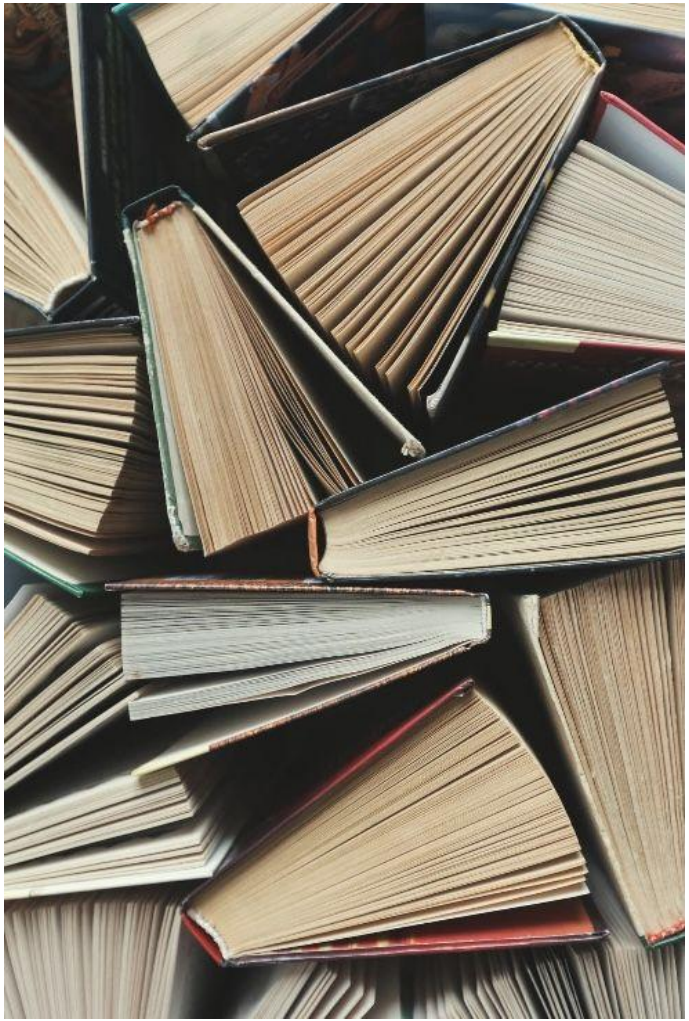
Book Recommendation System

Viktoria Gluhovskaya



Introduction

This project focuses on building a Book Recommendation System.

The goal is to apply data analysis and machine learning techniques to recommend books based on their similarity in title, author, and subject.

Through this project, I aim to demonstrate how recommendation systems work and how they can be implemented in practice

Data Collection

Books to Scrape (scraping)

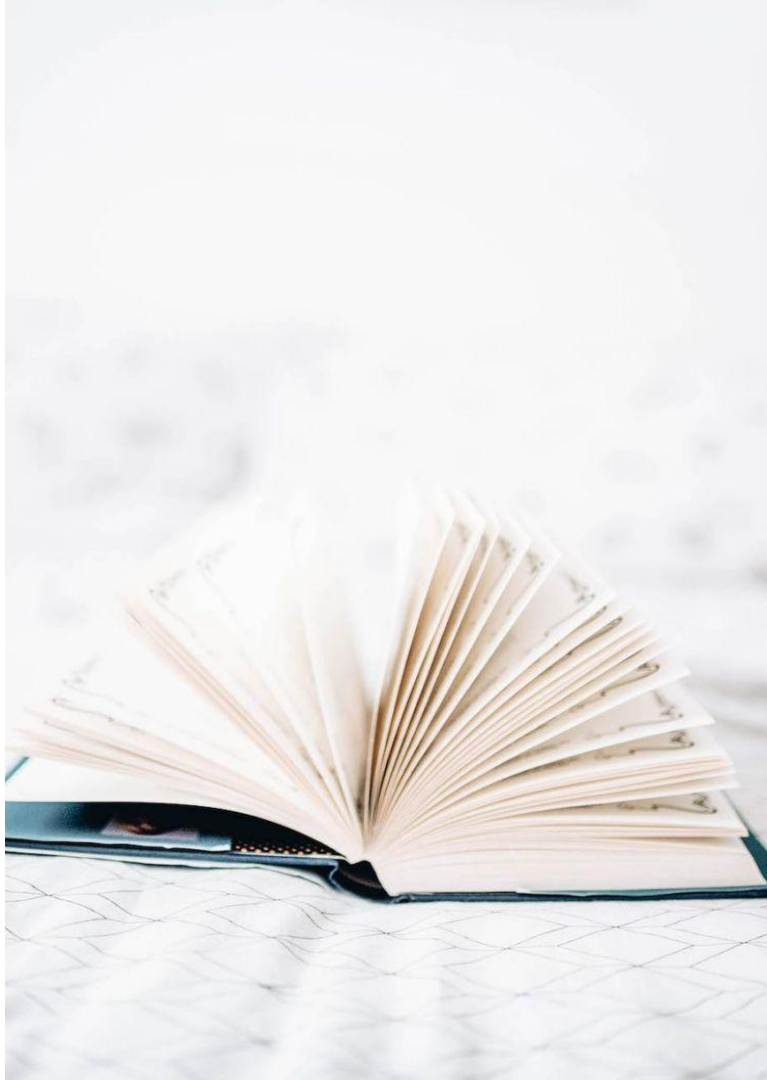
Open Library API(enrichment)

Final Dataset

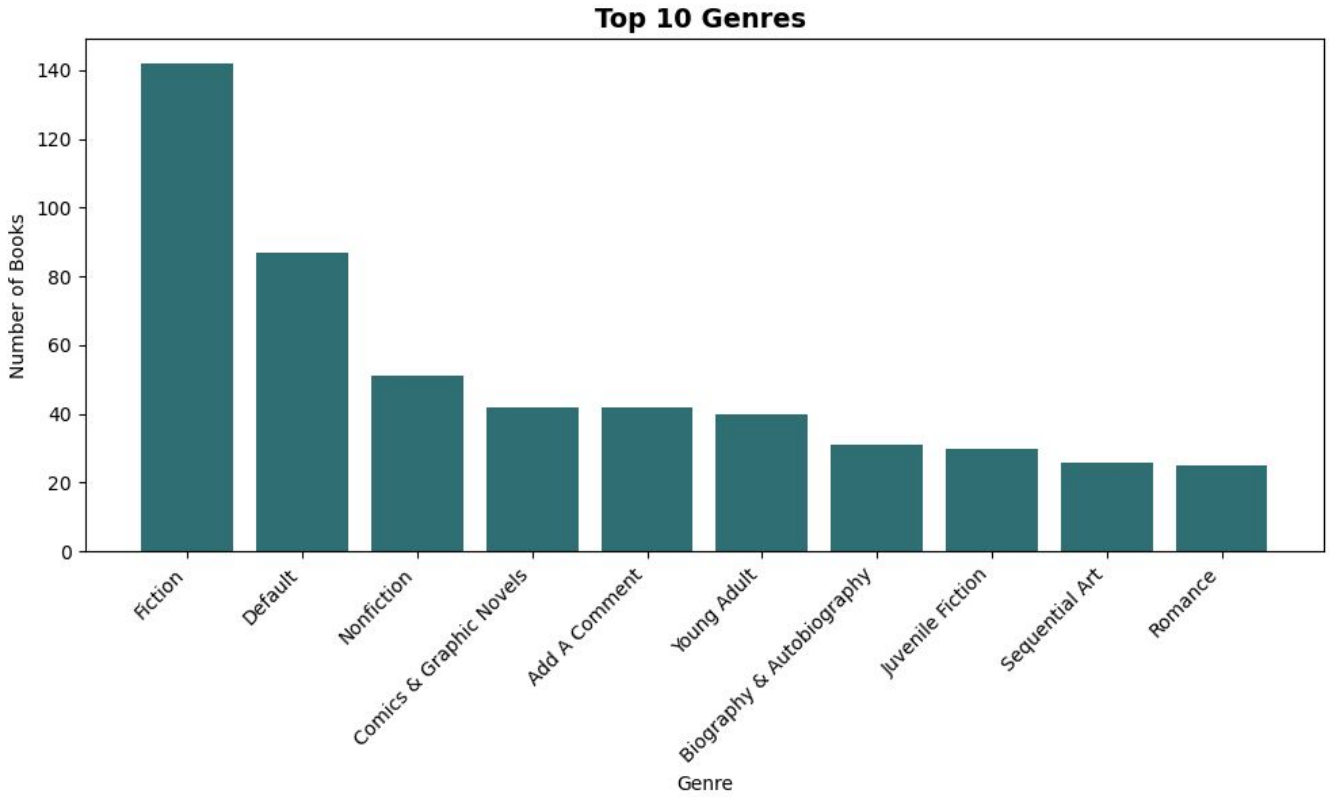
Dataset Overview

- Size: 999 books, 785 unique authors, 115 unique genres
- Unknown authors: 7.41% of the dataset

Insight: The dataset is diverse enough in terms of authors and genres, which makes it suitable for building a recommendation system



Exploratory Data Analysis (EDA)



Data Cleaning

Removed duplicates and null values

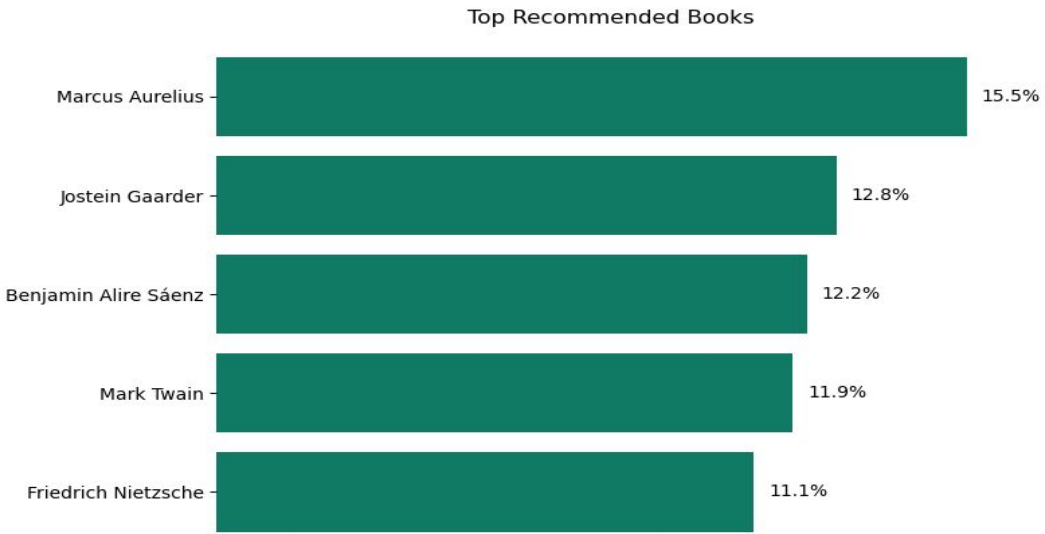
Standardized column names

Fixed invalid values (missing authors, inconsistent genres)

Cleaned & prepared dataset for analysis

Recommendation Methodology

- Convert book descriptions into vectors (TF-IDF)
- Measure similarity between books (Cosine Similarity)
- Recommend the most similar titles



Conclusion:

This comparison shows why I ultimately selected K-Means. It provided the best trade-off between:

- **Clarity** - clusters are visible and interpretable
- **Stability** - less sensitive to noise compared to DBSCAN in this dataset
- **Practical usefulness** - all books are included in the clustering, which is essential for a recommendation system

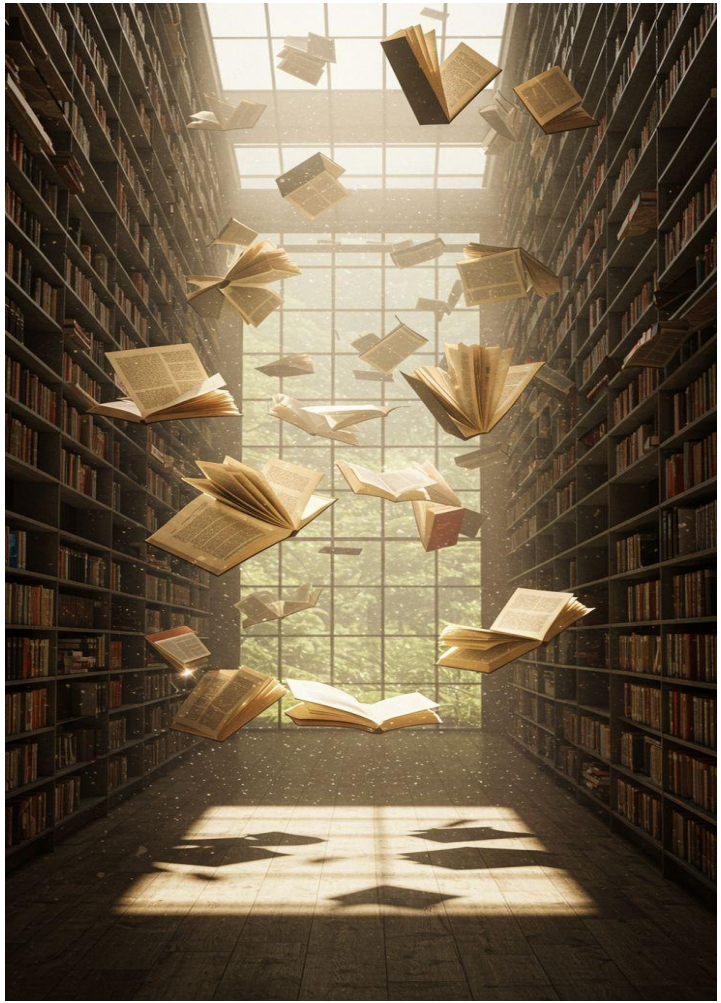
Therefore, **K-Means** was chosen as the **foundation** of my recommendation system.

Tools:

- **Trello** - project management
- **Pinterest** - design of slides
- **Python** - data processing & clustering
- **Streamlit** - interactive app
- **Google Slides** - presentation

Sources:

- **Books to Scrape** - dataset scraping
- **Open Library API** - enrichment with authors & genres



Thank you!