



Amazon Bookstore

Wide selection | 25M+ books
in 8+ languages

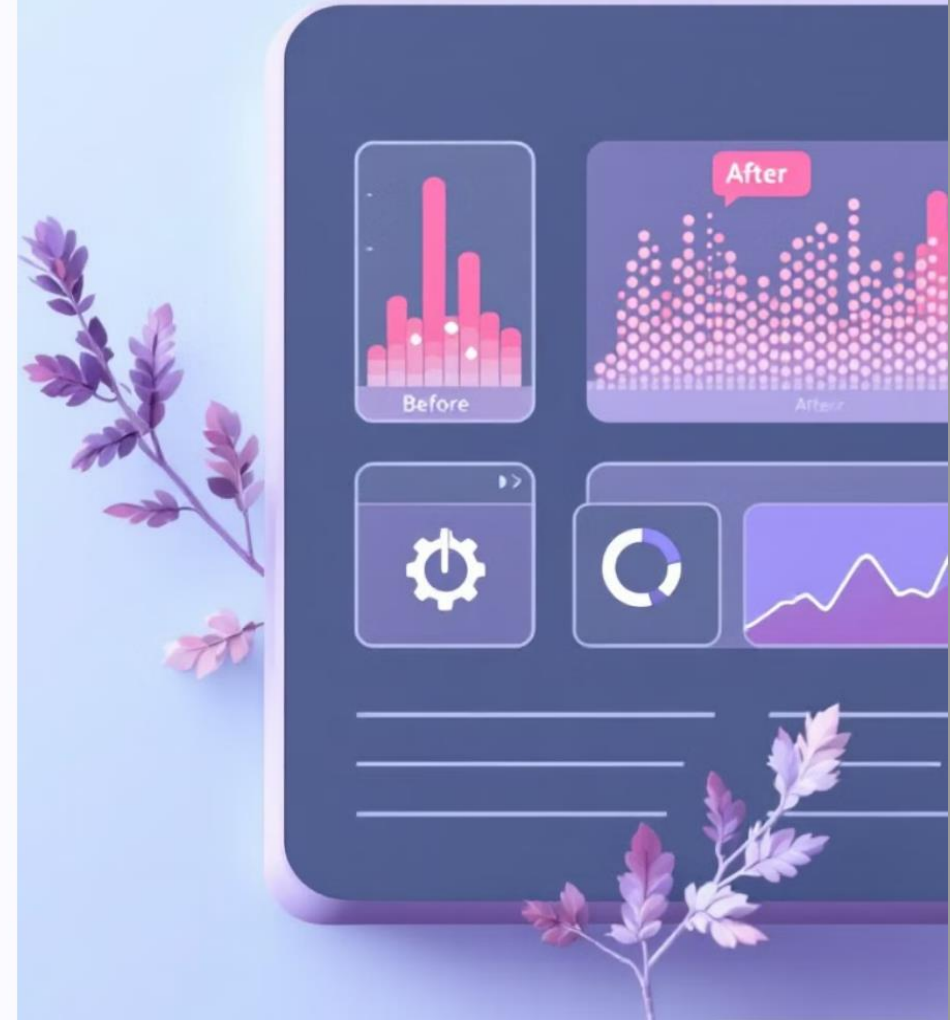
Amazon Bookstore Data Analysis with SQL

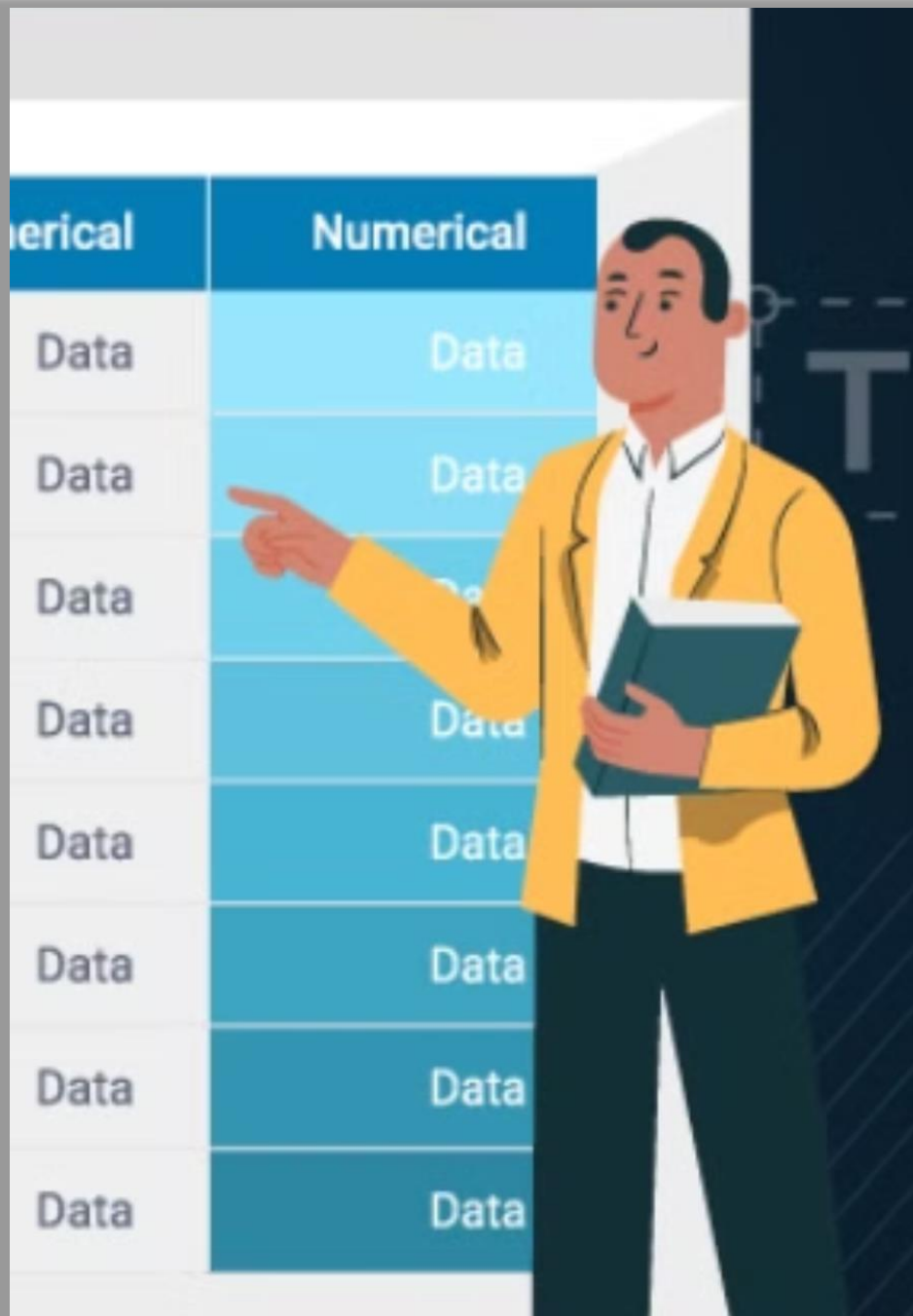
- This presentation outlines the steps to analyze bookstore data using SQL.
- We'll cover data cleaning, database creation, and insightful queries.

by Pavani K.V 

Data Cleaning

- 1 Remove NULL Values**
Eliminate NULL values from key columns.
- 2 Replace When Necessary**
Use defaults like 0 or UNKNOWN for unavoidable NULLs.





Database & Table Creation

```
CREATE DATABASE BookStore;  
USE BookStore;  
CREATE TABLE Books (  
    book_id INT AUTO_INCREMENT PRIMARY KEY,  
    book_name VARCHAR(255) NOT NULL,  
    author VARCHAR(255) NOT NULL,  
    pages INT NOT NULL,  
    language VARCHAR(50) NOT NULL,  
    ratings DECIMAL(2,1) NOT NULL DEFAULT 0.0,  
    total_ratings INT NOT NULL DEFAULT 0,  
    price DECIMAL(10,2) NOT NULL DEFAULT 0.0,  
    category VARCHAR(100) NOT NULL  
);
```

Inserting Cleaned Data

```
INSERT INTO Books (book_name, author, pages, language, ratings,  
total_ratings, price, category)  
VALUES  
('The Power of Habit', 'Charles Duhigg', 371, 'English', 4.8, 25000, 299.99, 'Self  
Improvement'),  
('Sapiens: A Brief History of Humankind', 'Yuval Noah Harari', 443, 'English',  
4.7, 50000, 499.99, 'Non-Fiction'),  
('The Silent Patient', 'Alex Michaelides', 336, 'English', 4.5, 15000, 399.99,  
'Thriller'),  
('Becoming', 'Michelle Obama', 426, 'English', 4.9, 30000, 599.99, 'Biography'),  
('The Alchemist', 'Paulo Coelho', 208, 'English', 4.6, 200000, 199.99, 'Fiction');
```



SQL Queries for Insights

Top 5 Highest-Rated Books

```
SELECT book_name, author, ratings
FROM Books
ORDER BY ratings DESC
LIMIT 5;
```

```
mysql> SELECT book_name, author, ratings FROM Books
-> ORDER BY ratings DESC
-> LIMIT 5;
```

book_name	author	ratings
The Power of Now	Eckhart Tolle	4.9
The Intelligent Investor	Benjamin Graham	4.8
The Life-Changing Magic of Tidying Up	Marie Kondo	4.8
The Art of Happiness	Dalai Lama	4.8
The 7 Habits of Highly Effective People	Stephen Covey	4.8

5 rows in set (0.09 sec)

Books with Most Reviews

```
SELECT book_name, author, total_ratings
FROM Books
ORDER BY total_ratings DESC
LIMIT 5;
```

```
mysql> SELECT book_name, author, total_ratings FROM Books
-> ORDER BY total_ratings DESC
-> LIMIT 5;
```

book_name	author	total_ratings
The Secret	Rhonda Byrne	290000
You Are a Badass	Jen Sincero	289432
The Obstacle Is the Way	Ryan Holiday	256874
Principles: Life and Work	Ray Dalio	239567
Principles: Life and Work	Ray Dalio	239567

5 rows in set (0.81 sec)

Average Price by Category

Average Price Query

```
SELECT category, AVG(price) AS avg_price  
FROM Books  
GROUP BY category;
```

```
mysql> SELECT category, AVG(price) AS avg_price FROM Books  
-> GROUP BY category;  
+-----+-----+  
| category | avg_price |  
+-----+-----+  
| Self Improvement | 344.734737 |  
| Spirituality | 299.990000 |  
| Business | 424.000000 |  
| Finance | 394.495000 |  
| Non-Fiction | 439.495000 |  
| Psychology | 401.000000 |  
| Philosophy | 199.000000 |  
| Leadership | 459.000000 |  
+-----+-----+  
8 rows in set (0.04 sec)
```

Most Popular Categories

```
SELECT category, SUM(total_ratings) AS total_reviews  
FROM Books  
GROUP BY category  
ORDER BY total_reviews DESC  
LIMIT 5;
```

```
mysql> SELECT category, SUM(total_ratings) AS total_reviews FROM Books  
-> GROUP BY category  
-> ORDER BY total_reviews DESC  
-> LIMIT 5;  
+-----+-----+  
| category | total_reviews |  
+-----+-----+  
| Self Improvement | 2362595 |  
| Business | 942653 |  
| Finance | 618771 |  
| Non-Fiction | 262349 |  
| Leadership | 229155 |  
+-----+-----+  
5 rows in set (0.00 sec)
```

Cheapest & Most Expensive

Cheapest Book

```
SELECT book_name, price
FROM Books
ORDER BY price ASC
LIMIT 1;
```

```
mysql> SELECT book_name, price FROM Books
-> ORDER BY price ASC
-> LIMIT 1;
```

book_name	price
The Four Agreements	149.99

1 row in set (0.00 sec)

Most Expensive Book

```
SELECT book_name, price
FROM Books
ORDER BY price DESC
LIMIT 1;
```

```
mysql>
mysql> SELECT book_name, price FROM Books
-> ORDER BY price DESC
-> LIMIT 1;
```

book_name	price
Principles: Life and Work	549.00

1 row in set (0.00 sec)



Distribution by Page Count

```
SELECT
CASE
  WHEN pages < 200 THEN 'Short (<200 pages)'
  WHEN pages BETWEEN 200 AND 400 THEN 'Medium (200-400 pages)'
  ELSE 'Long (>400 pages)'
END AS page_length_category,
COUNT(*) AS book_count
FROM Books
GROUP BY
  page_length_category
ORDER BY
  book_count DESC;
```

```
mysql> SELECT
-> CASE
->   WHEN pages < 200 THEN 'Short (<200 pages)'
->   WHEN pages BETWEEN 200 AND 400 THEN 'Medium (200-400 pages)'
->   ELSE 'Long (>400 pages)'
-> END AS page_length_category,
-> COUNT(*) AS book_count
-> FROM Books
-> GROUP BY page_length_category
-> ORDER BY book_count DESC;
+-----+-----+
| page_length_category | book_count |
+-----+-----+
| Medium (200-400 pages) | 38 |
| Long (>400 pages) | 6 |
| Short (<200 pages) | 4 |
+-----+-----+
3 rows in set (0.00 sec)
```



The image shows a section of a bookshelf with three rows of books. The books are mostly in Czech. The top row includes titles like 'Duané Elgin Živý v...', 'Ach Běda, přeběda', 'Encyklopedie mystiky', and 'Kalevala'. The middle row features 'Diz. Borů', 'Transformace vědomí - rozhovory', 'Kniha o štěstí', and 'Čínská gnoze'. The bottom row contains 'Ruiz & Ruiz', 'Karel Sládek', 'Mění věst', and 'Svět islamu'. A red bookmark is visible between the middle and bottom rows.

```
SELECT
CASE
    WHEN price < 200 THEN 'Below 200'
    WHEN price BETWEEN 200 AND 400 THEN '200-400'
    WHEN price BETWEEN 400 AND 600 THEN '400-600'
    ELSE 'Above 600'
END AS price_range,
AVG(ratings) AS avg_rating,
COUNT(*) AS book_count
FROM Books
GROUP BY
    price_range
ORDER BY
    price_range;
```

```
mysql> SELECT
->     CASE
->         WHEN price < 200 THEN 'Below 200'
->         WHEN price BETWEEN 200 AND 400 THEN '200-400'
->         WHEN price BETWEEN 400 AND 600 THEN '400-600'
->         ELSE 'Above 600'
->     END AS price_range,
->     AVG(ratings) AS avg_rating,
->     COUNT(*) AS book_count
-> FROM Books
-> GROUP BY price_range
-> ORDER BY price_range;
```

price_range	avg_rating	book_count
200-400	4.70357	28
400-600	4.68750	16
Below 200	4.65000	4

```
3 rows in set (0.05 sec)
```


A watercolor illustration featuring a central white rectangular area with the text 'Thank you!'. Surrounding this area are various artistic elements: a stack of books with green and brown covers on the left, some with yellow flowers; an open book with a green leaf on its page in the lower center; and other books and papers scattered on the right and bottom edges. The style is soft and painterly.

Thank you!

Any Questions?