

# Analysis of the Books Sale on Amazon.com

**Technology Used:** SQL

**Author:** Haider Ali

**Contact:** haiderali.jnr@gmail.com

---

## 1. Project Overview

This project involves a comprehensive analysis of book listings on **Amazon.com**, using SQL to derive insights from structured data. The goal was to explore what types of books are performing well, how prices vary across genres and formats, and how user ratings reflect customer engagement.

The dataset included details such as **book titles, genres, formats (Kindle, Paperback, Hardcover), prices, average ratings, and number of people who rated** each book.

---

## 2. Objective

The key objectives of this case study were:

- Identify top-selling books and authors
  - Explore pricing strategies by genre and format
  - Understand the distribution of book ratings
  - Determine the most preferred book types
  - Analyse user engagement through ratings count
- 

## 3. Dataset Description

The dataset used in this project contains information about books sold on Amazon.com. It includes:

- **Title:** Name of the book
  - **Author:** Name of the author or publishing house
  - **Main Genre and Sub Genre**
  - **Type:** Format (Paperback, Kindle, Hardcover, Audiobook)
  - **Price:** In INR
  - **Rating:** Average user rating (out of 5)
  - **No. of People Rated:** Count of users who rated the book
  - **URLs:** Link to Amazon product page
- 

## 4. SQL Analysis Performed

- Extracted genre-wise and type-wise average pricing
  - Identified top-selling books by estimating ratings count as proxy
  - Analysed rating distributions to gauge book quality
  - Queried for author performance within each genre
  - Compared engagement and popularity across book formats
  - Grouped and ranked genres by average price and engagement
- 

## 5. Key Insights

- **Top Selling Books:**  
"Where the Crawdads Sing" by Delia Owens was the most rated book (500k+), followed by "Girl on the Train" and "The Silent Patient".
- **Genre with Highest Price:**  
Medicine & Health Science had the highest average book price (~₹1850), although it had fewer listings.
- **Rating Distribution:**  
~62% of books were rated between 3 and 4.5. Less than 5% had ratings below 3.
- **Genre Engagement:**  
Genres with high rating counts reflect strong engagement and popularity. These should be targeted by publishers.

- **Top Authors by Genre:**  
High-volume, high-rating authors were identified genre-wise. Useful for partnerships and promotions.
  - **Most Preferred Book Type:**  
*Paperback* was the most preferred format, followed by *Kindle Edition* and *Hardcover*.
  - **Pricing by Book Type:**  
*Hardcover* editions were the most expensive and least preferred. *Paperback* offered the best balance of price and preference.
- 

## 6. Business Recommendations

- **Expand product offerings** in genres and formats with high user engagement.
  - Focus marketing on **top authors** and **popular sub-genres**.
  - Promote **paperbacks and Kindle editions** more than hardcovers.
  - Consider price adjustments for **underperforming book types**.
  - Use engagement metrics (like number of ratings) to tailor recommendations.
- 

## 7. Tools Used

- **SQL** for querying and analysis
  - **Excel** for formatting raw data
  - **Presentation tools** (PowerPoint/Canva) for visual storytelling
- 

## 8. Conclusion

This project successfully demonstrated how SQL can be used to extract actionable insights from an e-commerce dataset. The findings can help publishers, retailers, and marketers make data-driven decisions related to book pricing, marketing, and content strategy on Amazon.