

BOOK SALES AND RATINGS : EDA

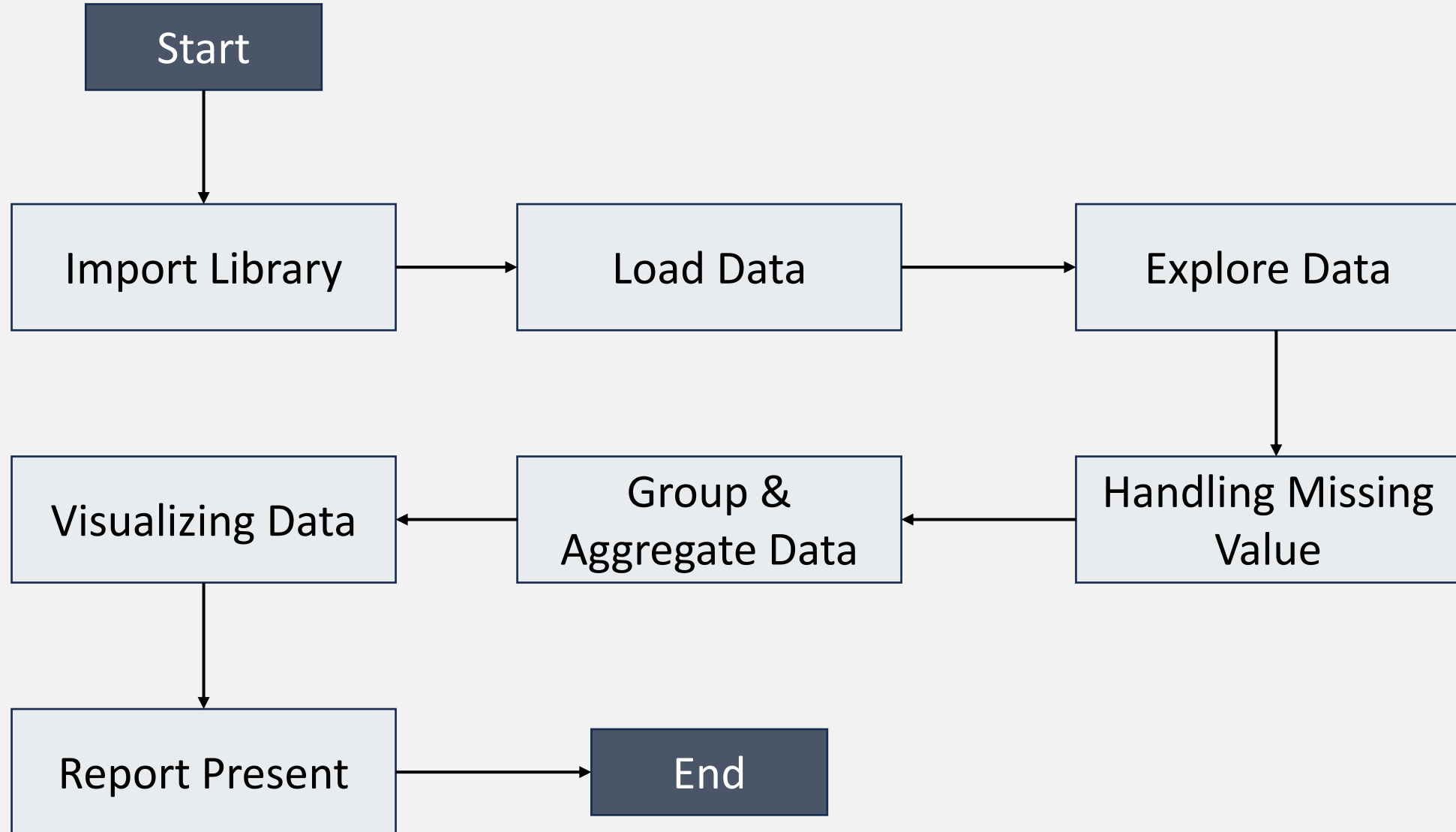
Data Description :

The dataset contains book-related information, including publishing details, author reputation, reader ratings, and financial sales metrics. Each row represents a single book, categorized by genre and other key attributes such as sales rank and units sold

Use Case:

This dataset is useful for analyzing trends in book publishing, evaluating author performance, understanding reader preferences, and assessing the financial success of books based on key metrics such as ratings, price, and units sold.

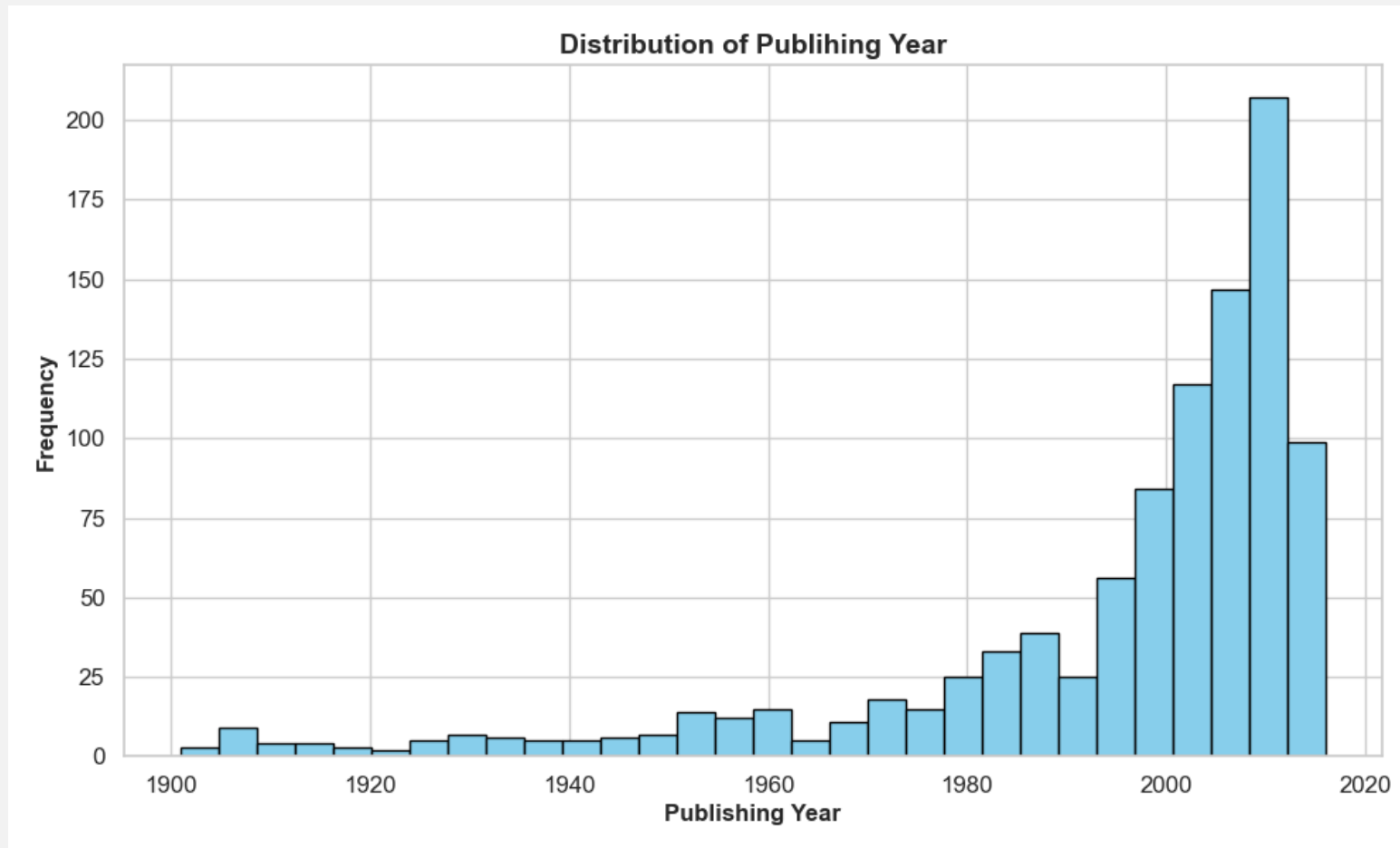
Flow Chart of This Book Sales Analysis



Outcomes: Book Sales & Ratings Analysis

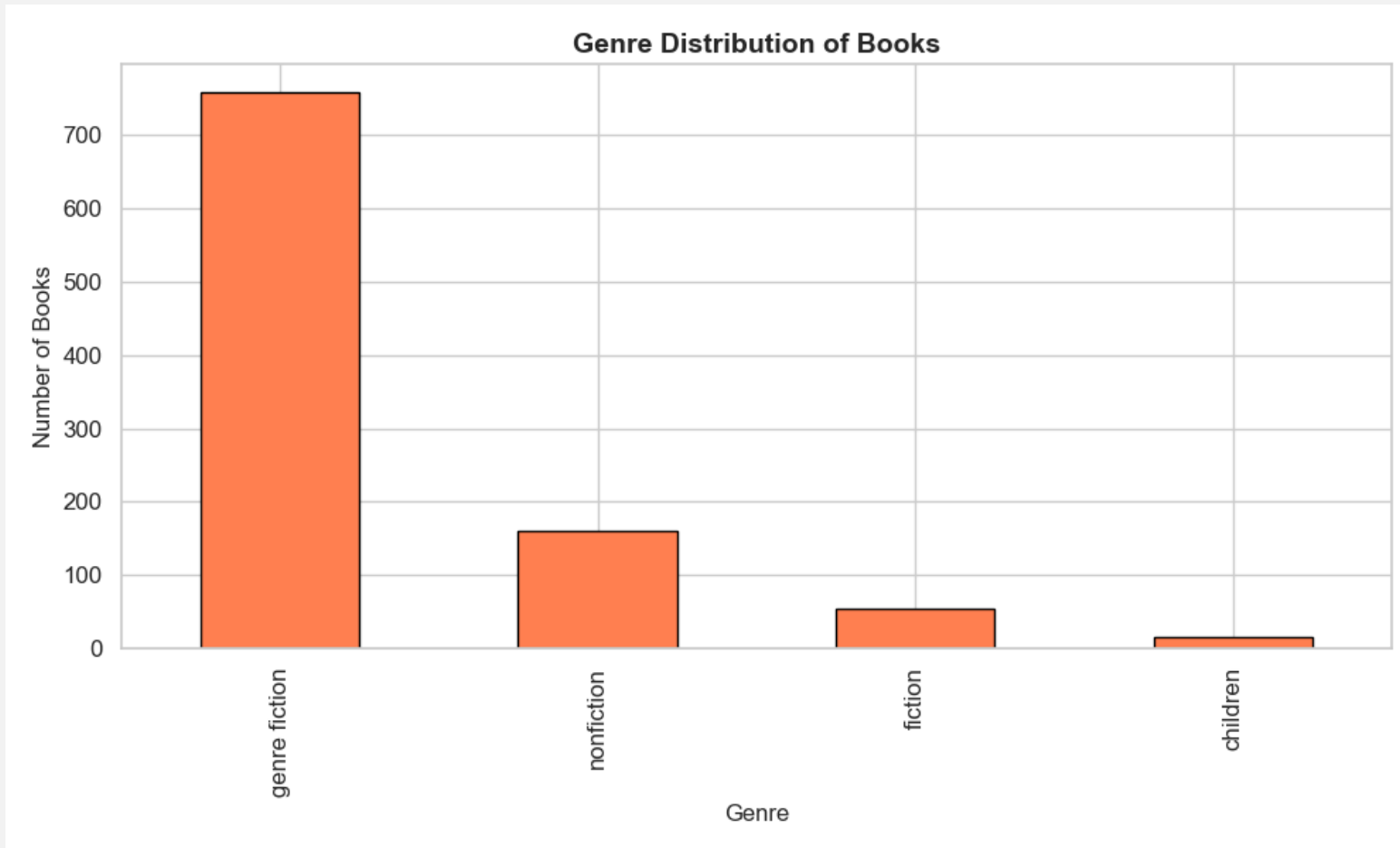
1. Publishing Year Distribution : -

- Most books are published **after 1900**, with a strong concentration between **2000 and 2020**.
- Indicates a focus on contemporary literature trends.



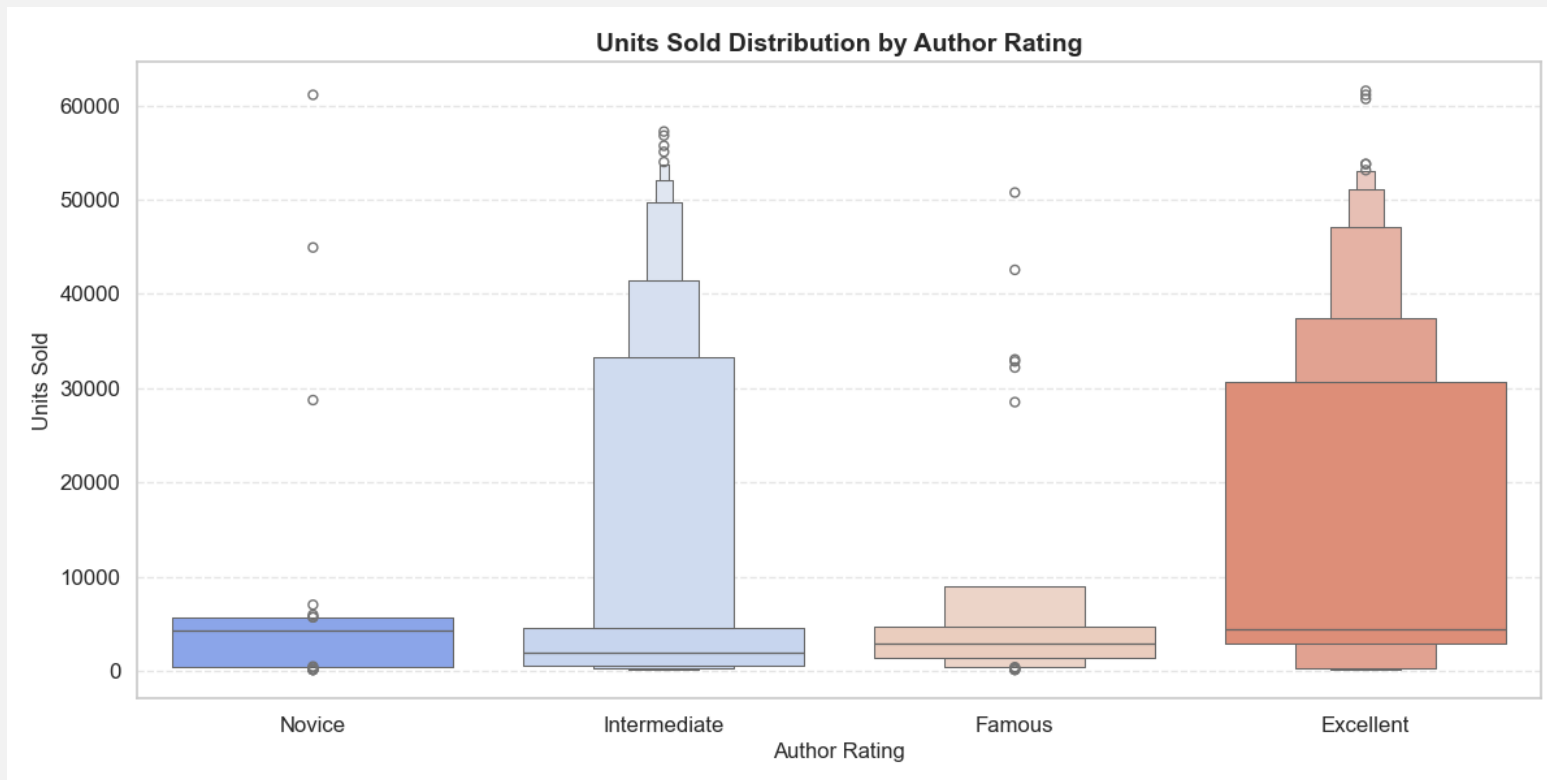
2. Genre Popularity

- Genres like **fiction** dominate the dataset.
- Variability in reader engagement and ratings count across genres.



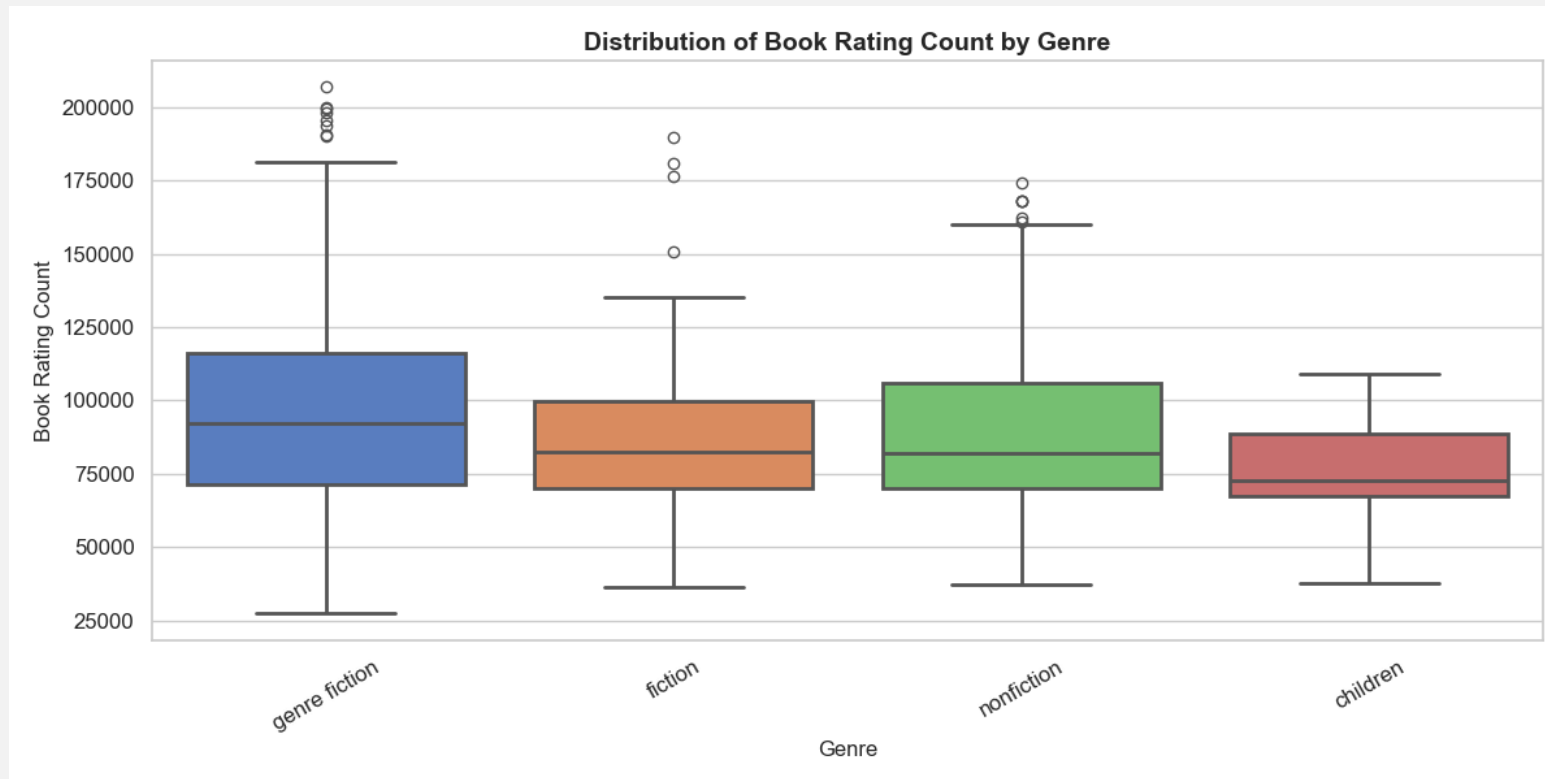
3. Units Sold Distribution by Author Ratings :

- The majority of units sold are concentrated among authors with "Intermediate" and "Excellent" ratings.
- Lesser-represented ratings like "Novice" and "Famous" indicate a skewed distribution of sales.



4. Book Ratings Count by Genre

- Box plots reveal wide variability in ratings count within genres.
- Some genres have standout books with exceptionally high ratings.



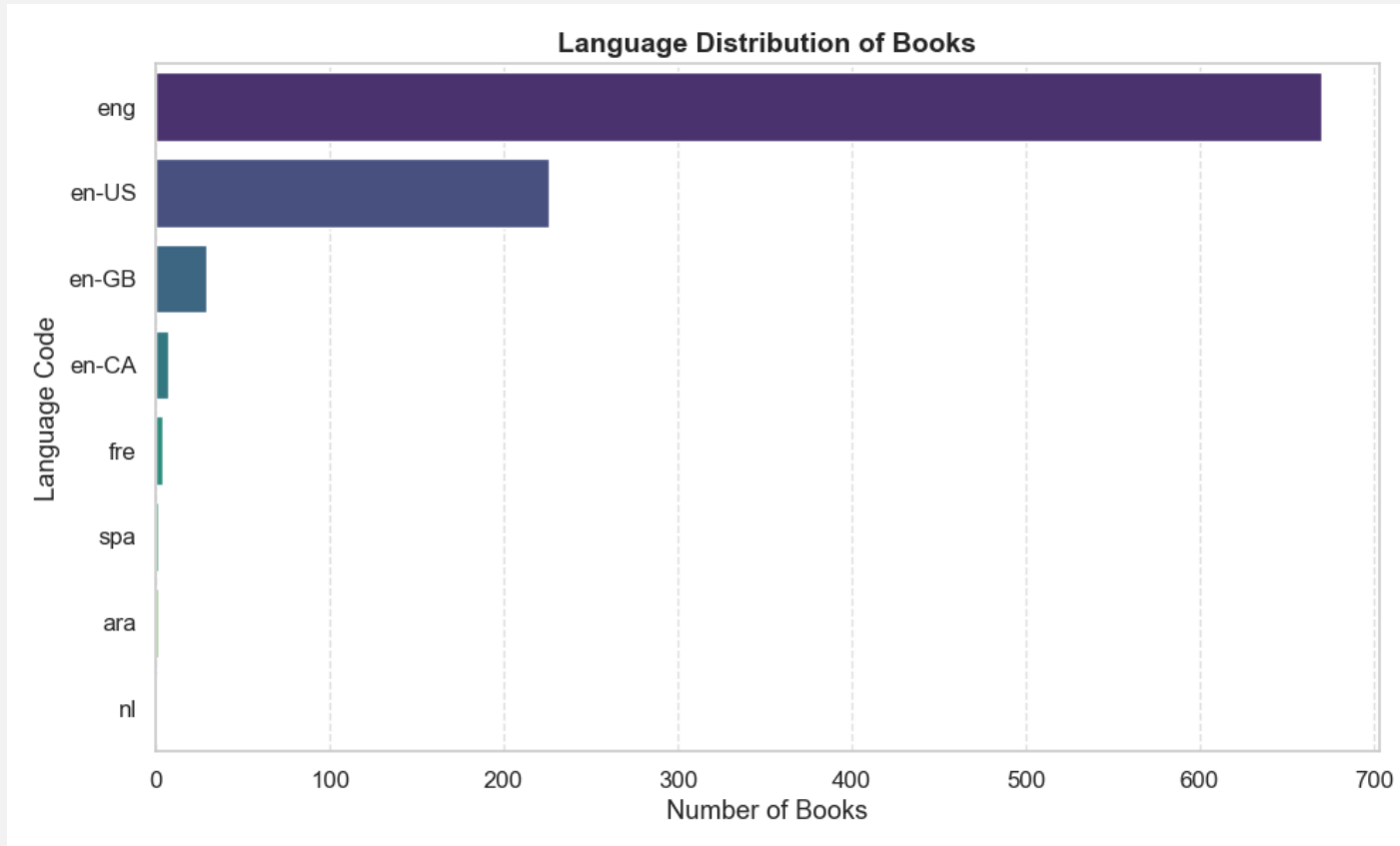
5. Sales Performance :

- Scatter plots indicate a trend where **lower sale prices tend to yield higher units sold**, with notable exceptions.
- Top authors dominate total gross sales.
- Units sold have increased over time, especially in recent decades.



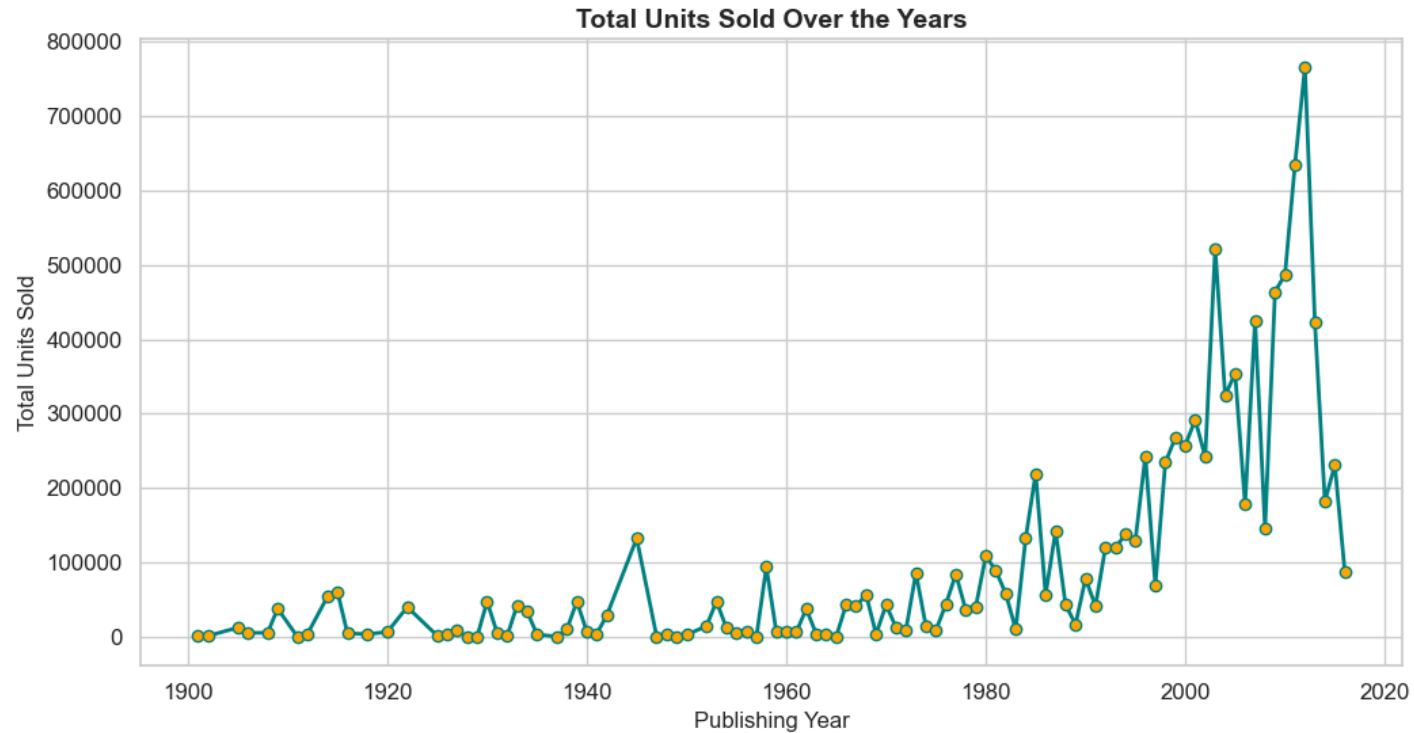
6. Language Distribution :

- The majority of books are published in the "eng" language, showing a dominant presence.
- Lesser-represented languages like "en-CA" and "fre" indicate a concentrated market distribution.



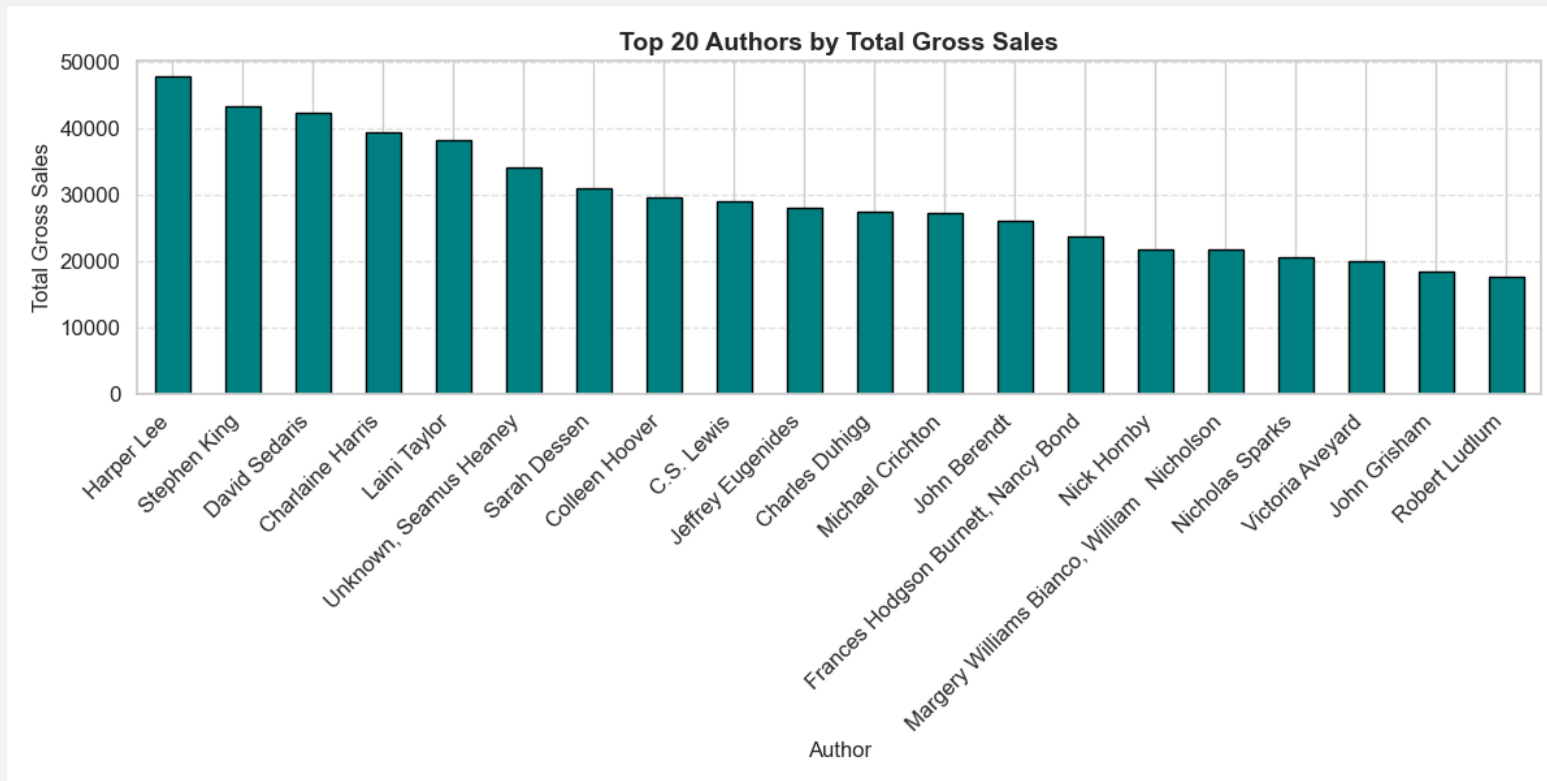
7. Sales Trend (Total Units Sold Over The Years) :

- Total units sold remained relatively low and stable from 1900 to the 1980s.
- A significant increase in sales occurred after 1980, peaking around 2020.



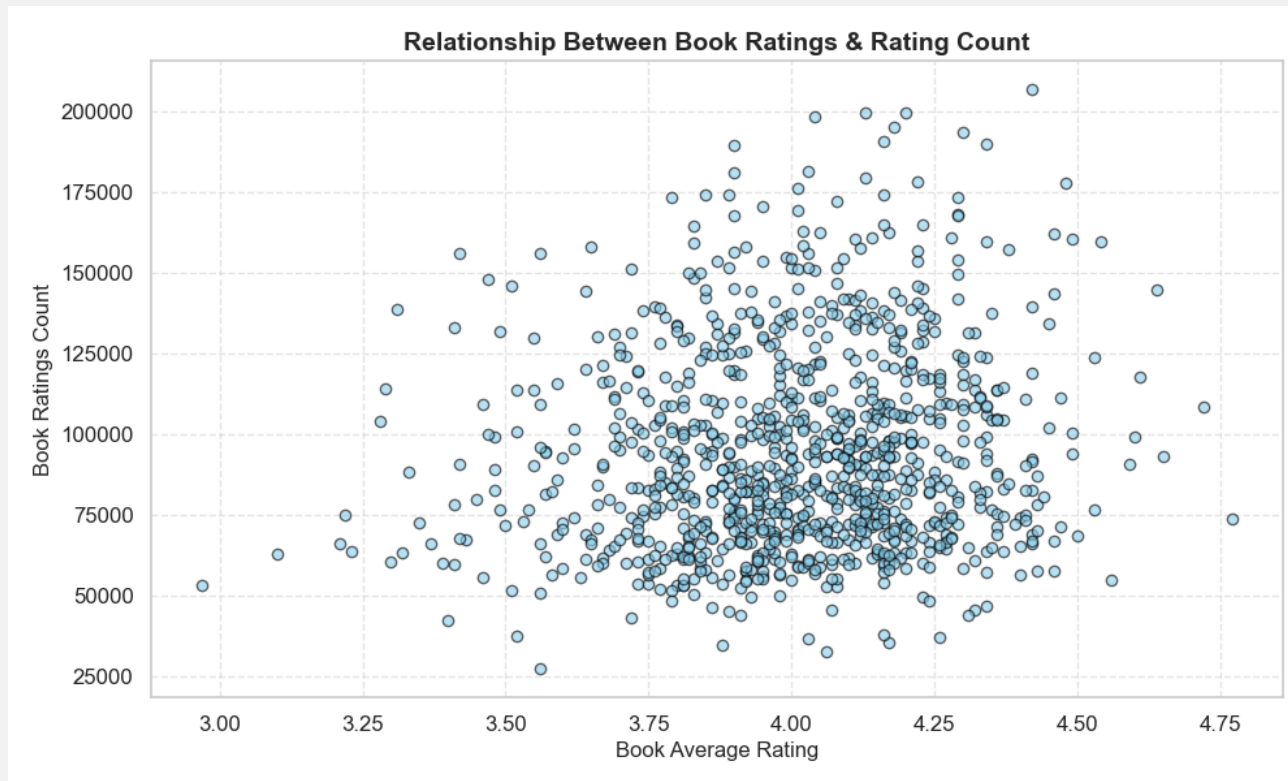
8. Top 20 Authors by Total Gross Sales Analysis :

- Harper Lee and Stephen King lead with the highest total gross sales among the top 20 authors
- Sales drop significantly for authors like Margery Williams and Robert Ludlum towards the lower end of the list



9. Relationship Between Book Ratings & Rating Count :

- The scatter plot shows a wide range of book rating counts, with most books having ratings between 3.5 and 4.5 and counts varying significantly from 25,000 to 200,000.
- There appears to be no strong linear correlation between average book ratings and the number of ratings, indicating that higher ratings do not necessarily correspond to higher rating counts.



Conclusion:

The analysis reveals a shift towards contemporary literature post-1900, with a peak between 2000 and 2020, and fiction dominating genres with varied reader engagement. Sales are concentrated among "Intermediate" and "Excellent" rated authors, with higher units sold at lower prices, driven by top authors like Harper Lee and Stephen King, and a significant sales increase since the 1980s, predominantly in English.