

ANÁLISIS GÉNEROS LITERARIOS

Mar Pérez Pulido



ÍNDICE

- 1. Introducción**
 - a. Contexto**
 - b. Hipótesis**
- 2. Exploración de datos**
- 3. Análisis Global de los datos**
 - a. Análisis univariante**
 - b. Análisis bivariante y multivariante.**
- 4. Conclusiones**
- 5. Referencias**

1. Introducción

En este proyecto vamos a analizar como afectan los géneros literarios para que los libros sean más leídos o mejor puntuados.

Para ello hemos tomado como referencia una muestra de datos de Kaggle que hace referencia a la red social literaria “Goodreads”.

Fuente:

Ishika Johari(2022). Best Books (10k) Multi-Genre Data.

<https://www.kaggle.com/datasets/ishikajohari/best-books-10k-multi-genre-data?resource=download>

a. Contexto

Se parte del dataset goodreads_data.csv que contiene una muestra de 10.000 entradas y 8 columnas. Aunque aparentemente parece limpio, más adelante se tuvieron que tratar los datos para poder realizar los análisis.

El estudio realizado comienza por un análisis general para encontrar los géneros más y menos consumidos del momento. De este modo podremos profundizar en cómo se ve afectada dicha distribución.

b. Objetivos e hipótesis

Antes de comenzar con el análisis debemos tener muy claros los objetivos y cuáles son nuestras hipótesis a la hora de realizar el análisis.

¿Cuál es el género literario más leído?

Según editoriales como Penguin, dentro del fenómeno Booktok existen géneros como Fantasía o Romance los cuáles son más promovidos.

¿Existe alguna relación entre los géneros más leídos y los más publicados?

Lo lógico es pensar que sí, ya que cuanto más leen un género más libros se publicarán.

¿Es el género literario más leído el mejor puntuado?

En este caso también se tiende a pensar que sí. Esto se debe a que si un género no está bien valorado no tendría sentido seguir produciéndolo.

¿Existe alguna relación entre los autores más publicados y los más leídos?

Tal y como sucede con las preguntas anteriores tendemos a pensar que sí por el mismo motivo. Al fin y al cabo, se tratan de empresas a las que les interesa obtener el mayor beneficio.

¿Qué autores destacan en los géneros más leídos?

Si seguimos con el pensamiento de que Fantasía y Romance son los géneros más leídos podríamos decir que Sarah J. Mass destacaría entre las más leídas ya que la promueven como autora número 1 del New York Times.

En cuanto al género de Romance, autoras como Megan Maxwell o Alice Kellen podrían ser las más populares, ya que incluso de la primera de ellas se ha estrenado una película recientemente.

¿Qué autores destacan en el género menos popular?

Si pensamos que dentro de los géneros más leídos el que menos triunfa es el género de poesía, podríamos establecer a Defreds como uno de los autores que más está triunfando gracias a las redes sociales.

2. Exploración de datos

Usando los métodos correspondientes hemos analizado de forma general el dataset.

De esta primera lectura observamos que la columna "Description", que es la que presenta nulos, podemos borrarla porque no aporta ninguna información necesaria a futuro. Del mismo modo podemos eliminar la columna URL.

En cuanto a la columna Num_Ratings debería ser de tipo “int” y no “object”, por lo que realizamos el cambio para posteriormente realizar un correcto análisis.

Además, al comprobar si había duplicados por título y autor vimos que también era necesario corregir estos datos y quedarnos con los registros más actuales.

Para terminar con la limpieza de datos, obtuvimos una columna nueva que reflejase el género principal de cada libro y tratamos de agrupar todo lo posible los géneros.

Por último, clasificamos las variables dándoles especial importancia a las variables numéricas y a la mencionada con anterioridad.

3. Análisis Global de los datos

a. Análisis univariante

Lo primero que necesitábamos saber para continuar con nuestro análisis era saber cuáles eran los géneros literarios más consumidos dentro de una muestra que abarcaba los 10 géneros más populares.

De este modo comprobamos nuestra hipótesis que situaba al género Fantasía entre los primeros obteniendo 167314412 lecturas totales seguido del género de Ficción con 163634569.

b. Análisis bivariante y multivariante

Dentro del análisis bivariante nos centramos en la segunda de nuestras preguntas.

Para responderla fue necesario hallar la correlación entre los géneros publicados y sus lecturas.

Obtuvimos una correlación situada en 0,9, lo que significa que existe una **fuerte correlación positiva**.

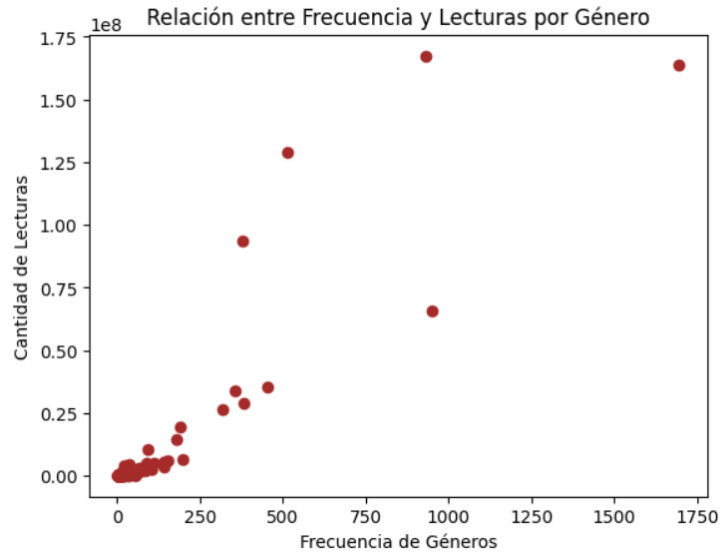


Fig. 1. Gráfico que indica la correlación existente entre la cantidad de lecturas y la frecuencia de géneros literarios.

Para responder la tercera de nuestras preguntas, el análisis se centró en analizar la relación existente dentro de los 10 géneros más leídos, sus puntuaciones y su frecuencia de lecturas dándonos las siguientes conclusiones:

- La mayoría de los libros se puntúan entre 3,5 y 4,5.
- Existen outliers con puntuaciones por debajo de 2,5 y cerca de 5.
- La mayoría de los libros se concentran en la parte izquierda y hacia abajo del gráfico lo que indica que la mayoría de los ellos tiene una puntuación relativamente baja.
- Existe un pequeño grupo de libros que tienen un alto número de calificaciones, por encima de 6 millones.
- A medida que aumenta el número de calificaciones, las puntuaciones comienzan a agruparse en torno a 3,5 y 4,5.
- Los libros con muy pocas calificaciones se ven más afectados por la influencia de opiniones individuales mostrando así más variabilidad.

Por otro lado, en cuanto a los géneros observamos que:

- Géneros como “Fantasy” y “Classics” presentan varios libros con un número muy alto de calificaciones.

- "Young Adult" y "Fantasy" parecen tener libros muy populares con buenas calificaciones promedio.
- Los géneros "Fiction" o "Nonfiction" parecen más dispersos, con más variedad en sus calificaciones promedio.

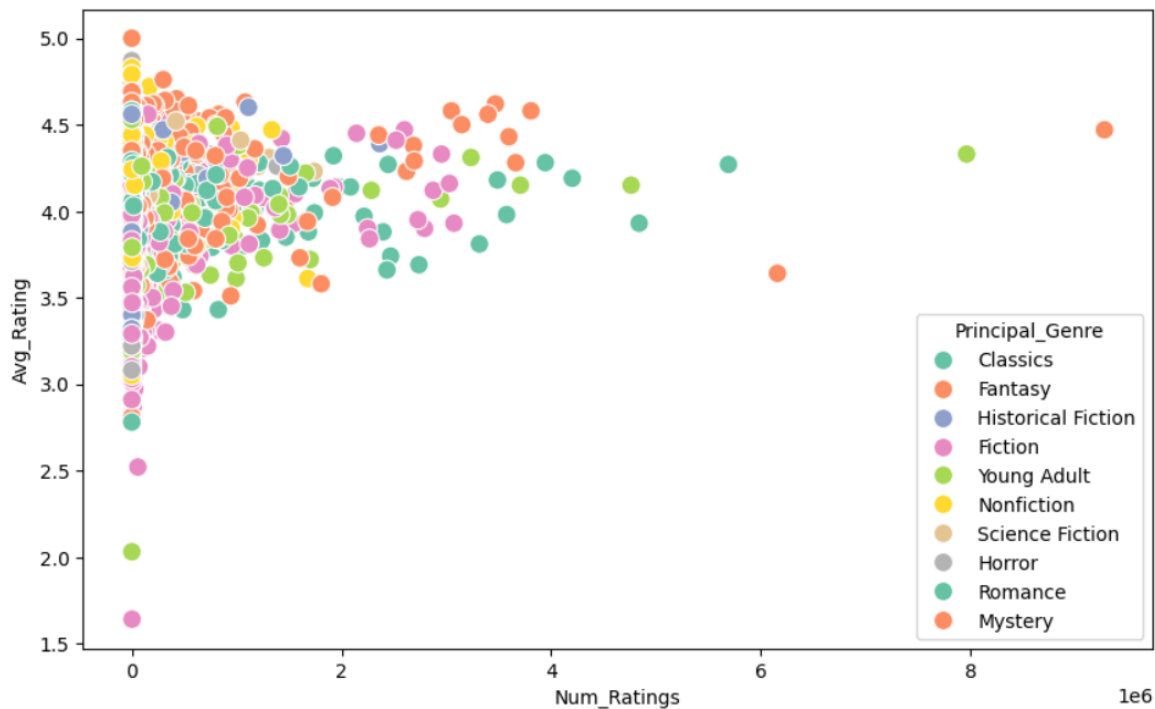


Fig. 2. Gráfico representativo de la correlación entre la cantidad de lecturas y las puntuaciones medias de los 10 géneros más leídos.

También obtuvimos las notas medias de los 10 género más leídos, pero, al rondar todos los 4 puntos se optó por relacionar las puntuaciones medias con el número de lecturas totales.

De este análisis podemos destacar que el género 'Fantasy' es el mejor valorado y el más leído, mientras que 'Fiction' a pesar de ser de los más leídos es de los peores puntuados.

También observamos que géneros como 'Horror' son menos leídos y también peor puntuados y otros como 'Young Adult' o 'Science Fiction' a pesar de tener gran popularidad no consiguen llegar a los 4 puntos de valoración.

Además, encontramos dos claros grupos, los que son leídos por más de 100 millones de personas y los que no, En el primer grupo destaca el género 'Horror' como el peor

puntuado y el menos leído. Por otro lado, 'Historical Fiction' y Nonfiction' son los más leídos y los mejores puntuados.

En el segundo grupo destaca 'Fantasy' como el género más leído y mejor puntuado. En este grupo encontramos que, excepto el género anterior, se puntúan mucho peor.

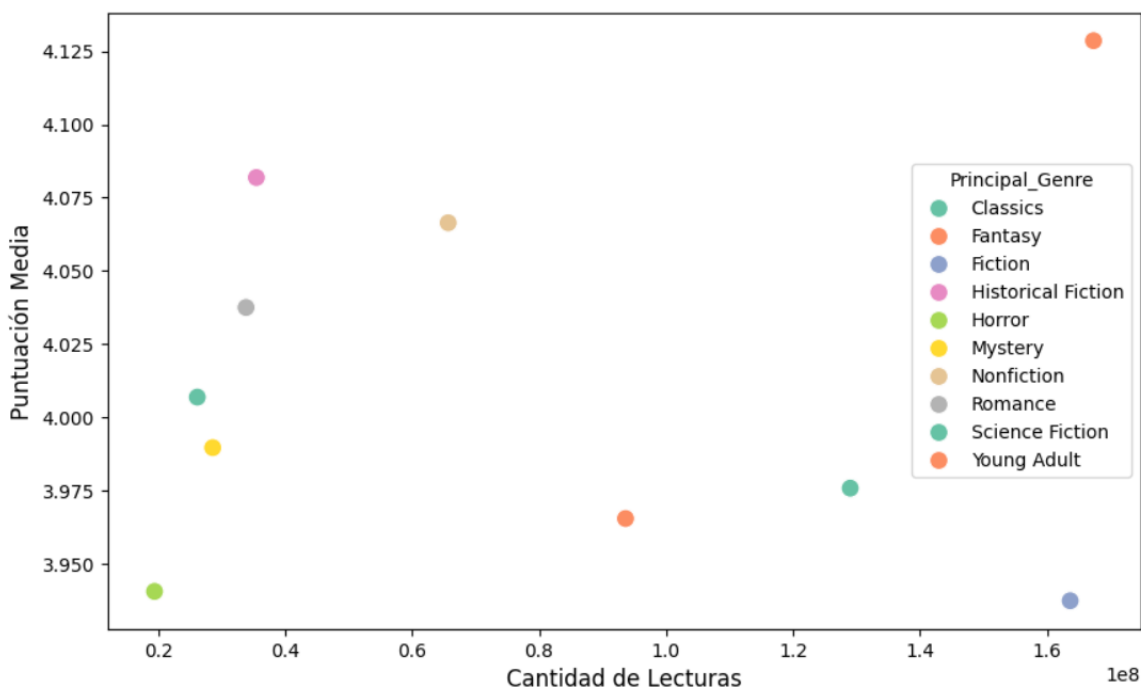


Fig 3. Gráfico que muestra la distribución de los 10 géneros más leídos en función de sus puntuaciones medias y sus lecturas totales.

Antes de analizar los géneros más destacables tuvimos que realizar un análisis más que hace referencia a la correlación entre el número de publicaciones y el número de lecturas.

La correlación obtenida fue de 0,51. Esto significa que existe una correlación positiva moderada. Es decir, podríamos determinar que cuanto más publiquen a un autor, más libros se van a leer. Sin embargo, al ser moderada no quiere decir que esto sea determinante, sino que existirán otras causas que hagan que los libros sean más leídos. Dichas causas podrían ser los géneros, la publicidad que les den, etc.

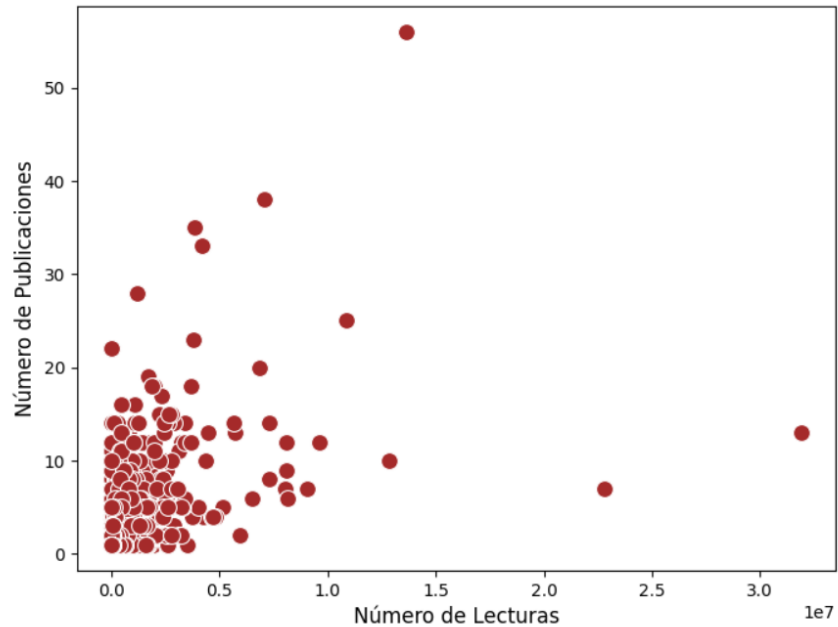


Fig 4. Correlación entre número de publicaciones y número de lecturas.

En la última parte del análisis nos centraremos en dos ramas. La primera de ella trata sobre los géneros más leídos y la segunda aborda un análisis dentro del género menos leídos.

Dentro de los géneros que más lectores presentan destacan 'Fantasy' y 'Fiction', pero observamos que cada uno de ellos tenía puntuaciones muy diferentes.

Es por ello, que procederemos a un análisis individual de cada tipo para ver sus autores y las puntuaciones pero lo primero que hemos hecho antes de profundizar es ver la distribución.

Análisis género 'Fantasy'

- La mayoría de los libros de este género tienen un número reducido de lectores, lo que indica que la mayoría de dichos libros no son muy populares en cuanto al número de calificaciones.
- La mayoría de las puntuaciones promedio oscilan entre 3,5 y 4,5, lo que implica que los libros de fantasía tienden a estar bien valorados.
- En cuanto a los outliers observamos que hay libros muy populares, con más de 6 millones de lectores los cuales presentan puntuaciones altas.

- Los libros con menos lectores tienen mayor dispersión oscilando sus puntuaciones entre 3 y 5, esto podría indicar que cuanto menos leído es un libro más influenciado puede estar por opiniones individuales extremas.
- En cuanto a su correlación existe una correlación positiva muy débil, lo que sugiere que haya otros factores que influyan en las puntuaciones promedio tales como opiniones individuales, campañas de marketing, calidad subjetiva, etc.

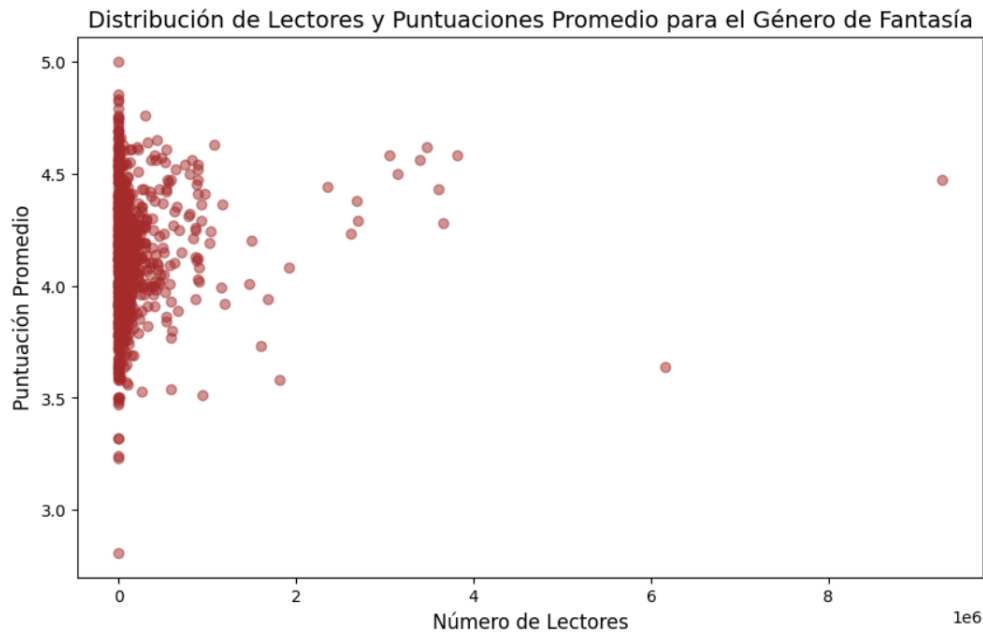


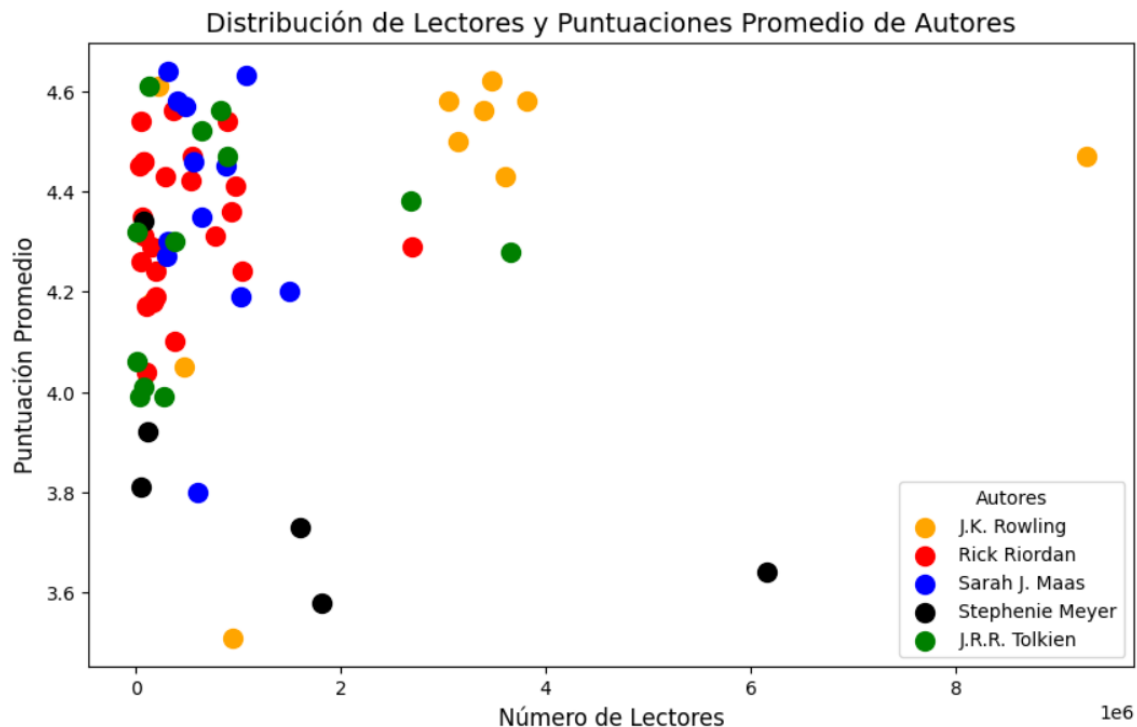
Fig. 5. Distribución de lectores y puntuaciones promedio de 'Fantasy'.

Profundizando más en este género obtenemos que 3 de los 5 libros de Fantasía más populares pertenecen a la autora J.K. Rowling, pero necesitamos saber si esto también hace que sea la más leída y la más puntuada.

Dentro de los autores más leídos han aparecido algunos como Rick Riordan y Sarah J. Mass. Sin embargo, J.K. Rowling muestra una gran ventaja respecto a los demás. Si además de ello tenemos en cuenta sus puntuaciones obtenemos que:

- J.K. Rowling se sitúa entre las mejor puntuadas, seguida por Sarah J. Mass.
- Rick Riordan es el que más libros ha publicado superando todos ellos la puntuación de 4.
- Stephanie Meyer tiene libros muy populares, pero no muy bien valorados.

- J.R.R. Tolkien ha escrito libros peor puntuados que los de Sarah J. Mass pero más populares.



- Se podría decir que no existe correlación alguna y que podría estar plenamente influenciado por juicios propios de valor, además de campañas de marketing, etc.

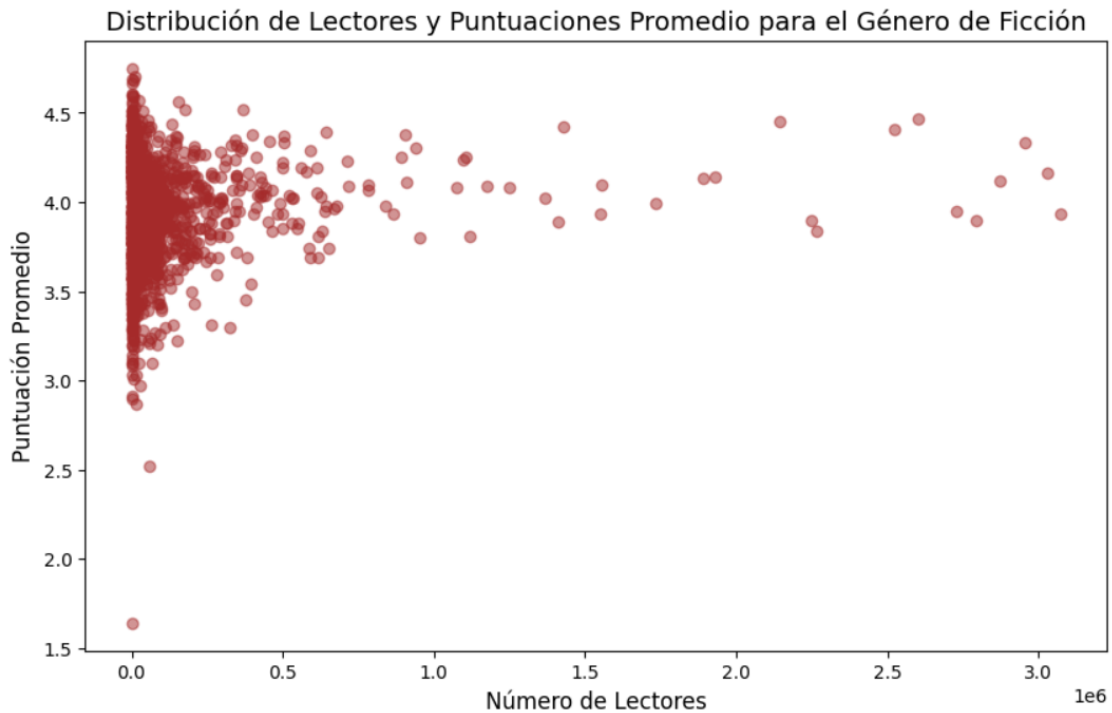


Fig. 7. Distribución de lectores y puntuaciones promedio de 'Fiction'

En este caso el género 'Fiction' presenta una mayor distribución de autores entre los libros más leídos siendo John Grisham el único autor entre los más leídos que no se encuentra entre el top 5 libros más leídos. Además de ello, se observa que:

- Las puntuaciones han disminuido de valores respecto a 'Fantasy', lo que indica que el género 'Fiction' presenta peores puntuaciones.
- John Grisham ha mantenido su público pero ha ido mejorando sus puntuaciones.
- Dan Brown ha ido creciendo en lectores pero no consigue llegar con ninguno de sus libros a 4 puntos.
- Stieg Larsson, Taylor Jenkins Reid y Khaled Hosseini tienen una distribución similar teniendo pocos libros publicados pero con gran público que les da unas buenas valoraciones.

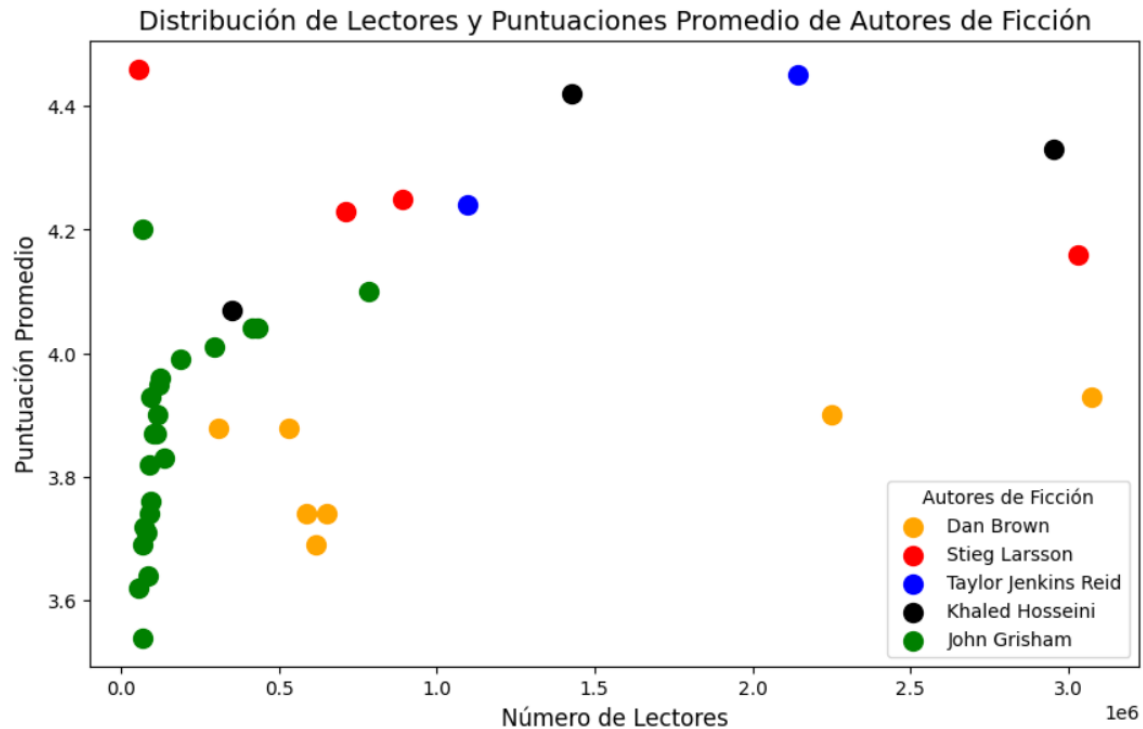


Fig. 8. Distribución de lectores y puntuaciones promedio de las publicaciones de los autores más leídos del género 'Fiction'.

Por último nos centramos en el género "Horror" como el género menos leídos dentro de los 10 géneros más populares.

En este caso obtuvimos que:

- Existe una alta concentración de libros con pocos lectores, teniendo la mayoría menos de 200 mil.
- La puntuación promedio oscila entre 3,5 y 4,25, aunque existen valores atípicos, generalmente se califican de manera favorable.
- Los libros que tienen más lectores presentan puntuaciones más altas.
- Los libros que tienen más de 600 mil lectores tienen de puntuación promedio 4 puntos, mientras que los de menos lectores sus puntuaciones son más dispersas.
- Apenas dos libros superan el millón de lectores y son puntuados con muy buena nota.
- La correlación obtenida implica que es muy débil o que no está muy clara la correlación entre ambas variables. Sin embargo, podemos concluir que existe un

mercado con una base muy sólida de obras populares pero con una amplia variedad de libros menos conocidos los cuales se ven más influenciados por las opiniones individuales.

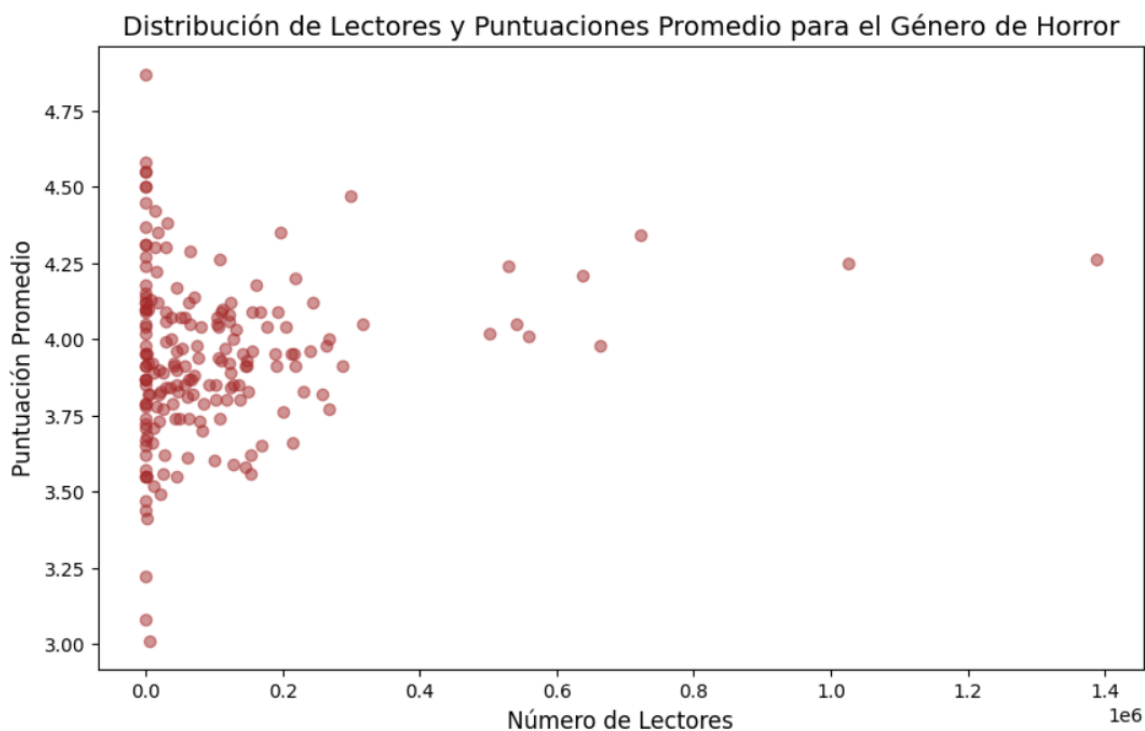


Fig. 9. Distribución de lectores y puntuaciones promedio de 'Horror'.

En cuanto al top 5 libros más leídos todos pertenecen a Stephen King, lo que explica la gran diferencia de lecturas existente entre él y el resto de autores del mismo género.

Además, también se observa que:

- Stephen King es el autor con más publicaciones, el cual posee libros con gran cantidad de lectores y con las mejores puntuaciones de todo el género.
- Anne Rice se mantiene entre los 50 mil y los 200 mil lectores, teniendo algún outlier de más de 500.000 lectores. Sus puntuaciones se mantienen entre los 3,8 y 4,3 puntos.
- De Thomas Harris observamos una clara correlación positiva en la que a medida que sus obras han ido recibiendo mejores puntuaciones, más lectores han tenido.

- La obra de Dean Koontz se concentra en no estar muy bien valorada y poseer una estable puntuación entre los 3,8 y 4 puntos, teniendo algún valor excepcional por encima de dichas puntuaciones.
- Max Brooks, es el autor que menos obras presenta concentrándose en puntuaciones entre 3,8 y 4 y no pasando de los 200 mil lectores, excepto con su Best Seller que llegó cerca de los 600 mil.

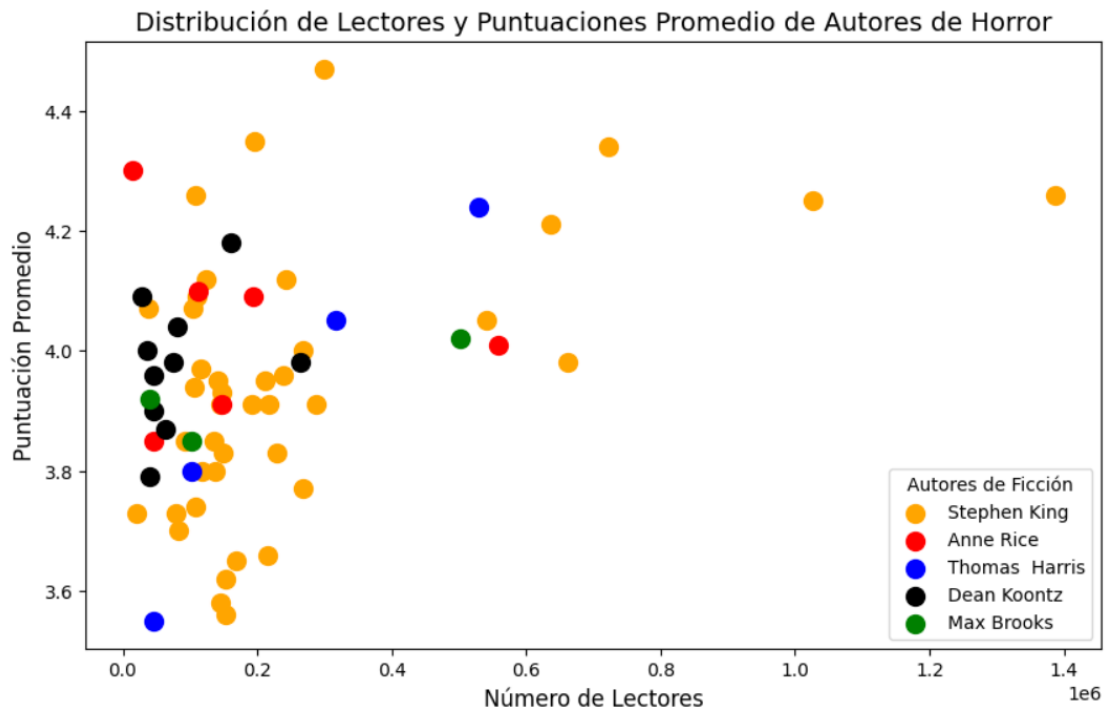


Fig. 8. Distribución de lectores y puntuaciones promedio de las publicaciones de los autores más leídos del género 'Horror'.

4. Conclusiones

Una vez realizados todos los análisis necesarios procedemos a recordar los objetivos y sus hipótesis y lo compararemos con los resultados objetivos.

- El primero de nuestros objetivos era saber cuál era el género literario más leído. Nuestra hipótesis se basaba en que sería Romance y Fantasía, pero tras el análisis hemos comprobado que los géneros más leídos son Fantasía y Ficción.
- El segundo de nuestros objetivos era saber si existe alguna relación entre los géneros más leídos y los más publicados. Según el análisis hemos coincidido con la

hipótesis que decía que sí existía, ya que la correlación se sitúa en 0,9, lo que implica que existe una fuerte correlación positiva entre ambas variables.

- Nuestro tercer objetivo se encargaba de saber si el género literario más leído es el mejor puntuado y aunque nuestra hipótesis se basaba en que sí, tras el análisis nos encontramos con dos situaciones diferentes. En el caso del género de Fantasía sí que coincide en ser el más leído y el mejor puntuado. Sin embargo, en el caso de Ficción no es así, ya que a pesar de ser el segundo género más leído es el peor valorado, dentro de los 10 géneros más leídos, con una puntuación media de 3,94.
- En cuanto a la cuarta pregunta que consistía en si existe alguna relación entre los autores más publicados y los más leídos, obtuvimos una correlación positiva moderada. Esto significa que, aunque sí que existe una relación tal y como pensábamos, hay otros factores que influyen en que un autor se lea en mayor medida.
- El quinto objetivo nos planteaba la pregunta sobre los autores más leídos en los géneros más populares. Según nuestra hipótesis pensábamos que sería Sarah J. Mass, la cual se encuentra en el top 5, pero se ve ampliamente superada en lecturas por la autora J.k. Rowling. A pesar de ello, tienen calificaciones similares. En cuanto al segundo género más leído es muy contrario a lo pensado, ya que establecimos a Megan Maxwell o Alice Kellen como las autoras de Romance más populares. Sin embargo, como hemos visto con anterioridad, el segundo género más popular es Ficción y su autor más popular es Dan Brown pero no coincide con el mejor valorado que es Taylor Jenkins Reid.
- Por último, tampoco coincidimos con la hipótesis de nuestro último objetivo que se centraba en saber el autor más popular del género menos consumido dentro del top 10 de los géneros más leídos. A pesar de que pensamos que el género sería poesía y que Defreds sería su autor más leído, es el género Horror el menos consumido y Stephen King como el autor más leído y casi el mejor valorado, pero le superan dos autores cuyas puntuaciones se ven beneficiadas al haber publicado escasos libros en comparación.

5. Referencias

Se adjunta adicionalmente una lista de links con los que hemos establecido algunas de las hipótesis.

https://www.penguinlibros.com/es/recomendaciones/descubre-algo-nuevo/tiktok-libros-mas-vendidos-fenomeno-booktok?recomendados=1&srsId=AfmBOopse5nn8xWvWZnFWBE9dUeUoDtPEN5EZ_QDT6DeEHny-GIviYRn

<https://www.editorialhidra.com/autor/sarah-j-maas/>

<https://www.planetadelibros.com/autor/sarah-j-maas/000045173>

<https://www.rtve.es/television/20220302/defreds-frases-libros-ojala-exito-redes-sociales/2300512.shtml>