GOODREADS

EDA & BASELINE MODELING

BY QUINCI BIRKER

Introduction

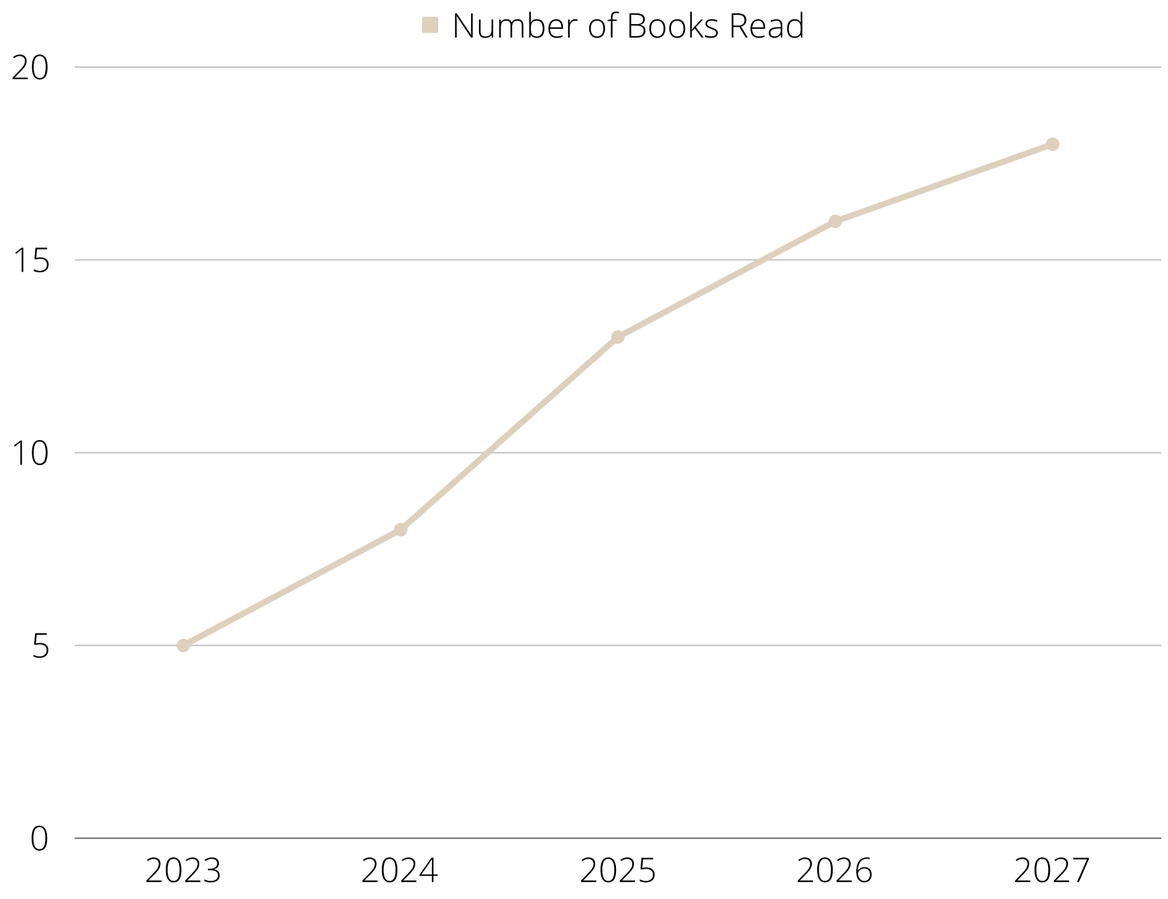
- Goodreads is one of the most popular book catalogue websites in America.

- Updates:
  - EDA
  - Cleaned the data
  - Preprocessed
  - Baseline Model

Problem & Solution

- Problem: If you pick up a book you don't like, it can put you off reading for a long time.

- Solution: Personalized book recommender system



Project Impact

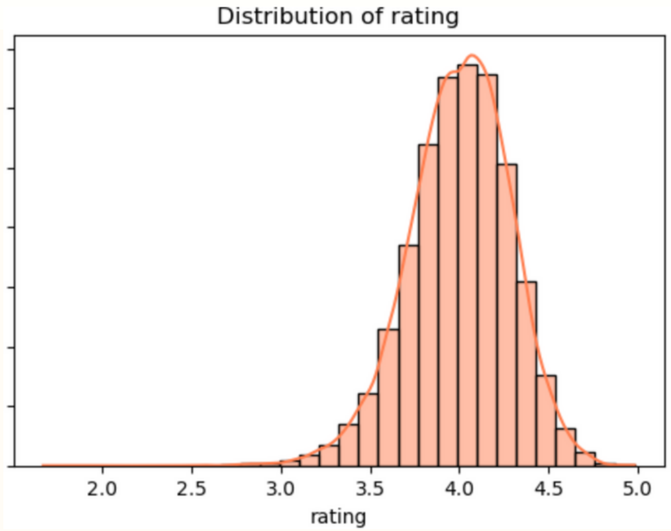
- Overall goal: Make reading more popular
- Target: Occasional and avid readers
- Impact: Increase revenue for bookstores

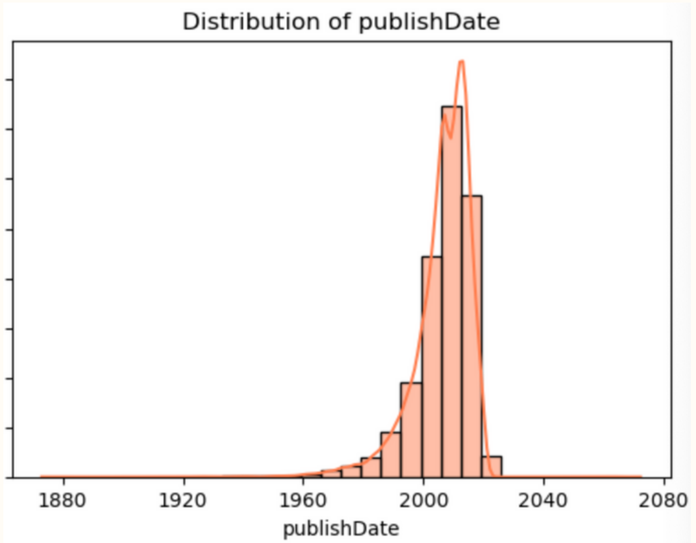
Dataset Overview w/ Preprocessing

- 52,400 rows → 41,300 rows
- 25 columns → 45 columns

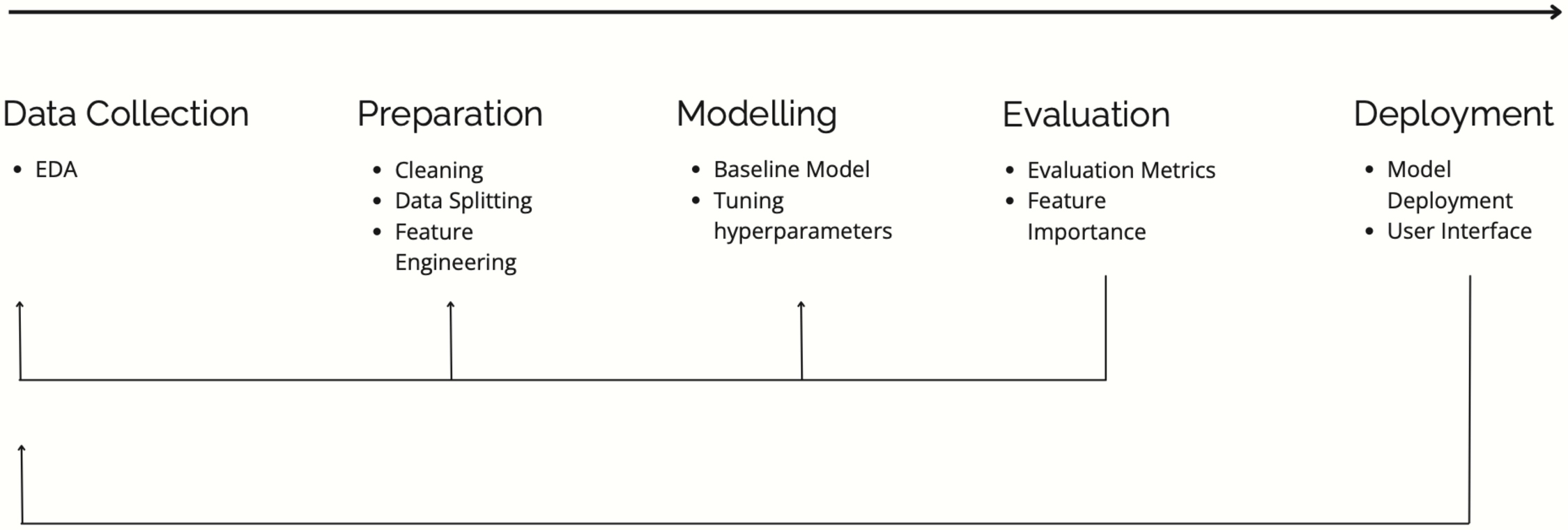
- Target variable: Rating Score (1 or 0)

- Independent variables:
  - Genres
  - Pages
  - Publish Date (Year & Month)
  - Number of Ratings
  - Series





Process Overview



Thank-you!