



Universidad de Costa Rica

Proyecto Individual

Curso:

Herramientas para Ciencias de Datos I
CA-0204

Tema:

Análisis descriptivo del impacto del presupuesto en el éxito cinematográfico

Profesor: Joshua Cervantes

Estudiante: Ashly Garro Villanueva

Carnet: C33198

II Ciclo 2025

Introducción

La industria cinematográfica internacional ha experimentado una evolución constante, esto debido a cambios tecnológicos, económicos y sociales que han transformado la manera en la que se producen, distribuyen y consumen las películas. En este contexto, el presupuesto asignado a la producción de las películas se ha convertido en un aspecto clave para la determinación de su éxito comercial y su alcance.

Las inversiones de la compañías suelen influir directamente en la rentabilidad de las películas. Sin embargo, la relación entre presupuesto y éxito no siempre es proporcional, hay casos donde películas con bajos presupuestos llegan a lograr un gran éxito, mientras que producciones grandes fracasan. Este trabajo busca analizar la relación entre presupuesto y éxito de las películas, tomando distintas variables como indicadoras de éxito. Este enfoque permite observar patrones de eficiencia en el uso del presupuesto.

Datos a utilizar

Para el presente análisis se utilizó el conjunto de datos "The Movies Dataset", disponible en la plataforma Kaggle (Banik, 2017). Este dataset contiene información sobre más de 45.000 películas, incluyendo aspectos financieros y de recepción del público.

Entre las variables disponibles se seleccionaron las más relevantes para analizar la relación entre el presupuesto de producción y el éxito obtenido, tanto en términos financieros como de aceptación del público. Las variables utilizadas fueron:

- **Budget:** Presupuesto de producción de la película (en dólares estadounidenses).
- **Revenue:** Ingresos totales obtenidos por la película (en dólares estadounidenses).
- **Vote_average:** Calificación promedio otorgada por usuarios de TMDB.
- **Runtime:** Duración de la película, utilizada como variable de control para explorar si la extensión de la película influye en su rendimiento

Desarrollo del análisis

Para el desarrollo del análisis se empezó identificando las variables dependientes e independientes. La variable que se tomó como independiente fue: budget (presupuesto), mientras que como variables dependientes oindicadores de éxito se tomaron las variables: revenue (ingresos), vote_average (promedio de calificación) y runtime (duración).

Para medir la rentabilidad de cada película, y por tanto, su éxito, se consideró una variable que permite medir la rentabilidad relativa de cada producción. Según Lee y Lim (2018) el ROI (Retorno de Inversión) está definido como el cociente entre la ganancia y la inversión, definiendo la ganacia como la el ingreso menos el costo (presupuesto). Al agregar esta variable al análisis, nos permite evaluar el éxito financiero no solamente en términos de ingresos, sino también en términos de rentalibilidad relativa, en nuestro caso el ROI está definido como:

$$ROI = \frac{revenue - budget}{budget}$$

Una vez definidas las variables, se procedió a limpiar y transformar los datos. Se eliminaron observaciones con valores faltantes o con presupuestos e ingresos nulos, y convirtiendo las variables numéricas a su formato correcto. Dado que los valores de presupuesto e ingresos presentaban una

alta dispersión, se aplicó una transformación logarítmica para estabilizar la varianza y facilitar la interpretación de las relaciones entre variables.

Se calculó también una versión logarítmica del ROI (ROI_log) para reducir la influencia de valores extremos y permitir un análisis más completo en los modelos de regresión. Además, se clasificó el presupuesto en cinco categorías: *Muy Bajo, Bajo, Medio, Alto, Muy Alto*, con el fin de visualizar mejor las diferencias entre los grupos de inversión.

Posteriormente, se realizaron estadísticas descriptivas y análisis de correlación entre presupuesto e indicadores de éxito, esto para determinar la relación entre el presupuesto y las variables indicadoras de éxito.

Seguidamente, se hicieron visualizaciones gráficas de dispersión y gráficos de barras para mostrar visualmente la relación entre las variables y finalmente, se aplicaron modelos de regresión múltiple para evaluar de forma cuantitativa el efecto del presupuesto sobre los indicadores de éxito. En estos modelos se utilizó log(budget) como predictor principal, junto con vote_average y runtime como variables de control.

Resultados

Estadísticas descriptivas y correlaciones

En esta sección se presenta los hallazgos obtenidos de la comparación entre presupuesto y las variables indicadoras de éxito:

Variable	Min	1er Cuartil	Mediana	Media	3er Cuartil	Max
Budget (USD)	1	2,500,000	10,000,000	23,318,930	28,000,000	380,000,000
Revenue (USD)	0	0	6,132,000	59,560,000	53,050,000	2,788,000,000
ROI	-1.0	-1.0	-0.3	3,337.5	1.8	12,396,382
Vote average	0.0	5.5	6.2	6.085	6.8	10
Runtime (min)	0.0	92.0	102.0	106.4	117.0	705.0

Table 1: Estadísticas descriptivas de las variables principales

Las estadísticas descriptivas muestran que los presupuestos y los ingresos presentan una alta dispersión, reflejando la diferencia entre producciones de un alto presupuesto y aquellas más pequeñas.

Variables	Correlación
Budget vs Revenue	0.736
Budget vs Vote Average	0.612
Budget vs ROI	-0.013
Budget vs Runtime	0.219

Table 2: Correlaciones entre presupuesto y variables de éxito

En cuanto a las correlaciones, se observa que el presupuesto está altamente relacionado con el ingreso de la producción y de una forma más estándar con el promedio de calificación. Sin embargo, el presupuesto no garantiza un alto ROI, lo que nos indica que el presupuesto no garantiza la rentabilidad relativa. Por otro lado, la duración muestra una baja correlación con el presupuesto, lo que nos indica que la duración de la producción no depende de gran forma del presupuesto de esta.

Visualización de comparaciones

Por otra parte, los gráficos a continuación muestran visualmente la relación entre el presupuesto y las variables indicadoras de éxito:

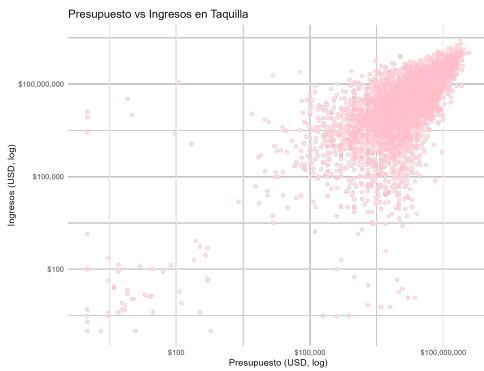


Figure 1: Presupuesto e Ingreso

Este primer gráfico¹ muestra la relación entre el presupuesto y los ingresos obtenidos. Para facilitar la visualización, ambos ejes se encuentran en escala logarítmica, esto permite apreciar la distribución de películas con presupuestos e ingresos muy dispares.

Se observa una clara tendencia positiva, entre mayor presupuesto mayor y mayor ingreso. Esto nos muestra que con más recursos económicos tienen una mayores oportunidades de éxito y nos destaca la importancia de la inversión por grandes empresas (Pangarker y Smit, 2013, como se citó en Viera, 2024).

Sin embargo, la dispersión de los datos indica que un presupuesto muy alto no siempre garantiza el éxito en taquilla de la película. Se pueden ver casos donde la película presenta un alto costo (presupuesto) pero ingresos muy bajos, al igual que producciones con presupuestos muy bajos que lograron altos ingresos.

Por lo tanto, el presupuesto tiene una influencia positiva pero no determinante sobre el éxito en taquilla. Existe una correlación general de mayor presupuesto y mayor ingreso, pero con una alta variabilidad.

El segundo gráfico² agrupa las películas por rangos de presupuesto para compararlos con el promedio de calificación, para esto se dividió el rango del presupuesto en 5 partes iguales y se clasificaron desde *Muy Bajo* a *Muy Alto*: Los resultados muestran un comportamiento no lineal. Las películas

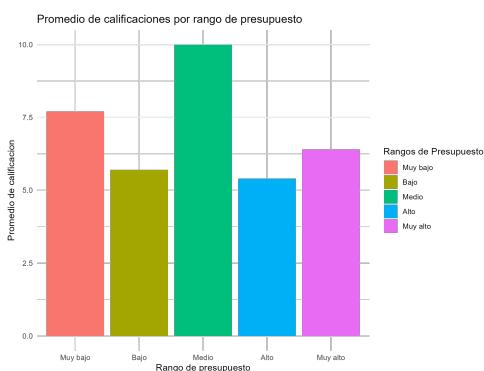


Figure 2: Presupuesto y promedio de calificaciones en TMDB

con presupuesto medio alcanzan un promedio de calificación muy alto, seguidas por aquellas con presupuestos bajos y muy altos.

Este patrón puede tener varias interpretaciones. Las películas de bajo presupuesto pueden llegar a un menor público lo que genera una menor cantidad de votos, y consecuentemente, un mayor promedio de calificación, este análisis aplica de igual forma a las películas de presupuesto medio, y inversamente a las de altos presupuestos.

Por último, el tercer gráfico ³ relaciona el presupuesto con el Retorno de Inversión (ROI). Al igual

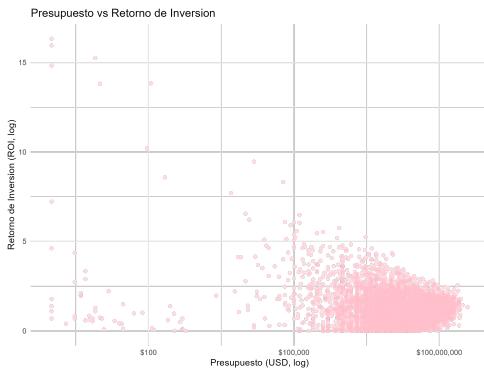


Figure 3: Presupuesto y ROI

que en el primer gráfico, se utiliza una escala logarítmica para observar mejor las variaciones.

Los resultados nos muestran una relación inversa, las películas con menor presupuesto tienden a tener retornos de inversión mucho más altos. Se pueden ver casos donde películas de bajo presupuesto logran multiplicar varias veces su inversión, mientras que producciones con altos presupuestos presentan un ROI más bajo debido a sus altos costos.

Esto nos muestra que las producciones más pequeñas pueden ser proporcionalmente más rentables, incluso si sus ganancias son bajas. Desde un punto de vista económico, un alto presupuesto no garantiza una mayor rentabilidad.

Modelos de regresión múltiple

A continuación se presentan los resultados obtenidos a partir de tres modelos de regresión lineal múltiple, formulados para analizar cómo se diferencia la influencia del presupuesto a comparación de la calificación promedio y la duración de las películas sobre diferentes indicadores de éxito: el Retorno de Inversión (ROI), los ingresos en taquilla (*revenue*) y la calificación promedio del público (*vote_average*).

Modelo 1: Análisis de la rentabilidad (ROI)

El primer modelo tiene como variable dependiente el retorno de inversión en escala logarítmica (*ROI_log*).

Variable	Coeficiente	Error estándar	Significancia
(Intercept)	1.913	0.1373	***
log(budget)	-0.1390	0.0062	***
vote_average	0.2543	0.0147	***
runtime	-0.00009	0.0006	ns

Table 3: Resultados del modelo de regresión para *ROI_log*

El coeficiente de determinación ajustado fue $R^2_{aj} = 0.141$, lo que indica que el modelo explica un

14.1% de la variación en el retorno de inversión. A pesar de que el poder explicativo es moderado, todos los predictores, excepto la duración, resultan importantes.

El signo negativo de $\log(\text{budget})$ sugiere que a medida que aumenta el presupuesto, el retorno proporcional de la inversión disminuye, lo que muestra los rendimientos decrecientes en grandes producciones. Por su parte, la calificación promedio (*vote_average*) tiene un efecto positivo fuerte: un aumento de un punto en la calificación se asocia con un incremento aproximado del $e^{0.2543} - 1 \approx 28.9\%$ en el ROI. La duración no muestra un efecto significativo.

Modelo 2: Análisis de ingresos en taquilla

El segundo modelo busca explicar los ingresos totales (*revenue*) a partir del presupuesto y de la calificación promedio, ambos en escala logarítmica.

Variable	Coeficiente	Error estándar	Significancia
(Intercept)	-1.1481	0.2591	***
$\log(\text{budget})$	0.8908	0.0117	***
<i>vote_average</i>	0.5416	0.0265	***

Table 4: Resultados del modelo de regresión para $\log(\text{revenue})$

Este modelo presenta un coeficiente de determinación ajustado de $R_{aj}^2 = 0.532$, lo cual indica que explica un 53.2% de la variabilidad en los ingresos en taquilla. El coeficiente de $\log(\text{budget})$ implica que un aumento del 1% en el presupuesto se asocia con un incremento promedio de 0.8908% en los ingresos, es decir, casi una relación proporcional. La calificación promedio también tiene un efecto positivo, un punto adicional en *vote_average* aumenta los ingresos esperados en aproximadamente $e^{0.5416} - 1 \approx 71.8\%$.

Modelo 3: Análisis de percepción del público

En este modelo, la variable dependiente es la calificación promedio (*vote_average*), buscando determinar si el presupuesto, los ingresos o la duración influyen en la valoración del público.

Variable	Coeficiente	Error estándar	Significancia
(Intercept)	5.930	0.1087	***
$\log(\text{budget})$	-0.0727	0.0061	***
<i>revenue</i>	9.667e-10	7.759e-11	***
<i>runtime</i>	0.0131	0.0005	***

Table 5: Resultados del modelo de regresión para *vote_average*

El modelo alcanza un $R_{aj}^2 = 0.135$, lo que significa que explica un 13.5% de la variabilidad en la percepción del público. El presupuesto presenta un efecto negativo y significativo, aunque de magnitud reducida, es decir, películas más costosas tienden a recibir ligeramente peores calificaciones promedio. Por otro lado, los ingresos (*revenue*) y la duración (*runtime*) muestran efectos positivos. En particular, cada 10 minutos adicionales de duración aumentan la calificación promedio en torno a 0.13 puntos. Un incremento de 100 millones de dólares en los ingresos se asocia con un aumento de aproximadamente 0.1 puntos en la valoración del público.

La comparación de los tres modelos se resume en la Tabla 6.

Los resultados muestran que el presupuesto y la valoración del público son factores centrales en el éxito comercial, mientras que los rendimientos decrecientes se hacen evidentes en la rentabilidad

Variable dependiente	Predictores más influyentes	Tipo de relación	R^2 ajustado
ROI_log	vote_average (+), log(budget) (-)	Positiva moderada, rendimientos decrecientes	0.141
log(revenue)	log(budget), vote_average	Positiva fuerte	0.532
vote_average	revenue, runtime (positivos), log(budget) (negativo)	Mixta (leve)	0.135

Table 6: Comparación general de modelos de regresión

relativa (ROI). Por su parte, la percepción del público se ve levemente influida por la duración y los ingresos, pero no necesariamente por la cantidad del presupuesto, lo que indica que la calidad percibida no depende del gasto sino de otros factores, probablemente creativos.

Recomendaciones

El análisis realizado permitió examinar cómo el presupuesto de producción influye en diversas variables que reflejan el éxito cinematográfico, entre ellas los ingresos (*revenue*), la rentabilidad (*ROI*) y la calificación promedio del público (*vote_average*). A partir de los resultados obtenidos en los modelos de regresión y en los gráficos exploratorios, es posible establecer una serie de conclusiones y recomendaciones.

En general, se observó una relación positiva entre el presupuesto y los ingresos, lo cual indica que las producciones con mayores inversiones tienden a generar más ganancias. Sin embargo, la dispersión de los datos sugiere que no todas las películas con alto presupuesto logran grandes beneficios, lo cual también fue confirmado por el modelo de regresión múltiple. Aunque el efecto del presupuesto sobre los ingresos fue altamente significativo ($p < 0.001$), el coeficiente de determinación ($R^2 = 0.14$) muestra que este factor explica solo una fracción del éxito comercial. Por tanto, se concluye que invertir más dinero no garantiza el éxito financiero absoluto, sino que los recursos deben destinarse estratégicamente a aspectos como la promoción, la distribución y la calidad narrativa.

En relación con la rentabilidad (*ROI*), los resultados mostraron una tendencia menos clara. Aunque el presupuesto resultó ser significativo en el modelo ($p < 0.001$), su efecto fue más débil que el de la calificación promedio del público. Esto evidencia que el retorno de inversión depende más de la percepción y aceptación del espectador que del monto invertido. De este modo, las producciones que logran una buena recepción tienden a ser más rentables, incluso con presupuestos más bajos. Por lo tanto, se recomienda que los estudios prioricen la calidad del producto final —guion, dirección, elenco y coherencia narrativa— antes que el simple incremento del gasto.

Por otra parte, el análisis entre el presupuesto y la calificación promedio (*vote_average*) reveló que no existe una relación significativa entre ambas variables. Los resultados del modelo indicaron que el coeficiente de *log(budget)* no fue estadísticamente significativo ($p = 0.195$), lo que implica que un mayor presupuesto no necesariamente se traduce en una mejor valoración del público o de la crítica. En cambio, tanto los ingresos como la duración de la película mostraron una relación positiva con la calificación promedio, lo cual sugiere que las películas más extensas y exitosas tienden a recibir una mejor evaluación. En consecuencia, el éxito crítico y artístico no depende del dinero invertido, sino de la solidez del contenido y de la forma en que este se presenta al espectador.

En conjunto, los resultados muestran que, aunque el presupuesto tiene un papel importante en el desempeño comercial, su influencia sobre la rentabilidad y la valoración del público es limitada. El éxito cinematográfico depende de una interacción entre inversión, calidad del contenido y recepción del espectador. Por ello, se recomienda adoptar un enfoque de inversión equilibrado, priorizando la eficiencia en el uso de los recursos y fomentando la innovación creativa. Asimismo, se sugiere incorporar en futuros análisis variables adicionales como el género, la procedencia, el año de lanzamiento o la estrategia de distribución, con el fin de obtener una visión más completa de los factores que

determinan el éxito en la industria cinematográfica.

References

- [1] Banik, R. (2017). *The Movies Dataset*. <https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset/data>
- [2] Viera, A. (2024). *Determinantes de la recaudación en taquilla como indicador de éxito cinematográfico*. <http://hdl.handle.net/10334/9660>
- [3] Lee, Y., & Lim, R. (2018). *At the movies: Some stylized facts on investment returns and consumption patterns*. *International Journal of Business and Economics*, 17(2), 123–142. https://ijbe.fcu.edu.tw/assets/ijbe/past_issue/No.17-2/pdf/vol_17-2-3.pdf