

Análisis de Reseñas de Series y Películas

Por Juan Solinas

Motivación y audiencia

En el siguiente notebook se hará una revisión de una base de datos con reseñas correspondientes a diversas series y películas. El objetivo es diagramar un análisis que pueda describir ciertos patrones en cuanto al género de series y películas que tienden a ser calificadas de manera positiva.

Además, se buscarán patrones en los elementos que más importan a la hora de evaluar los contenidos audiovisuales en las series de los últimos tiempos.

Finalmente, se desarrollará un modelo para predecir calificaciones en función de lo indicado en las reseñas, con miras a automatizar el proceso de evaluación de las críticas a las series y películas.

Este notebook podría ser conveniente para asesorar a plataformas de streaming, como forma de asesorarlos con respecto a lo más conveniente a la hora de invertir en nuevas series y películas, así como también podría ser útil para investigaciones de mercado.

El dataset es una versión reducida manualmente de la siguiente fuente:

<https://www.kaggle.com/andrsmosquera/crticas-pelculas-filmaffinity-en-espaol-netflix?resource=download>

Resumen de los datos

El DataFrame reviews, que contiene las reseñas de las series y películas, se compone de la siguiente estructura

Columna	Valor	Tipo de dato
film_name	Nombre de serie o película	String
gender	Género de la serie o película	String
film_avg_rate	Promedio de puntuación de la serie o película	Float
review_rate	Puntuación de la reseña	Int
review_title	Título de la reseña	String
review_text	Cuerpo textual de la reseña	String

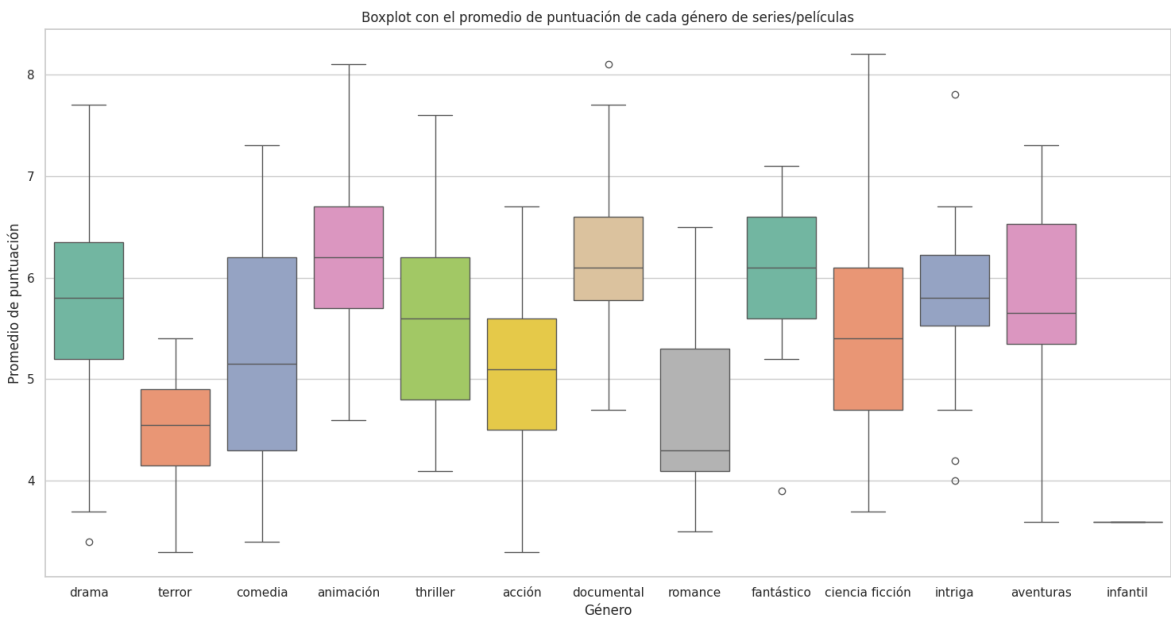
Hipótesis

La hipótesis de este notebook es que las películas y series dramáticas tienden a ser mejor valoradas que otros géneros.

Adicionalmente, se intentará comprobar si hay ciertos actores que por sí mismos puedan justificar valoraciones positivas.

Principales insights del análisis de datos

Promedio de puntuación de los diferentes géneros de series y películas:



Las puntuaciones del género terror son excesivamente bajas, con un 75% de los casos por debajo de un promedio de 5 puntos. Aunque esto puede ser debido a la poca cantidad de reseñas que tiene este género.

Algo similar se podría decir del género Drama y del género comedia, que tienen un 50% de los casos por debajo de los 6 puntos, aunque estos géneros si tienen un importante caudal de casos.

Cabe resaltar además el puntaje **de las series y películas del género de Animación y del género Fantástico**, que tienen un 50% de los casos por encima de los 6 puntos.

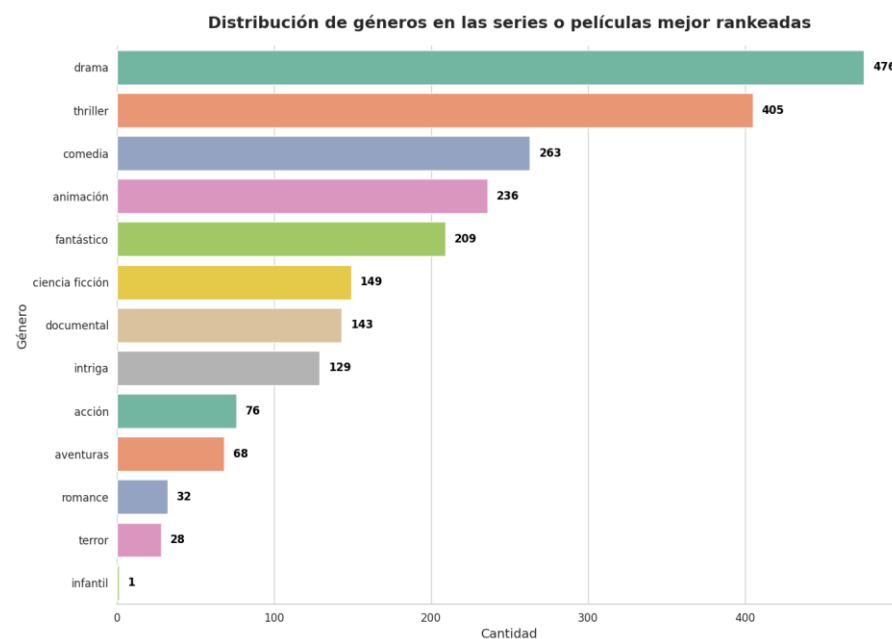
Películas y series con mejor promedio de puntuación

	film_name	film_avg_rate	Género 1
429	Black Mirror (Serie de TV)	8.2	ciencia ficción
274	Samurái de ojos azules (Serie de TV)	8.1	animación
431	Nuestro planeta II (Miniserie de TV)	8.1	documental
129	Ripley (Miniserie de TV)	7.8	intriga
435	Este mundo no me hará mala persona (Miniserie de TV)	7.8	animación
266	The Crown (Serie de TV)	7.7	drama
239	La II Guerra Mundial: Desde el frente (Miniserie de TV)	7.7	documental
213	La sociedad de la nieve	7.6	drama
68	As bestas	7.6	thriller
311	JoJo's Bizarre Adventure: Golden Wind (Serie de TV)	7.6	animación
241	Hilda (Serie de TV)	7.5	animación
344	Top Boy (Serie de TV)	7.5	drama
291	Pluto (Miniserie de TV)	7.4	animación
293	La vida en nuestro planeta (Serie de TV)	7.4	documental
392	La inspiración más profunda	7.3	documental

Películas y series con peor promedio de puntuación

	film_name	film_avg_rate	Género 1
372	Atrapaalmas	3.3	acción
271	V de venganza	3.3	terror
292	El lado dulce de la traición	3.4	drama
37	La familia Benetón	3.4	comedia
225	Mi cuñado es un vampiro	3.4	comedia
298	Disco Inferno (C)	3.5	terror
413	A través del mar	3.5	romance
398	Don Cochecito y los caballeros templarios	3.6	aventuras
325	Spy Kids: El armagedón	3.6	infantil
97	La madre de la novia	3.6	romance
264	Mejor Navidad, ¡imposible!	3.6	comedia
178	Mátame si te atreves, cariño	3.6	comedia
276	Vacaciones de verano	3.7	comedia
165	Mea culpa	3.7	drama
350	El club de los lectores criminales	3.7	terror

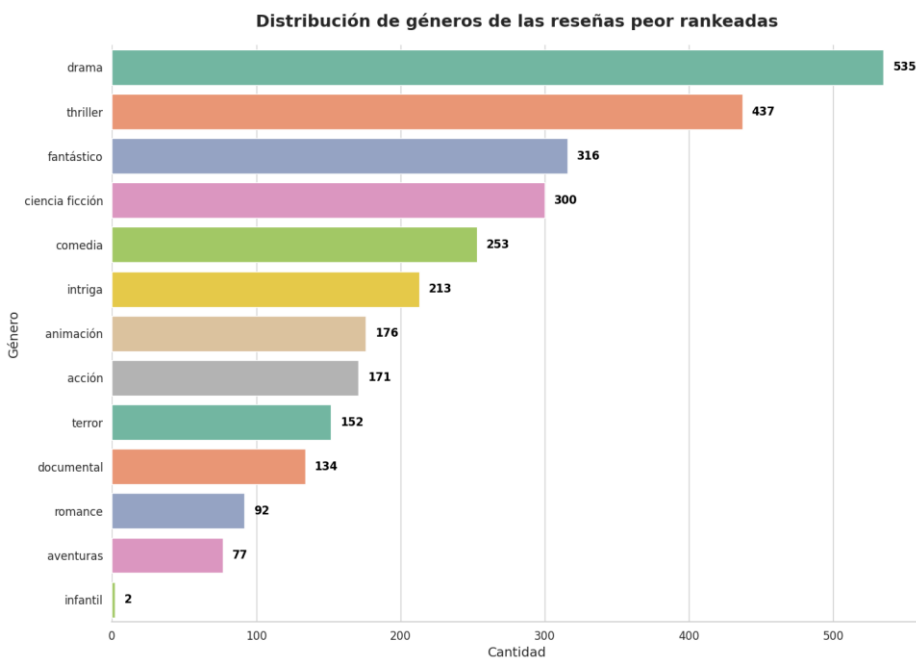
Distribución de series y películas con calificación positiva



Observando la cantidad de reseñas, se puede decir **que casi la mitad de las series o películas del género "Drama" obtuvieron reseñas con un puntaje mayor a 7**. Números similares obtuvo la categoría de Thriller, de Comedia y del género fantástico.

Cabe resaltar que **el género de Animación obtuvo más de la mitad de las reseñas con un puntaje mayor a 7 puntos**, así como también sucedió lo mismo con el género documental.

Distribución de series y películas con calificación negativa



En este ranking de reseñas con puntajes bajos, podemos observar la predominancia cuantitativa del Drama, del Thriller y del género fantástico. Pero teniendo en cuenta los números en general, prácticamente las reseñas están repartidas en mitades.

En cambio, géneros como la Ciencia Ficción (con 300 reseñas negativas sobre un total de 449), así como la categoría intriga (con 213 sobre un total de 342 reseñas) y el género Terror (con 152 reseñas negativas sobre un total de 180 películas) parecieran ser las categorías peor rankeadas.

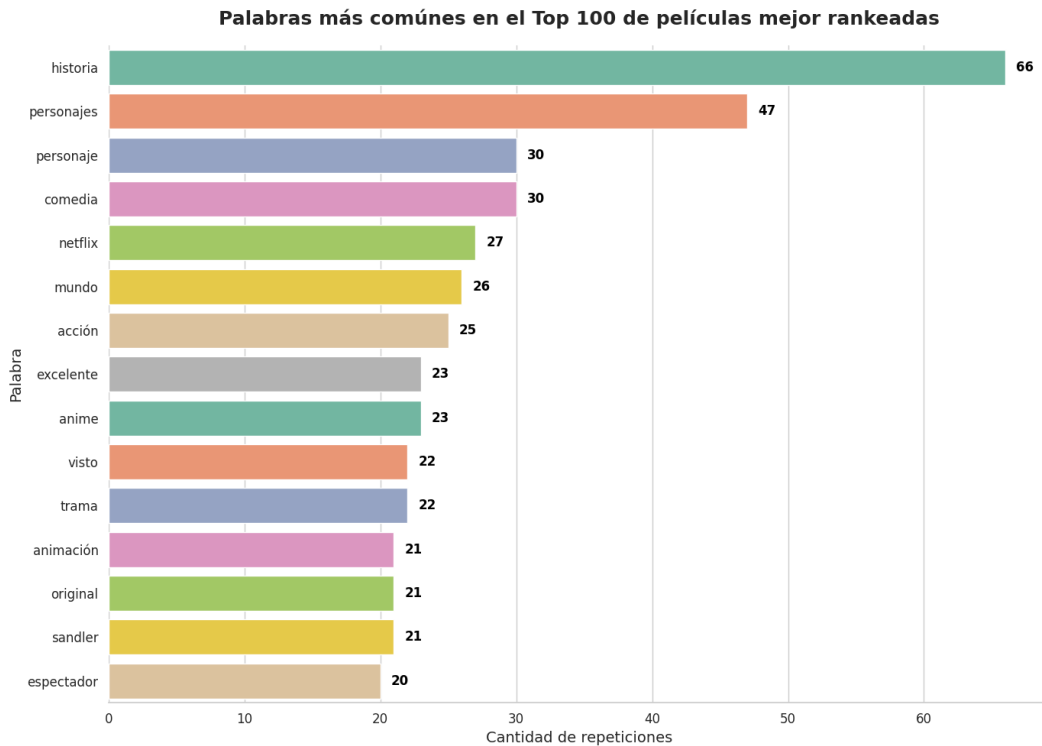
Wordcloud de las 100 reseñas con mejor puntuación



En este wordcloud de bigramas de las series mejor rankeadas, **podemos observar la relevancia que tienen los efectos especiales, la banda sonora, la acción y la comedia, así como el drama.**

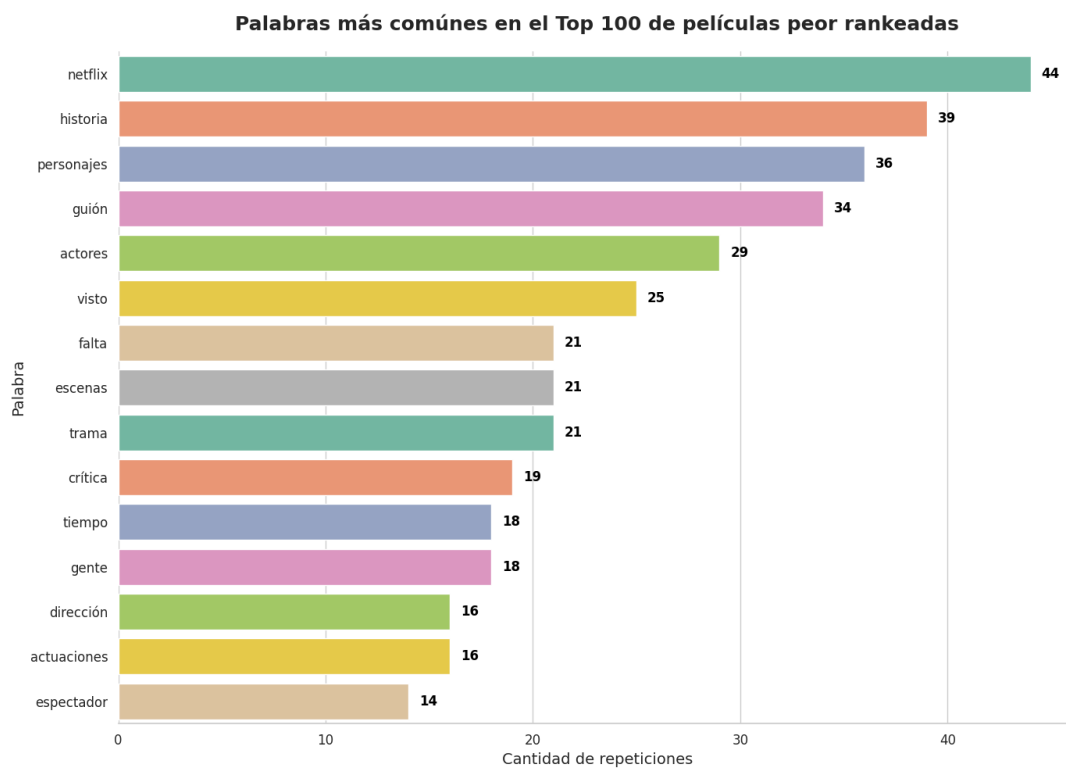


En este Word Cloud, quizás lo más relevante sea la aparición de la idea de pérdida de tiempo, como señal inequívoca de repudio en las reseñas. No pareciera demasiado relevante la aparición de una serie como Black Mirror, que tiene un buen promedio de calificación.



En el conjunto de reseñas del top 100, **parecieran ser muy relevantes la historia, los personajes, la comedia y el anime**. Además, resalta la palabra Sandler, probablemente por el apellido del actor Adam Sandler.

Palabras que más aparecen en el top 100 de reseñas negativas



La aparición de Netflix como la palabra más mencionada pareciera ser claramente un indicio de una crítica a la plataforma de series y películas en las reseñas.

Conclusiones

Como conclusión de este Notebook, se puede decir que se han detectado ciertos patrones a la hora del análisis de los datos.

En primer lugar, **se ha observado cierta conveniencia para apostar por series y películas de comedia o del género fantástico, así como la conveniencia de invertir para tener una importante cantidad de series o películas animadas (entre las cuales se puede incluir el anime japonés),** que tienden a tener una mayor aprobación que otro tipo de series y películas.

Si bien cabe rechazar la hipótesis desde la cual se inició este notebook, pues no se ha comprobado que el género dramático tenga mejores calificaciones que otros géneros, si es posible decir que este género tiene cerca de un 50% de sus reseñas aprobadas por la crítica, por lo cual apostar en este género no sería inadecuado.**

Además, luego del análisis desarrollado, **se puede recomendar no invertir en películas o series de terror,** ya que resulta hartamente dificultoso que tengan una reseña positiva, posiblemente por la dificultad de efectuar adecuadamente este tipo de contenidos.

Además, a lo largo del análisis **se ha podido observar el peso de ciertos directores, de los cuales sería recomendable tener su contenido para congraciarse con el público de habla española** que realiza estas reseñas, con el fin de fidelizarlos como público.

Con respecto a los límites de este notebook, **es evidente que se necesitarán más reseñas para mejorar la eficacia de los dos modelos de regresión logística**, los cuales **presentaron un importante overfitting** así como también tuvieron problemas para poder detectar las reseñas positivas adecuadamente.

Además, **habría que seguir eliminando palabras inadecuadas de la tokenización**, con el fin de quedarse solamente con palabras que denoten una actitud positiva o negativa con respecto a las películas y de esa manera poder mejorar la precisión de un modelo de Machine Learning