

Project Report

TASK 2: ANALYZING PUBLIC LIBRARY E-BOOK LENDING TRENDS

Introduction:

In this project, we analyzed e-book lending trends using a public dataset from Kaggle. The dataset contains information about free e-books, including their titles, authors, genres, and user ratings. We cleaned and analyzed the data to identify popular genres and download trends over time, and presented our findings in a clear and visually engaging manner.

Data Cleaning:

We started by loading the dataset and handling missing values in the Book-Author column. We then extracted genre information from the Book-Title column and dropped rows with missing values in the Year-Of-Publication column. Finally, we converted the Year-Of-Publication column to datetime format.

Data Analysis:

We analyzed the data to identify popular genres and download trends over time. We found that the dataset contained 40,823 unique genres, with the most common genre being "Unknown" (195,886 books). We also found that there were significantly more non-fiction books (271,345) than fiction books (15).

We then focused on analyzing the fiction books in the dataset. We found that the most common fiction genres were Fiction (15 books), Romance (16 books), and Mystery (10 books). We also analyzed the download trends over time and found that the average Book-Rating was highest in the early 1900s and decreased over time.

Visualizations:

We created several visualizations to help illustrate our findings. We used a bar chart to show the top 10 genres in the dataset, a pie chart to compare the proportion of fiction and non-fiction books, and a scatter plot to show the relationship between the Year-Of-Publication and Book-Rating.

Conclusion:

In this project, we explored e-book lending trends using a public dataset. We found that the dataset contained significantly more non-fiction books than fiction books, and that the most common fiction genres were Fiction, Romance, and Mystery. We also found that the average Book-Rating was highest in the early 1900s and decreased over time.

Limitations and Future Work:

This project has several limitations, including the lack of information about borrowing durations and user demographics. In the future, we could explore these factors to gain a more complete understanding of e-book lending trends.

Summary:

- Popular genres: The most common genres in the dataset were "Unknown" (195,886 books), Fiction (15 books), and Romance (16 books).
- Download trends: The average Book-Rating was highest in the early 1900s and decreased over time.
- Limitations: The dataset did not include information about borrowing durations or user demographics.
- Future work: We could explore these factors to gain a more complete understanding of e-book lending trends.

Recommendations:

Based on our findings, we recommend that libraries and publishers focus on promoting fiction and romance books, as these genres were the most popular in the dataset. We also recommend that libraries and publishers consider offering more e-books with high average ratings, as these books are likely to be more popular with users.