

Proyecto SQL

El coronavirus tomó al mundo entero por sorpresa, cambiando la rutina diaria de todos y todas. Los habitantes de las ciudades ya no pasaban su tiempo libre fuera, yendo a cafés y centros comerciales; sino que más gente se quedaba en casa, leyendo libros. Eso atrajo la atención de las startups (empresas emergentes) que se apresuraron a desarrollar nuevas aplicaciones para los amantes de los libros.

Te han dado una base de datos de uno de los servicios que compiten en este mercado. Contiene datos sobre libros, editoriales, autores y calificaciones de clientes y reseñas de libros. Esta información se utilizará para generar una propuesta de valor para un nuevo producto.

Descripción de los datos

books:

Contiene datos sobre libros:

- book_id — identificación del libro
- author_id — identificación del autor o autora
- title — título
- num_pages — número de páginas
- publication_date — fecha de la publicación
- publisher_id — identificación de la editorial

authors:

Contiene datos sobre autores:

- author_id — identificación del autor o autora
- author — el autor o la autora

publishers:

Contiene datos sobre editoriales:

- publisher_id — identificación de la editorial
- publisher — la editorial

ratings:

Contiene datos sobre las calificaciones de usuarios:

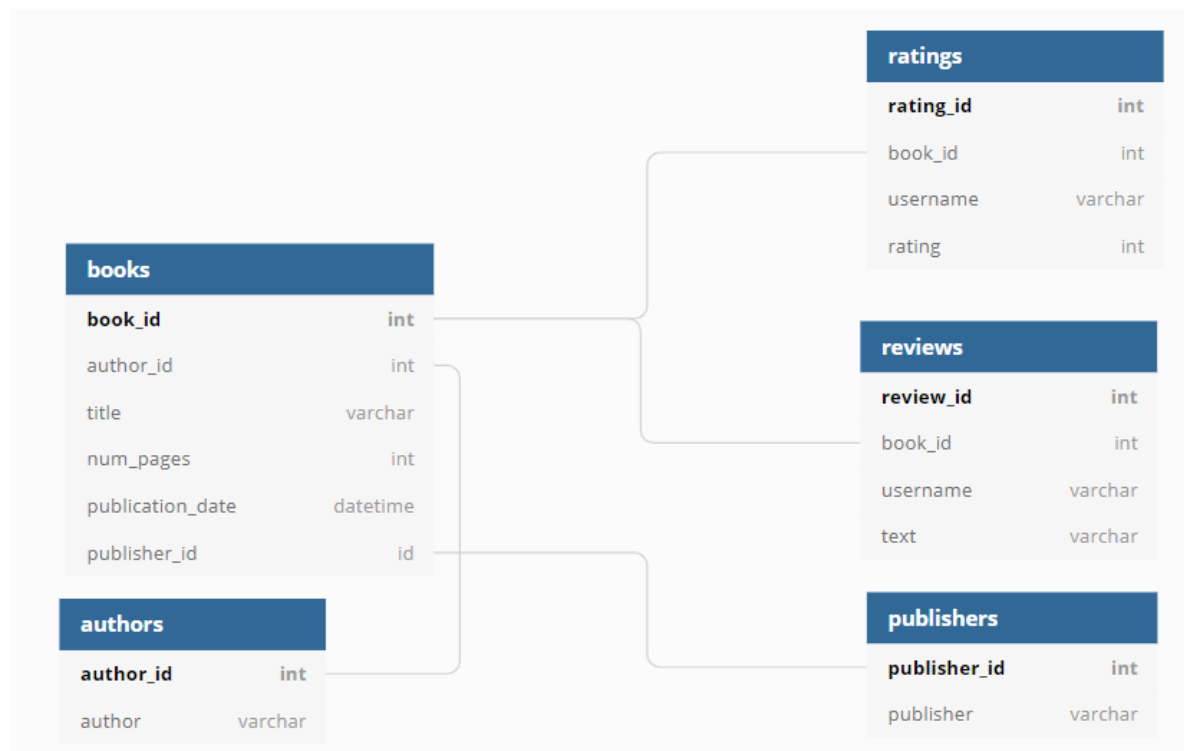
- rating_id — identificación de la calificación
- book_id — identificación del libro
- username — el nombre del usuario que revisó el libro
- rating — calificación

reviews:

Contiene datos sobre las reseñas de los y las clientes:

- review_id — identificación de la reseña
- book_id — identificación del libro
- username — el nombre del usuario que revisó el libro
- text — el texto de la reseña

Diagrama de datos



Ejercicio

- Encuentra el número de libros publicados después del 1 de enero de 2000.
- Encuentra el número de reseñas de usuarios y la calificación promedio para cada libro.
- Identifica la editorial que ha publicado el mayor número de libros con más de 50 páginas (esto te ayudará a excluir folletos y publicaciones similares de tu análisis).
- Identifica al autor que tiene la más alta calificación promedio del libro: mira solo los libros con al menos 50 calificaciones.
- Encuentra el número promedio de reseñas de texto entre los usuarios que calificaron más de 50 libros.

Instrucciones para completar la tarea

- Describe los objetivos del estudio.
- Estudia las tablas (imprime las primeras filas).
- Realiza una consulta SQL para cada una de las tareas.
- Genera los resultados de cada consulta en el Notebook.
- Describe tus conclusiones para cada una de las tareas.

Acceso a la base de datos

Conéctate a la base de datos siguiendo estas [instrucciones](#).

Notas

- ¡No te olvides de las funciones! Pueden facilitar considerablemente tu vida y la ejecución de consultas.
- Tus resultados deben ser obtenidos con SQL. Usa pandas solamente para imprimir y almacenar los resultados de la consulta.