



Data Science II

Alumna:

Luz Eva Morales

Profesor:

Jorge Ruiz

Comisión:

#61605

Introducción

En los últimos años, el arte cinematográfico se fue consolidando como mercado, donde se mueven grandes cantidades de dinero de recaudación que significan ganancias para sus empresas distribuidoras. Esto se debe a la masificación y globalización de las películas hollywoodenses.

La audiencia de estos films es poco predecible, ya que tiene que ver con muchos factores sociales, económicos y contextuales. La mejor herramienta que tenemos para traducir las preferencias de las audiencias es la recaudación que estas películas tienen y cómo difieren los datos entre las preferencias del público doméstico (U.S.A), y el resto del mundo.

Los críticos de las películas no siempre suelen coincidir con los niveles de audiencia, pero podemos inferir que intervienen cuando una persona elige ver tal o cual película.



En este trabajo vamos intentar determinar si las críticas especializadas coinciden con las recaudaciones totales de los films, así como la relación con las críticas del público y los diferentes géneros de las mismas.

Para ello usaremos diferentes variables de dos datasets diferentes, construiremos hipótesis y llegaremos a insights que nos ayudarán a concluir sobre las hipótesis creadas.



Descripción de los datasets utilizados

Para la comparativa, vamos a utilizar dos datasets descargados de Kaggle. El primero trata sobre críticas subidas a la reconocida página “**Rotten**



Tomatoes” y tiene una gran cantidad de películas con su información correspondiente y, sobre todo, los datos en los que más nos enfocaremos que son las críticas especializadas y las críticas del público para poder compararlas con los datos extraídos de nuestro segundo dataset.



El otro dataset que usaremos es sobre la información de recaudaciones de diferentes películas que van del período 1990-2020, donde contiene varias informaciones de los films pero nos centraremos en las recaudaciones domésticas, internacionales, mundiales y en el género principal de cada film. Estos datos son recopilados y extraídos de la web de “**Box Office Mojo by IMDbPro**”, quienes son unos grandes recopiladores de data sobre los films.

La conjunción que haremos entre los dos datasets es a partir del nombre de las películas, es por ello que quedaran muchos films del DF1 de Rotten Tomatoes, ya que nos enfocaremos en el período propuesto por el DF2.

Creamos el DF3 entre ambos DataFrames para poder tener los datos de forma clara y poder compararlos mejor. Este DF3 se obtiene eliminando la información irrelevante del DF1 y DF2, eliminando algunos datos nulos, transformando algunos de estos datos nulos y manteniendo varios de ellos, sobre todo en columnas con datos de tipo string, por ejemplo, nombre de Director del film, ya que no modificara las variables que analizaremos.

Los datos nulos que modificamos a 0 son las recaudaciones domésticas e internacionales, ya que tenemos los datos de la mundial e inferimos que ambos datos sean igual a 0, es decir, sin recaudación en esos territorios.

Los datos nulos en la columna del puntaje de la audiencia los convertimos en una media entre todos los datos de la columna para asignarles un dato que sea neutral pero que sea nulo ni 0.

Cambiamos nombres de las columnas a español y modificamos los tipos de datos para que queden más prolijos y legible.

Así queda el nuevo DF:

Película	string[python]
Clasificación	string[python]
Director	string[python]
Fecha de lanzamiento	datetime64[ns]
Aprobación de la critica	string[python]
Puntaje medio críticas	float64
Puntaje de audiencia	float64
TBC	int64
TMC	int64
Duración	string[python]
Distribuidor	string[python]
Presupuesto	float64
RD	float64
RI	float64
RT	float64
Género	string[python]
dtype: object	

Referencias:

TBC	'Total Buenas Criticas'
TMC	'Total Malas Criticas'
RD	'Recaudación Doméstica'
RI	'Recaudación Internacional'
RT	'Recaudación Total'

Hipótesis

Al analizar el rendimiento de las películas de Hollywood, podemos plantear diversas hipótesis para explorar y validar, especialmente enfocadas en la relación entre las críticas recibidas y los resultados financieros de las películas. Estas hipótesis son suposiciones preliminares que se pueden someter a pruebas utilizando los datos disponibles en los DataFrames. A continuación, presentamos algunas posibles hipótesis:

1. Impacto de las Críticas en la Recaudación: Se podría hipotetizar que las películas con mejores críticas, tanto especializadas como del público, tienen una mayor recaudación en taquilla. Esta hipótesis podría analizarse comparando las calificaciones de las críticas con las cifras de recaudación, tanto doméstica como internacional, para ver si existe una relación entre la recepción crítica y el rendimiento económico de las películas.

2. Relación entre Presupuesto y Recaudación: Se podría investigar si las películas con mayores presupuestos tienden a generar mayores ingresos, tanto en el mercado doméstico como internacional. Este análisis ayudaría a comprender cómo la inversión en producción de una película se traduce en ingresos en taquilla.

3. Críticas vs. Presupuesto: Podría hipotetizarse que el presupuesto de una película está relacionado con la recepción crítica de la misma. Este análisis permitiría explorar si las películas con un presupuesto más alto tienden a recibir mejores críticas, tanto de los especialistas como del público, independientemente de su desempeño en taquilla. Además, se podría investigar si el mayor presupuesto se traduce en una mejor calidad percibida por los críticos y la audiencia, ya sea por los efectos visuales, la calidad de la producción o el elenco.

4. Críticas vs. Recaudación según Género: Se podría hipotetizar que la relación entre las críticas y la recaudación de una película varía según el género al que pertenece. Este análisis permitiría explorar si ciertos géneros, como los de acción o ciencia ficción, tienden a generar mayores ingresos en taquilla independientemente de las críticas recibidas, mientras que otros géneros, como los dramas o películas independientes, pueden estar más influenciados por las críticas especializadas y del público. Así, se podría investigar si las películas de ciertos géneros son más sensibles a las opiniones críticas y cómo esto afecta su desempeño en taquilla tanto a nivel doméstico como internacional.