

# Análisis de datos sobre películas, sus metadatos y premios

Javier Pozueco, Technical Analytics Director, Jellyfish,

29 de junio de 2017

## Resumen

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

**Keywords:** IMDb, Kaggle

## 1. Introducción y objeto del análisis

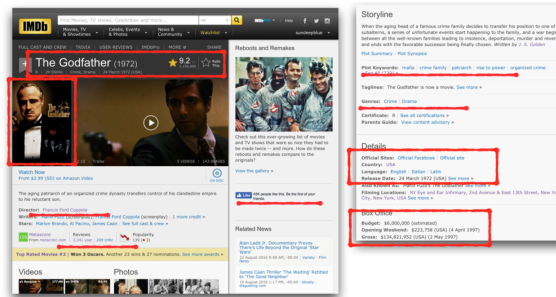
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

## 2. Carga de los datos y análisis descriptivo

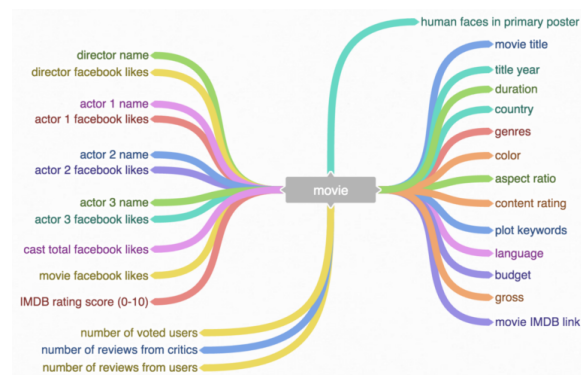
Existen multitud de bases de datos de libre disposición con información general relacionada con películas. Una de estas bases de datos es La Internet Movie Database (o IMDb de manera abreviada), que muestra toda su información a través del portal web <http://www.imdb.com>.

Para esta base de datos en particular, el usuario tiene acceso al título de las películas, el nombre del director, las personas que han participado en su realización, el género, su presupuesto o el dinero recaudado, como se puede ver a continuación:



A pesar IMDb sólo presenta su información a través de su portal web, hay trabajos como [1] que ya han extraído los datos más relevantes utilizando, en este caso en particular, Python y su librería [scrapy](#), siguiendo los siguientes pasos:

- Se ha utilizado scrapy para obtener un listado de 5043 películas desde el portal [The Numbers](#).
- Se ha realizado una búsqueda en IMDb para obtener el enlace a cada una de las películas.
- Para cada uno de los enlaces, se ha descargado y parseado su contenido, con el fin de obtener los datos más relevantes que se pueden ver en la siguiente figura:



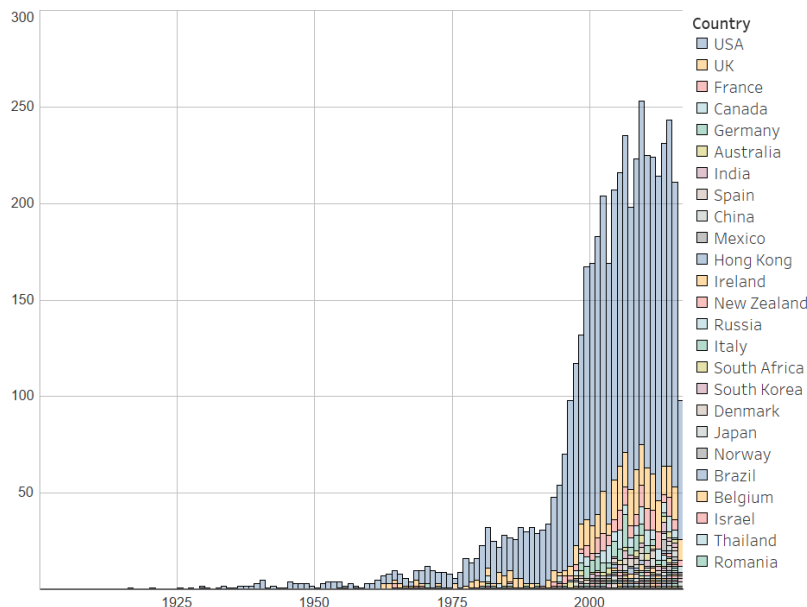
- Mediante el reconocimiento de imágenes, también se sabe para cada película el número de caras que aparecen en su póster.

A continuación se incluyen notas importantes del autor, sobre cada uno de los atributos extraídos de las películas, que deberemos tener en cuenta en el análisis que se realizará más adelante:

- En aproximadamente 800 películas las ganancias son 0, debido a que esta información no está disponible o porque la herramienta utilizada para extraer el dato no devolvió ninguna respuesta en un tiempo razonable.
- En 908 directores de las películas descargadas, el número de likes en Facebook es 0 debido, como en el caso anterior, a que los valores aparecen en un marco que no se carga junto con el resto de la página.
- Hay películas para las que no se ha tenido en cuenta la moneda del país que las ha producido, y aunque en los datos se muestran dólares, en realidad es la moneda del país correspondiente.
- Para calcular los actores principales de cada película, se han tenido en cuenta todos los actores y actrices del reparto, y aquellos tres con mayor número likes en Facebook, son considerados los principales.
- Por último, cabe destacar también que el presupuesto y las ganancias de las películas no tienen en cuenta la inflación o el cambio de moneda que había en el año de su realización.

## 2.1. Datos generales sobre el dataset

Como primer paso vamos a analizar el dataset extraído de [Kaggle](#), que será el utilizado en este trabajo. Para ello se muestra en la siguiente figura el número de películas por país en nuestro dataset cada año disponible en [Tableau Public](#):



A continuación se muestra el número total de películas, incluyendo también series, según IMDb[2]. El número de películas descargado resulta ser inferior al 3 % del total, pero al ser un listado con los datos disponibles en [The Numbers](#), podemos considerar que son las que tienen ganancias conocidas y mayores:



## 2.2. Rating otorgado por los usuarios

En el listado de películas considerado se incluye también el rating otorgado por los usuarios según IMDb. Este valor proviene de las puntuaciones que dan los usuarios con valores comprendidos entre 0 y 10, y que a través de diferentes métodos se pondera para evitar que un mismo usuario pueda votar varias veces.

Para que una película pueda estar dentro del top 250 de IMDb sólo se tienen en cuenta las votaciones de los usuarios más frecuentes, aunque para evitar cualquier tipo de engaño los requisitos necesarios para ser un usuario frecuente no se han publicado:

$$W = \frac{R_v + C_m}{v + m} \quad (1)$$

donde:

$W$  = rating ponderado

$R$  = media de las votaciones de 0 a 10

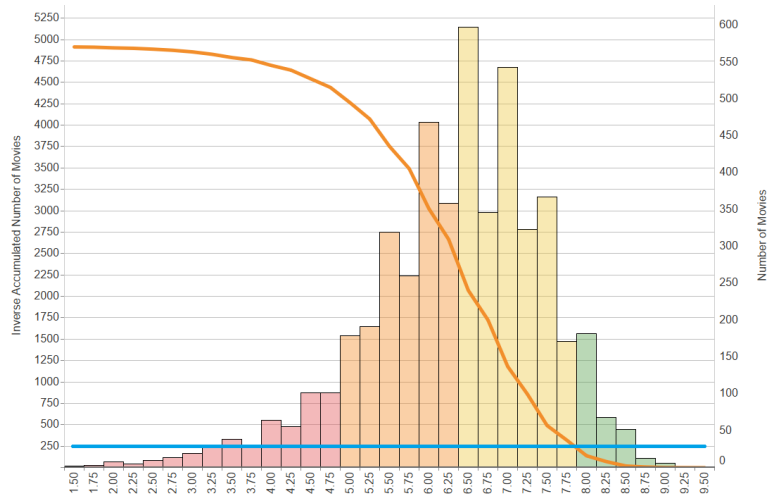
$v$  = número de votos

$m$  = mínimo número de votos para estar en el top 250

$C$  = la media de todos los votos

En la figura que se muestra a continuación se puede observar la distribución del rating otorgado por los usuarios en el dataset considerado. En este gráfico se han destacado las películas según su calidad, en base a los siguientes criterios disponible en [Tableau Public](#):

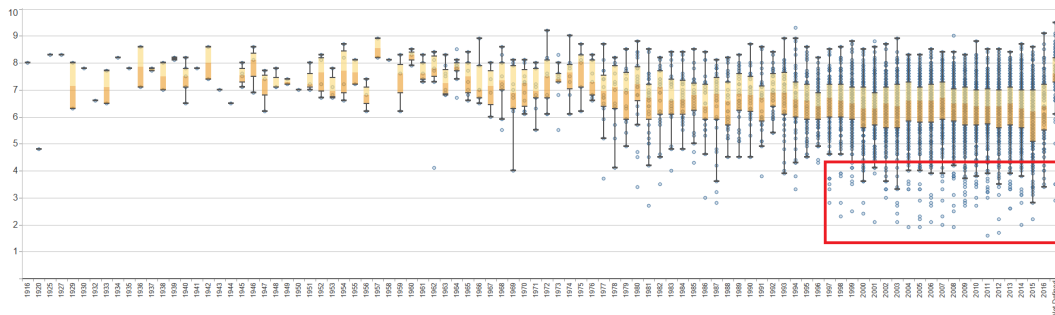
- Verde, si las películas son muy buenas y superan una puntuación de 8.
- Amarillo, si las películas son buenas y tienen una puntuación entre 6,5 y 8.
- Naranja, si las películas son regulares y tienen una puntuación entre 5 y 6,5.
- Rojo, si las películas son malas y bajan del 5.



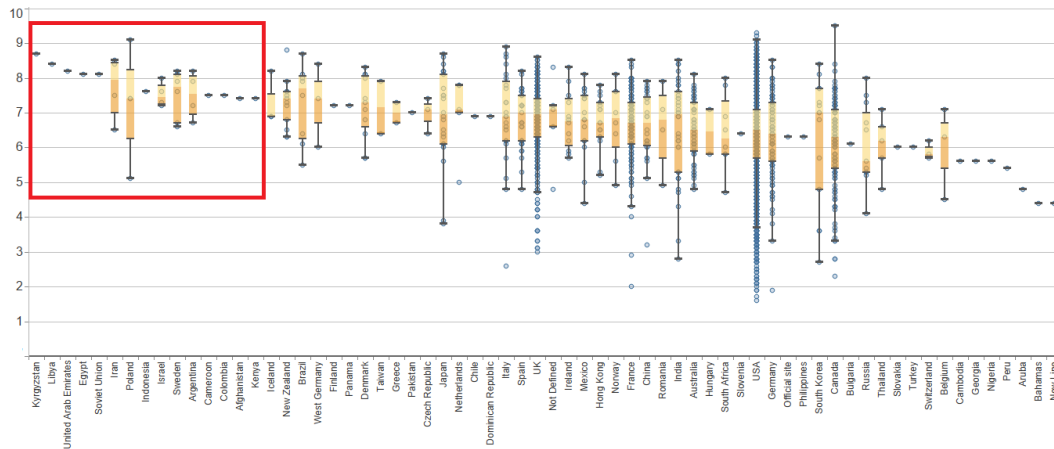
Cabe destacar también como el top 250 de IMDb se corresponde con una puntuación ligeramente inferior a 8, según el gráfico anterior.

### 2.3. Distribucion de las puntuaciones

Desde finales de la segunda guerra mundial, se ha producido un incremento gradual de las películas producidas debido al avance de la tecnología para la realización de las mismas. Además, como se pudo ver anteriormente en el número de películas producidas cada año, desde finales del siglo pasado el avance de los medios digitales ha facilitado y abaratado su distribución para multitud de productoras independientes[3]. Por otro lado, en la siguiente figura también se puede observar que aunque el número de trabajos ha aumentado, también ha disminuido la calidad en algunos casos disponible en [Tableau Public](#):

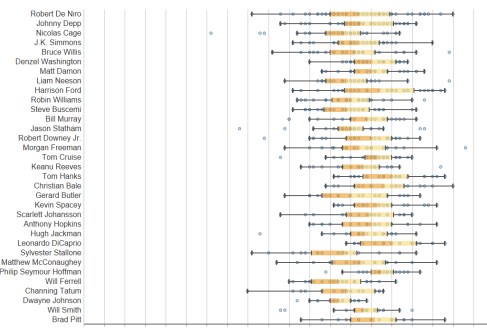


Hay países como Kirguistán, Libia, Egipto o Irán con un reducido número de películas pero cuya puntuación es superior al resto. Por otro lado Estados Unidos, Reino Unido y Francia son los países con más películas, pero cuya calidad no siempre es mejor que las demás. Cabe destacar que en el dataset que estamos estudiando aparecen películas de la India pero no en gran cantidad, siento este país uno de los productores actuales más grande disponible en [Tableau Public](#):

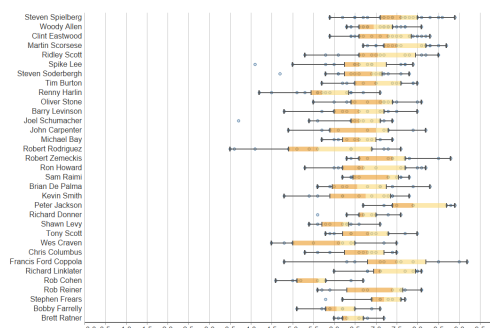


## 2.4. Análisis de actores y directores

Resulta interesante también analizar las puntuaciones de los actores y directores con mayor número de películas. Se puede ver a continuación como la variación de las puntuaciones de los actores es mayor que la de los directores más populares. En el caso de los actores, algunos nombre como Robert De Niro, Sylvester Stallone o Channing Tatum con buenas películas, pero también con películas de calidad mediocre. Otros actores como Matt Damon destacan por tener un gran número de películas de calidad notable y Nicolas Cage, por ejemplo, por haber realizado algunos papeles de mala calidad. Con respecto a los directores destaca Steven Spielberg como el actor mas prolífico y siempre de gran calidad y Robert Rodríguez con diferentes producciones pésimas disponible en [Tableau Public 1](#) y [Tableau Public 2](#).

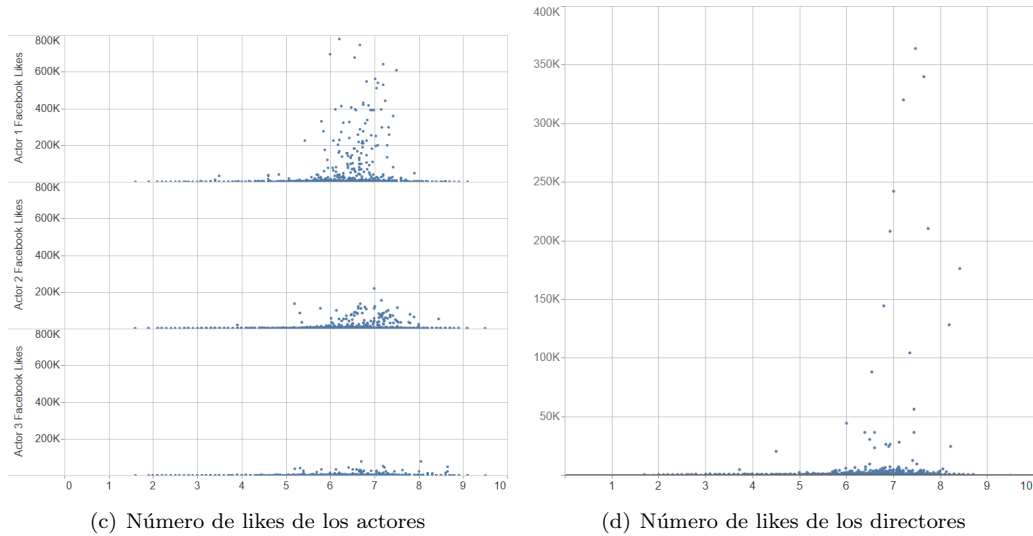


(a) Actores más populares



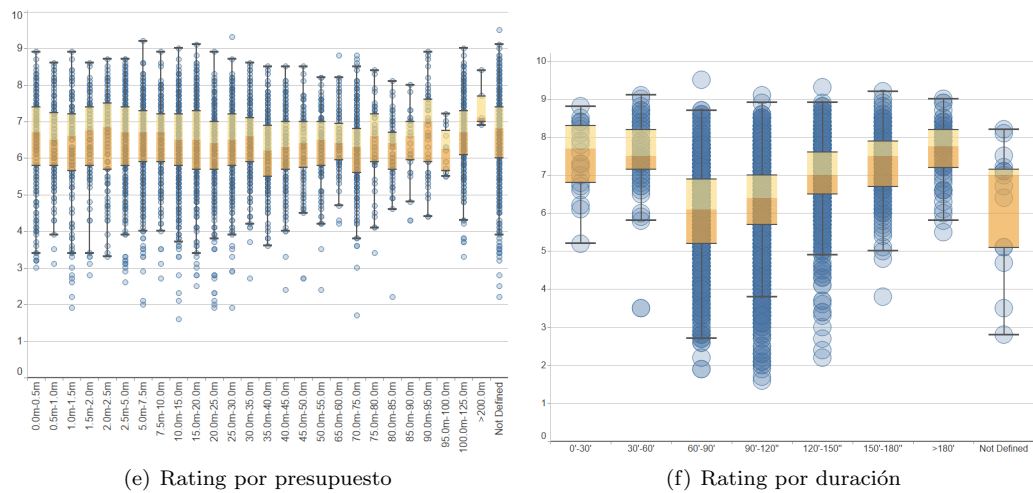
(b) Directores más populares

Para saber si influye el número de likes en la puntuación de las películas, podemos ver en los siguientes gráficos como es más importante en los actores tener más likes para obtener mejores resultados, y sobre todo el en actor principal de la película, pero en los directores aunque con menos likes, las puntuaciones son superiores disponible en [Tableau Public 1](#) y [Tableau Public 2](#):



## 2.5. Análisis de presupuesto y duración

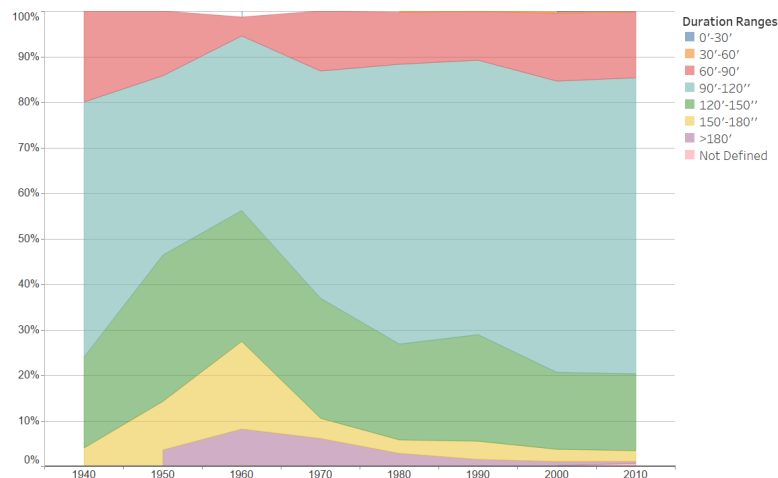
Para saber si influye el presupuesto y la duración de las películas en la puntuación obtenida en IMDb, se puede ver como a mayor dinero invertido menor es la dispersión en las puntuaciones, y aquellas películas que son mejores tienen una duración de menos de una hora o de longitud más larga posible disponible en [Tableau Public 1](#) y [Tableau Public 2](#):



La combinación de estos dos factores nos muestra que una gran duración y presupuesto moderado resulta en mejores películas y como las películas de entre media hora y una hora también tienen gran aceptación entre el público disponible en [Tableau Public](#):

Budget Ranges	0'-30'	30'-60'	60'-90'	90'-120"	120'-150"	150'-180"	>180'	Not Defined
0.0m-0.5m	6.80	7.50	6.49	6.45	5.97	8.30		5.10
0.5m-1.0m			6.56	6.37	7.50	8.30	8.10	
1.0m-1.5m		8.05	5.59	6.46	6.74	6.20		7.00
1.5m-2.0m	6.80		6.32	6.49	7.44	4.80		
2.0m-2.5m		8.20	6.28	6.43	7.25	7.83	8.70	
2.5m-5.0m	6.60	7.80	5.64	6.56	7.64	7.60	8.25	
5.0m-7.5m		6.47	5.62	6.50	7.08	7.20	7.90	6.70
7.5m-10.0m			6.33	6.32	7.13	7.44	7.80	
10.0m-15.0m			5.84	6.25	7.04	7.60	8.00	
15.0m-20.0m			5.46	6.29	7.28	7.56	8.08	
20.0m-25.0m			5.16	6.19	6.79	6.43	7.80	7.20
25.0m-30.0m			5.38	6.20	7.21	7.70	7.40	
30.0m-35.0m	8.10		6.18	6.26	7.07	7.00	7.90	
35.0m-40.0m			5.12	6.03	6.80	7.93	7.83	
40.0m-45.0m			5.99	5.92	7.02	7.20	7.57	
45.0m-50.0m		7.30	5.75	6.17	6.73	7.00	7.60	
50.0m-55.0m			5.50	6.15	6.61	7.90	7.43	
55.0m-60.0m			6.37	6.24	6.85	6.35	6.47	
60.0m-65.0m			5.55	5.97	6.71	7.43	7.55	5.10
65.0m-70.0m			6.00	6.15	6.79	7.20		
70.0m-75.0m			6.17	6.07	6.57	6.00		
75.0m-80.0m			5.53	6.18	7.01			
80.0m-85.0m			6.77	6.21	7.19	8.00	8.90	
85.0m-90.0m			6.73	6.13	5.95			
90.0m-95.0m			5.86	6.48	6.83	7.13	7.13	
95.0m-100.0m				7.20	7.75			
100.0m-125.0m								
>200.0m								
Not Defined	7.65	7.52	6.47	6.37	6.81	7.15	7.93	6.25

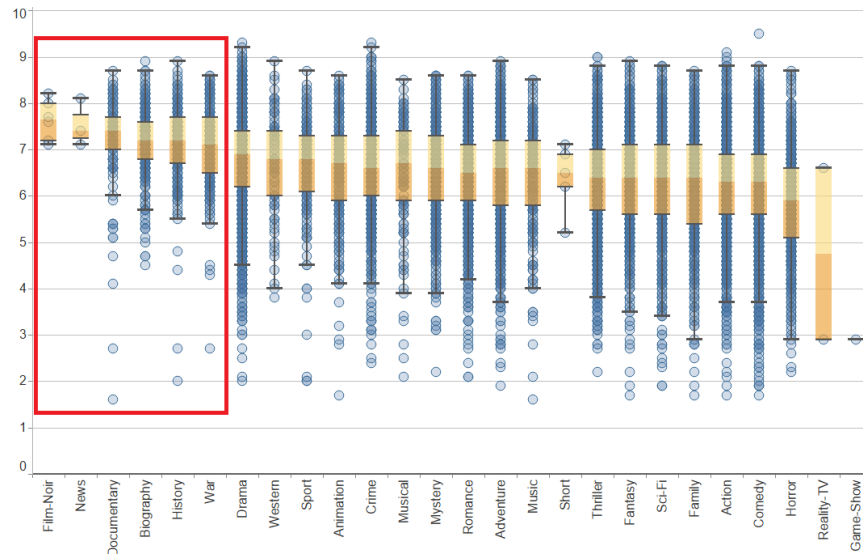
Como curiosidad se puede ver como ha variado la duración de las películas a lo largo del tiempo[4], podemos representar los rangos de duración anualmente y ver como son cada vez más populares las películas con una duración de entre 90 y 120 minutos disponible en [Tableau Public](#):



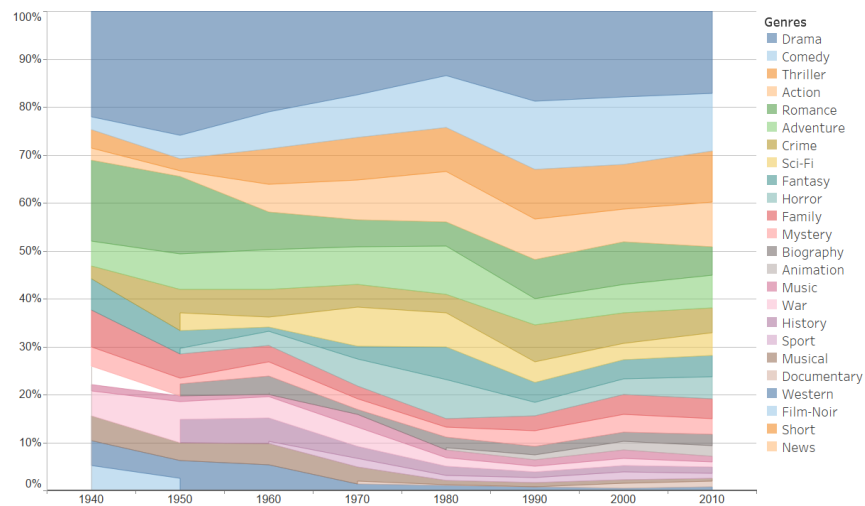


## 2.6. Análisis del género de las películas

Para saber si el género de las películas influye en el rating de las mismas se muestra a continuación como aquellos generos menos populares son los que mayor puntuación reciben y el drama, que es el género más popular, se situa tras de estos, disponible en [Tableau Public](#) y el script utilizado para dividir el género en [R Markdown](#):



Es interesante también ver como han evolucionado los géneros más populares a lo largo del tiempo, para ver como la acción, el thriller o la comedia son cada vez más populares y otros como el drama y el romance, ya no tanto, disponible en [Tableau Public](#):



### 3. Preparación y limpieza del dataset

En primer lugar se ha realizado un análisis de la calidad del dataset para saber la cantidad de datos no disponible para cada uno de los atributos. Como se puede ver en el anexo [A](#), el presupuesto de las películas y las ganancias en mayor medida, arrojan una mayor falta de datos que el resto de características del dataset. Como las ganancias es un dato conocido únicamente después de haberse realizado la película, no será tenido en cuenta en la creación de los modelos detallados en las siguientes secciones.

#### 3.1. Modelo no supervisado

A continuación se seleccionan los datos para la construcción del modelo no supervisado solamente con las siguientes columnas, y se eliminan las filas con datos no definidos con [R Markdown](#):

- Actor 1 Facebook Likes, Actor 2 Facebook Likes y Actor 3 Facebook Likes
- Director Facebook Likes
- Duration
- Budget

#### 3.2. Modelo supervisado

Para la construcción de los modelos de regresión se utilizarán las siguientes columnas y eliminándose como en el caso anterior las filas con datos no definidos con [R Markdown](#). Cabe destacar que se ha creado una columna por género de la película con el valor 0 o 1 si la película pertenece a ese género, una columna con el valor 0 o 1 si la película es el blanco y negro o en color, y una columna para el director, el actor 1, el actor 2 y el actor 3, con el número de películas en las que participan:

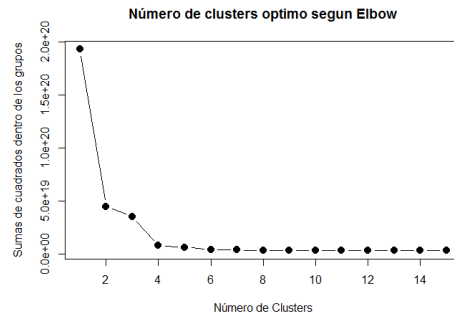
- Actor 1 Facebook Likes, Actor 2 Facebook Likes y Actor 3 Facebook Likes
- Director Facebook Likes
- Duration
- Budget
- Title Year
- Face Number in Poster
- Number Critic for Reviews
- Color
- Genre
- Director Movies
- Actor 1 Movies, Actor 2 Movies y Actor 3 Movies

### 3.3. Recomendador

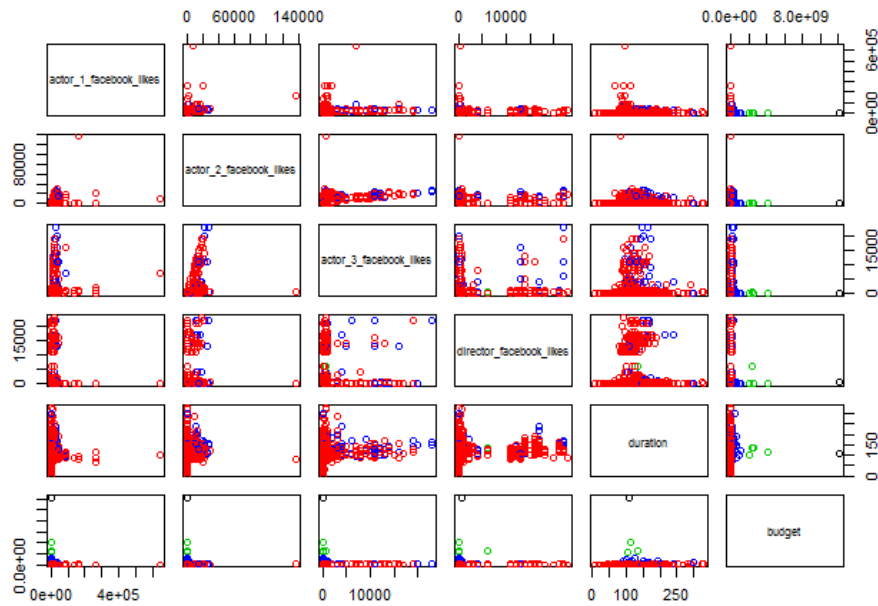
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

## 4. Análisis exploratorio apoyado en algún método NO supervisado

Para la creación del modelo no supervisado se utilizará el conjunto de datos generado anteriormente y se calculará el número óptimo de clusters según [R Markdown](#):



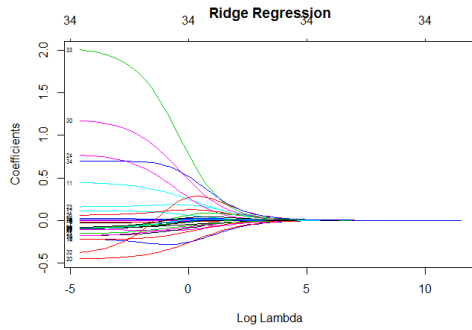
En la siguiente imagen se puede ver la relación entre los distintos clusters generados según el número óptimo calculado previamente, que en este caso es 4:



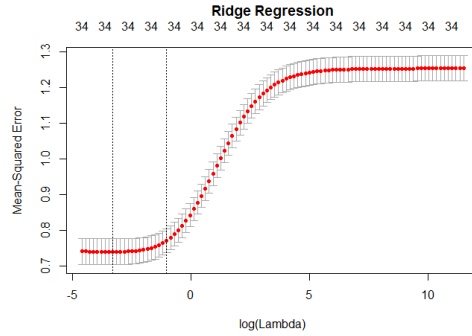


Según la matriz de correlación anterior, existe una gran influencia entre los atributos Cast Total Facebook Likes y Actor 1 Facebook Likes, por lo que se ha eliminado esta primera para la construcción de los siguientes modelos de regresión.

El primero de ellos será un modelo de regresión Ridge[5] el cual será optimizado para obtener los mejores parámetros que minimicen el error del modelo obtenido con [R Markdown](#):

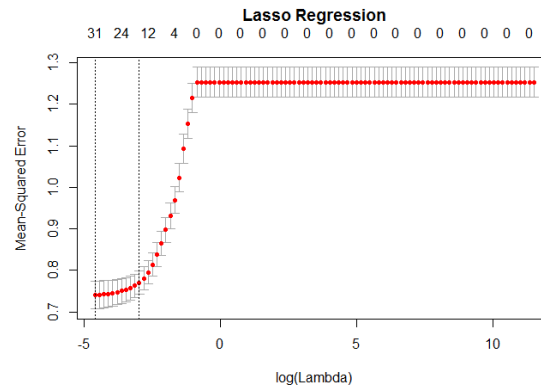


(g) Número de likes de los actores



(h) Número de likes de los directores

El segundo de ellos será un modelo de regresión Lasso[5] el cual también será optimizado para obtener los mejores parámetros que minimicen el error del modelo obtenido con [R Markdown](#):



## 6. Creación de un sistema de recomendación

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

## 7. Conclusiones y próximos pasos

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



## Referencias

- [1] C. Sun, *IMDB 5000 Movie Dataset*, [5000+ movie data scraped from IMDB website](#), Kaggle, 2016.
- [2] Z. Brown, [How many films are produced each year?](#), Quora, 2016.
- [3] M. Lanzagorta, [Horror Cinema By the Numbers](#), PopMatters, 2007.
- [4] Z. Brown, [Watching all the movies ever made](#), Justgeek, 2014.
- [5] T. Hastie and J. Qian, [Glmnet Vignette](#), Stanford University, 2014.

## A. Calidad del dataset

	Actor 1	Actor 2	Actor 3	Director	Genre	CR	Language	Title	Keywords	Length
1916	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%
1920	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%	0.00%	0.00%	0.00%
1925	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%
1927	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1929	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1930	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1932	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1933	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1934	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1935	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1936	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1937	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1938	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1939	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1940	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1941	0.00%	0.00%	20.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1942	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1943	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1944	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1945	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1946	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1947	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1948	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1949	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1950	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1951	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1952	0.00%	0.00%	0.00%	0.00%	0.00%	25.00%	0.00%	0.00%	0.00%	0.00%
1953	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1954	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1955	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1956	0.00%	0.00%	0.00%	0.00%	0.00%	33.33%	0.00%	0.00%	0.00%	0.00%
1957	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1958	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1959	0.00%	0.00%	0.00%	0.00%	0.00%	33.33%	0.00%	0.00%	0.00%	0.00%
1960	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1961	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1962	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1963	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1964	0.00%	0.00%	0.00%	0.00%	0.00%	10.00%	0.00%	0.00%	0.00%	0.00%
1965	0.00%	0.00%	0.00%	0.00%	0.00%	12.50%	0.00%	0.00%	0.00%	0.00%
1966	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1967	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1968	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1969	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1970	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1971	0.00%	0.00%	9.09%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1972	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1973	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1974	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1975	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1976	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	10.00%	0.00%	0.00%	0.00%
1977	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1978	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1979	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1980	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1981	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1982	0.00%	0.00%	0.00%	0.00%	0.00%	3.33%	0.00%	0.00%	3.33%	0.00%
1983	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1984	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1985	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1986	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1987	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1988	0.00%	0.00%	0.00%	0.00%	0.00%	3.23%	0.00%	0.00%	0.00%	0.00%
1989	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1990	0.00%	0.00%	0.00%	0.00%	0.00%	3.33%	0.00%	0.00%	0.00%	0.00%
1991	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1992	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1993	0.00%	0.00%	0.00%	0.00%	0.00%	2.08%	0.00%	0.00%	0.00%	0.00%
1994	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1995	0.00%	0.00%	0.00%	0.00%	0.00%	2.86%	0.00%	0.00%	0.00%	0.00%
1996	0.00%	0.00%	0.00%	0.00%	0.00%	1.01%	0.00%	0.00%	0.00%	0.00%
1997	0.00%	0.00%	0.00%	0.00%	0.00%	3.39%	0.00%	0.00%	0.00%	0.00%
1998	0.00%	0.00%	0.00%	0.00%	0.00%	2.24%	0.00%	0.00%	0.00%	0.00%
1999	0.00%	0.00%	0.00%	0.00%	0.00%	1.19%	0.00%	0.00%	0.00%	0.00%
2000	0.00%	0.00%	0.00%	0.00%	0.00%	2.34%	0.00%	0.00%	0.00%	0.58%
2001	0.53%	0.53%	1.06%	0.00%	0.00%	1.06%	0.00%	0.00%	0.00%	0.00%
2002	0.48%	0.48%	0.96%	0.00%	0.00%	2.87%	0.00%	0.00%	0.48%	0.00%
2003	0.00%	0.00%	0.00%	0.00%	0.00%	3.55%	0.00%	0.00%	0.00%	0.00%
2004	0.00%	0.00%	0.00%	0.00%	0.00%	3.27%	0.00%	0.00%	1.40%	0.00%
2005	0.45%	0.45%	0.90%	0.00%	0.00%	4.52%	0.00%	0.00%	1.81%	0.45%
2006	0.00%	0.00%	0.42%	0.00%	0.00%	5.02%	0.42%	0.00%	0.42%	0.42%
2007	0.00%	0.00%	0.00%	0.00%	0.00%	6.86%	0.49%	0.00%	1.47%	0.00%
2008	0.00%	0.00%	0.44%	0.00%	0.00%	3.11%	0.00%	0.00%	0.89%	0.00%
2009	0.00%	0.00%	0.00%	0.00%	0.00%	7.31%	0.00%	0.00%	1.54%	0.38%
2010	0.00%	0.00%	0.00%	0.00%	0.00%	3.48%	0.00%	0.00%	2.17%	0.87%
2011	0.89%	0.89%	0.89%	0.00%	0.00%	2.67%	0.00%	0.00%	2.67%	0.44%
2012	0.00%	0.00%	0.00%	0.00%	0.00%	5.43%	0.00%	0.00%	5.88%	0.45%
2013	0.42%	1.27%	1.27%	0.00%	0.00%	11.39%	0.00%	0.00%	7.59%	0.42%
2014	0.00%	0.00%	0.00%	0.00%	0.00%	15.87%	0.79%	0.00%	10.71%	0.40%
2015	0.44%	0.88%	1.33%	0.00%	0.00%	19.91%	0.00%	0.00%	14.60%	0.88%
2016	0.00%	0.00%	0.00%	0.00%	0.00%	13.21%	0.94%	0.00%	17.82%	0.00%
Not Defined	0.00%	2.78%	4.63%	86.30%	0.00%	38.89%	2.78%	0.00%	12.04%	2.78%

	Actor 1 Likes	Actor 2 Likes	Actor 3 Likes	Director Likes	Cast Likes	Critic Reviews	Movie Likes	User Reviews	Votes	Faces	Rating	Film Budget	Film Gross
1916	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
1920	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	10.00%
1925	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
1927	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1929	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	50.00%	0.00%
1930	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
1932	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
1933	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	50.00%
1934	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
1935	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	10.00%
1936	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	50.00%
1937	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	50.00%	50.00%
1938	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
1939	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	33.33%
1940	0.00%	0.00%	20.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	20.00%	60.00%
1941	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
1942	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	50.00%	50.00%
1943	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
1944	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
1945	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	25.00%	100.00%
1946	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	33.33%
1947	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	66.67%
1948	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	33.33%	66.67%
1949	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
1950	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1951	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
1952	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	25.00%	75.00%
1953	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	50.00%
1954	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	60.00%
1955	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	50.00%	100.00%
1956	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
1957	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	50.00%
1958	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
1959	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	66.67%
1960	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	66.67%
1961	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	80.00%
1962	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	75.00%
1963	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	62.50%
1964	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	40.00%
1965	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	12.50%	37.50%
1966	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	83.33%
1967	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	25.00%	75.00%
1968	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	9.09%	81.82%
1969	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	10.00%	70.00%
1970	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	66.67%
1971	0.00%	0.00%	9.09%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	9.09%	63.64%
1972	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	77.78%
1973	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	44.44%
1974	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	22.22%
1975	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	50.00%
1976	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	10.00%	80.00%
1977	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	56.25%
1978	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	6.25%	37.50%
1979	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	12.50%	56.25%
1980	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	8.33%	37.50%
1981	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.03%	48.48%
1982	0.00%	0.00%	0.00%	0.00%	0.00%	3.33%	0.00%	0.00%	0.00%	0.00%	0.00%	3.33%	46.67%
1983	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	9.09%	36.36%
1984	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	25.81%
1985	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	48.28%
1986	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.85%
1987	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.13%	3.13%
1988	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.23%
1989	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1990	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.33%	6.67%
1991	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.23%
1992	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.94%	0.00%
1993	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.08%	4.17%
1994	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.85%	1.85%
1995	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.29%	1.43%
1996	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.01%	4.04%
1997	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	8.47%	1.69%
1998	0.00%	0.00%	0.00%	0.00%	0.00%	1.49%	0.00%	0.00%	0.00%	0.00%	0.00%	8.21%	3.73%
1999	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.38%	1.19%
2000	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.51%	2.34%
2001	0.53%	0.53%	1.06%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.19%	0.00%
2002	0.48%	0.48%	0.96%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.35%	4.31%
2003	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	7.69%	5.92%
2004	0.00%	0.00%	0.00%	0.00%	0.00%	0.93%	0.00%	0.00%	0.00%	0.00%	0.00%	7.94%	6.54%
2005	0.45%	0.45%	0.90%	0.00%	0.00%	0.90%	0.00%	0.00%	0.00%	0.00%	0.00%	8.14%	8.14%
2006	0.00%	0.00%	0.42%	0.00%	0.00%	0.42%	0.00%	0.00%	0.00%	0.00%	0.00%	10.46%	9.62%
2007	0.00%	0.00%	0.00%	0.00%	0.00%	1.47%	0.00%	0.98%	0.00%	0.00%	0.00%	9.80%	12.25%
2008	0.00%	0.00%	0.44%	0.00%	0.00%	0.44%	0.00%	0.44%	0.00%	0.00%	0.00%	10.67%	10.67%
2009	0.00%	0.00%	0.00%	0.00%	0.00%	1.54%	0.00%	0.38%	0.00%	0.00%	0.00%	16.92%	17.31%
2010	0.00%	0.00%	0.00%	0.00%	0.00%	0.87%	0.00%	0.43%	0.00%	0.43%	0.00%	14.35%	13.48%
2011	0.89%	0.89%	0.89%	0.00%	0.00%	0.89%	0.00%	0.00%	0.00%	0.00%	0.00%	8.89%	15.11%
2012	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.90%	0.00%	0.00%	10.41%	18.10%
2013	0.42%	1.27%	1.27%	0.00%	1.69%	0.00%	0.84%	0.00%	0.42%	0.00%	0.00%	9.70%	21.10%
2014	0.00%	0.00%	0.00%	0.00%	0.00%	2.38%	0.00%	0.79%	0.00%	1.19%	0.00%	9.13%	33.73%
2015	0.44%	0.88%	1.33%	0.00%	0.00%	3.98%	0.00%	2.21%	0.00%	1.33%	0.00%	8.85%	36.73%
2016	0.00%	0.00%	0.00%	0.00%	0.00%	1.89%	0.00%	0.94%	0.00%	2.83%	0.00%	12.26%	30.19%
Not Defined	0.00%	2.78%	4.63%	96.30%	0.00%	8.33%	0.00%	5.56%	0.00%	0.00%	0.00%	92.59%	97.22%