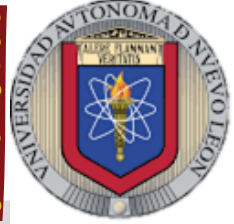




# CINE 2018

1884093 GARCÍA ROSAS TANIA DESSIRÉ  
1724741 VALERO FUENTES MARICELA ALEJANDRA



## 1. INTRODUCCIÓN

El interés que entraña el cine, su magia y su belleza, la versatilidad de sus técnicas y la infinita gama de contenidos es, en muchas ocasiones, la clave de la investigación sobre otras épocas, historias, relatos o documentos, o sobre el mismo cine, su lenguaje y su tecnología. **El cine es lenguaje vivo.**

Un lenguaje que no estuviera en constante adaptación y crecimiento, moriría. Por ello avanza, se enriquece, mira al pasado, busca futuro, enriqueciendo su propio lenguaje y el de la sociedad, acrecentando y mejorando así su propio lenguaje. El lenguaje del cine es interactivo, enriquecido por el avance imparable y vertiginoso de la nueva tecnología, por lo que se hace imprescindible en el debate social y se convierte en inexcusable vehículo de cultura.

## 4. METODOLOGÍA

Se tiene una base de datos en la que son clasificadas 52845 películas. Se clasifican de manera cualitativa por género, país, descripción, notas, quien la dirigió, los protagonistas. De manera cuantitativa se contempla la duración, el año, los votos y el promedio de evaluación. Estamos evaluando 10 variables para cada película. .

Se toman como significativas las variables de año, género, duración, puntuación promedio y nacionalidad. Esto debido a que se los actores y directores cambian constantemente, las variables de descripción y notas abarcan muchos valores nulos con los que nos es imposible trabajar, y el id de la película es específicamente de la variable de datos por lo que no afectaría en nada eliminarla. De manera que, la base de datos tiene una dimensión de 512445x6. Ver Figura 1.

	title	year	genre	duration	country	avg_vote
0	Hot Water	1924.0	Comedy	59	United States	10.0
1	Gabriel Over the White House	1933.0	Drama	87	United States	10.0
2	Zuiderzee	1934.0	Documentary	54	Holland	10.0
3	Intermezzo	1936.0	Drama	93	Sweden	10.0
4	Fire Over England	1937.0	Adventure	92	Great Britain	10.0

FIGURA 1. Previsualización de la base de datos.

## 5. ARQUITECTURA

Para realizar una predicción se ocupa una serie de tiempo, lo cual es un registro periódico de una o más variables cuantitativas, ya sea anual, mensual, diaria, etc.

La regresión inicia al evaluar estas variables en torno al tiempo y puede o no tener tendencia. En la figura 2 se presenta un ejemplo de una serie de tiempo con una tendencia lineal.



FIGURA 2. La gráfica muestra un ejemplo de movimientos estacionales.

## 6. PARÁMETROS

Se analizan nuestros datos para obtener la cantidad de películas estrenadas por año, de modo que se obtenga una fórmula de regresión para poder realizar la una gráfica donde se verán proyectados dichos datos y el resultado, el cual se verá reflejado en la gráfica. Nuestra serie de tiempo va de manera anual empezando en el año 1909.

En el gráfico de la Figura 3 se muestra la predicción con la línea roja y los datos reales con los puntos en azul. La línea de regresión tiene un margen de error 12.58%, esto debido a que un principio la producción de películas iba de manera exponencial, hasta hace unas décadas donde se aprecia que la producción sigue un patrón diferente. Sin embargo, dicho margen de error puede variar según la amplitud del tiempo que manejes y la variabilidad de los datos.

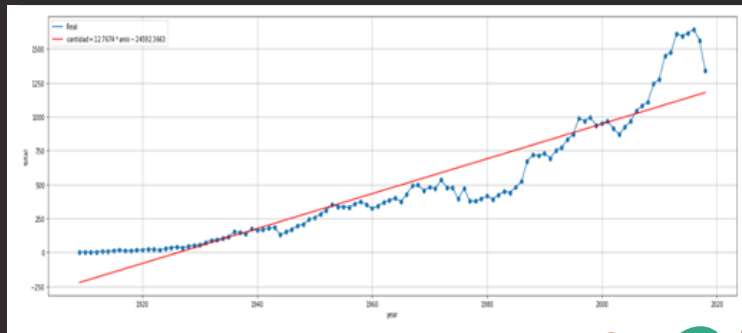


FIGURA 3. Gráfico de regresión.

## 7. EXPERIMENTACIÓN

Los géneros de las películas fueron evaluados por cantidad de producciones anuales y promedio de votos.

En la proyección del gráfico de barras (Figura 4) se ve la cantidad de películas que se han realizado por género. Nuestro top cinco de producciones por género es de Drama (15861), Comedia (11600), Suspense (4868), Documentales (3153) y Acción (2868). Apreciamos que los primeros géneros con mayor producción difieren mucho entre sí. Después de los primeros diez, la producción por género se distribuye de manera más uniforme.

En base al BoxPlot (Figura 5) nos damos cuenta de la relación del promedio de votos por género. Son pocos los géneros en donde los valores no están sesgados, es decir, donde nuestra mediana, media y moda coinciden. Si se tuvieran los mejor promediados se tendrían representados el género Noir, Mélo, Experimental, Gangster y documentales.

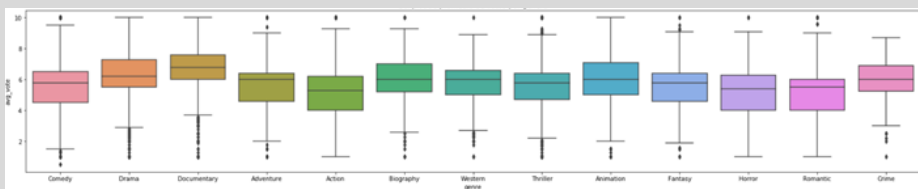


FIGURA 5. Boxplot con el promedio de votos de los trece géneros que más se producen.

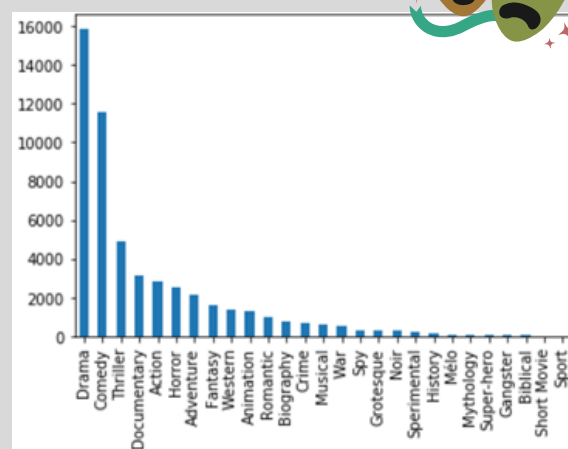


FIGURA 4. Cantidad de películas por género.

Los géneros que tienen mayor relación entre promedio de votos y cantidad de películas son los Documentales, las películas de Drama, de Animación, Biográficas y de Crimen.

## 8. CONCLUSIONES

En base a lo anterior, se concluye que no por producirse más un género se va a tener mejor puntaje. Se producen más debido a la demanda que tiene el género, sin embargo, eso no nos asegura que sean buenas o que sean bien evaluadas.

Por otro lado, consideramos que el hecho de tener más plataformas streaming para ver películas ha provocado que la producción se reduzca. Esto debido a la propiedad que se tienen que prescindir dichas plataformas, aparte de que la visualización en ellas es más accesible para el público en general.



FIGURA 6. Plataformas de streaming de video más utilizadas

## 9. REFERENCIAS

• Python: <https://www.python.org/>:  
Lenguaje de programación interpretado.

• Base de datos digital: filmsmtv movies  
<https://www.kaggle.com/stefanoleone992/filmtv-movies-dataset>

