BOOK SALES TRENDS

Remas Al-Subaie, Mona Alnajjar, Jana Alruzuq, Rima Alsonbul, Razan Aldosari

Supervised By: Dr. Mashael Aldayel

King Saud University · جامعة الملك سعود · 1957

Introduction

With the rapid expansion of the book market and the increasing diversity of genres, there is a growing need to understand the factors that contribute to a book's success. Our project addresses this need by analyzing book data such as the number of reviews and bestseller status to uncover patterns that reveal the most in-demand genres and the elements that attract readers. Through this analysis, we aim to provide valuable insights that can help publishers and authors enhance their marketing strategies and boost the success of their books.

Objectives

The analysis will focus on understanding:

- How do ratings and the number of reviews vary among bestsellers?
- Are certain authors more likely to have their books become bestsellers?
- What visual design patterns are consistently associated with bestseller book covers?
- What genres are most represented among bestsellers?
- What is the relationship between price and bestseller books? What is the price range of bestseller books?

Data Collection

Sources:

- Amazon.sa and Jarir.com bestsellers (top 100–200 books per site).

Tools:

- Web Scraper (Chrome extension) – for Amazon.
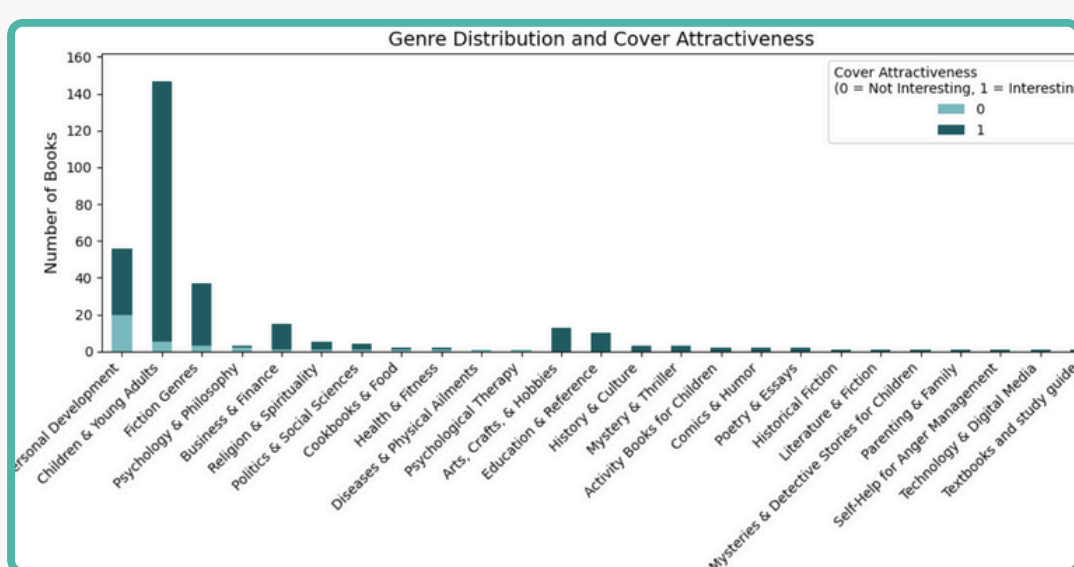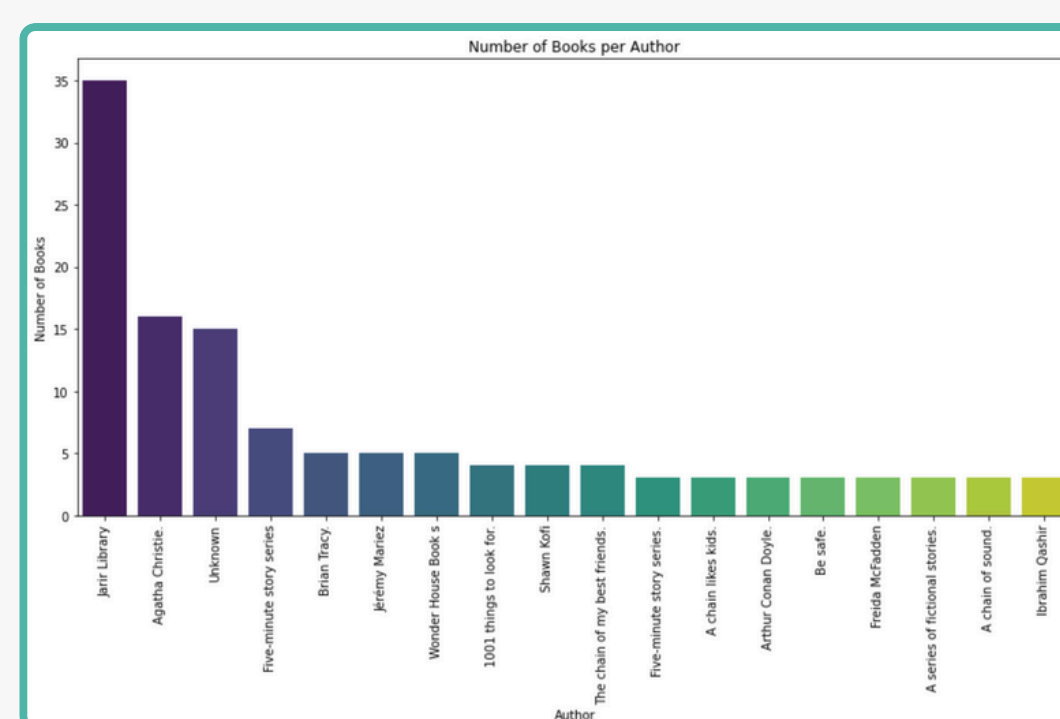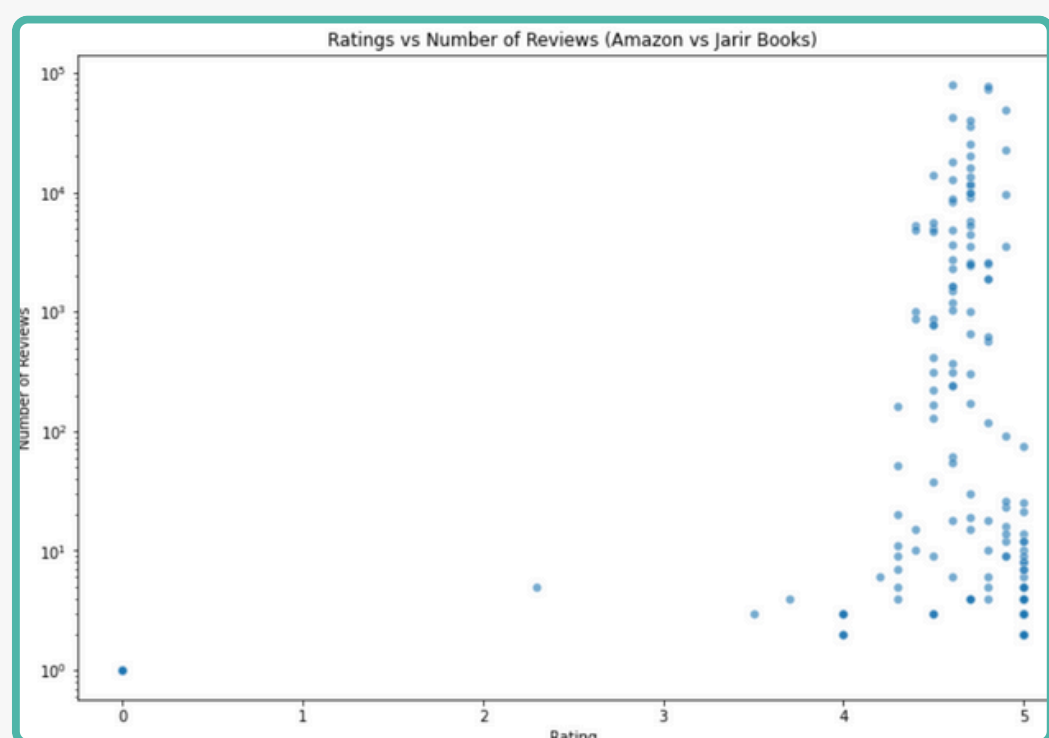- Instant Data Scraper – for Jarir.

Technologies:

- Python libraries: requests, BeautifulSoup, Selenium, pandas.
- Purpose: Scraped book info like title, genre, rating , reviews

Process:

- Navigated bestseller pages.
- Extracted and stored data in structured formats.

DATA ANALYSIS

MODELS AND FINDINGS

We aimed to classify bestseller book covers based on visual features to uncover design patterns associated with successful books.

Feature Extraction:

Extracted features: Contrast, Edge Complexity, Color Saturation, Color Variance.

Clustering Algorithms Applied:

K-Means Clustering (k=2), Spectral Clustering (k=2).

Evaluation Method:

Used Silhouette Score to measure clustering quality.
Calculated the mean value of features within each cluster to understand the visual characteristics.

- K-Means: 0.25
- Spectral Clustering: 0.22

Clustering Conclusion:

K-Means was selected as the better method for its easier interpretation, slightly better performance, and clearer feature separation.

Key Findings and Insights:

- Two main cover styles emerged:
  - Bold, high-contrast, vibrant designs (~50%)
  - Subtle, muted designs (~50%)
- High contrast and vivid color saturation strongly correlate with attractive book covers.
- Edge complexity and sharpness had much less influence.

Conclusion

Bestseller success is driven by a combination of factors authorship, bold and vibrant cover design, dominant genres, and affordable pricing not a single element. Popular authors with loyal audiences, eye-catching visuals, and strategic pricing all contribute to greater market success.

These findings empower authors, publishers, and designers to make data-driven decisions to optimize content, design, and marketing strategies.