A logo with text on it

Description automatically generatedDepartamento de Sistemas de información

Centro de Ciencias Básicas  
Visualización de Datos Empresariales

Documentación de Proyecto Final

Alumno: Rodolfo Eduardo Cordova Perez (334876)

Docente: Lizeth Itziguery Solano Romo

Fecha: 11/12/2024

**Descripción de la Problemática del Proyecto**

En este proyecto, se analizarán los datos de diversas películas a lo largo de 40 años con el objetivo de tomar decisiones informadas sobre el tipo de películas en las que un estudio cinematográfico debería especializarse para maximizar su éxito y rentabilidad. Se busca identificar patrones y características comunes en las películas exitosas, en términos de ingresos y calificación, con el fin de orientar la estrategia del estudio hacia la producción de ciertos géneros o tipos de películas que optimicen su rendimiento financiero.

A lo largo del análisis, se pretende entender cómo los diferentes géneros y características de las películas influyen en su éxito comercial. Se explorarán tendencias en los ingresos brutos, el retorno de inversión (ROI) y la evolución de los presupuestos de las producciones, lo que permitirá formular recomendaciones concretas sobre en qué tipo de cine un estudio debería centrarse. Este enfoque se apoya en la premisa de que, en un mercado altamente competitivo y volátil, especializarse en géneros o nichos específicos puede proporcionar una ventaja significativa a los estudios más pequeños o nuevos, permitiéndoles reducir riesgos y maximizar sus oportunidades de éxito.

**KPIs Utilizados:**

* **Promedio de Ingresos Brutos por Género**
  + Utilidad: Este KPI muestra qué géneros generan más ingresos en promedio, proporcionando una base para decidir en qué géneros especializarse.
* **Retorno de Inversión (ROI) por Género**
  + Fórmula: Promedio de (gross - budget) / budget por género.
  + Utilidad: Permite evaluar qué géneros ofrecen mejores retornos sobre la inversión, un factor crucial para determinar la rentabilidad de las producciones.
* **Distribución de Géneros en el Catálogo de Películas**
  + Utilidad: Este KPI indica qué porcentaje del catálogo de películas corresponde a cada género, lo que ayuda a evaluar si el estudio debe diversificar sus producciones o especializarse más en ciertos géneros para optimizar sus resultados.

**Fuente de datos:**  
El dataset utilizado proviene de Kaggle en el siguiente enlace:

[**https://www.kaggle.com/datasets/danielgrijalvas/movies**](https://www.kaggle.com/datasets/danielgrijalvas/movies)

Los datos han sido recopilados desde la página IMDB, una fuente de alta credibilidad. Aunque los datos son bastante completos, se identificaron y corrigieron varios problemas a lo largo del proceso de limpieza, los cuales se detallan a continuación.

**Descripción de los Datos:**

El archivo CSV **movies.csv** contiene información sobre una variedad de películas. Los datos incluyen varias columnas relacionadas con los detalles de cada película, tales como:

* **name**: El nombre de la película.
* **rating**: La calificación que la película ha recibido
* **released**: La fecha de estreno de la película.
* **genre**: El género de la película (por ejemplo, acción, comedia, drama, etc.).
* **director**: El director o los directores de la película.
* **writer**: El guionista o los guionistas de la película.
* **star**: Los actores principales o las estrellas de la película.
* **country**: El país o los países en los que se produjo la película.
* **company**: La productora de la película.
* **budget**: El presupuesto de la película.
* **gross**: Los ingresos brutos de la película.

Cada fila representa una película diferente y contiene estos atributos. Sin embargo, se presentaron varios problemas en los datos que requerían ser corregidos.

**Problemas Encontrados en los Datos:**

1. **Fechas inconsistentes**
   * La columna released contenía fechas en formatos no uniformes, como "January 1983", "March 1992", y "1981". Este tipo de inconsistencias dificulta el análisis de tendencias a lo largo del tiempo.
2. **Valores nulos o vacíos**
   * Varias columnas, como name, rating, director, writer, star, country, y company, tenían valores nulos o vacíos, lo que podría afectar los análisis y cálculos. Estos valores vacíos necesitan ser gestionados adecuadamente para evitar errores en el análisis posterior.
3. **Espacios innecesarios y formatos inconsistentes**
   * Algunas columnas de texto, como name, director, genre, country, y company, tenían espacios innecesarios o formatos inconsistentes que dificultaban su uso en análisis posteriores.
4. **Datos no estandarizados**
   * Algunos valores de fecha no seguían el formato estándar, lo que podía causar problemas al intentar procesar o analizar las fechas.

**Proceso de Preparación de Datos:**

A continuación, se detallan los pasos realizados para limpiar y preparar los datos:

1. **Eliminación de filas con valores nulos o vacíos**:
   * Se recorrió todo el DataFrame y se eliminaron las filas que contenían **valores nulos o vacíos** en cualquier columna. Esto se realizó mediante la función dropna(), que asegura que solo se mantengan las filas con datos completos.

df = df.dropna()

1. **Reformateo de fechas**:
   * La columna released, que contenía fechas con formatos inconsistentes, fue procesada utilizando la función formatear\_fecha(date\_str), que usa la librería dateutil.parser para manejar una variedad de formatos de fecha.

**Función formatear\_fecha**:

* + Esta función intenta convertir las fechas a un formato estándar, manejando correctamente las fechas con formatos no estandarizados.

def formatear\_fecha(date\_str):

try:

return parser.parse(date\_str, fuzzy=True)

except (ValueError, TypeError):

return pd.NaT

1. **Eliminación de espacios innecesarios**:
   * Se eliminaron los espacios al principio y al final de las cadenas de texto en varias columnas, como name, rating, genre, director, writer, star, country, y company, para asegurar la consistencia de los datos. Esto también incluye la normalización de los textos.

df[column] = df[column].str.strip()

1. **Eliminación de filas con valores nulos en cualquier columna**:
   * Después de la eliminación de espacios y la corrección de fechas, se verificó nuevamente si había valores nulos y se eliminó cualquier fila que contuviera **valores nulos o vacíos** en cualquier columna. Este paso es crítico para asegurarse de que el DataFrame resultante no contenga ningún dato faltante.

df = df.dropna()

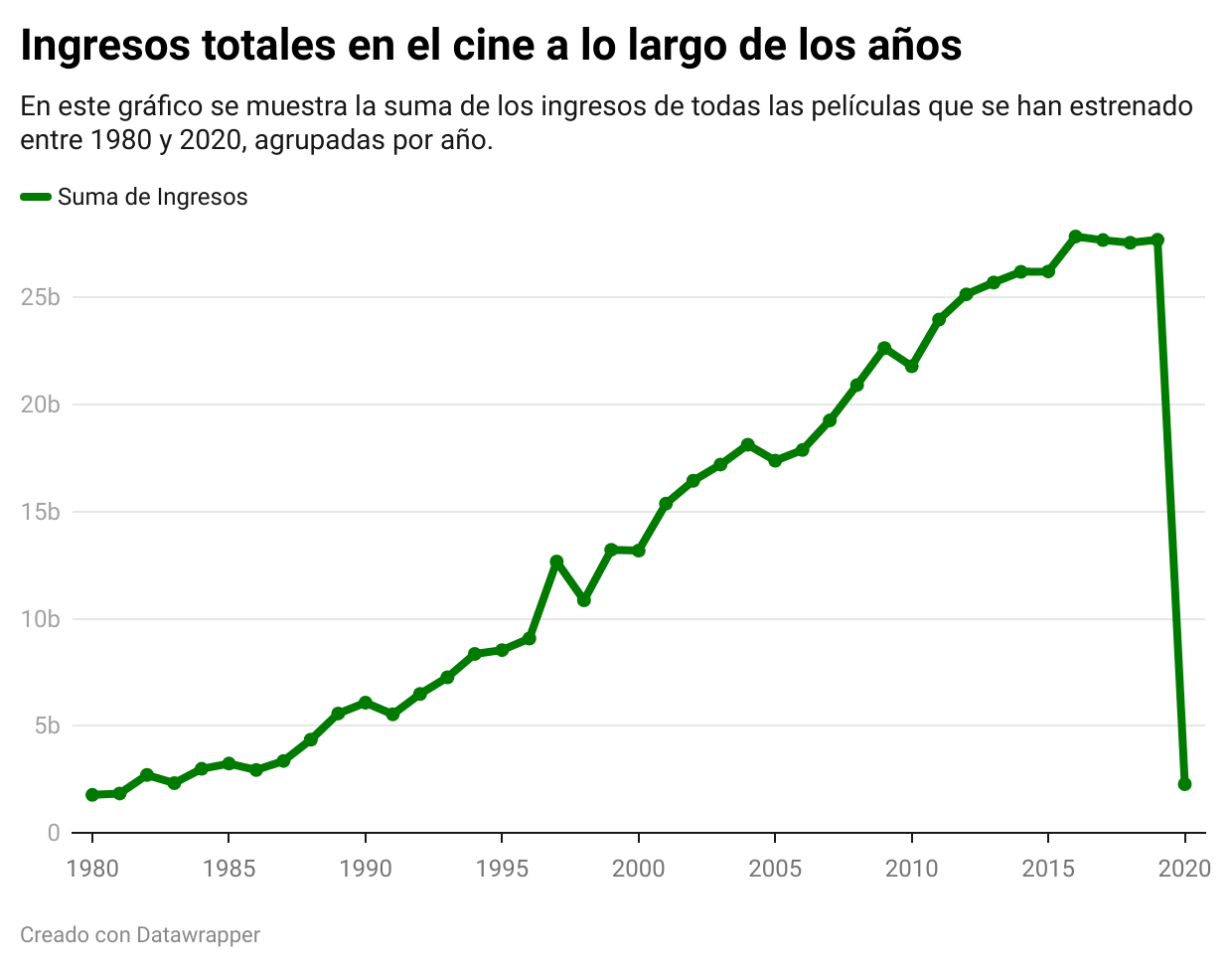
1. **Adición de columna con el ROI (Retorno de Inversión)**:
   * Cuando los datos ya están completamente listos y todos los casos de datos vaciós o nulos han sido removidos, se puede calcular el ROI de manera automático. Este cálculo es bastante importante para el objetivo de este proyecto ya que es lo que determina si una película se cuenta como haber sido exitosa o un fallo.

df['ROI'] = ((df['gross'] - df['budget']) / df['budget']).round(4)

**Análisis del problema**

Antes de abordar cualquier solución sobre qué tipo de cine debería especializarse un nuevo estudio, es fundamental entender el estado actual de la industria cinematográfica. Este análisis previo es esencial para identificar las tendencias, desafíos y oportunidades que los estudios enfrentan hoy en día. La industria cinematográfica se caracteriza por su volatilidad, evidenciada en la evolución de los ingresos y la cantidad de películas exitosas a lo largo de los años. Factores como los cambios en los hábitos de consumo, la creciente competencia de plataformas de streaming, y eventos globales como la pandemia de COVID-19 han tenido un impacto directo en la rentabilidad de los estudios. Por lo tanto, antes de tomar decisiones estratégicas sobre qué géneros o enfoques de producción adoptar, es crucial comprender cómo estas dinámicas influyen en el mercado actual y cómo un estudio puede adaptarse a ellas para maximizar su éxito.

<https://datawrapper.dwcdn.net/MS23L/1/>

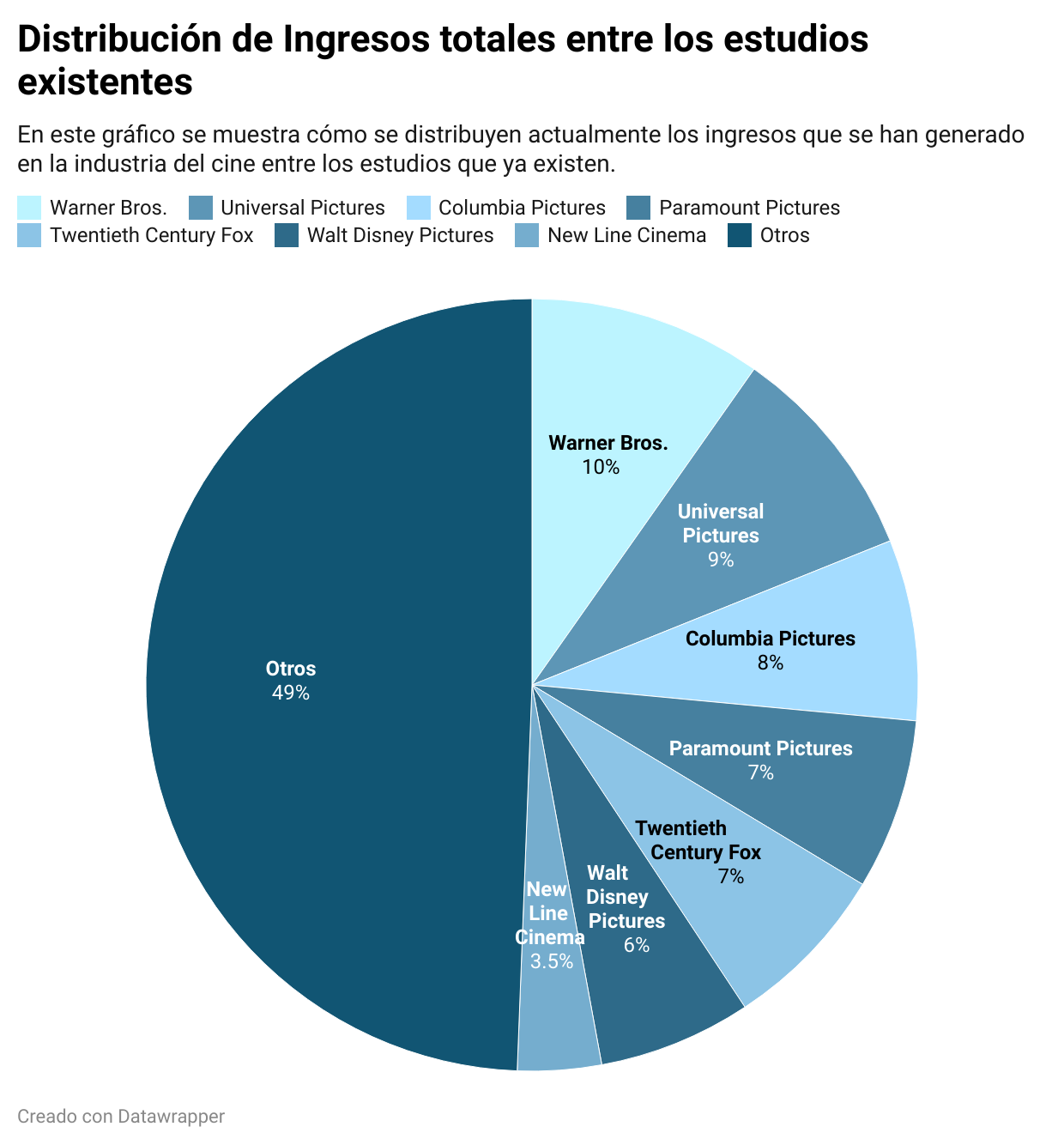


En este primer gráfico, que muestra la evolución de los ingresos brutos por año, se observa cómo los ingresos crecieron de manera sostenida hasta alcanzar un pico significativo en 2019, seguido por una caída abrupta en 2020. Esta disminución, probablemente causada por factores externos como la pandemia de COVID-19, resalta la volatilidad de la industria cinematográfica y subraya la necesidad de que los estudios se adapten para mantener la rentabilidad en tiempos inciertos. Esta caída abrupta también pone de manifiesto la dependencia de la industria de factores externos, como la disponibilidad de los cines, lo que refuerza la importancia de diversificar las estrategias y estar preparados ante posibles disrupciones globales. A graph with a line going up

Description automatically generated

En el segundo gráfico, que presenta el número de películas exitosas por año, se observa que, al igual que los ingresos, las películas exitosas experimentaron un crecimiento continuo hasta 2019, seguido de un descenso abrupto en 2020. Es importante destacar que, para considerar una película como exitosa, debe generar al menos un 2.25 veces su presupuesto en ingresos (es decir, un ROI superior a 2.25). Este patrón refleja la fragilidad de la industria cinematográfica ante eventos inesperados, como lo sucedido en 2020, lo que resalta la importancia de la capacidad de adaptación de los estudios ante crisis globales u otros factores disruptivos. La industria debe ser capaz de ajustarse rápidamente a cambios en las circunstancias para mantener su rentabilidad y continuidad.

<https://datawrapper.dwcdn.net/3tVhD/1/>



En este tercer gráfico de pie, se observa cómo los ingresos de la industria cinematográfica se distribuyen entre los 1,472 estudios existentes, con solo 7 de ellos concentrando el 51% de los ingresos totales. Este hecho resalta la centralización del poder económico en unos pocos grandes estudios, mientras que los más pequeños enfrentan mayores dificultades para acceder a las ganancias del mercado. Sin embargo, también muestra que existe espacio para la innovación y la especialización de los estudios pequeños, que pueden identificar y aprovechar nichos específicos del mercado. Este análisis pone de manifiesto la necesidad de que un estudio de cine tenga una visión clara y una especialización bien definida. En un mercado tan competitivo, destacarse es crucial. Al enfocarse en un nicho específico o tipo de cine, un estudio puede atraer a un público fiel y generar una mayor rentabilidad. Especializarse optimiza recursos y reduce riesgos, lo cual es esencial en un entorno dominado por los grandes estudios.

**Análisis para una recomendación**

**Tendencias a lo largo del tiempo**

<https://datawrapper.dwcdn.net/dQH1p/3/>

A graph of green and white lines

Description automatically generated

La gráfica de Evolución de Ingresos por Género muestra cómo han evolucionado los ingresos brutos generados por las películas de diferentes géneros desde 1980 hasta 2020. A través de las líneas que representan cada género, podemos observar que las películas de acción han sido consistentemente las más rentables, dominando el panorama desde 1990 hasta la fecha sin ser superadas por ningún otro género. Le sigue el género de animación, que ha mostrado un crecimiento constante, con algunos altibajos en la última década. En tercer lugar se encuentra el género de comedia, que experimentó un aumento significativo en los años 2000, pero que en los últimos años ha mostrado una leve disminución, lo que sugiere que podría ser superado por el género de aventuras en el futuro cercano. Esta información es útil para un estudio de cine al evaluar en qué géneros invertir dependiendo de la rentabilidad histórica, con la acción y la animación como opciones estables y con gran potencial.

<https://datawrapper.dwcdn.net/PIBSA/3/> A graph of red lines and numbers

Description automatically generated

La gráfica de *Evolución de Presupuesto por Género* muestra la cantidad promedio de dinero invertido en películas de cada género entre 1980 y 2020. En esta gráfica se observan tendencias similares a las de la gráfica anterior, lo que tiene sentido, ya que si un género demuestra generar altos ingresos, es probable que se le asigne un mayor presupuesto. El género que ha recibido más inversión es **acción**, seguido por **animación**, con una diferencia notable. Un hallazgo interesante en esta gráfica es que el género de **comedia** ha experimentado una disminución en el presupuesto en comparación con **aventura**, lo cual podría explicarse por la caída en los ingresos generados por las películas de comedia en los últimos años.

Con base en las observaciones de ambas gráficas, se pueden ofrecer dos recomendaciones. Para un estudio con un mayor presupuesto, sería viable invertir en **animación**, ya que este género ha mostrado una tendencia creciente en inversión, o en **acción**, que constantemente ha generado altos ingresos. Para estudios más pequeños, se recomienda centrarse en géneros de menor presupuesto, como **crimen**, **terror** o **drama**, los cuales continúan siendo rentables sin la necesidad de grandes inversiones. En cualquier caso, se desaconseja la inversión en **comedias**, ya que el rendimiento de este género ha estado disminuyendo en términos de ingresos en los últimos años.

**Éxito según el genero**A graph with green and red lines

Description automatically generated

La gráfica de *Géneros con Mayor Éxito* muestra los diferentes géneros cinematográficos junto con su retorno de inversión (ROI) promedio. Las barras están coloreadas para diferenciar si el ROI es considerado exitoso (mayor a 2.25) o no exitoso (menor a 2.25), donde el color verde indica un ROI positivo y el rojo uno negativo. Según esta gráfica, el género de **terror** destaca como el más exitoso en términos de ROI, lo cual tiene sentido al considerar que las películas de terror suelen tener presupuestos relativamente bajos pero pueden generar grandes ingresos, especialmente en la última década. Este género se ve acompañado por las **películas de familia**, que también presentan un ROI alto debido a la demanda constante de películas aptas para niños y familias, un público recurrente en las salas de cine. Con esta información, se recomienda que un estudio nuevo considere especializarse en la producción de películas de terror o familiares, ya que ambos géneros presentan un alto potencial de rentabilidad.

**Éxito según los involucrados**

A graph with numbers and text

Description automatically generated with medium confidence

En la gráfica de Distribución de Ingresos Brutos de los Top 10 Directores, se muestran los 10 directores más destacados según los ingresos generados por las películas que han dirigido. Entre estos nombres figuran algunos de los más grandes de la industria, como Steven Spielberg, James Cameron, Christopher Nolan y J.J. Abrams, entre otros. Esta gráfica no solo muestra el promedio de los ingresos generados, sino también la dispersión de esos ingresos, lo que es clave para evaluar el riesgo financiero al trabajar con estos directores.

La dispersión en los ingresos nos indica el nivel de riesgo que podríamos asumir. Por ejemplo, James Cameron puede generar hasta 2.8B USD, pero también puede caer hasta 78M USD. Esto sugiere que hay una alta variabilidad en el éxito financiero de las películas dirigidas por él. En cambio, directores como Tim Burton, Ron Howard y Steven Spielberg tienen un rango más compacto, lo que implica un riesgo financiero menor, aunque sus ingresos mínimos sean inferiores a los de otros directores más volátiles.

En base a esta información, se recomienda que un estudio de cine considere trabajar con directores como Tim Burton, Ron Howard y Steven Spielberg, ya que su rango más predecible de ingresos reduce el riesgo financiero asociado a sus proyectos.

A graph with numbers and a number of characters

Description automatically generated with medium confidence

En la gráfica de **Distribución de Ingresos Brutos de los Top 10 Actores**, se presentan los 10 mejores actores según los ingresos brutos generados en las películas donde han sido la estrella. Al igual que en la gráfica anterior, esta información es útil para evaluar qué actores podrían representar un riesgo financiero, como es el caso de **Robert Downey Jr.**, cuyo rango de ingresos varía considerablemente entre un máximo de **2.79B USD** y un mínimo de **435K USD**.

En este contexto, los actores con un rango más compacto y estable, como **Johnny Depp**, **Leonardo DiCaprio** y **Tom Hanks**, son opciones recomendables para trabajar, ya que su desempeño financiero tiende a ser más consistente, similar a los directores previamente mencionados. Otra opción sería **Bruce Willis**, quien también tiene un rango relativamente estable. Sin embargo, es importante señalar que Bruce Willis ya no está activo debido a su jubilación, tras haber recibido un diagnóstico médico de demencia.

En base a estos análisis, se recomienda que un estudio de cine considere trabajar con actores como **Johnny Depp**, **Leonardo DiCaprio** y **Tom Hanks**, quienes ofrecen un menor riesgo financiero gracias a su consistencia en el éxito de taquilla.

A graph with text and numbers

Description automatically generated with medium confidence

A graph with many names

Description automatically generated with medium confidence

En estas dos gráficas se muestran los 10 mejores actores y directores según el retorno de inversión (ROI) que han generado en sus proyectos. Los nombres de los directores y actores se presentan junto con el color de la barra, que indica en qué género de películas predominan. Al analizar estas gráficas, se observa una historia distinta en comparación con otras, ya que aquí se trabaja sobre el ROI en lugar de los ingresos brutos. En ambas gráficas se destaca un género con un alto potencial de ROI: el de Terror. Los cuatro primeros puestos en ambas gráficas corresponden a este género, seguidos por una mezcla de Comedia, Drama, Acción y Aventuras.

Otro aspecto interesante es que los miembros destacados en estas gráficas provienen principalmente de producciones pequeñas, es decir, de películas con presupuestos muy bajos en comparación con el resto de la industria y creadas por cineastas con relativamente poca experiencia. Los ejemplos más representativos de esto son "The Blair Witch Project" y "Paranormal Activity", que tenían presupuestos de 60K USD y 15K USD, respectivamente, y fueron creadas por personas que eran principiantes en el mundo del cine. Estas observaciones son clave para un estudio pequeño, ya que podría ser más conveniente enfocarse en proyectos más pequeños con presupuestos reducidos. De esta forma, el riesgo de pérdidas se minimiza, y a la vez, se abre la posibilidad de recuperar la inversión de forma más rápida.

**Éxito según características de la película**

A colorful circle with numbers and a number of numbers

Description automatically generated with medium confidence

La gráfica de Películas Exitosas por Tipo de Clasificación en formato de dona muestra la proporción de películas que se consideran exitosas por tener un ROI de 2.25 o mayor, agrupadas según su clasificación. Al observar la distribución, se puede notar que las películas con clasificación R ocupan una parte significativa del gráfico, lo que indica que son las más exitosas en términos de rentabilidad, seguidas por las de clasificación PG-13, PG, G y, finalmente, NC-17, que tiene la menor proporción de películas exitosas.

Este patrón sugiere que las películas con clasificación R tienen un alto potencial de éxito financiero, probablemente debido a su atractivo para audiencias maduras que buscan contenido más provocador o controvertido. Las películas PG-13 también muestran un buen rendimiento, logrando un equilibrio entre una amplia audiencia y la posibilidad de generar ingresos considerables. Por otro lado, las clasificaciones PG y G tienen una menor proporción de películas exitosas, lo que podría reflejar un público más limitado, mientras que NC-17 tiene la menor proporción de éxito, dado su enfoque en una audiencia mucho más restringida.

Con base en esta información, se recomienda que un estudio de cine se especialice en la producción de contenido con clasificación R y PG-13. Estas clasificaciones presentan un mayor potencial para generar ingresos, con las películas R orientadas a un público adulto, y las PG-13 ofreciendo un alcance más amplio, atrayendo tanto a jóvenes como a adultos.

A graph with blue dots

Description automatically generated

La gráfica de Dispersión entre Duración y ROI para películas con ROI mayor a 2.25 muestra la relación entre la duración de las películas y su retorno de inversión (ROI). Al observar la distribución, se puede notar que la mayoría de las películas con un ROI exitoso tienen una duración que se encuentra entre los 70 y 150 minutos, lo que indica que estas películas son las más exitosas en términos de rentabilidad. En contraste, las películas con duraciones menores a 70 minutos o mayores a 150 minutos representan una proporción menor de las películas exitosas.

Este patrón sugiere que las películas con una duración moderada tienen un mayor potencial de éxito financiero, posiblemente debido a que estas logran un equilibrio adecuado entre contenido atractivo y costos de producción. Las películas más cortas, aunque pueden ser más económicas de producir, podrían no generar suficientes ingresos debido a la falta de desarrollo o profundidad. Por otro lado, las películas muy largas podrían enfrentar un mayor riesgo financiero debido a los altos costos de producción y la posibilidad de que su extensión no atraiga a toda la audiencia.

Con base en esta información, se recomienda que los estudios cinematográficos consideren una duración de 70 a 150 minutos para maximizar las posibilidades de obtener un ROI exitoso. Este rango parece ser el más equilibrado en términos de costos de producción y retorno económico.

**Conclusión y Recomendación Clave**

El análisis de la industria cinematográfica y las tendencias actuales muestra un panorama desafiante pero lleno de oportunidades para los nuevos estudios. La volatilidad de los ingresos, exacerbada por eventos globales como la pandemia de COVID-19, subraya la necesidad de que los estudios adapten sus estrategias para mantenerse rentables. A pesar de la concentración de ingresos en un pequeño número de grandes estudios, existen nichos rentables donde los estudios más pequeños pueden especializarse y competir eficazmente. Especializarse en géneros y enfoques cinematográficos puede ser la clave para destacar en un mercado saturado.

A lo largo de los años, géneros como acción, animación, terror y comedia han demostrado ser rentables, aunque con fluctuaciones en la rentabilidad. Para los estudios con mayores recursos, los géneros de acción y animación siguen siendo apuestas seguras, con una fuerte tendencia a generar ingresos elevados. Para los estudios pequeños, géneros como el terror y la familia presentan un alto retorno de inversión (ROI), ya que a menudo requieren presupuestos más bajos mientras siguen siendo populares y rentables. Además, las películas de clasificación R y PG-13 ofrecen un potencial de éxito financiero considerable, ya que tienden a generar altos ingresos en comparación con otras clasificaciones.

Otro factor crucial en la rentabilidad es la elección de los colaboradores, como directores y actores. Trabajar con directores y actores con una trayectoria más estable y un rango predecible de ingresos puede reducir significativamente el riesgo financiero. Por ejemplo, directores como Tim Burton y actores como Leonardo DiCaprio tienen una consistencia en sus resultados financieros, lo que los convierte en opciones menos arriesgadas para los estudios.

Finalmente, la duración de las películas también influye en el éxito. Las películas con una duración entre 70 y 150 minutos parecen ser las más rentables, ya que ofrecen un buen equilibrio entre los costos de producción y el atractivo para la audiencia.

En resumen, un nuevo estudio de cine debería considerar especializarse en géneros como el terror o la familia, trabajar con colaboradores de éxito comprobado, y producir películas de duración moderada. Al enfocarse en estos aspectos, el estudio puede minimizar riesgos y maximizar las posibilidades de éxito financiero en un mercado altamente competitivo.