

Book-store Analysis

Objective:

Imagine you run a chain of bookstores. Create a sample dataset of your favorite books by noting the following details for each book in a table: Book_ID, Book Name, Publisher, ISBN, Edition, Number of pages, Sales, City, and Price.

SQL Data Analysis Project Idea: Use the dataset to answer the following questions:

Task # 01

Which book is the most expensive?

Task # 02

What are the most popular books in each city?

Task # 03

Top 5 most bought books.

Task # 04

Calculate the total revenue generated by each city.

Task # 05

Write a SQL query that categorizes books based on their price range into three categories: "Affordable" for books priced below \$15.00, "Moderate" for books priced between \$15.00 and \$25.00, and "Expensive" for books priced \$25.00 and above.

Task # 06

Which three books are less preferred by reader?

Database: Bookstore

Table: Books

The following table contains detailed information about the books available in the bookstore chain, including identifiers, titles, publication details, physical attributes, sales figures, and pricing.

Column Name	Data Type	Description
Book_ID	INT	Unique identifier for each book.
Book_Name	VARCHAR	Name of the book.
Publisher	VARCHAR	Publisher of the book.
ISBN	VARCHAR	International Standard Book Number.
Edition	VARCHAR	Edition of the book.
Number_of_Pages	INT	Total number of pages in the book.
Sales	INT	Number of copies sold.
City	VARCHAR	City where the book is sold.
Price	DECIMAL	Price of the book in USD.

Primary Key: **Book_ID** is the primary key for the **Books** table, ensuring each book entry is unique.

Here is the sample data used to generate insights into book sales and distribution across multiple cities. This dataset includes key details such as book titles, publishers, ISBNs, editions, page counts, sales figures, city distributions, and pricing information.

Book_ID	Book Name	Publisher	ISBN	Edition	Number of Pages	Sales	City	Price (USD)
1	To Kill a Mockingbird	J.B. Lippincott	978-0061120084	1st	324	1500	New York	18.99
2	1984	Secker & Warburg	978-0451524935	1st	328	1700	Los Angeles	15.99
3	The Great Gatsby	Scribner	978-0743273565	1st	180	1300	Chicago	14.99
4	The Catcher in the Rye	Little, Brown	978-0316769488	1st	214	1400	Houston	10.99
5	Pride and Prejudice	T. Egerton	978-1503290563	2nd	279	1200	New York	9.99
6	The Hobbit	Allen & Unwin	978-0547928227	2nd	310	1800	Philadelphia	14.99
7	Fahrenheit 451	Ballantine Books	978-1451673319	1st	158	1600	Los Angeles	13.99
8	Jane Eyre	Smith, Elder	978-0141441146	3rd	500	1100	San Diego	11.99
9	Brave New World	Chatto & Windus	978-0060850524	1st	268	1250	Philadelphia	12.99
10	The Lord of the Rings	Allen & Unwin	978-0618640157	4th	1178	1900	Chicago	25.99
11	Animal Farm	Secker & Warburg	978-0451526342	1st	112	1550	Austin	9.99
12	Gone with the Wind	Macmillan	978-1416548942	1st	1037	1400	San Antonio	24.99
13	The Chronicles of Narnia	Geoffrey Bles	978-0066238500	3rd	768	1300	San Francisco	22.99
14	The Book Thief	Picador	978-0375842207	1st	552	1200	Columbus	12.99
15	The Hunger Games	Scholastic Press	978-0439023481	1st	374	1700	San Diego	10.99
16	Harry Potter and the Sorcerer's Stone	Bloomsbury	978-0590353427	1st	309	2000	Fort Worth	16.99
17	The Fault in Our Stars	Dutton Books	978-0525478812	1st	313	1450	Austin	11.99
18	The Alchemist	HarperOne	978-0062315007	2nd	208	1600	Seattle	14.99
19	Moby Dick	Harper & Brothers	978-1503280786	1st	635	1100	Denver	12.99
20	War and Peace	The Russian Messenger	978-0199232765	1st	1225	1000	Washington	29.99
21	Great Expectations	Chapman & Hall	978-0141439563	1st	505	1200	San Antonio	13.99
22	The Shining	Doubleday	978-0307743657	1st	659	1450	Denver	15.99
23	Crime and Punishment	The Russian Messenger	978-0486415871	1st	430	950	Columbus	11.99
24	The Da Vinci Code	Doubleday	978-0307474278	1st	489	1750	Seattle	19.99
25	The Catch-22	Simon & Schuster	978-1451626650	1st	453	1600	San Francisco	14.99

Task 1: Which book is the most expensive?

The SQL query to determine the most expensive book is given by:

```
SELECT Book_ID, Book_Name, Price
FROM bookstore.books
ORDER BY Price DESC
LIMIT 1;
```

Output:

	Book_ID	Book_Name	Price
▶	20	War and Peace	29.99
•	NULL	NULL	NULL

The book "War and Peace" with a price of \$29.99 is the most expensive book in this dataset.

Task 2: What are the most popular books in each city?

```
WITH Max_Sales_By_City AS (  
  SELECT City, MAX(Sales) AS Max_Sales  
  FROM bookstore.Books  
  GROUP BY City  
)  
SELECT Books.Book_ID, Books.Book_Name, Books.City, Books.Sales  
FROM bookstore.Books  
INNER JOIN Max_Sales_By_City  
ON Books.City = Max_Sales_By_City.City AND Books.Sales = Max_Sales_By_City.Max_Sales;
```

Output:

	Book_ID	Book_Name	City	Sales
▶	1	To Kill a Mockingbird	New York	1500
	2	1984	Los Angeles	1700
	10	The Lord of the Rings	Chicago	1900
	4	The Catcher in the Rye	Houston	1400
	6	The Hobbit	Philadelphia	1800
	15	The Hunger Games	San Diego	1700
	11	Animal Farm	Austin	1550
	12	Gone with the Wind	San Antonio	1400
	25	The Catch-22	San Francisco	1600
	14	The Book Thief	Columbus	1200
	16	Harry Potter and the Sorcerer's Stone	Fort Worth	2000
	24	The Da Vinci Code	Seattle	1750
	22	The Shining	Denver	1450
	20	War and Peace	Washington	1000

Task 3: Top 5 most bought books.

```
SELECT Book_ID, Book_Name, Sales
FROM bookstore.Books
ORDER BY Sales DESC
LIMIT 5;
```

Output:

	Book_ID	Book_Name	Sales
	16	Harry Potter and the Sorcerer's Stone	2000
	10	The Lord of the Rings	1900
	6	The Hobbit	1800
	24	The Da Vinci Code	1750
	15	The Hunger Games	1700
*	NULL	NULL	NULL

Task 4: Calculate the total revenue generated by each city.



```
SELECT DISTINCT City, SUM(Sales * Price) AS Total_Revenue
FROM bookstore.Books
GROUP BY City
ORDER BY Total_Revenue;
```

Output:

	City	Total_Revenue
►	Houston	15386.00
	Columbus	26978.50
	Washington	29990.00
	San Diego	31872.00
	Austin	32870.00
	Fort Worth	33980.00
	Denver	37474.50
	New York	40473.00
	Philadelphia	43219.50
	Los Angeles	49567.00
	San Antonio	51774.00
	San Francisco	53871.00
	Seattle	58966.50
	Chicago	68868.00

Task 5:

Write a SQL query that categorizes books based on their price range into three categories: "Affordable" for books priced below \$15.00, "Moderate" for books priced between \$15.00 and \$25.00, and "Expensive" for books priced \$25.00 and above.

```
SELECT Book_Name, Price,
CASE
    WHEN Price < 15.00 THEN "Affordable"
    WHEN Price BETWEEN 15.00 AND 25.00 THEN "Moderate"
    WHEN Price > 25.00 THEN "Expensive"
END AS Price_Category
FROM Books
LIMIT 10;
```

Output:

	Book_Name	Price	Price_Category
►	To Kill a Mockingbird	18.99	Moderate
	1984	15.99	Moderate
	The Great Gatsby	14.99	Affordable
	The Catcher in the Rye	10.99	Affordable
	Pride and Prejudice	9.99	Affordable
	The Hobbit	14.99	Affordable
	Fahrenheit 451	13.99	Affordable
	Jane Eyre	11.99	Affordable
	Brave New World	12.99	Affordable
	The Lord of the Rings	25.99	Expensive

Task 6: Which three books are less preferred by reader?

```
SELECT Book_ID, Book_Name, Sales
FROM bookstore.Books
ORDER BY Sales ASC
LIMIT 3;
```

Output:

	Book_ID	Book_Name	Sales
▶	23	Crime and Punishment	950
	20	War and Peace	1000
	8	Jane Eyre	1100
•	NULL	NULL	NULL