

MEMORIA

Objetivo del análisis

Un negocio editorial propone un análisis acerca de los títulos que han obtenido algún reconocimiento del público, recibiendo una valoración positiva con los votos emitidos en las diferentes plataformas.

Partiendo de la premisa de que ese reconocimiento obtenido conducirá a una buena cifra de ventas del título, el trabajo consiste en extraer las características externas generales que conducen a lograrlo, más allá de la calidad del contenido, que se da por supuesto.

Para ello se van a analizar las correspondencias que puedan existir entre la variable principal que expone ese reconocimiento ("Reconocimiento") y el resto de variables que puedan intervenir en él.

Dataset de trabajo

El punto de partida es un dataset que contiene una recomendación de títulos. De todo el campo, el análisis se va a centrar en los títulos publicados entre los años 2000 y 2021.

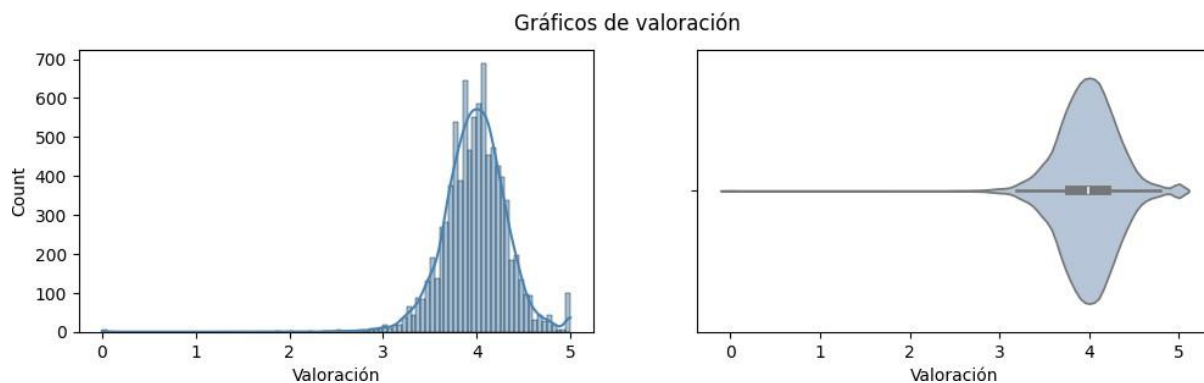
Aquí se muestra la variación, prácticamente inapreciable, entre los títulos totales y los publicados en los años establecidos.

Nombre	Tipo	Descripción
Título	<i>string</i>	Contiene el título del libro
Seriado	<i>boolean</i>	Booleano que indica si el título pertenece a una saga
Serie	<i>string</i>	Variable del nombre de la saga a la que pertenece el título
Género	<i>string</i>	Género literario al que pertenece la obra
Año de publicación	<i>int</i>	Año de primera publicación del título
Páginas	<i>int</i>	Número de páginas que contiene el libro en su primera edición
BestSeller	<i>boolean</i>	Booleano que indica si se considera un título
Premiado	<i>boolean</i>	Booleano que indica si el libro ha recibido algún premio literario
Premios	<i>string</i>	Premios literarios que ha recibido el título
Mejor Valorado	<i>boolean</i>	Booleano que indica si el título se encuentra entre los más valorados por los lectores
Valoración	<i>float</i>	Valoración media de los lectores

Variable directora, “Mejor Valorados”

La variable directora del estudio será la columna "Mejor Valorados", que incluye los títulos con una nota media superior a 4.0 y un mínimo de 1000 votos.

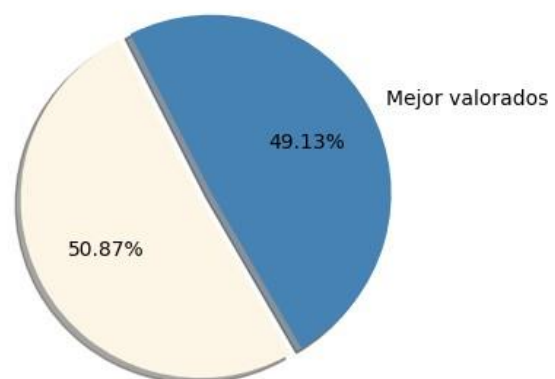
Prácticamente la mitad de los títulos se encuentra en ese rango, y veremos qué características tienen.



Pero antes, para contrastar la discriminación numérica que vamos a hacer, se revisa la información de la columna "Valor" y se comprueba que el grueso de los títulos guarda una distribución simétrica en torno a la valoración de 4, con algunos títulos destacados en el 5.

Porcentaje de libros mejor valorados

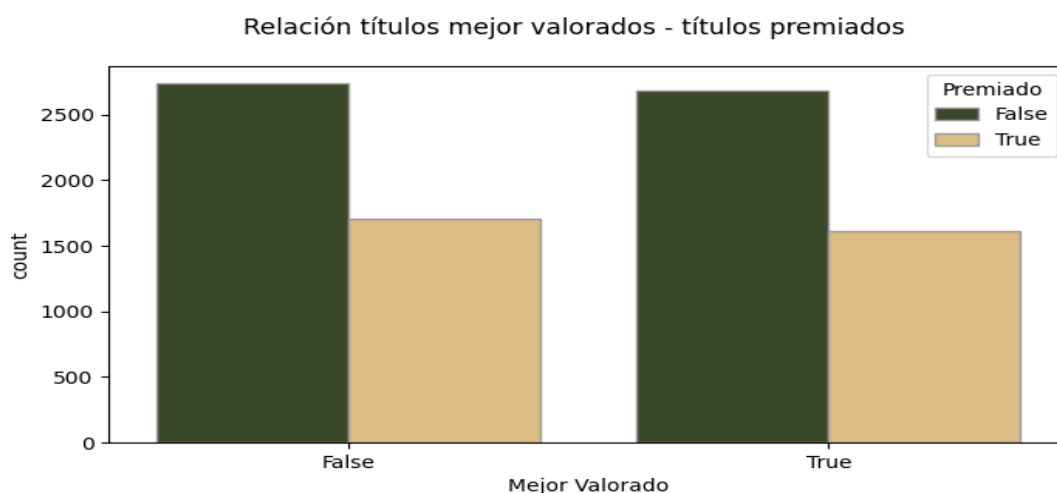
count	8734.000000
mean	3.986061
std	0.344889
min	0.000000
25%	3.790000
50%	3.990000
75%	4.190000
max	5.000000



Relación con variable “Premiados”

¿Tiene este valor alguna relación con la variable de premiados?

Si existiera relación, podría concluirse que los títulos galardonados en algún certamen literario tienen mejor recepción en el público en general, y entonces una de las respuestas a la consulta sería encontrar títulos premiados que aún no tengan editorial con la que publicar el libro (improbable, porque el objetivo de los títulos es precisamente encontrar esos libros de calidad que lanzar al mercado, conforme a los criterios que se establezcan para cada editorial y su certamen), o recomendar la inversión en el lanzamiento de un premio propio de la editorial.



Aparentemente guardan la misma relación entre los datos, así que se puede esperar que las variables son independientes entre sí.

Se visualiza con un gráfico de valores relativos y se aplica la **prueba de chi cuadrado** para concluir que, efectivamente, ambas variables son independientes.

Conclusión: Que un título esté premiado no conduce a una mejor valoración del público, por lo que se descarta proponer la búsqueda de títulos premiados no editados o la creación de un premio editorial perteneciente al sello.

Valor Chi-Cuadrado: 0.8306762485835706

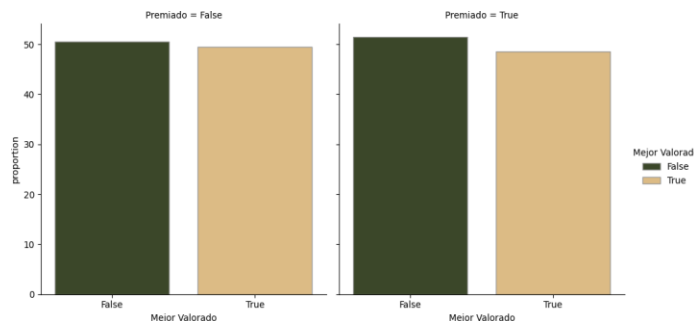
P-Value: 0.3620770575573812

Grados de Libertad: 1

Tabla de Frecuencias Esperadas:

[[2757.162812 1685.837188]

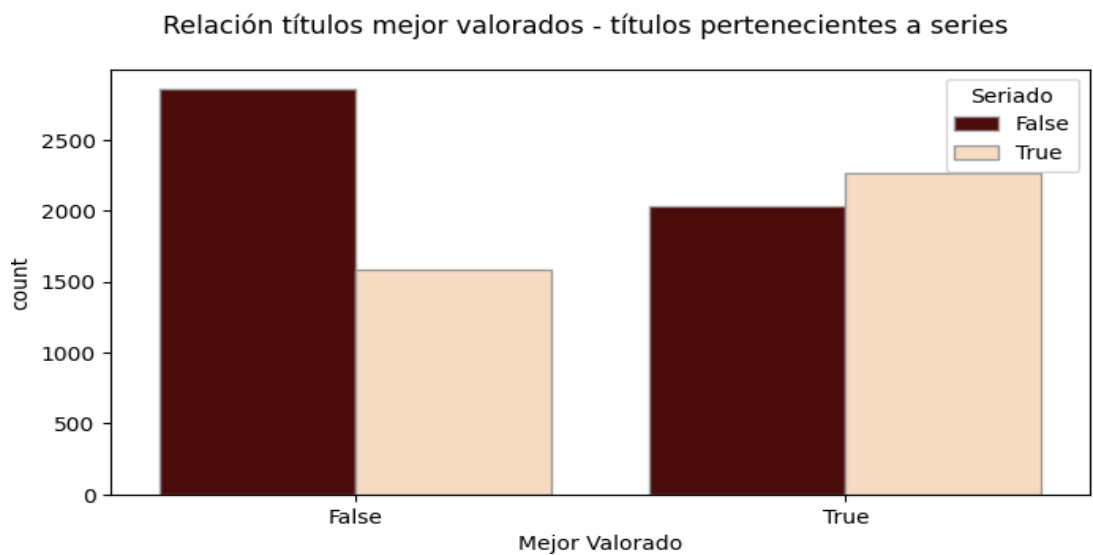
[2662.837188 1628.162812]]



Relación con variable "Seriado"

Otra variable a tener en cuenta en el estudio es la posibilidad de que se mejore la consideración de los títulos pertenecientes a una serie cuyo éxito inicial alce el resultado en el reconocimiento de todos los títulos de la serie, debido a la creación de una base de lectores fieles que disculpen los altibajos de las diferentes entregas.

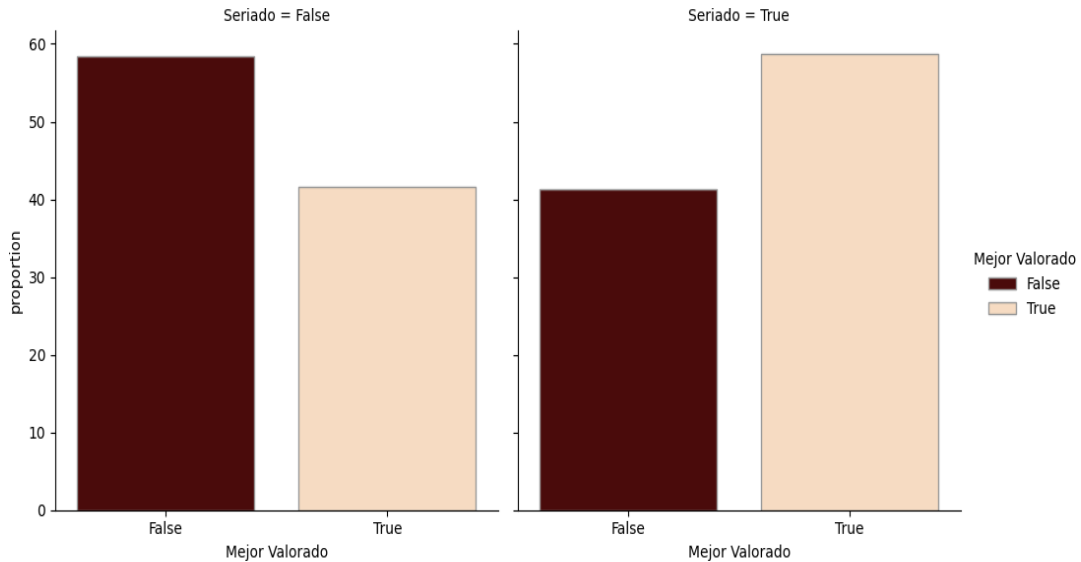
Si existiera relación entre ambas, la recomendación sería considerar como objeto de publicación los libros pertenecientes a una serie de cualquier extensión, para asegurar las ventas con las entregas posteriores de la saga.



En esta relación si hay unos valores que se salen de lo corriente; los títulos que se encuentran entre los mejor valorados tienen un valor alto de ellos que pertenecen a una serie, lo que refuerza el argumento propuesto.

Como en el caso anterior, se muestra un gráfico con los valores relativos y se analizan los datos con una prueba de chi cuadrado.

```
Valor Chi-Cuadrado: 254.43377112205144
P-Value: 2.8045993522103574e-57
Grados de Libertad: 1
Tabla de Frecuencias Esperadas:
[[2484.49874055 1958.50125945]
 [2399.50125945 1891.49874055]]
```



Conclusión: El '*valor p*' en la '*prueba de chi cuadrado*' concluye que las variables sí guardan relación entre sí.

Se recomienda a la editorial buscar títulos que formen parte (o vayan a formar parte) de una serie que amplíe el relato, para garantizar un mayor número de ventas en caso de éxito.

A partir de este resultado surgen otras preguntas:

¿Hay algún dato que nos indique que un título será un éxito? o

¿Se puede extraer más información sobre estos libros para ajustar los parámetros de búsqueda?

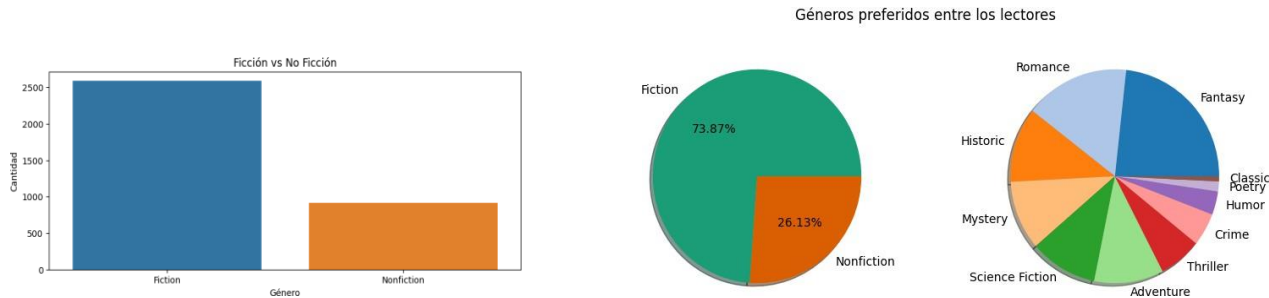
La primera pregunta recae de nuevo en la calidad del contenido, que es un parámetro fuera de este análisis. Sin embargo, sí resultaría interesante indagar en lo propuesto en la segunda pregunta.

Estudio de variable "Género"

¿Hay alguna preferencia entre los géneros de ficción o no ficción?

¿Qué género es el que más títulos tiene entre los mejor valorados?

El gráfico demuestra que se prefieren los títulos de ficción sobre los títulos de no ficción, y los libros con temática histórica, romance o fantasía, suponen más de la mitad de los títulos mejor valorados.



Conclusión: Los títulos de ficción pertenecientes a los géneros fantasía, romance o histórico son los más valorados por los lectores.

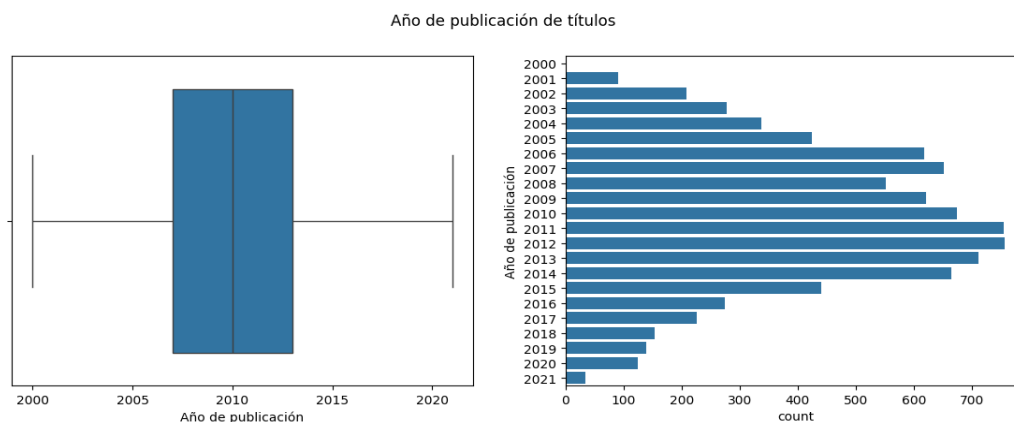
Se recomienda incidir en la búsqueda y publicación de esos títulos.

Géneros presentes en los títulos mejor valorados



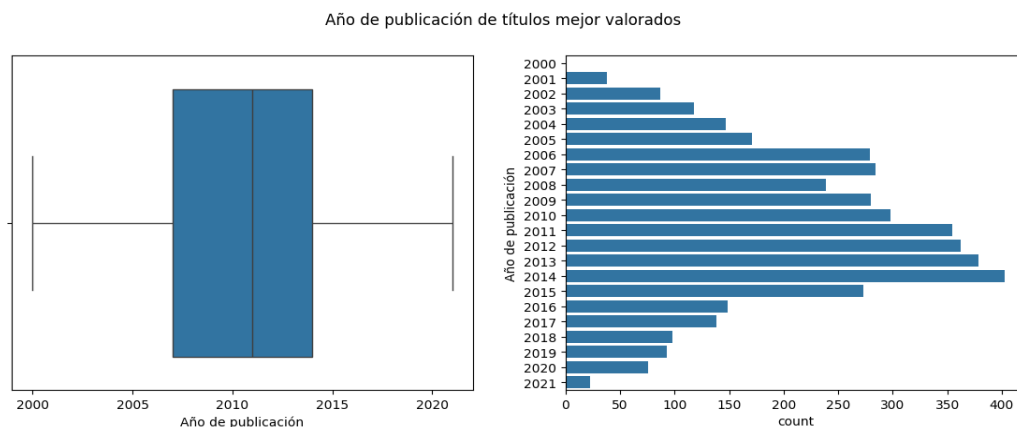
Estudio de variable "Año de publicación"

¿Se puede determinar que pasan unos años desde la publicación del libro hasta que puede considerarse como un éxito?



Revisando los datos, puede decirse que es positivo que pasen ****entre 7 y 14 años**** para que un título se cuente entre los mejor valorados.

Aunque existen un incremento de valores para los años 2005 y 2006, se puede concluir que los títulos más antiguos quedan relevados por los nuevos, mientras que los más nuevos no cuentan aún con suficiente recorrido editorial como para entrar dentro de la lista de mejores obras.



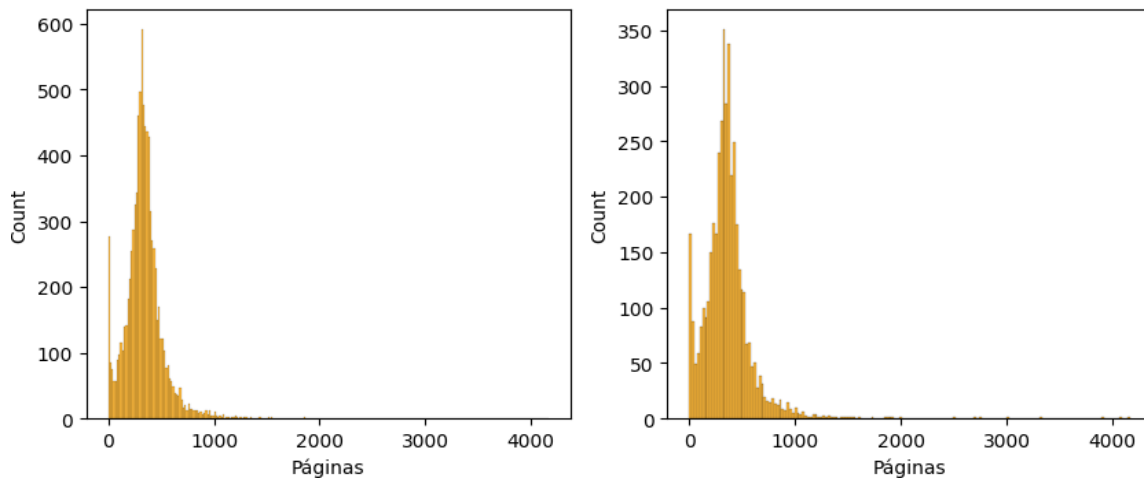
Conclusión: Se recomienda a la editorial que o bien publique títulos ya asentados en el mercado y que no tengan un acuerdo previo con otra compañía, o bien importe títulos que ya tengan algún bagaje desde su publicación y que aún no se encuentren en el ámbito geográfico de la editorial.

En cualquier caso, debe tener en cuenta que pasarán años entre la primera fecha de publicación y el reconocimiento que se busca con su edición.

Variable "Páginas"

¿Hay alguna preferencia entre los lectores acerca del número de páginas que tiene un título bien valorado?

Conclusión: El gráfico del número de páginas de todos los títulos del estudio no varía respecto al de los títulos mejor valorados, de lo que se infiere que no existe ninguna relación entre el volumen del libro y su reconocimiento.



Conclusiones finales

Con el análisis de los datos de la tabla se puede confirmar que el esfuerzo de la editorial debería dirigirse hacia los títulos de ficción, con temática de fantasía, romance o historia preferiblemente.

Los títulos que pertenecen a alguna saga de libros (títulos seriados) cuentan con mejor acogida entre el público, que genera una base de seguidores fieles a los lanzamientos posteriores, y más receptivos para la acogida y valoración en los títulos subsiguientes.

Habría que lograr representar a los títulos que se encuentren en proceso de cambio editorial o sin representación en el ámbito geográfico de la compañía que solicita el informe. En caso de que no fuera posible y se lanzara un título de nueva publicación, hay que tener en cuenta que el tiempo necesario para alcanzar el reconocimiento mayoritario del público se encuentra en torno a los 8-12 años desde la fecha de lanzamiento.