

Tarea 1: Cine - Análisis de Ingresos de Taquilla

Link Repositorio: https://github.com/vlepin/INF379/tree/Tarea_1

Link Datos - Movie Dataset kaggle:

<https://www.kaggle.com/datasets/utkarshx27/movies-dataset?resource=download>

Integrantes:

Ignacio Arcos 202073083-0

Valentina Lepin 201904113-4

Matías Godoy 202073082-2

Análisis de Ingresos de Taquilla

En el presente informe se analizan seis visualizaciones generadas a partir de un dataset de películas que incluye información como ingresos, presupuesto, géneros, fecha de estreno, reparto, entre otros.

Criterio 1 - Valentina: Relación entre taquilla y el género principal

Datos Movie Dataset kaggle: Ingresos (*revenue*), géneros (*genres*)

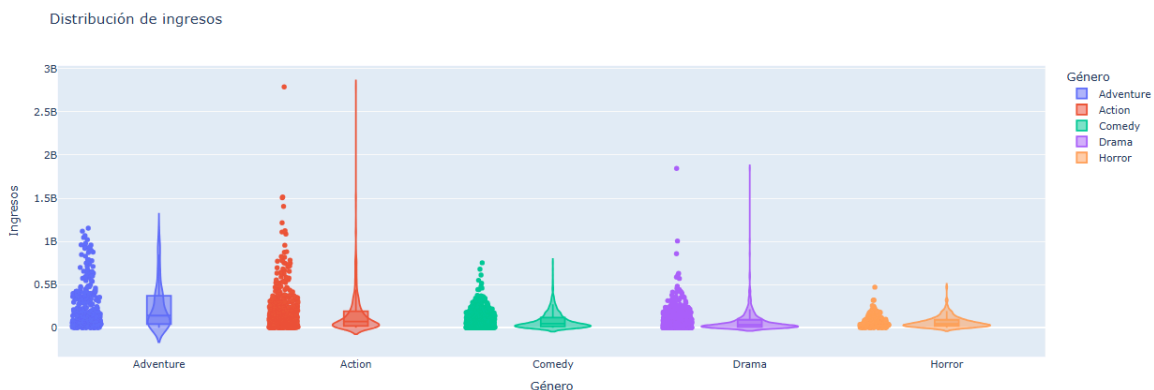


Figura 1: Violin plot polarizado

El primer gráfico corresponde a un diagrama de violín que muestra la distribución de los ingresos según los géneros más representados en el dataset. Para construir esta visualización, se seleccionaron las cinco categorías de género con mayor número de películas debido a la alta frecuencia de estas y se limitó la muestra a un máximo de 800 películas por género, seleccionadas aleatoriamente. Esta decisión permite balancear el conjunto de datos y evitar sesgos derivados de la presencia desigual de géneros, lo que garantiza una comparación más justa entre categorías.

El uso del diagrama de violín resulta útil en este contexto al combinar las ventajas del gráfico de caja mostrando la mediana, los cuartiles y los valores atípicos, más una estimación de la densidad de probabilidad, que permite visualizar la forma de la distribución de los datos. Esta representación facilita la identificación de géneros que presentan

ingresos consistentemente altos o bajos, así como aquellos con gran dispersión. Además, permite detectar outliers, es decir, películas que han tenido un rendimiento extraordinario en taquilla. Así, esta visualización ofrece una perspectiva clara sobre qué géneros tienden a ser más rentables y cuáles presentan mayor riesgo en términos de retorno económico.

Del gráfico de violín se puede concluir que el género Adventure es el más representado en términos de ingresos promedio, sin embargo Action contiene la película con mayores ingresos de toda la muestra, la cual se identifica claramente como un outlier. Asimismo, el segundo mayor ingreso corresponde al género Drama, también marcado como un outlier en su categoría.

Por otro lado, el género Drama presenta una distribución más concentrada, lo que sugiere que los ingresos de sus películas tienden a ser más estables y consistentes. En contraste, el violín de Adventure es el más estrecho y con mayor dispersión, lo cual indica una mayor variabilidad en los ingresos de las películas de ese género.

Finalmente, el género Adventure muestra un violín con un ancho significativo en la parte superior, lo que indica que, si bien no alcanza los ingresos más altos absolutos, mantiene un nivel elevado de ingresos de manera más frecuente en comparación con los otros géneros analizados.

Criterio 2 - Valentina: Comparación de ingresos en el tiempo

Datos Movie Dataset kaggle: Fecha de estreno (`release_date`), ingresos (`revenue`).

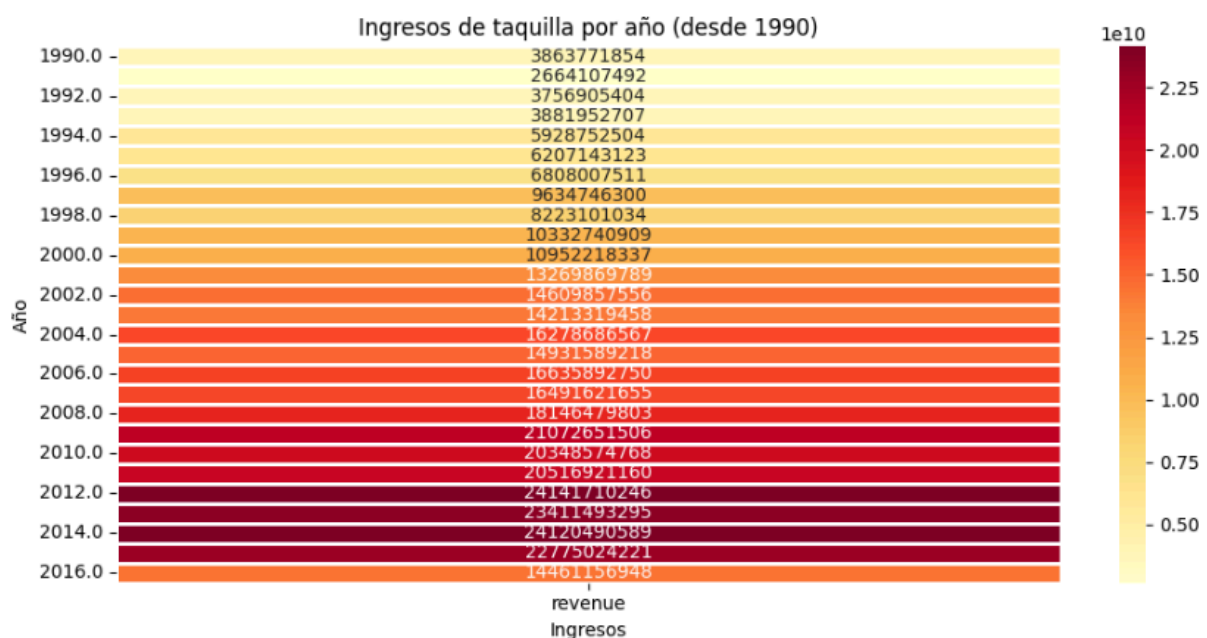


Figura 2: Calendar heatmap

El segundo gráfico corresponde a un heatmap que ilustra la suma total de ingresos anuales de todas las películas desde el año 1990, se limitó la muestra ya que no resulta provechoso

analizar esos años debido a la inflación, acceso al cine entre otros. Para su elaboración, se procesó la fecha de lanzamiento de cada película y se agregaron los ingresos por año. Este tipo de visualización resulta útil para captar patrones globales de comportamiento a lo largo del tiempo, permitiendo identificar rápidamente los años con mayores o menores niveles de recaudación gracias al uso de un gradiente de colores.

El heatmap permite observar de forma efectiva la evolución temporal de los ingresos en la industria cinematográfica. A través del color, es posible identificar años particularmente exitosos en términos de taquilla, los cuales podrían estar asociados con el estreno de grandes producciones o tendencias de consumo específicas. Asimismo, esta visualización permite reconocer años de bajo desempeño, posiblemente relacionados con factores económicos, sociales o eventos extraordinarios que hayan afectado la industria. De este modo, el gráfico proporciona un contexto histórico valioso para comprender cómo ha cambiado la rentabilidad del cine en las últimas décadas.

Del gráfico podemos observar que durante la década de los 90, los ingresos de taquilla fueron considerablemente más bajos en comparación con los años posteriores al 2000. A partir de este nuevo milenio, se aprecia una tendencia general al alza, con un pico significativo entre los años 2012 y 2015, lo que indica un período de alto rendimiento en la industria cinematográfica. Posteriormente, en el año 2016, los ingresos presentan una disminución notable. Por otro lado, el año con menores ingresos dentro del rango analizado es 1991, lo que lo posiciona como uno de los períodos menos exitosos en términos de recaudación.

Criterio 3 - Matías: Comparación de presupuesto vs ingresos

Datos Movie Dataset kaggle: Presupuesto (budget), Ingresos (revenue).

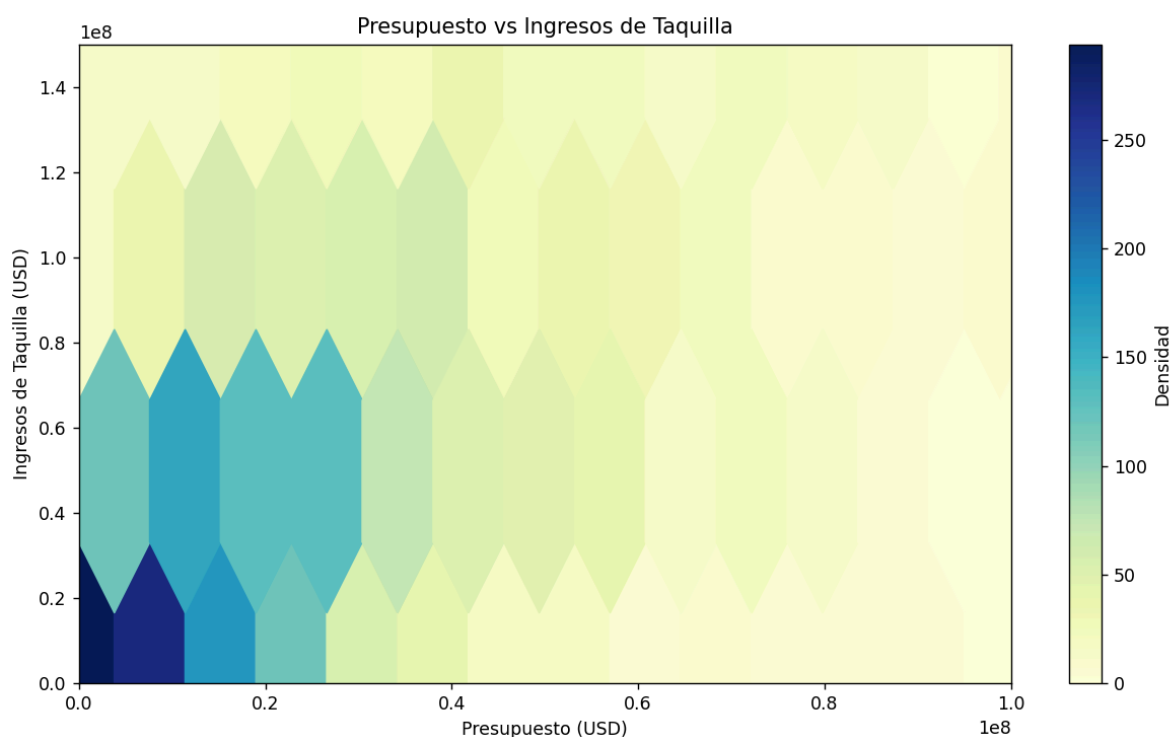


Figura 3: Gráfico de hexágonos

Este análisis busca visualizar la relación entre el presupuesto de una película y sus ingresos de taquilla. Se eligió un gráfico de hexágonos debido a su capacidad para mostrar de manera efectiva la densidad de las observaciones en el espacio bidimensional, especialmente cuando los datos están dispersos.

El gráfico de hexágonos permite ver en qué rangos de presupuesto se concentran las películas con mayores ingresos. Además, al ser una visualización de densidad, destaca las áreas con más películas que tienen un presupuesto similar y generan ingresos dentro de esos rangos. Esto ayuda a identificar patrones y tendencias, como si un mayor presupuesto está relacionado con mayores ingresos o si existen películas con bajo presupuesto que lograron gran éxito de taquilla.

Criterio 4 - Matías: Comparación de idioma original vs ingresos

Datos Movie Dataset kaggle: Idioma original (`original_language`), Ingresos (`revenue`).

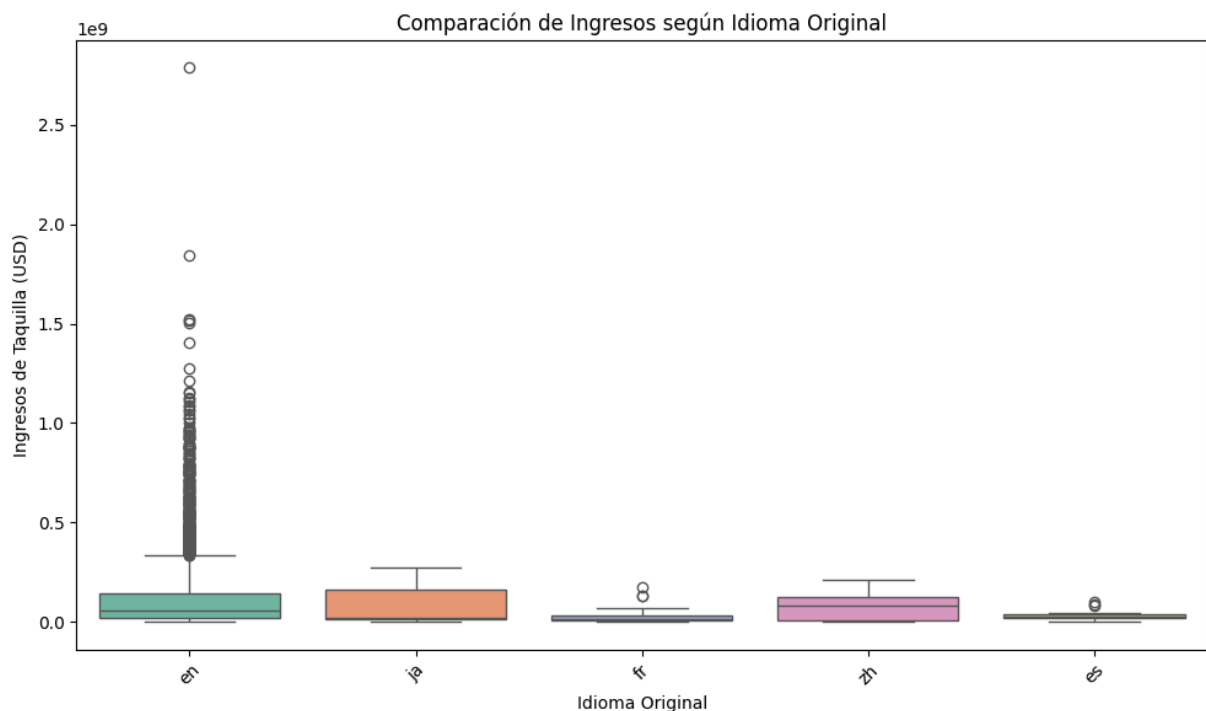


Figura 4: Gráfico de caja

Este análisis busca identificar si el idioma original de una película tiene algún impacto en los ingresos de taquilla. Para este propósito, se eligió un gráfico de caja, que es útil para visualizar la distribución de los ingresos de taquilla en diferentes idiomas originales, mostrando los valores centrales, la dispersión y la existencia de posibles valores atípicos.

El gráfico de caja permite comparar de manera clara y efectiva cómo se distribuyen los ingresos de taquilla en los diferentes idiomas, destacando el rango intercuartílico, los valores atípicos y la mediana de cada idioma. Esto puede ayudar a identificar si ciertos

idiomas tienden a generar más ingresos que otros, o si existen algunas películas en idiomas menos comunes que lograron un éxito significativo.

Criterio 5 - Ignacio: Cantidad de películas y cantidad de ganancias por país

Datos Movie Dataset kaggle: Ingresos (`revenue`), Países de producción (`production_countries`).

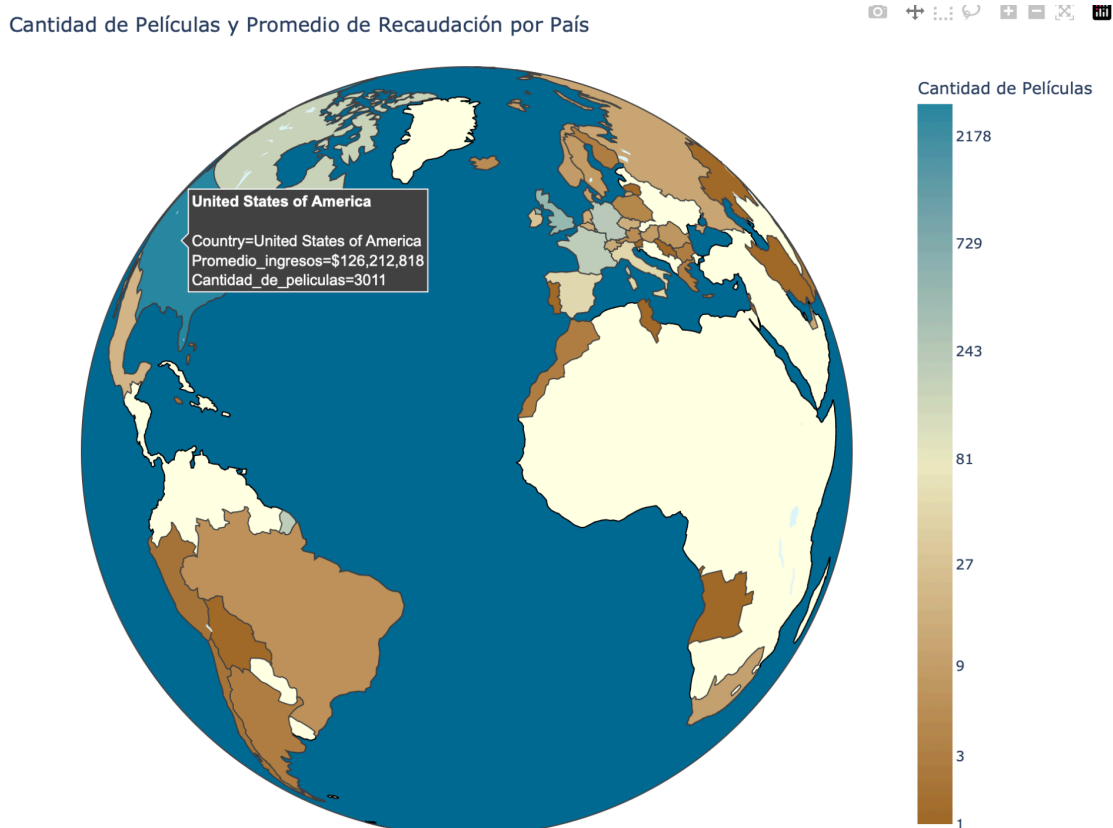


Figura 5: Gráfico de mapa

El quinto gráfico busca comparar las ganancias de países que posean cantidades similares de películas, utilizar este criterio nos brinda información sobre la forma en que influye la zona, además de, considerar la cantidad de películas que han sido lanzadas en el país correspondiente buscando relaciones entre estas informaciones. Este tipo de gráfico es útil para encontrar relaciones entre diferentes Zonas geográficas, Países y Continentes.

Se utilizó una escala de potencias de 3, ya que, el dataset posee una diferencia significativa entre la cantidad de películas producidas en los Estados Unidos y los otros países, el realizar esto permite la diferenciación de color en una cantidad baja de películas por ejemplo en el caso de Chile y Bolivia, se observa un cambio de color siendo que la

diferencia es tan solo de un par de películas, si esto no se hiciera de esta forma , estados unidos seria el unico que poseyera un color distinto.

Podemos concluir a partir del gráfico que gran parte de la concentración de películas e ingresos corresponde, Reino Unido, Francia, Alemania y Canadá , esto nos indica que son los países mas representativos del dataset en caracteres de ingresos y cantidad de películas .

Algunos países con **pocas películas** muestran un **promedio de ingresos elevado**, lo que podría indicar **producciones selectas pero exitosas**.

Criterio 6 - Ignacio: Productoras y cantidad de ingresos por película históricos

Datos Movie Dataset kaggle: Ingresos (`revenue`), géneros (`genres`), fecha de lanzamiento(`release_date`) ,nombre de la película (`original_title`) y `original_title`) y compañías de producción (`Production_companies`)

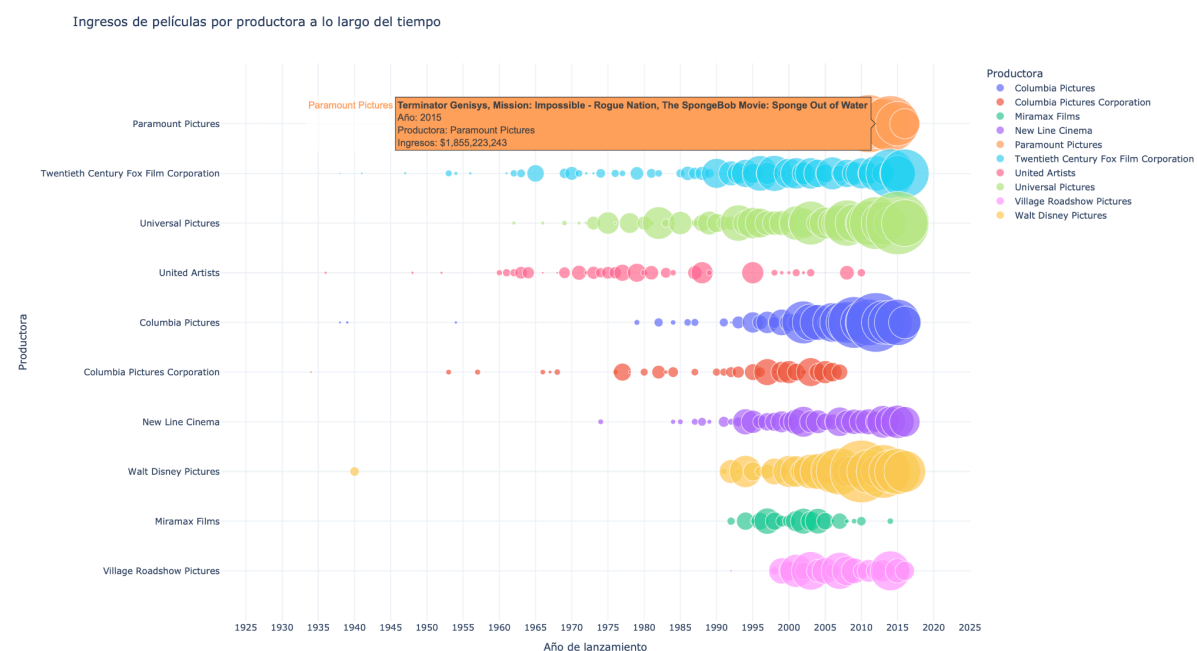


Figura 6: Gráfico de burbujas

El sexto gráfico busca comparar las ganancias históricamente según la productora de la película , este tipo tiene como finalidad buscar una relación entre el crecimiento histórico de las productoras de películas y las ganancias generadas .



Se limitó la cantidad de productoras a las 10 más frecuentes, ya que son aquellas que pueden proveer una mayor cantidad de información históricamente , esto para disminuir la saturación visual.

En conclusión, algunas productoras como **Universal Pictures**, **Warner Bros.** y **Paramount Pictures** dominan en términos de producción y generación de ingresos durante múltiples años .