Segunda entrega proyecto de curso

Angie Dayana Rincón Mandón Fonsy Johan Mercado Agudelo Orlando José Salazar Polo

Introducción a la Inteligencia Artificial para las Ciencias e Ingenierías

Raúl Ramos Pollan



UNIVERSIDAD DE ANTIOQUIA FACULTAD DE INGENIERÍA MEDELLÍN 2023

Exploración de datos

Para esta exploración sólo se analiza la información de las películas, explorando el presupuesto, los ingresos, popularidad, cantidad de votos, la media de votos, géneros cinematográficos, entre otros.

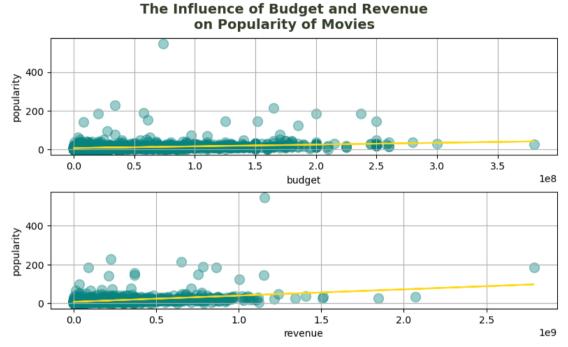


Figura 1. Influencia del presupuesto y los ingresos sobre la popularidad de las películas.

De la Figura 1 se puede ver que el presupuesto que se tiene para la realización de una película no está directamente relacionado con su popularidad, mientras que los ingresos generados por estas películas si tienen un poco más de relación con la popularidad de estas.

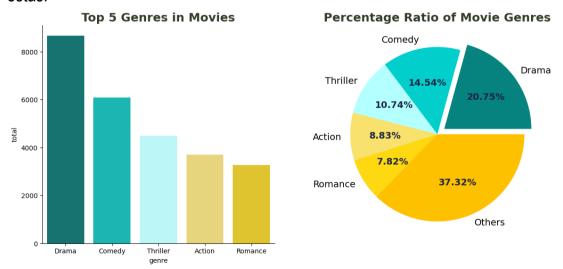


Figura 2. Top de géneros cinematográficos.

El top de géneros cinematográficos son drama, comedia, terror, acción y romance, siendo el género de drama el género más usado para la realización de las películas, como se puede ver de la Figura 2, aunque cabe resaltar que las películas suelen ser etiquetadas por más de un género. En los otros géneros se tienen etiquetas como guerra, animación, historia, documental, familia, entre otros.

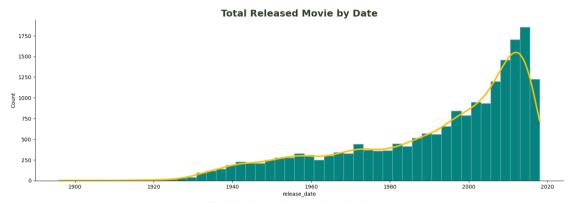


Figura 3. Total de lanzamientos de películas por año.

La Figura 3 presenta un análisis de la cantidad de lanzamientos de películas por año. Al examinar la gráfica, es evidente que la industria cinematográfica ha experimentado un aumento significativo en la cantidad de películas lanzadas en los últimos años. En particular, se destaca una concentración de estrenos en el período que abarca desde el año 2004 hasta el 2020.

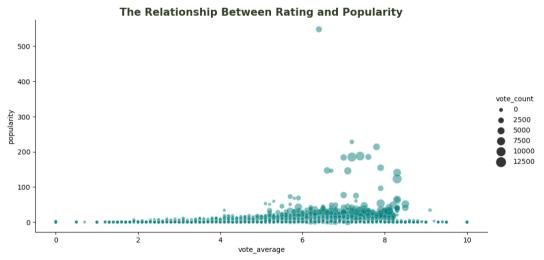


Figura 4. Relación entre rating y popularidad.

La Figura 4 presenta una relación significativa entre la popularidad de las películas y la media de votos emitidos por los usuarios, además de una fuerte dependencia con respecto a la cantidad de votos recibidos por cada película en particular.

Data Distribution Across Top 5 Genres

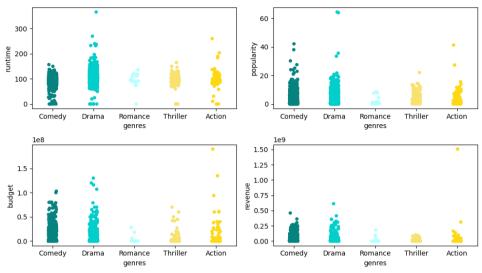


Figura 5. Distribución de datos en las películas del top 5 de géneros cinematográficos.

La Figura 5 revela varios datos relevantes. Entre ellos, se destaca que el género cinematográfico con la mayor duración media es el drama, mientras que el género menos popular dentro del top 5 es el romance. Además, se puede observar que las películas de acción gastaron más dinero que el resto de los géneros y que una de ellas logró una gran ganancia en comparación con las demás.

Correlation of Movie Features



Figura 6. Correlación entre las características cinematográficas.

La Figura 6 muestra una representación visual de la relación entre el costo de producción de una película, sus ingresos en la taquilla, la cantidad de votos emitidos por los usuarios y los costos de marketing asociados. A partir de esta figura, se puede notar una buena correlación entre estos indicadores.

endarrive followcalled tries Story tries S

The Most Common Word in Movie Overviews

Figura 7. Nube de palabras más comunes en la descripción de películas.

En la Figura 7 se puede observar una nube con las palabras más usadas para la descripción de las películas en la base de datos analizada. Esta nube de palabras es una técnica de visualización que permite clasificar las palabras más comunes en un texto y representarlas gráficamente en función de su frecuencia de uso.

A partir de esta figura, se pueden identificar algunas de las palabras más utilizadas para describir las películas en la base de datos. Algunas de las palabras más frecuentes son "vida", "amor", "historia", "mundo", "familia", "amigos", entre otras. Estas palabras dan una idea general de los temas y motivos recurrentes en las películas que conforman la base de datos.