

PM3 - Business Insights

Book.next

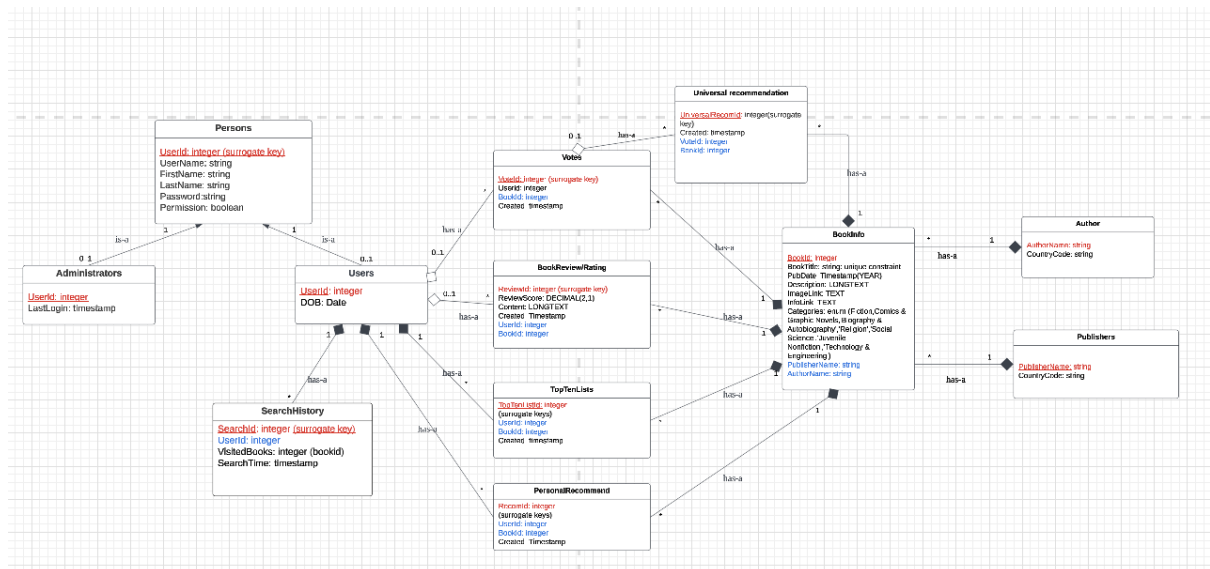
Team Name: List never ends.

Team Members : Jae Butler, Dae Hyun Chung, Yang Chen, Shuang Du, Hui Zou, Jiena Wu, Yunfei Zhang

Our data holds valuable insights that can help us improve our offerings both now and in the future. To unlock these insights, we developed 10 questions that our data can answer. These questions touch on topics ranging from universal and personal recommendations for our current proposition to analyzing user behavior and habit for future proposition.

To ensure that we can efficiently and effectively query our data, we made a small change to the search History table in our original design. Specifically, we modified the visitedBooks field to store the book ID rather than the book's link. This change will simplify table joins and make our data analysis more efficient.

Here is our new UML.



Here is the new version of searchHistory table.

```
CREATE TABLE SearchHistory (
  SearchId INT AUTO_INCREMENT,
  UserId INT,
  VisitedBooks INT NOT NULL,
  SearchTime TIMESTAMP, #DEFAULT CURRENT_TIMESTAMP,
  CONSTRAINT Pk_SearchHistory_SearchId PRIMARY KEY (SearchId),
  CONSTRAINT Fk_SearchHistory_UserId FOREIGN KEY (UserId)
    REFERENCES Users(UserId)
    ON UPDATE CASCADE ON DELETE CASCADE
);
```

10 Questions.

1. Which 10 books have top 10 votes ? Include the book ID, book title, and average voting rating in the result set. (universal recommendation).

```
SELECT BookInfo.BookId AS book_id, BookInfo.BookTitle AS book_title, COUNT(BookInfo.BookTitle) as total_vote_count
FROM BookInfo

INNER JOIN Votes ON BookInfo.BookId = Votes.BookId

GROUP BY book_title

ORDER BY total_vote_count DESC
```

```
LIMIT 10;
```

book_id	book_title	total_vote_co...
44979	A Guidebook to Learning: For a Lifelong Pursuit...	5
100208	What Do You Know	3
20532	The Ghosts of Cape May: book 1	3
131601	Acts (Believers Church Bible Commentary)	3
145793	BLACK REBELLION: Five Slave Revolts (Cosi...	3
89034	The life and times of Joseph E. Clark: From slav...	3
57249	Dragonseye (Dragonriders of Pern Series)	3
124692	Sappho to Valery: Poems in Translation.	3
135352	Mary's Song: Living Her Timeless Prayer	3
6289	Kengo Kuma	3

2. Count the frequency of categories appearing in user's top 10 lists for each user? Include the book ID, book title, book author, and book category.(personal recommendation).

(We did not show all the query result for this problem since it is too long and difficult to show it all.)

```
SELECT Users.UserId, BookInfo.Categories, COUNT(Users.UserId) as CNT
FROM TopTenLists
INNER JOIN Users ON TopTenLists.UserId = Users.UserId
INNER JOIN BookInfo ON TopTenLists.BookId = BookInfo.BookId
WHERE NOT BookInfo.Categories = ''
GROUP BY Users.UserId, BookInfo.Categories
ORDER BY Users.UserId, CNT DESC
```

21	Fiction	1
21	Biography & Autobiography	1
21	Business & Economics	1
22	Religion	2
22	Juvenile Nonfiction	1
23	Fiction	3
23	Religion	1
23	Business & Economics	1
24	Fiction	2
24	Comics & Graphic Novels	1
24	Religion	1
25	Fiction	3
25	Architecture	1
25	Business & Economics	1
26	Juvenile Nonfiction	2
26	Art	1
27	Fiction	1
28	Juvenile Nonfiction	1
28	Fiction	1
29	Fiction	1
29	Business & Economics	1
29	Art	1
30	Technology & Engineering	1
30	Religion	1
30	Fiction	1
31	Religion	2
31	Social Science	1
31	Fiction	1
32	Juvenile Nonfiction	2
32	Fiction	2
32	Social Science	1
32	Religion	1
33	Religion	1
33	Technology & Engineering	1
33	Social Science	1
34	Fiction	1
34	Business & Economics	1
34	Religion	1
35	Fiction	3
35	Religion	1
35	Architecture	1
35	Business & Economics	1
36	Fiction	2

3. Which author appears the most frequently in user's top 10 lists for each user? (personal recommendation).

```
SELECT ttl.UserId, bi.AuthorName, COUNT(*) AS Count
FROM BookInfo AS bi
JOIN TopTenLists AS ttl ON bi.BookId = ttl.BookId
GROUP BY ttl.UserId, bi.AuthorName
ORDER BY ttl.UserId, Count DESC
```

Since each author appears once in the top 10 list for each user, we will recommend the first book shown in the list and its category to the user for now. If we update the TopTenLists table in the future, we will recommend the book with the highest frequency instead.

UserId	AuthorName	Count
22	Karen Salmansohn	1
22	Shirley Harrison, Lynn F	1
22	Judy Donnelly	1
22	Donald A. Hagner	1
22	Bertrand Mathieu	1
22	Editors Of Writers Diges	1
23	Sharon Kathryn Walsh, I	1
23	Sir Ernest Henry Shackl	1
23	Saint Jerome, Malcolm I	1
23	Brian Herbert, Kevin J. /	1
23	Phyllis Reynolds Naylor,	1
23	Gary Jobson	1
23	David Partington	1
23	Larry Evans	1
23	Regina Scott	1
23	Bryan G. Norton, Profes	1
24	Aylmer Haldane, Sir Jan	1
24	Arthur Robert Harding	1
24	Muncy Christian, Sage \	1
24	Brady Udall	1
24	Holman Bible Editorial, \	1

4. Which top 3 books are visited the most by each user? Include the book ID, book title, book author, and book category.(personal recommendation).

(USING MYSQL 8.0 Version Feature)

```
WITH user_search_count AS ( SELECT UserId, VisitedBooks, COUNT(*) AS search_count FROM SearchHistory GROUP BY UserId, VisitedBooks )
SELECT
BookInfo.BookTitle,
BookInfo.AuthorName,
BookInfo.Categories,
t.*
FROM
(
SELECT
t1.*,
( SELECT count(*) + 1 FROM user_search_count t2 WHERE UserId = t1.UserId AND t2.search_count > t1.search_count ) top
FROM
user_search_count t1
) t
INNER JOIN BookInfo ON BookInfo.BookId = t.VisitedBooks
WHERE
top <= 3
ORDER BY
t.UserId,
top;
```

BookTitle	AuthorName	Categories	UserId	VisitedBooks	search_count	top	
Barnaby Rudge; and Hard Times. In Two Volumes. Library Edition	Charles Dickens		21	4	9	1	
Portals to freedom	Howard Colby Ives		21	9	8	2	
Ways of Escape	Graham Greene		21	7	6	3	
Kurt Weill: Composer in a Divided World	Ronald Taylor		21	5	6	3	
Life on a Pig Farm (Carolrhoda Photo Books)	Judy Wolfman	American literature	21	1	6	3	
Kurt Weill: Composer in a Divided World	Ronald Taylor		22	5	10	1	
Deaf Heritage: A Narrative History of Deaf America	Jack R. Gannon		22	3	9	2	
SCENTED GARDENS FOR THE BLIND a Novel	Janet Frame		22	6	5	3	
Portals to freedom	Howard Colby Ives		22	9	5	3	
Ways of Escape	Graham Greene		23	7	10	1	
Key to the Sinai: The battles for Abu Ageila in the 1956 and 1967 Arab	George Walter Gawrych		23	10	8	2	
Barnaby Rudge; and Hard Times. In Two Volumes. Library Edition	Charles Dickens		23	4	7	3	

5. What are the top 10 average review ratings for books? Include the book ID, book title, and average rating in the result set.(personal recommendation).

```

SELECT
AVG( ReviewScore ) AS AverageRating,
BookTitle,
AuthorName
FROM
BookReview
INNER JOIN BookInfo ON BookReview.BookId = BookInfo.BookId
GROUP BY
BookReview.BookId
ORDER BY
AverageRating DESC
LIMIT 10

```

	AverageRating	BookTitle	AuthorName
▶	5.00000	Spectrum Math, Grad...	Spectrum
●	5.00000	Staying Power : 30 S...	Thomas A. Schweich
●	5.00000	Patti Smith : An Unau...	Victor Bockris, Roberta Bayley
●	5.00000	Call of Duty Official S...	Bart Farkas
●	5.00000	Ferrari F1	Stuart Codling
●	5.00000	Seven Minutes to Noon	Katia Lief
●	5.00000	Falling Petals: A Coll...	Joyce Ann Edmondson
●	5.00000	Moondream	Victor Osborne
●	5.00000	The Great Physician'...	Jordan Rubin
●	5.00000	The TALE OF THE VI...	Ted Pedersen

6. Analyze users' behavior

- 1) Which users have created more than one review ?
 - 2) Which users have created more than one vote?
 - 3) Which users have not created any review nor vote?
- 1)

```

SELECT Persons.UserId, Persons.UserName, COUNT(*) as review_count

FROM BookReview

INNER JOIN Users

ON Users.UserId = BookReview.UserId

INNER JOIN Persons

ON Persons.UserId = Users.UserId

GROUP BY UserId

HAVING review_count > 1

ORDER BY review_count DESC;

```

UserId	UserName	review_count
102	lgoodwins2t	3
103	dkenset2u	3
99	cplum2q	2
100	csothcott2r	2
101	agirardini2s	2
31	aslineyu	2
32	stellingv	2
33	calabastarw	2
34	rhymex	2
98	ggypps2p	2

2)

```

SELECT Persons.UserId, Persons.UserName, COUNT(*) as vote_count

FROM Votes

INNER JOIN Users

ON Users.UserId = Votes.UserId

INNER JOIN Persons

ON Persons.UserId = Users.UserId

GROUP BY Persons.UserId

HAVING vote_count > 1

ORDER BY vote_count DESC;

```

UserId	UserName	vote_count
25	rmartusovo	4
26	dcoffinp	4
27	tclemmeyq	4
32	stellingv	4
35	mmacmurrayy	3
36	hpiscotz	3
37	rmould10	3
38	mrummying11	3
39	sbarbery12	3
28	vkilleenr	3
29	dsaverys	3
30	urobeiro	3
31	aslineyu	3
33	calabastarw	3
34	rhymex	3
21	cstihlk	2
22	aweymouthl	2
23	jsparshottm	2
24	flimern	2
56	mchastand1j	2

3)

```
SELECT Users.UserID as UserIDNoReviewAndVote
FROM Users
WHERE Users.UserID
NOT IN (SELECT DISTINCT Users.UserId
FROM Users
RIGHT JOIN Votes ON Users.UserId = Votes.UserId)
AND Users.UserID
NOT IN (SELECT DISTINCT Users.UserId
FROM Users
RIGHT JOIN BookReview ON Users.UserId = BookReview.UserId);
```

UserIDNoReviewAndVote
14

7. List the top 3 categories users reviewed with the review counts for each category

```
Select BookInfo.Categories, Count(BookReview.reviewID) AS ReviewCNT
FROM BookInfo
INNER JOIN BookReview ON BookInfo.Bookid = BookReview.Bookid
WHERE BookInfo.categories <>''
Group by BookInfo.categories
Order by ReviewCNT DESC
LIMIT 3
```

Categories	ReviewCNT
Fiction	27
Religion	9
Biography & Autobiography	8

8. Find out the average rating of books published by different publishers. List the top 20 publishers with their name, country Code, and average rating. (relationship between book review and book publisher, it help publishers identify their marketing strategies)

```
SELECT Publishers.PublisherName, CountryCode, AVG(ReviewScore) AS Avg_Rating
FROM Publishers
INNER JOIN BookInfo ON Publishers.PublisherName = BookInfo.PublisherName
INNER JOIN BookReview ON BookInfo.BookId = BookReview.BookId
WHERE Publishers.PublisherName IS NOT NULL
GROUP BY Publishers.PublisherName
ORDER BY Avg_Rating DESC
LIMIT 20;
```

PublisherName	CountryCode	Avg_Rating
Carson	BR	5.00000
McGraw Hill Professional	US	5.00000
Simon and Schuster	CO	5.00000
Gale / Cengage Learning	PG	5.00000
Motorbooks	US	5.00000
Blue Table Books	US	5.00000
Publishamerica Incorporated	CD	5.00000
Lothrop Lee & Shepard	AU	5.00000
Thomas Nelson	ET	5.00000
Simon & Schuster/Paula Wiseman Books	AF	5.00000
John Wiley & Sons	US	5.00000
Brady	KE	5.00000
AuthorHouse	MM	5.00000
Berkley	TN	5.00000
Our Sunday Visitor	GY	5.00000
Penguin UK	AR	5.00000
HarperCollins Australia	US	5.00000
University of Washington Press	US	5.00000
Boldwood Books Ltd	AU	5.00000
Routledge	US	5.00000

9. What is the user's range(Generation) of ages of their favorite genre of the book?(does age differentiate different genre of the book) (relationship between favorite book genre and user's age.)

Consider:

Baby Boomer : 1946 -1964

Gen X : 1965-1979

MIL : 1980 - 1995

Gen Z = 1996 - 2012

```
SELECT GenCat.Generation AS Geno, GenCat.Categories AS GenCategory, COUNT(GenCat.UserId) AS CNT
FROM (
SELECT BookInfo.Categories,Users.UserId,
CASE
WHEN YEAR(DOB) BETWEEN 1946 AND 1964 THEN "BABYBOOMER"
WHEN YEAR(DOB) BETWEEN 1965 AND 1979 THEN "GEN_X"
WHEN YEAR(DOB) BETWEEN 1980 AND 1995 THEN "MILLENNIAL"
WHEN YEAR(DOB) BETWEEN 1996 AND 2012 THEN "GEN_Z"
END AS Generation
FROM TopTenLists
INNER JOIN Users on TopTenLists.UserId = Users.UserId
INNER JOIN BookInfo ON TopTenLists.BookId = BookInfo.BookId
WHERE BookInfo.Categories<>'')
```



```

) AS GenCat
GROUP BY Geno, GenCategory
ORDER BY Geno, CNT DESC

```

	Geno	GenCategory	CNT	
►	GEN_Z	Fiction	86	
	GEN_Z	Religion	38	
	GEN_Z	Biography & Autobiography	27	
	GEN_Z	Business & Economics	25	
	GEN_Z	Juvenile Nonfiction	19	
	GEN_Z	Social Science	13	
	GEN_Z	Art	9	
	GEN_Z	Technology & Engineering	7	
	GEN_Z	Comics & Graphic Novels	5	
	GEN_Z	Architecture	3	
	GEN_Z	American literature	2	
	GEN_Z	Antiques & Collectibles	2	
	GEN_Z	Aeronautics	1	
	MILLENNIAL	Fiction	159	
	MILLENNIAL	Religion	55	
	MILLENNIAL	Business & Economics	39	
	MILLENNIAL	Biography & Autobiography	35	
	MILLENNIAL	Social Science	34	
	MILLENNIAL	Juvenile Nonfiction	27	
	MILLENNIAL	Comics & Graphic Novels	9	
	MILLENNIAL	Architecture	9	
	MILLENNIAL	Art	9	
	MILLENNIAL	Technology & Engineering	6	
	MILLENNIAL	American literature	4	
	MILLENNIAL	American fiction	2	
	MILLENNIAL	Aeronautics	1	
	MILLENNIAL	African Americans	1	
	MILLENNIAL	Antiques & Collectibles	1	

10. What are the top 5 most searched and most reviewed books? (relationship between search history and book reviews, which can provide insight into recent trends)

```

SELECT BookId, Count(*) AS COUNT
FROM BookReview
GROUP BY BookId
ORDER BY COUNT DESC
LIMIT 5;

```

	BookId	COUNT
►	75553	3
	102930	3
	25824	3
	114632	3
	16376	3

```

SELECT VisitedBooks, Count(*) AS COUNT
FROM SearchHistory
GROUP BY VisitedBooks
ORDER BY COUNT DESC

```

(Only part of results (top 15 books) is presented here.)

	VisitedBooks	COUNT	
▶	3	11	
	7	10	
	35	9	
	12	8	
	28	7	
	26	7	
	1	7	
	5	6	
	46	6	
	29	6	
	48	6	
	14	6	
	25	6	
	11	6	
	40	6	
	2	6	