

```
En [11]: # importar bibliotecas
import pandas as pd from sqlalchemy import create_engine db_config = { 'usuario' :
db_config [ 'contraseña' ],
db_config [ 'host' ],
db_config [ 'puerto' ],
db_config [ 'db' ])

motor = crear_motor ( cadena_de_conexión , argumentos_de_conexión = { 'sslmode'
```

Objetivo general

Analizar el comportamiento de lectura y participación de los usuarios en una plataforma de libros digitales, utilizando información sobre libros, autores, editoriales, calificaciones y reseñas, con el fin de identificar patrones clave que apoya la creación de una propuesta de valor para un nuevo producto orientado a lectores.

Objetivos específicos

Identificar el nivel de participación de los usuarios a través de calificaciones y reseñas de texto, para comprender cómo interactúan con los libros dentro de la plataforma.

Analizar la relación entre la cantidad de reseñas y la calificación promedio de los libros, con el objetivo de detectar títulos con mayor impacto y aceptación entre los usuarios.

Determinar qué editoriales publican la mayor cantidad de libros relevantes (excluyendo folletos o publicaciones breves), para identificar actores clave dentro del mercado editorial.

Identificar a los autores con mejor calificación promedio, considerando únicamente libros con un número significativo de calificaciones, para asegurar resultados estadísticamente representativos.

Evaluar el comportamiento de los usuarios más activos, analizando cuántas reseñas de texto realizadas aquellos que califican un alto número de libros, con el fin de comprender el nivel de compromiso del lector avanzado.

```
En [24]: consulta = """
SELECT *
FROM libros
```

```
LIMIT 5;

"""
pd . read_sql ( consulta , con = motor )
```

Salida[...]

	id del libro	id del autor	título	núm_páginas	fecha de publicación	ID del editor
0	1	546	'Salem's Lot	594	1 de noviembre de 2005	93
1	2	465	1.000 lugares que ver antes de morir	992	22 de mayo de 2003	336
2	3	407	13 pequeños sobres azules (Pequeños sobres azules...	322	21 de diciembre de 2010	135
3	4	82	1491: Nuevas revelaciones de las Américas antes...	541	10 de octubre de 2006	309
4	5	125		1776	386	4 de julio de 2006

En [26]:

```
consulta = """
SELECCIONAR *
DE autores
LÍMITE 5;
"""

pd . read_sql ( consulta , con = motor )
```

Salida[...]

	id del autor	autor
0	1	AS Byatt
1	2	Esopo/Laura Harris/Laura Gibbs
2	3	Agatha Christie
3	4	Alan Brennert
4	5	Alan Moore y David Lloyd

En [27]:

```
consulta = """
SELECCIONAR *
DE calificaciones
LÍMITE 5;
"""

pd . read_sql ( consulta , con = motor )
```

Salida[...]

	id de calificación	id del libro	nombre de usuario	clasificación
0	1	1	Ryan Franco	4
1	2	1	GrantPatricia	2
2	3	1	brandtandrea	5
3	4	2	lorichen	3
4	5	2	Mariokeller	2

En [28]:

```
consulta = """
SELECCIONAR *
DE revisiones
LÍMITE 5;
"""

pd.read_sql(consulta, con=engine)
```

Salida[...]

	id de revisión	id del libro	nombre de usuario	texto
0	1	1	brandtandrea	Mencione sociedad diga enviar profesor análisis. ...
1	2	1	Ryan Franco	El público se golpeó el pie de cristal. Amo...
2	3	2	lorichen	Escucha, trata de mantener la preocupación. Señorita, impuestos, pero...
3	4	3	Johnson Amanda	Por fin mes interesante azul podría naturaleza cu...
4	5	3	Scottamara	Nación propósito pesado dar espera canción voluntad. Lista...

En [29]:

```
consulta = """
SELECCIONAR *
DE publicadores
LÍMITE 5;
"""

pd.read_sql(consulta, con=engine)
```

Salida[...]

	ID del editor	editor
0	1	As
1	2	Libro Ace
2	3	Libros Ace
3	4	Ace Tapa dura
4	5	Compañía editorial Addison Wesley

En [12]:

```
consulta = """
SELECT *
FROM libros
"""
pd.read_sql(consulta, con=engine)
```

Salida[...]

	id del libro	id del autor	título	núm_páginas	fecha de publicación	ID del editor
0	1	546	'Salem's Lot	594	1 de noviembre de 2005	93
1	2	465	1.000 lugares que ver antes de morir	992	22 de mayo de 2003	336
2	3	407	13 pequeños sobres azules (Pequeños sobres azules...)	322	21 de diciembre de 2010	135
3	4	82	1491: Nuevas revelaciones de las Américas antes...	541	10 de octubre de 2006	309
4	5	125		1776	386	4 de julio de 2006
...
995	996	571	Hermanas Wyrd (Mundodisco #6; Brujas #2)	265	6 de febrero de 2001	147
996	997	454	Xenocidio (La saga de Ender #3)	592	15 de julio de 1996	297
997	998	201	Año de las Maravillas	358	30 de abril de 2002	212
998	999	94	Eres un desastre (Una historia de amor #2)	328	16 de enero de 2007	331
999	1000	509	Zen y el arte del mantenimiento de la motocicleta: Un...	540	25 de abril de 2006	143

1000 filas × 6 columnas

En [13]:

```
query = """
SELECT *
FROM books
WHERE publication_date > '2000-01-01'
"""
pd.read_sql(query, con=engine)
```

Out[13]:

	book_id	author_id		title	num_pages	publication_date	publisher_id
0	1	546	'Salem's Lot	594	2005-11-01	93	
1	2	465	1 000 Places to See Before You Die	992	2003-05-22	336	
2	3	407	13 Little Blue Envelopes (Little Blue Envelope...	322	2010-12-21	135	
3	4	82	1491: New Revelations of the Americas Before C...	541	2006-10-10	309	
4	5	125	1776	386	2006-07-04	268	
...
814	994	408	World War Z: An Oral History of the Zombie War	342	2006-09-12	68	
815	996	571	Wyrd Sisters (Discworld #6; Witches #2)	265	2001-02-06	147	
816	998	201	Year of Wonders	358	2002-04-30	212	
817	999	94	You Suck (A Love Story #2)	328	2007-01-16	331	
818	1000	509	Zen and the Art of Motorcycle Maintenance: An ...	540	2006-04-25	143	

819 rows × 6 columns

Me parece interesante que viendo el resultado de la cantidad de filas después de la consulta sigue siendo muy alto, resaltando que la mayoría de los libros que están en este data set fueron publicados después de Enero del 2000.

In [14]:

```
query = """
SELECT
    a.book_id,
    COUNT(DISTINCT b.review_id) AS "Número de reseñas",
    AVG (a.rating) AS "Calificación promedio"
FROM ratings a
LEFT JOIN
reviews b
ON a.book_id = b.book_id
GROUP BY a.book_id
ORDER BY a.book_id
"""
pd.read_sql(query, con=engine)
```

Out[14]:

	book_id	Número de reseñas	Calificación promedio
0	1	2	3.666667
1	2	1	2.500000
2	3	3	4.666667
3	4	2	4.500000
4	5	4	4.000000
...
995	996	3	3.666667
996	997	3	3.400000
997	998	4	3.200000
998	999	2	4.500000
999	1000	4	3.833333

1000 rows × 3 columns

Los resultados muestran que existe variabilidad tanto en el número de reseñas como en la calificación promedio entre los distintos libros, lo que indica diferentes niveles de participación de los usuarios y percepción de calidad. Algunos libros cuentan con pocas reseñas pero calificaciones altas, mientras que otros presentan mayor volumen de reseñas con promedios más moderados.

In [15]:

```
query = """
SELECT
a.publisher,
COUNT (*) AS total_libros
FROM publishers a
LEFT JOIN
books b
ON a.publisher_id = b.publisher_id
WHERE
b.num_pages > 50
GROUP BY
a.publisher
ORDER BY total_libros;

"""
pd.read_sql(query, con=engine)
```

Out[15]:

		publisher	total_libros
0	Alfred A. Knopf Borzoi Books	1	
1	Arthur A. Levine Books / Scholastic Inc.	1	
2	Dante University of America Press	1	
3	Red Fox	1	
4	Citadel	1	
...	
329	Ballantine Books	19	
330	Penguin Classics	24	
331	Grand Central Publishing	25	
332	Vintage	31	
333	Penguin Books	42	

334 rows × 2 columns

Se identificó la editorial que ha publicado el mayor número de libros con más de 50 páginas utilizando las tablas publishers y books. Para ello, se filtraron los libros con un número de páginas superior a 50 y se agruparon por editorial, contabilizando el total de libros publicados por cada una.

Los resultados muestran que varias editoriales presentan un número similar de libros que cumplen con este criterio. Sin embargo, al ordenar los resultados por el total de libros publicados, se puede identificar la editorial con mayor producción de obras de mayor extensión, lo que permite excluir folletos u otro tipo de publicaciones breves y enfocarse en editoriales con un catálogo más robusto.

In [16]:

```
query = """
SELECT
    au.author_id,
    au.author,
    AVG(r.rating) AS avg_rating
FROM authors au
JOIN books b ON au.author_id = b.author_id
JOIN ratings r ON b.book_id = r.book_id
WHERE b.book_id IN (
    SELECT
        book_id
    FROM ratings
    GROUP BY book_id
    HAVING COUNT(*) >= 50
)
GROUP BY au.author_id, au.author
```

```

ORDER BY avg_rating DESC
LIMIT 1;

"""

pd.read_sql(query, con=engine)

```

	author_id	author	avg_rating
0	236	J.K. Rowling/Mary GrandPré	4.287097

Los resultados muestran que J.K. Rowling / Mary GrandPré presenta la calificación promedio más alta entre los autores que cumplen con este criterio, con un promedio aproximado de 4.29. Esto indica que, además de contar con una alta participación de usuarios, sus libros mantienen una percepción de calidad consistentemente positiva entre los lectores.

```

In [23]: query = """
SELECT
    AVG("Review por usuario") AS avg_reviews
FROM (
    SELECT
        username,
        COUNT(review_id) AS "Review por usuario"
    FROM reviews
    WHERE username IN (
        SELECT
            username
        FROM ratings
        GROUP BY username
        HAVING COUNT(DISTINCT book_id) >= 50
    )
    GROUP BY username
) t;
"""

pd.read_sql(query, con=engine)

```

	avg_reviews
0	24.222222

El resultado muestra que los usuarios que calificaron más de 50 libros escriben, en promedio, aproximadamente 24 reseñas de texto. Esto sugiere que, aunque estos usuarios son altamente activos en términos de calificaciones, solo una parte de su interacción se traduce en reseñas escritas, lo que indica que la acción de calificar requiere menor esfuerzo que la elaboración de una reseña textual.