad entre el público. En el trabajo de Lozano et al (2012) las variables determinantes del éxito de una película vienen dadas por la calidad y las expectativas del público. Considerando variables claves de calidad, la calidad técnica, el presupuesto de producción, las críticas y reseñas y los premios y reconocimientos. Midiendo las expectativas a través del presupuesto en marketing y promoción, la dirección y el reparto (actores y actrices). Además, en este estudio se controlan variables como el género, la estacionalidad y la competencia. desempeñando cada una de estas variables un papel único en el ecosistema del cine y determinando en última instancia el éxito o fracaso de una película en las taquillas de todo el mundo.

El Modelo Basado en la Crítica, desarrollado por Terry et al. (2004), revela que la crítica de una película tiene un impacto significativo y positivo en su recaudación cinematográfica.

En cuanto a las técnicas utilizadas en la predicción o caracterización del éxito de las películas medidas a través de los ingresos en taquilla podemos destacar el trabajo reciente de Xu (2024) que utiliza Machine Learning haciendo un análisis detallado de la correlación entre variables como la popularidad, el presupuesto y la duración de una película observando que películas con mayor popularidad, presupuestos más altos y una duración óptima tienden a generar mayores ingresos en taquilla. Estos factores se consideran cruciales para atraer audiencias y lograr el éxito deseado.

En el mismo sentido, Bhadrashetty y Patil (2024) abordan el tema del éxito en taquilla desde la perspectiva del análisis predictivo utilizando modelos predictivos abordando un modelo ARIMA (AutoRegressive Integrated Moving Average) como herramienta para predecir el éxito y la calificación de las películas. Este modelo es ampliamente utilizado en el análisis de series temporales y puede ser aplicado al análisis de datos relacionados con películas para pronosticar su desempeño en términos de taquilla y evaluación crítica.

El Modelo de Regresión Múltiple basado en la identificación del éxito medido como la recaudación en taquilla y múltiples variables independientes, como género, costos de producción, críticas y premios, es el enfoque adoptado por Litman (1983), que permite identificar variables significativas que influyen en la recaudación y predecir el éxito financiero de una película.

Por otro lado, el Modelo de Duración, propuesto por Sawhney y Eliashberg (1996), se enfoca en las diferentes etapas del ciclo de vida de una película para determinar la duración óptima de su exhibición en cines. Este enfoque considera la temporalidad en la recaudación y la respuesta del público a lo largo del tiempo.

## 3.- METODOLOGÍA

### 3.1.- La base de datos.

Para el análisis de la identificación de los determinantes del éxito en taquilla de las películas con participación española en la producción, en territorio español, se ha confeccionado una base de datos que combina datos de películas proyectadas en el año 2022, proporcionados por el Ministerio de Cultura del Gobierno de España, con información detallada de variables recopiladas de FilmAffinity.

El Ministerio de Cultura del Gobierno de España publica un listado de películas proyectadas en el país cada año, que incluye información básica como la compañía distribuidora, la fecha de estreno, la recaudación y el número de espectadores. Por lo tanto, se ha hecho uso de esta fuente de datos oficial, para elaborar una lista exhaustiva de las películas con participación española en la producción, estrenadas durante el año 2022.

Con el objetivo de enriquecer la base de datos y profundizar en el análisis econométrico de las películas, se ha complementado la información proporcionada por el Ministerio de Cultura con datos adicionales obtenidos del sistema de recomendación, FilmAffinity; una plataforma en línea que ofrece una amplia gama de información sobre el mundo del cine, incluyendo valoraciones de usuarios, críticas, elenco y género. Se ha utilizado esta plataforma para recopilar datos detallados sobre las características y atributos de cada película, enriqueciendo así la base de datos con información relevante para la investigación.

### 3.2.- Los datos.

La base de datos la constituye una muestra formada por 497 películas con participación española en la producción, que, como se mencionó anteriormente, fueron proyectadas en cines españoles en el año 2022. La muestra cuenta también con largometrajes que permanecieron en las salas desde años anteriores y reestrenos.

Para este estudio, se creó una base de datos teniendo como punto de partida el listado de películas que brinda cada año el Ministerio de Cultura del Gobierno en su página web contando con datos como: título de la película, distribuidora, fecha de estreno, recaudación y espectadores. Para la confección se utilizó el lenguaje de programación Python, que permitió la recopilación, limpieza

y estructuración de los datos de manera eficiente, asegurando la calidad y la consistencia de la información utilizada en el análisis.

La información recopilada fue extraída principalmente de la plataforma FilmAffinity, y se utilizaron fuentes auxiliares como la página del Ministerio de Cultura del Gobierno e IMDb (Internet Movie Database). El algoritmo se basó en una serie de peticiones que se hicieron a los sitios mencionados para de esta manera, identificar y extraer las variables clave que pudieran influir en la recaudación en taquilla y ser relevantes en el análisis según los estudios que antecedieron.

Las variables que se extrajeron de estos sitios incluyen las críticas, que recoge el número de críticas que ha recibido la película tanto de usuarios registrados en el sitio como de críticos profesionales; la duración, que indica la duración de la película en minutos y el rating, que refleja la valoración promedio otorgada por los usuarios.

Además, se consideraron variables relacionadas con premios y nominaciones, como el premio Goya a mejor película, que indica si la película ganó el premio, y varias variables que registran nominaciones en diferentes festivales y premios relevantes en España, como: el Festival de Málaga, Festival de San Sebastián, Premios Feroz, Premios Platino, Festival de Valladolid-Seminci, Premios Gaudí, Premios Gijón, Premios Forqué y festival de Sevilla.

Otra variable incluida fueron los directores, donde se tuvo en cuenta directores populares y galardonados de España (Pedro Almodóvar, Isabel Coixet, Jaume Balagueró, Fernando León de Aranoa, Álex de la Iglesia, Icíar Bollaín, Pablo Berger, Santiago Segura, Fernando Colomo, Enrique Gato, Javier Fesser, Imanol Uribe, Montxo Armendáriz, Albert Serra, Carlos Vermut), en esta variable se almacenaron las películas que fueron dirigidas por los nombres mencionados; igual tratamiento recibió el reparto, con los actores y actrices más populares y mejor pagados (Paz Vega, Alicia Borrachero, Santiago Segura, Astrid Bergès-Frisbey, Óscar Jaenada, Elena Anaya, Jordi Mollá, Javier Botet, José Coronado, Javier Bardem, Penélope Cruz, Antonio Banderas, Paco León, Carmen Machi, Luis Tosar, Javier Gutiérrez, Blanca Suárez, Miguel Ángel Muñoz, Leo Harlem, Miren Ibarguren)

También se tuvieron en cuenta las películas que pertenecieran a las distribuidoras más importantes (Warner Bros, Sony, Paramount Int, Walt Disney, Filmax, DeAPlaneta, Entertainment One Films,

European Dreams Factory, A Contracorriente, Alfa Pictures, Avalon Distribución, Syldavia, Bteam Pictures, Wanda, Universal), y se separaron individualmente las pertenecientes a las distribuidoras Sony y Warner\_Bros por el marcado renombre que tienen en la industria cinematográfica.

Teniendo en cuenta el contexto del estreno se consideraron variables como los estrenos de fin de semana, semana santa, semana de navidad, meses de verano, semana de la fiesta del cine y el puente de la Constitución, fechas que pueden repercutir en la afluencia de espectadores a las salas de cine.

Por último, se incluyeron variables para identificar el género de la película, como animación, fantástico, thriller, drama, romance, documental, terror, aventuras, comedia y acción. Una vez identificadas las variables, se integraron en una base de datos unificada. Este proceso implicó la normalización de datos para asegurar la consistencia entre diferentes fuentes, la eliminación de duplicados y el manejo de valores faltantes. El uso de Python fue crucial en esta etapa, permitiendo la automatización de tareas repetitivas y complejas de manipulación de datos.

### 3.3.- Las variables.

En el marco del análisis econométrico del éxito en taquilla de las películas exhibidas en España en 2022, las variables desempeñan un papel crucial permitiendo explorar y cuantificar las relaciones entre los diferentes factores. Estas variables proporcionan datos fundamentales que permiten explorar las relaciones causales y predictivas y el rendimiento financiero de las películas en la taquilla. A continuación, se presenta una descripción detallada de las variables utilizadas en este estudio, destacando su relevancia en el análisis y su fuente de información correspondiente:

Tabla 2. Descripción de las variables.

VARIABLE	FUENTE	DESCRIPCIÓN
ESPECTADORES	Página web del Ministerio de Cultura	Cantidad de espectadores que visitaron las salas de cine en España.
FECHA_ESTRENO	Página web del Ministerio de Cultura	Fecha de estreno en cines españoles, se tuvieron en cuenta las estrenadas en los últimos 4 meses del 2021.
RATING	Filmaffinity	Calificación promedio otorgada por los usuarios de la plataforma. Valores 1-10
CRITICAS	Filmaffinity	Críticas realizadas por críticos profesionales y usuarios registrados en la plataforma.
RECAUDACION	Página web del Ministerio de Cultura	Recaudación en salas de cine.

<b>GOYA_MEJOR_PEL</b>	Filmaffinity	Películas ganadoras de Premios Goya, mejor película. Valor 1
<b>NOM_GOYA_MEJOR_PEL</b>	Filmaffinity	Películas nominadas a los Premios Goya, mejor película. Valor 1
<b>NOM_FEST_MALAGA</b>	Filmaffinity	Películas nominadas en el Festival de Málaga. Valor 1
<b>NOM_SAN_SEBASTIAN</b>	Filmaffinity	Películas nominadas en el Festival de San Sebastián. Valor 1
<b>NOM_FEROZ</b>	Filmaffinity	Películas nominadas a los Premios Feroz. Valor 1
<b>NOM_PLATINO</b>	Filmaffinity	Películas nominadas a los Premios Platino. Valor 1
<b>NOM_VALLADOLID</b>	Filmaffinity	Películas nominadas en el Festival de Valladolid-Seminci. Valor 1
<b>NOM_GAUDI</b>	Filmaffinity	Películas nominadas a los Premios Gaudí. Valor 1
<b>NOM_GIJON</b>	Filmaffinity	Películas nominadas a los Premios Gijón. Valor 1
<b>NOM_FORQUE</b>	Filmaffinity	Películas nominadas a los Premios Forqué. Valor 1
<b>NOM_SEVILLA</b>	Filmaffinity	Películas nominadas en el Festival de Sevilla. Valor 1
<b>DIRECTORES</b>	IMDb	Películas dirigidas por los directores españoles más relevantes Valor 1
<b>DISTRIB_MAS_IMP</b>	Página web del Ministerio de Cultura	Películas asociadas a las distribuidoras de cine más importantes Valor 1
<b>SONY</b>	Página web del Ministerio de Cultura	Películas asociadas a la distribuidora de cine Sony Pictures Valor 1
<b>WARNER_BROS</b>	Página web del Ministerio de Cultura	Películas asociadas a la distribuidora de cine Warner Bros Valor 1
<b>REPARTO</b>	IMDb	Películas donde participan actores importantes españoles Valor 1
<b>EST_FIN_DE_SEMANA</b>	Página web del Ministerio de Cultura	Películas estrenadas un fin de semana Valor 1
<b>EST_SEMANA_SANTA</b>	Página web del Ministerio de Cultura	Películas estrenadas en semana santa Valor 1
<b>EST_SEMANA_NAVIDAD</b>	Página web del Ministerio de Cultura	Películas estrenadas en la semana de navidad Valor 1
<b>EST_VERANO</b>	Página web del Ministerio de Cultura	Películas estrenadas en las vacaciones de verano Valor 1
<b>EST_FIESTA_DEL_CINE</b>	Página web del Ministerio de Cultura	Películas estrenadas durante la fiesta del cine Valor 1
<b>EST_PTE_CONSTITUCION</b>	Página web del Ministerio de Cultura	Películas estrenadas durante el puente de la Constitución Valor 1
<b>ANIMACION</b>	Filmaffinity	Películas que pertenecen al género animación Valor 1
<b>FANTASTICO</b>	Filmaffinity	Películas que pertenecen al género fantástico Valor 1
<b>THRILLER</b>	Filmaffinity	Películas que pertenecen al género thriller Valor 1
<b>DRAMA</b>	Filmaffinity	Películas que pertenecen al género drama Valor 1
<b>ROMANCE</b>	Filmaffinity	Películas que pertenecen al género romance Valor 1
<b>DOCUMENTAL</b>	Filmaffinity	Películas que pertenecen al género documental Valor 1
<b>TERROR</b>	Filmaffinity	Películas que pertenecen al género terror Valor 1
<b>AVENTURAS</b>	Filmaffinity	Películas que pertenecen al género aventuras Valor 1
<b>COMEDIA</b>	Filmaffinity	Películas que pertenecen al género comedia Valor 1
<b>ACCION</b>	Filmaffinity	Películas que pertenecen al género acción Valor 1

### 3.4.- Estadísticos descriptivos. Uso de STATA

Teniendo en cuenta la versatilidad, facilidad y el excelente soporte de datos con que cuenta el software STATA se ha empleado como herramienta para el análisis estadístico de los datos recopilados.

En este contexto, Stata permite calcular medidas centrales como la media, la mediana y la moda, revelando tendencias centrales cruciales para entender la distribución de las variables analizadas. De manera general los estadísticos descriptivos proporcionan información clave sobre la distribución, centralidad y dispersión de los datos, lo que ayuda a comprender mejor la naturaleza y las propiedades del conjunto de datos en estudio.

Para reducir la influencia de valores extremos, haciendo que los datos sean más manejables y menos sensibles a diferencias absolutas, se decide usar en el análisis el logaritmo de la recaudación en lugar de la recaudación misma como indicador de éxito, esto proporciona una medida más robusta y equitativa del rendimiento relativo de películas, ayudando a evaluar de manera más precisa el éxito percibido en términos financieros y su impacto en el público.

Como se puede observar, existen 37 variables de las cuales solo utilizaremos en el modelo aquellas que se han utilizado previamente como indicadores de éxito en estudios similares, mencionados en el apartado dedicado a la revisión de la literatura anterior. Teniendo en cuenta que existen variables individuales que pertenecen a la misma categoría (las nominaciones y los estrenos en fechas especiales), se decide incluir cada grupo en una única variable (nominaciones, est\_fechas\_esp) para así simplificar el modelo, reducir la presencia de multicolinealidad y poder hacer una mejor comparación con el resto de las variables, quedando de la siguiente forma:

Tabla 3. Estadísticos descriptivos variable log\_recaudacion

<i>Variable</i>	<i>Obs</i>	<i>Mean</i>	<i>Std. dev.</i>	<i>Min</i>	<i>Max</i>
<i>log_recaudacion</i>	497	7.283937	3.026485	1.098612	16.56323

La variable log\_recaudación, muestra que la recaudación de las películas tiene una media de aproximadamente 7.28 y una desviación estándar de 3.03. El rango va desde 1.10 hasta 16.56 en la escala logarítmica. Esto sugiere que las películas de la muestra tienen una variabilidad

significativa en términos de recaudación, con algunos éxitos notables (valores altos de log\_recaudación) y otros menos exitosos (valores bajos).

Tabla 4. Estadísticos descriptivos variables discretas

<i>Variable</i>	<i>Freq.</i>	<i>Percent.</i>	<i>Cum.</i>
<i>nominaciones</i>			
0	325	65.39	65.39
1	172	34.61	100.00
<i>est_fechas_esp</i>			
0	334	67.20	67.20
1	163	32.80	100.00
<i>reparto</i>			
0	449	90.34	90.34
1	48	9.66	100.00
<i>directores</i>			
0	468	94.16	94.16
1	29	5.84	100.00
<i>distrib_mas_imp</i>			
0	350	70.42	70.42
1	147	29.58	100.00
<i>goya mejor pel</i>			
0	481	96.78	96.78
1	16	3.22	100.00
<i>rating</i>			
0	111	22.33	22.33
2	1	0.20	22.54
2.3	2	0.40	22.94
2.5	2	0.40	23.34
2.8	1	0.20	23.54
2.9	1	0.20	23.74
3.1	1	0.20	23.94
3.2	2	0.40	24.35
3.3	1	0.20	24.55
3.4	2	0.40	24.95
3.5	1	0.20	25.15
3.6	1	0.20	25.35
3.7	2	0.40	25.75
3.8	3	0.60	26.36
3.9	4	0.80	27.16
4	2	0.40	27.57
4.1	2	0.40	27.97
4.2	6	1.21	29.18
4.3	2	0.40	29.58
4.4	3	0.60	30.18
4.5	10	2.01	32.19
4.6	8	1.61	33.80
4.7	3	0.60	34.41
4.8	4	0.80	35.21
4.9	4	0.80	36.02
5	10	2.01	38.03
5.1	11	2.21	40.24
5.2	10	2.01	42.25
5.3	11	2.21	44.47
5.4	14	2.82	47.28
5.5	10	2.01	49.30
5.6	12	2.41	51.71
5.7	11	2.21	53.92
5.8	20	4.02	57.95
5.9	12	2.41	60.36
6	20	4.02	64.39
6.1	14	2.82	67.20
6.2	16	3.22	70.42
6.3	11	2.21	72.64
6.4	14	2.82	75.45
6.5	20	4.02	79.48
6.6	16	3.22	82.70
6.7	7	1.41	84.10
6.8	16	3.22	87.32
6.9	7	1.41	88.73
7	11	2.21	90.95
7.1	14	2.82	93.76
7.2	7	1.41	95.17
7.3	1	0.20	95.37
7.4	6	1.21	96.58
7.5	6	1.21	97.79
7.6	5	1.01	98.79

7.7	3	0.60	99.40
7.8	2	0.40	99.80
8	1	0.20	100.00
<i>accion</i>			
0	492	98.99	98.99
1	5	1.01	100.00
<i>comedia</i>			
0	405	81.49	81.49
1	92	18.51	100.00
<i>drama</i>			
0	310	62.37	63.37
1	187	37.63	100.00
<i>thriller</i>			
0	448	90.14	90.14
1	49	9.86	100.00
<i>romance</i>			
0	474	95.37	95.37
1	23	4.63	100.00
<i>documental</i>			
0	291	58.55	58.55
1	206	41.45	100.00
<i>fantastico</i>			
0	479	96.38	96.38
1	18	3.62	100.00
<i>terror</i>			
0	473	95.17	95.17
1	24	4.83	100.00
<i>animacion</i>			
0	478	96.18	96.18
1	19	3.82	100.00
<i>aventuras</i>			
0	490	98.59	98.59
1	7	1.41	100.00

Las variables nominaciones, *est\_fecha\_estreno*, *reparto*, *directores*, *distrib\_mas\_imp* y *goya\_mejor\_pel*, acción, comedia, drama, thriller, romance, documental, fantastico, terror, animacion y aventuras al ser binarias sugieren 0 como valor mínimo para determinar ausencia de características y 1 como valor para indicar presencia de alguna de estas categorías en las películas de la muestra.

La variable *rating* indica una medida de la calificación promedio de cada película, varía en términos de la calificación recibida, con un rango desde 0 hasta 8.

### 3.5.- El modelo.

Para analizar de una manera adecuada y eficiente los factores que influyen en el éxito de las películas seleccionadas, definido en términos de la recaudación en taquilla, se utilizará un modelo de regresión lineal múltiple. Esta metodología, entre otras cosas, permite la inclusión de múltiples variables independientes, la interpretación clara de los coeficientes, eficiencia en la predicción de la variable dependiente cuando se dispone de múltiples predictores y proporciona una base sólida para identificar y cuantificar relaciones lineales.

La selección de la variable dependiente y las variables independientes para este estudio se basó en una revisión exhaustiva de la literatura existente sobre los factores que influyen en el éxito de las películas, medido en términos de recaudación en taquilla. Como ya mencionaba en capítulos anteriores, diversos estudios han identificado una serie de variables que se consideran determinantes para el éxito financiero de una película. En particular, se han tomado en cuenta las siguientes variables:

Variable dependiente: log\_recaudación

Variables independientes: nominaciones, est\_fechas\_esp, reparto, directores, distrib\_mas\_imp, goya\_mejor\_pel, rating, accion, comedia, drama, thriller, romance, documental, fantastico, terror, animacion, aventuras

La regresión lineal se fundamenta en la siguiente expresión matemática:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

Donde  $Y$  representa la variable dependiente,  $\beta_0$  es el término de intersección o constante,  $\beta_1, \beta_2, \dots, \beta_p$  son los coeficientes de regresión que cuantifican la influencia de cada variable independiente  $X_1, X_2, \dots, X_p$  sobre  $Y$ , y  $\epsilon$  representa el término de error que captura la variabilidad no explicada por las variables independiente.

Adecuando la expresión al estudio en cuestión quedaría de la siguiente forma:

$$\begin{aligned} \log(\text{recaudacion}) = & \beta_0 + \beta_1(\text{nominaciones}) + \beta_2(\text{est_fechas_esp}) + \beta_3(\text{reparto}) + \\ & \beta_4(\text{directores}) + \beta_5(\text{distrib_mas_imp}) + \beta_6(\text{goya_mejor_pel}) + \beta_7(\text{rating}) + \\ & \beta_8(\text{accion}) + \beta_9(\text{comedia}) + \beta_{10}(\text{drama}) + \beta_{11}(\text{thriller}) + \beta_{12}(\text{romance}) + \\ & \beta_{13}(\text{documental}) + \beta_{14}(\text{fantastico}) + \beta_{15}(\text{terror}) + \beta_{16}(\text{animacion}) + \\ & \beta_{17}(\text{aventuras}) + \epsilon \end{aligned}$$

## 4.- RESULTADOS

Teniendo en cuenta las variables mencionadas en el apartado anterior, analizamos el modelo de regresión lineal múltiple:

Tabla 5. Estadísticas del modelo de regresión

<i>Number of obs</i>	<i>F(8, 488)</i>	<i>Prob &gt; F</i>	<i>R-squared</i>	<i>Adj R-squared</i>
497	10.73	0.0000	0.2757	0.2500

El valor del coeficiente de determinación (R-cuadrado) indica que aproximadamente el 27.57% de la variabilidad en la recaudación de taquilla (*log\_recaudacion*) puede explicarse a través de variables independientes incluidas en el modelo, y a su vez el R-cuadrado ajustado (0.2500) sugiere una ligera reducción cuando se ajusta por el número de predictores en el modelo, lo que es normal y muestra una cantidad similar de explicación de la variabilidad; al final como el objetivo del análisis es caracterizar y describir, contrastar con estudios anteriores, no predecir, el valor de este coeficiente solo tendría carácter informativo. En términos de significancia global se puede apreciar por el valor del estadístico F (10.73) y su correspondiente valor p (0.0000) que el modelo es estadísticamente significativo, lo que implica que al menos una de las variables independientes tiene un efecto significativo sobre la variable dependiente.

Tabla 6. Modelo de regresión lineal múltiple en STATA

<i>log_recaudacion</i>	<i>Coefficient</i>	<i>Std. err.</i>	<i>t</i>	<i>P&gt; t </i>	<i>[95% conf. interval]</i>	
<i>nominaciones</i>	-0.0562099	0.3038052	-0.19	0.853	-0.6531655	0.5407457
<i>est_fechas_imp</i>	-0.7862868	0.254982	-3.08	0.002	-1.287308	-0.2852653
<i>reparto</i>	1.458574	0.4410731	3.31	0.001	0.591897	2.325251
<i>directores</i>	0.2962423	0.5467324	0.54	0.588	-0.7780479	1.370532
<i>distrib_mas_imp</i>	2.240328	0.2868282	7.81	0.000	1.676731	2.803925
<i>goya_mejor_pel</i>	0.2872194	0.7243415	0.40	0.692	-1.13606	1.710499
<i>rating</i>	0.118145	0.05128	2.30	0.022	0.0173835	0.2189064
<i>accion</i>	0.7905806	1.23293	0.64	0.522	-1.632039	3.2132
<i>comedia</i>	0.8840989	0.4055839	2.18	0.030	0.0871553	1.681042
<i>drama</i>	-0.4723587	0.3994603	-1.18	0.238	-1.25727	0.3125523
<i>thriller</i>	0.9088306	0.4446456	2.04	0.042	0.0351337	1.782527
<i>romance</i>	0.0627418	0.5867321	0.11	0.915	-1.090145	1.215629
<i>documental</i>	-0.2375938	0.4567002	-0.52	0.603	-1.134977	0.6597895
<i>fantastico</i>	-0.0889922	0.6618051	-0.13	0.893	-1.389392	1.211408
<i>terror</i>	0.0701446	0.6108562	0.11	0.909	-1.130144	1.270433
<i>animacion</i>	0.6668746	0.7127923	0.94	0.350	-0.7337116	2.067461
<i>aventuras</i>	1.522364	1.062297	1.43	0.152	-0.5649752	3.609703
<i>_cont</i>	6.161055	0.4960748	12.42	0.000	5.186303	7.135806

## 4.1 Interpretación de los coeficientes de regresión.

est\_fechas\_esp: Aunque pueda parecer contradictorio que el valor p (0.002) indique un efecto estadísticamente significativo teniendo un coeficiente de -0.7862, donde cada unidad de aumento en la variable de fechas de estreno específicas se relaciona con una disminución promedio de 0.6745 unidades en la recaudación en taquilla; podríamos pensar que más importante que la elección de la fecha de estreno sería el presupuesto gastado en promoción de la película antes del estreno, aunque esta variable no ha sido considerada en el análisis por las dificultades de obtener datos realistas de las productoras acerca del presupuesto en marketing para cada una de las películas.

reparto: El coeficiente de 1.4585, con un valor p de 0.001, sugiere que la presencia de actores populares y reconocidos en el reparto tiene un efecto positivo y significativo sobre la recaudación logarítmica, aumentándola en promedio en 1.4585 unidades.

distrib\_mas\_imp: Con un coeficiente de 2.2403, cada incremento en la estrategia de distribución está asociado con un aumento promedio de 2.2403 unidades en la recaudación en taquilla (logaritmo). Este efecto es altamente significativo (valor p = 0.000), indicando que una distribución efectiva y una sólida estrategia de marketing pueden tener un impacto significativo en el éxito financiero de una película.

rating: Con un coeficiente de 0.1181, indica que por cada unidad adicional en la calificación, el logaritmo de la recaudación aumenta. Es estadísticamente significativo al nivel convencional del 5%, teniendo un valor p de 0.022, lo que sugiere que las películas mejor calificadas tienden a recaudar más, posiblemente porque los espectadores confían en estas calificaciones para tomar decisiones.

comedia: Siendo una variable significativa con un coeficiente de 0.8844, sugiere que las películas de comedia parecen ser populares y una opción recurrente entre el público, incrementando así su éxito en taquilla.

thriller: Con un coeficiente de 0,9088 y un valor p de 0.042 esta variable indica que las personas que asisten a las salas de cine están interesadas en este género y que podría influir en la decisión de ver una película, aumentando el logaritmo de la recaudación en 0.9088 unidades.

## 4.2 Diagnóstico del modelo. Evaluación de los supuestos de la regresión lineal.

Una vez estimado el modelo de regresión lineal para investigar por qué utilizar la taquilla como indicador de éxito, se evalúa la validez de los supuestos subyacentes para garantizar la robustez de los resultados estadísticos obtenidos. Los supuestos críticos evaluados incluyen la normalidad de los residuos, la homocedasticidad (varianza constante) y la ausencia de multicolinealidad entre las variables independientes.

### 4.2.1 Normalidad de los residuos.

Para verificar la normalidad de los residuos se realizó el test de Shapiro-Wilk; como se muestra en la tabla, el valor de  $W$  (0.99782) cercano a 1 sugiere que los residuos se ajustan razonablemente bien a una distribución normal. El valor  $p$  (0.77668) asociado a la prueba indica que no hay suficiente evidencia para rechazar la hipótesis nula de normalidad de los residuos a un nivel de significancia del 5% ( $\alpha = 0.05$ ). Por lo tanto, basado en este análisis, se puede aceptar que los residuos son aproximadamente normales, lo cual es consistente con los supuestos de la regresión lineal.

Tabla 7. Test de Shapiro-Wilk

<i>Variable</i>	<i>Obs</i>	<i>W</i>	<i>V</i>	<i>z</i>	<i>Prob &gt; z</i>
<i>residuos</i>	497	0.99782	0.729	-0.761	0.77668

### 4.2.2 Homocedasticidad.

Para evaluar la homocedasticidad de los residuos en el modelo utilizado, se empleó el test de Breusch-Pagan/Cook-Weisberg, resultando lo siguiente:

H0: Constant variance

$\chi^2(1) = 53.92$

Prob >  $\chi^2 = 0.0000$

Aunque el test demostró presencia de heterocedasticidad con un valor de p muy pequeño rechazando la hipótesis nula de varianza constante, no afecta significativamente la validez del modelo teniendo en cuenta los siguientes factores:

- La inspección visual de los residuos frente a los valores ajustados no revela patrones claros de variación de varianza, sugiriendo que la heterocedasticidad presente no es severa.
- Las principales variables independientes mostraron coeficientes altamente significativos, incluso bajo la presencia de heterocedasticidad.
- Estudios previos en el ámbito del análisis como el de Vany y Walls (2006). han demostrado que aunque existe heterocedasticidad, los modelos utilizados han mantenido su validez y las conclusiones derivadas han sido fiables.
- Considerando el tamaño de la muestra (497 observaciones) las estimaciones de los coeficientes son menos susceptibles a la influencia de heterocedasticidad.

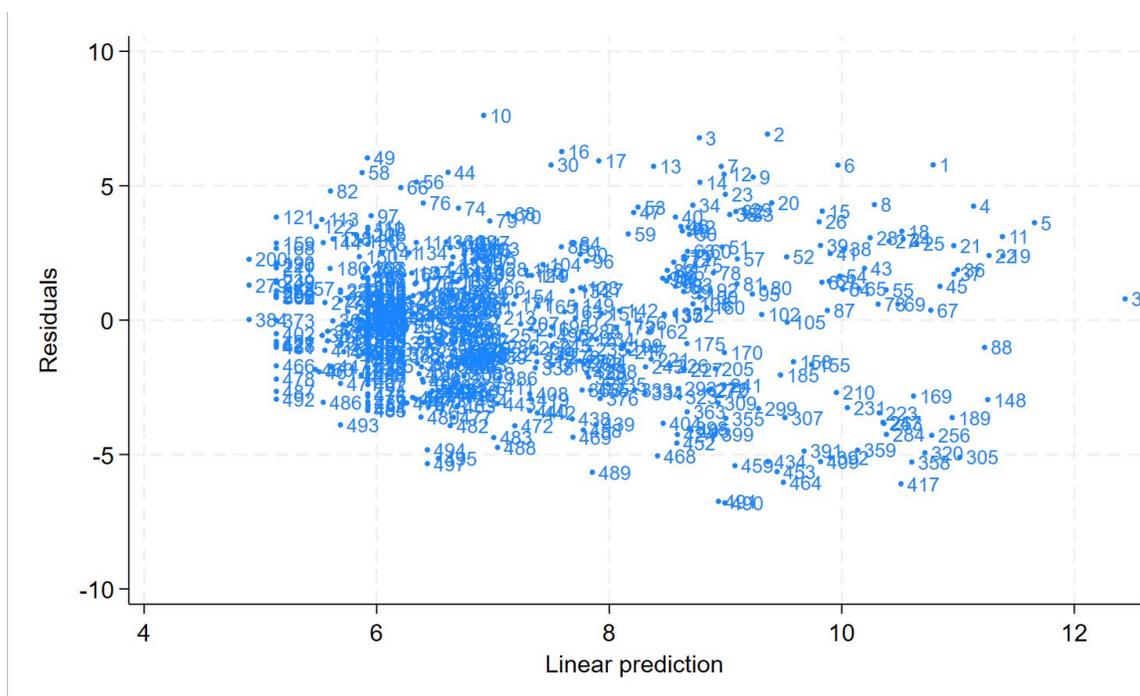


Figura 1. Gráfico de dispersión de residuos

### 4.2.3 Multicolinealidad entre variables independientes.

Se calculó el factor de inflación de la varianza (VIF) para todas las variables independientes del modelo, como se puede observar los valores son cercanos a 1, indicando que la multicolinealidad no es un problema entre las variables predictoras, asegurando la fiabilidad de las estimaciones de los coeficientes.

Tabla 8. Factor de inflación de la varianza (VIF)

<i>Variable</i>	<i>VIF</i>	<i>1/VIF</i>
<i>documental</i>	3.66	0.273072
<i>drama</i>	2.71	0.369105
<i>comedia</i>	1.80	0.557052
<i>nominaciones</i>	1.51	0.661756
<i>animacion</i>	1.35	0.739933
<i>rating</i>	1.30	0.771965
<i>thriller</i>	1.27	0.786680
<i>terror</i>	1.24	0.806027
<i>distrib_mas_imp</i>	1.24	0.806623
<i>reparto</i>	1.23	0.814313
<i>directores</i>	1.19	0.841601
<i>goya_mejor_pel</i>	1.18	0.845567
<i>aventuras</i>	1.13	0.882091
<i>fantastico</i>	1.11	0.904131
<i>romance</i>	1.10	0.909733
<i>accion</i>	1.10	0.913036
<i>est_fechas_esp</i>	1.04	0.964599
<b>Mean VIF</b>	<b>1.48</b>	

Los resultados obtenidos revelan una caracterización detallada de los factores que influyen en la recaudación cinematográfica medida en logaritmo. Variables como las fechas especiales de estreno, el reparto, la distribución y géneros como la comedia y el thriller juegan roles fundamentales para determinar el éxito de una película. Las fechas especiales de estreno representan momentos estratégicos para maximizar la asistencia a las salas de cine, derivando en un impacto negativo si se escogen fechas posiblemente menos favorables o competitivas. La reputación y popularidad del reparto pueden ser factores decisivos para persuadir a los espectadores, desempeñan un papel crucial en la percepción de calidad de la película. Respecto a las distribuidoras, una distribución adecuada no solo amplía la accesibilidad de la película, sino que también optimiza su visibilidad y capacidad para mantener una presencia mantenida en el mercado, maximizando así las oportunidades de generar ingresos a largo plazo. Los géneros no solo definen el tipo de experiencia que ofrece la película al público, sino que también influyen en su atractivo y en la base de seguidores potenciales. En conjunto, estos elementos interactúan de manera sinérgica, definiendo el potencial de éxito financiero de una película. Entender cómo cada uno de estos factores contribuye a la experiencia general del público y a los resultados económicos es crucial para desarrollar estrategias efectivas que incrementen tanto la rentabilidad como el impacto de una película en el mercado global.

## 5.- CONCLUSIONES

El estudio ha demostrado que varios factores juegan roles significativos en la determinación del éxito financiero de una película en taquilla. En productos como las películas la percepción del consumidor es vital, de ahí la importancia de tener una distribuidora con un profundo conocimiento de las tendencias del mercado y del comportamiento del público, una correcta distribución elige una fecha de estreno óptima para evitar la competencia directa con otras películas importantes, una campaña de marketing bien ejecutada para generar anticipación y la atracción de grandes audiencias, tiene la capacidad de asegurar una amplia red de salas de cine para que la película esté disponible en más ubicaciones y llevarla a mercados internacionales, logrando así maximizar la visibilidad, multiplicar las oportunidades de ingresos y el éxito en taquilla.

Un reparto de actores y actrices reconocidos y talentosos aumenta el prestigio de la película y atrae a audiencias que buscan contenido de alta calidad; el reconocimiento de un nombre con marcada trayectoria aumenta la visibilidad y genera interés incluso antes de su estreno. Además, un elenco fuerte puede atraer inversiones adicionales y facilitar acuerdos de distribución más ventajosos. Este componente presenta ventajas sobre otras variables como los directores, que a pesar de que juegan un papel crucial en la realización de una película, están detrás de cámara y el espectador tiende a relacionarse y conectar más fácilmente con las personas que ven en la pantalla.

El rating actúa como un indicador rápido y accesible de la calidad de un filme, ya que una alta calificación es sinónimo de que ha sido bien recibido por los críticos, esto genera confianza en los espectadores, que ven esta medida como una garantía de que su tiempo y dinero serán bien invertidos; funciona como un sello de aprobación que puede influir significativamente en la decisión del público de ver una película en el cine.

Aunque los premios Goya gozan de prestigio y reconocimiento dentro de la industria, el público de manera general puede no considerarlos como un factor decisivo al elegir que película ver. El impacto de ganar un premio de esta dimensión se ve diluido en el mercado general, donde las audiencias se guían más por factores inmediatos y tangibles como el elenco y la promoción, en lugar de los galardones.

Este trabajo proporciona una comprensión de los elementos críticos que afectan la recaudación en taquilla de las películas, ofreciendo valiosa información para la industria cinematográfica en la formulación de estrategias efectivas de producción, distribución y marketing que pueden maximizar tanto el rendimiento económico como el impacto cultural de las películas lanzadas al mercado. La coherencia de estos resultados con estudios previos subraya la relevancia y aplicabilidad de estos hallazgos en el contexto actual de la industria cinematográfica.

Sugerencias para futuros estudios:

- Realizar un análisis longitudinal para observar cómo evolucionan las tendencias a lo largo del tiempo a través de un análisis de supervivencia como factor de éxito.
- Comparar los resultados obtenidos entre diferentes mercados cinematográficos internacionales.
- Con la creciente popularidad de las plataformas de streaming, futuros estudios podrían investigar cómo la disponibilidad simultánea de películas en el cine y plataformas digitales afectan la recaudación en taquilla.
- Emplear otros métodos estadísticos y de análisis para descubrir patrones ocultos y relaciones complejas entre las variables estudiadas.

## REFERENCIAS BIBLIOGRÁFICAS

- Bhadrashetty, A., & Patil, S. (2024). Movie Success and Rating Prediction Using Data Mining. *Journal of Scientific Research and Technology*, 1-4.
- Carballo Sánchez, A. (2022). La industria del cine en España. Estado de la cuestión en tiempos de pandemia y pospandemia en los ámbitos de producción, distribución y exhibición. *Revista panamericana de comunicación*, 4(1), 70-83.
- De Vany, A. (2006). The Movies. In D. Throsby & V. Ginsburgh (Eds.), *Handbook of the economics of art and culture* (pp. 615-667). Amsterdam: North Holland
- De Vany, A., & Walls, W. D. (1999). Incertidumbre en la industria del cine: ¿El poder de las estrellas reduce el terror de la taquilla? *Revista de economía cultural*, 23, 285-318.
- Hennig-Thurau, T., Houston, M. B., & Walsh, G. (2007). Determinantes de la taquilla cinematográfica y la rentabilidad: un enfoque de interrelación. *Revista de Ciencias Gerenciales*, 1, 65-92.
- Litman, B. R. (1983). Predicción del éxito de las películas teatrales: un estudio empírico. *Revista de cultura popular*, 16(4), 159
- Lozano Treviño, D. F., Barragán Codina, J. N., Guerra Moya, S. A., Treviño Ayala, M. E., & Villalpando Cadena, P. (2012). Factores narrativos utilizados por las organizaciones de producción cinematográfica y su impacto en los ingresos en taquilla. *Innovaciones de negocios*, 18(9), 279-317.
- Pangarker, N. A., & vdM Smit, E. (2013). Revisión de los determinantes del rendimiento de taquilla en la industria cinematográfica. *Revista Sudafricana de Administración de Empresas*, 44(3), 47-58.
- Sawhney, M. S., & Eliashberg, J. (1996). Un modelo parsimonioso para pronosticar los ingresos brutos de taquilla de las películas. *Ciencia del Marketing*, 15(2), 113-131.
- Selvaretnam, G., & Yang, J. Y. (2015). Factores que afectan el éxito financiero de las películas: ¿cuál es el papel del poder de las estrellas?

Terry, N., Butler, M., & De'Armond, D. (2004). Critical acclaim and the box office performance of new film releases. *Academy of Marketing Studies Journal*, 8(1), 61-73.

Vany, A. S. D., y Walls, W. D. (1997). El mercado de las películas: rango, ingresos y supervivencia. *Investigación económica*, 35(4), 783-797.

Wang, Z., Zhang, J., Ji, S., Meng, C., Li, T., & Zheng, Y. (2020). Predicting and ranking box office revenue of movies based on big data. *Information Fusion*, 60, 25-40

Xu, J. (2024). Predicción de taquilla de películas a largo plazo basada en el aprendizaje automático. *Destacados en Ciencia, Ingeniería y Tecnología*, 92, 308-315.

## ANEXOS

Anexo 1. Carga de películas y procesamiento de cada fila.

```

1 data = pd.read_csv('movies.csv', delimiter=';')
2 print(data.columns)
3 data = data.apply(procesar_fila, axis=1)

```

Anexo 2. Método principal.

```

1 def procesar_fila(fila):
2     rank = fila['RANK']
3
4     ultimas_seis = fila['TITULO'].strip().replace(" ", "+")[-6:]
5     patron = re.compile(r'\d{4}\d{2}')
6
7     url_search = fila['TITULO'].strip().replace(" ", "+")[:7] if patron.match(ultimas_seis) else fila['TITULO'].strip().replace(" ", "+")
8     release_year = fila['FECHA_ESTRENO'][-4:]
9     print(rank, '-----', url_search, release_year)
10    valores = getMovieInfo(url_search, release_year, fila['TITULO'].strip()[:7] if patron.match(ultimas_seis) else fila['TITULO'].strip())
11    if valores:
12        print(valores)
13        fila['TITULO'] = valores[0]
14        fila['AÑO'] = valores[1]
15        fila['DURACION'] = valores[2]
16        fila['PAIS'] = valores[3]
17        fila['DIRECTORES'] = valores[4]
18        fila['REPARTO'] = valores[5]
19        fila['GENEROS'] = valores[6]
20        fila['RATING'] = valores[7]
21        fila['PREMIOS'] = valores[8]
22        fila['MEJOR_PELICULA'] = valores[9]
23        fila['CRITICAS'] = valores[10]
24
25    return fila

```

### Anexo 3. Búsqueda de películas del listado en Filmaffinity.

```

1 def getMovieInfo(url, release_year, titulo):
2     movie = []
3     movieItem = None
4     response = requests.get(
5         f'https://www.filmaffinity.com/es/search.php?stype=title&stext={url}')
6     soup = BeautifulSoup(response.text, 'html.parser')
7
8     flag_movie = soup.find("div", {"class": "z-movie"})
9     flag_movie_list = soup.find("div", {"class": "z-search"})
10
11     if flag_movie_list :
12         movieItem = getMovieFromList(soup, titulo, release_year)
13     if flag_movie :
14         movieItem = soup.find("div", {"class": "z-movie"})
15
16     if not movieItem : return None
17
18     movie = getMovieInfoDetail(movieItem)
19
20     return movie

```

### Anexo 4. Búsqueda de la información de cada película.

```

1 def getMovieInfoDetail(movie):
2     result = []
3
4     title = movie.find_all("dd")[0].getText().strip()
5     year = movie.find("dd", {"itemprop": "datePublished"}).getText()
6     duration = movie.find("dd", {"itemprop": "duration"}).getText() if movie.find("dd", {"itemprop": "duration"}) else None
7     country = movie.find("img", {"class": "nflag"}).get("alt")
8     directors = movie.find("dd", {"class": "directors"}).getText().strip() if movie.find("dd", {"class": "directors"}) else None
9     cast = getCast(movie.find("dd", {"class": "card-cast-debug"})) if movie.find("dd", {"class": "card-cast-debug"}) else None
10    genders = getGenders(movie.find("dd", {"class": "card-genres"}))
11    rating = movie.find("div", {"id": "movie-rat-avg"}).getText().strip() if movie.find("div", {"id": "movie-rat-avg"}) else None
12    awards = getAwards(movie.find("dd", {"class": "award"})) if movie.find("dd", {"class": "award"}) else []
13    best_movie = [movie for movie in awards if "Mejor película" in movie]
14    pro_reviews = movie.find("div", {"class": "total-abs"}).get("title").strip() if movie.find("div", {"class": "total-abs"}) else None
15    colaboration = getColaboration(movie.find("dd", {"class": "card-producer"})) if movie.find("dd", {"class": "card-producer"}) else None
16
17    result.append(title)
18    result.append(year)
19    result.append(duration)
20    result.append(country)
21    result.append(directors)
22    result.append(cast)
23    result.append(genders)
24    result.append(rating)
25    result.append(awards)
26    result.append(best_movie)
27    result.append(pro_reviews)
28    result.append(colaboration)
29
30    return result

```

## Anexo 5. Definición de variables cualitativas

```

1 directores = ['Pedro Almodóvar','Isabel Coixet','Jaume Balagueró','Fernando León de Aranoa',
2              'Álex de la Iglesia','Iciar Bollain','Pablo Berger', 'Santiago Segura', 'Fernando Colomo',
3              'Enrique Gato','Javier Fesser','Imanol Uribe','Montxo Armendáriz', 'Albert Serra','Carlos Vermut']
4
5 actores = ['Paz Vega','Alicia Borrachero', 'Santiago Segura', 'Astrid Bergès-Frisbey','Óscar Jaenada',
6           'Elena Anaya', 'Jordi Mollà', 'Javier Botet', 'José Coronado', 'Javier Bardem', 'Penélope Cruz',
7           'Antonio Banderas', 'Paco León', 'Carmen Machi', 'Luis Tosar', 'Javier Gutiérrez', 'Blanca Suárez',
8           'Miguel Ángel Muñoz', 'Leo Harlem', 'Miren Ibarguren']
9
10 distribuidoras = ['Warner Bros','Sony','Paramount Int','Walt Disney','Filmax', 'DeAPlaneta','Entertainment One Films',
11                 'European Dreams Factory','A Contracorriente','Alfa Pictures','Avalon Distribución','Syldavia',
12                 'Bteam Pictures','Wanda', 'Universal']
13
14 generos_guardar = ['Comedia','Animación','Aventuras','Thriller','Drama','Terror','Fantástico','Acción','Documental','Romance']
15

```

## Anexo 6. Estructuración y limpieza de la Base de Datos

```

1 def procesar_fila(fila):
2     fila['PREMIOS_GOYA_MEJOR_PELICULA'] = premioGoyaMejorPelicula(fila['MEJOR_PELICULA'])
3     fila['NOMINADA_PREMIOS_GOYA_MEJOR_PELICULA'] = premioGoyaMejorPelicula(fila['PREMIOS'])
4     fila['NOMINADA_FESTIVAL_MALAGA'] = nominadaPremios(fila['PREMIOS'],'málaga')
5     fila['NOMINADA_FESTIVAL_SAN_SEBASTIAN'] = nominadaPremios(fila['PREMIOS'],'festival de san sebastián')
6     fila['NOMINADA_PREMIOS_FEROZ'] = nominadaPremios(fila['PREMIOS'],'premios feroz')
7     fila['NOMINADA_PREMIOS_PLATINO'] = nominadaPremios(fila['PREMIOS'],'premios platino')
8     fila['NOMINADA_FESTIVAL_VALLADOLID_SEMINCI'] = nominadaPremios(fila['PREMIOS'],'seminci')
9     fila['NOMINADA_PREMIOS_GAUDI'] = nominadaPremios(fila['PREMIOS'],'premios gaudí')
10    fila['NOMINADA_FESTIVAL_GIJON'] = nominadaPremios(fila['PREMIOS'],'festival de gijón')
11    fila['NOMINADA_PREMIOS_FORQUE'] = nominadaPremios(fila['PREMIOS'],'forqué')
12    fila['NOMINADA_FESTIVAL_SEVILLA'] = nominadaPremios(fila['PREMIOS'],'festival de sevilla')
13    fila['DIRECTORES_TMP'] = mejoresDirectores(fila['DIRECTORES'])
14    fila['DISTRIBUIDORA_MAS_IMPORTANTES'] = mejoresDistribuidoras(fila['DISTRIBUIDORA'])
15    fila['SONY'] = sony_warner(fila['DISTRIBUIDORA'],'Sony')
16    fila['WARNER_BROS'] = sony_warner(fila['DISTRIBUIDORA'],'Warner Bros')
17    fila['REPARTO_TMP'] = mejoresActores(fila['REPARTO'])
18    fila['ESTRENO_FIN_DE_SEMANA'] = finde(fila['FECHA_ESTRENO'])
19    fila['ESTRENO_SEMANA_SANTA'] = rango_fechas(fila['FECHA_ESTRENO'],'10/04','17/04')
20    fila['ESTRENO_SEMANA_NAVIDAD'] = rango_fechas(fila['FECHA_ESTRENO'],'19/12','25/12')
21    fila['ESTRENO_VERANO'] = rango_fechas(fila['FECHA_ESTRENO'],'21/06','23/09')
22    fila['ESTRENO_SEMANA_FIESTA_DEL_CINE'] = rango_fechas(fila['FECHA_ESTRENO'],'03/10','06/10')
23    fila['ESTRENO_PUENTE_CONSTITUCION'] = rango_fechas(fila['FECHA_ESTRENO'],'06/12','11/12')
24    fila['RECAUDACION'] = recaudaciones_acumuladas(fila['TITULO'],fila['RECAUDACION'])
25    fila['ESPECTADORES'] = espectadores_acumuladas(fila['TITULO'],fila['ESPECTADORES'])
26
27
28    return fila

```

Anexo 7. Ejemplos de métodos para convertir las variables cualitativas a binarias.

```
1 def nominadaPremios(columna_premios, premios):  
2     return 1 if premios in str(columna_premios).lower() else ''  
3  
4 def rango_fechas(columna_fecha, inicio, fin):  
5     fecha = datetime.datetime.strptime(columna_fecha, '%d/%m/%Y')  
6     inicio = datetime.datetime.strptime(inicio, '%d/%m')  
7     fin = datetime.datetime.strptime(fin, '%d/%m')  
8     return 1 if inicio <= fecha.replace(year=1900) <= fin else ''
```