

The background features two large, overlapping, curved lines. One line is a light green color and the other is a light blue color. They are positioned in the top-left and bottom-right corners of the slide, framing the central text.

AudioRama

**A content-based audiobook recommendation engine
for Audible and audiobook listeners**

Audible's recommendation system is not useful to me as a subscriber

- Audible has an extensive library but the in-house recommendations are heavily focused on genre, ratings, and authors.
- Users might want narrator-centric or author-centric book recommendations that are both popular, highly rated, and similar.

Data Wrangling

Scrape 80GB data from Audible and LibriVox

Data Preprocessing

Vectorize all relevant features and metadata for model building

Model Building

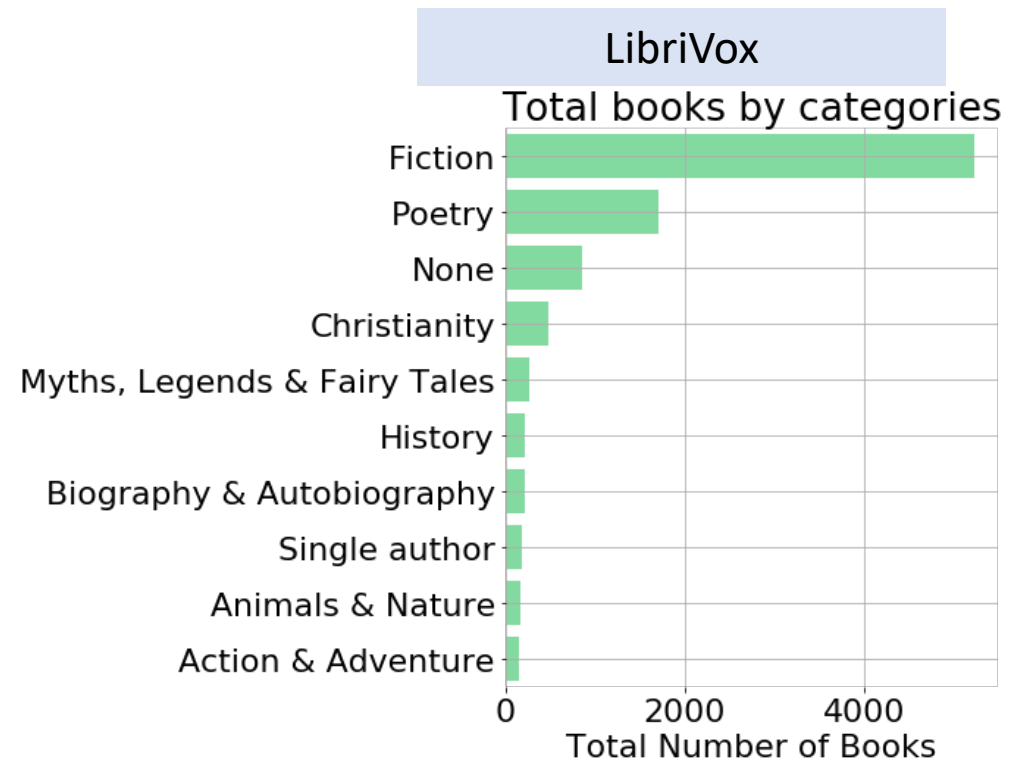
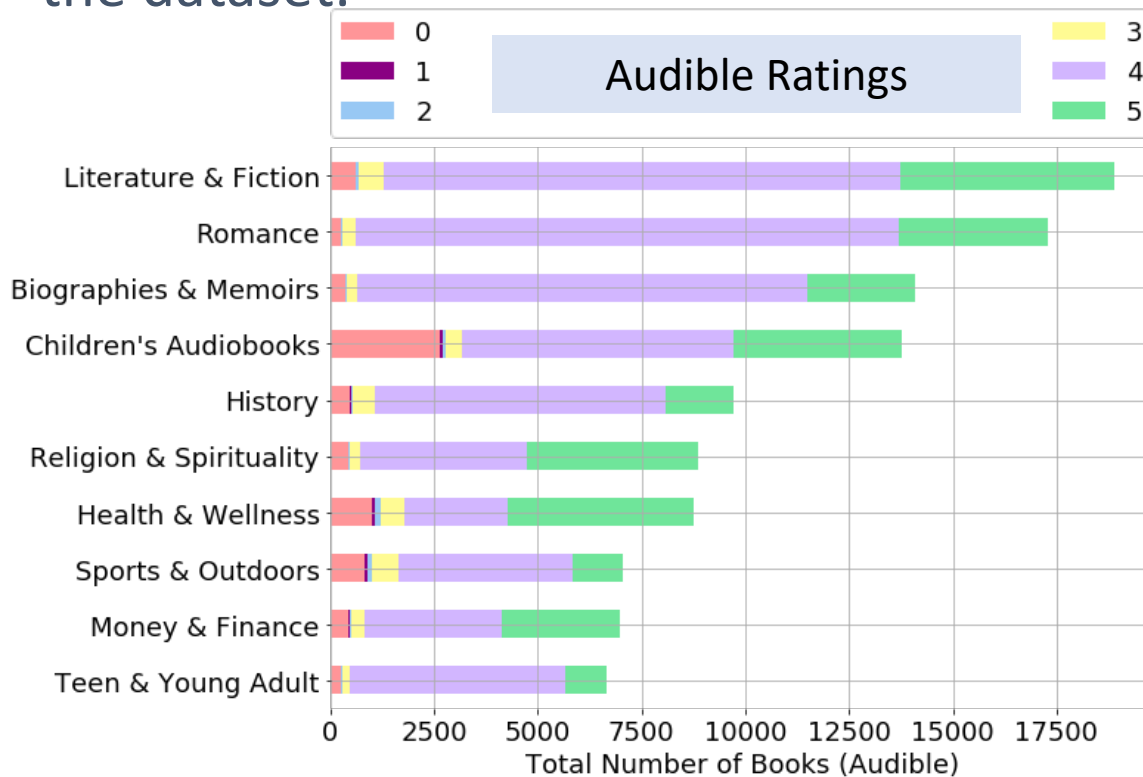
Use BERT and a ratings classifier to create an author-centric or narrator-centric recommendation

Model Deployment

Deploy app on AWS using Flask and other tools

Scrapped 80GB of synopsis and metadata from Audible and LibriVox (>100k books)

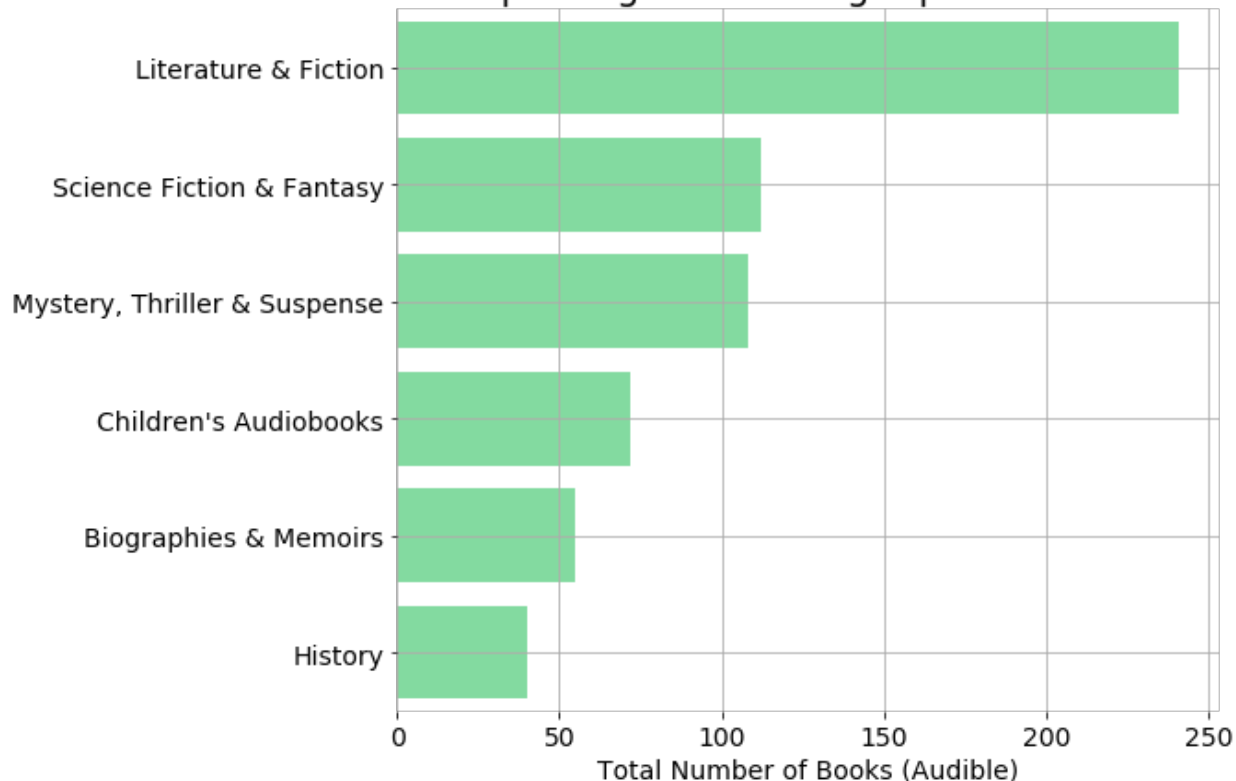
- A significant number of books in the dataset are rated because I scraped top 1200 books from each sub-category.
- LibriVox is a free crowdsourced platform without any ratings and contributes to 12% of the dataset.



Narrator-centric recommendations will differ from author-centric ones

- There is a significant difference in genres and number of books contributed by top authors and top narrators.
- The top 5 narrators contribute to 100% more topics than authors.

Top categories among top 5 narrators



Top categories among top 5 authors

