Synthetic Book Price Predictor

This app demonstrates a simple machine learning model built on a synthetic dataset of 5,000 books. Use the sidebar to explore the data and predict prices.

Project Presentation

Synthetic Book Price Predictor

Demonstrating an end-to-end pipeline: scraping → processing → ML → web app.

- Context: no Internet access → synthetic dataset of an online bookstore.
- Scraping: extracted from local HTML pages using BeautifulSoup.
- Data prep: cleaning, normalization, categorical encoding.
- ML: simple regression model to estimate book prices.
- App: demo UI for exploration and predictions.

Date: Sep 07, 2025

1. Context & Pipeline

Goal

Show a full workflow from data collection to delivery via a web app.

Synthetic dataset

• Generated to mimic a bookstore (titles, authors, categories, prices, ratings, stock).

• Exported as local HTML pages to simulate a live website.

Scraping & preparation

• BeautifulSoup used to extract key fields.

• Cleaning: formats, missing values, and categorical encoding.

	id	title	category	rating	price	availability
0	1	Curious Legend 1	Historical Fiction	2	30.78	8
1	2	Shadowy Sky 2	Womens Fiction	4	38.26	3
2	3	Golden Adventure 3	Autobiography	3	26.2	3
3	4	Mysterious Legend 4	Science Fiction	2	34.01	13
4	5	Majestic Whisper 5	Childrens	2	18.87	14
5	6	Majestic Empire 6	Sports and Games	4	25.94	17
6	7	Incredible Quest 7	Philosophy	1	23.1	14
7	8	Whispering Legend 8	Autobiography	2	23.23	8
8	9	Silent Sky 9	Young Adult	1	21.45	5
9	10	Hidden Legacy 10	Science	2	31.25	4

Preview of structured data after scraping.

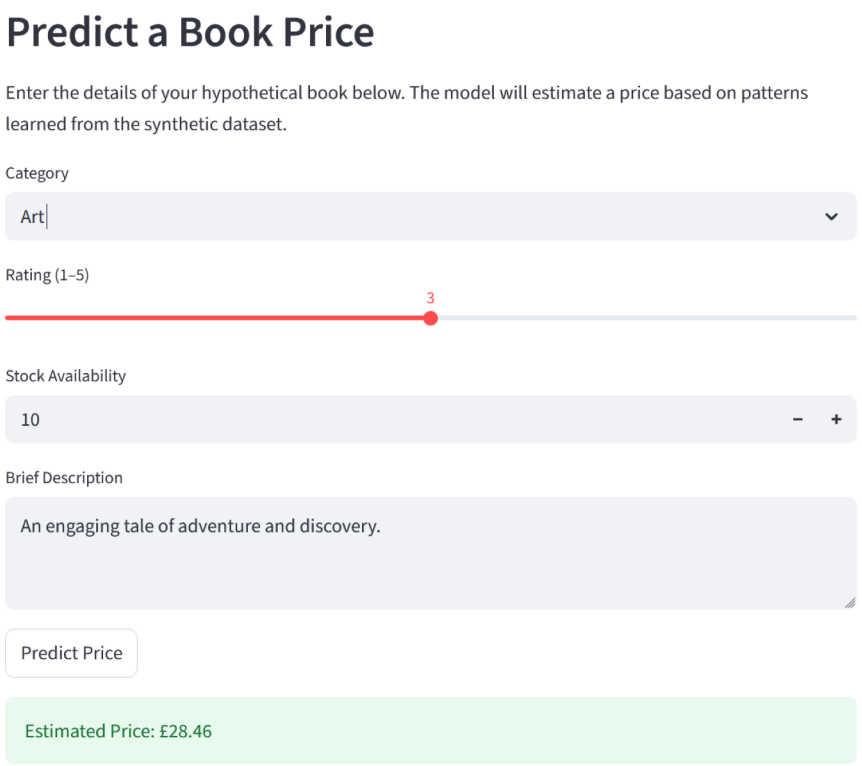
2. Modeling

Task: predict a book’s price from its features.

Features used (examples):

• Category, rating (1–5), availability, short blurb, etc.

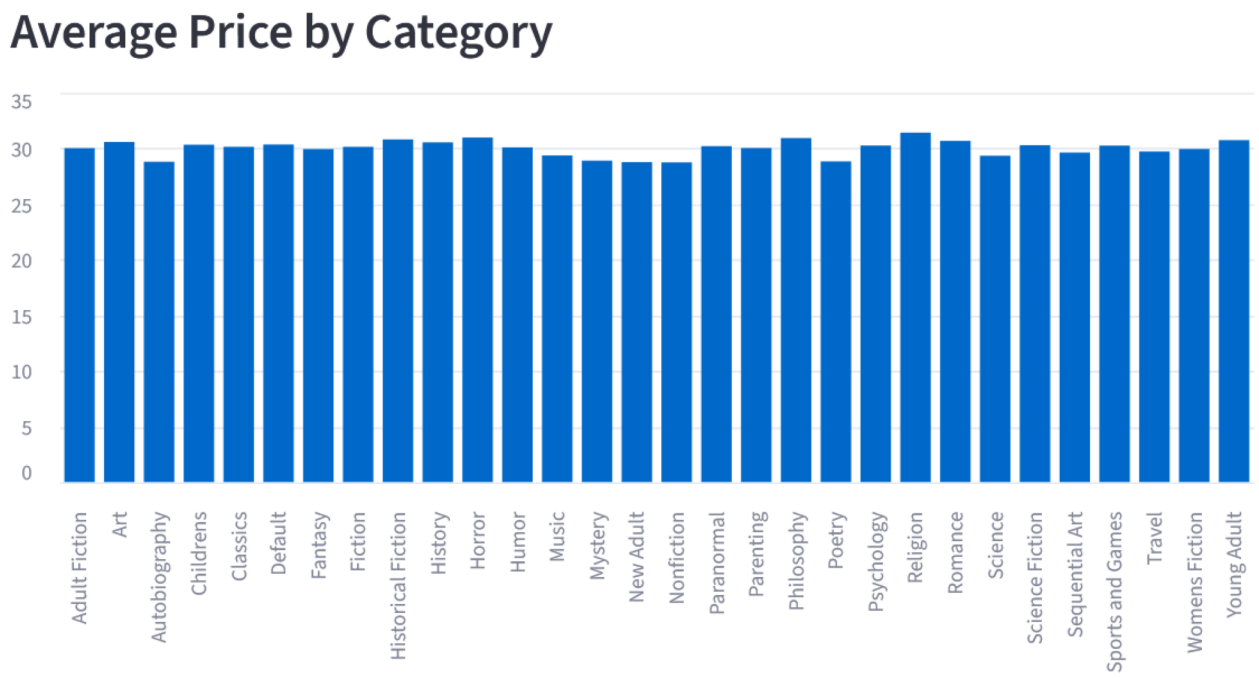
• Simple regression model for demonstration purposes.



Prediction UI: input features → estimated price.

3. Exploratory Analysis

Example metric: average price by category.



Visualization from the synthetic dataset.

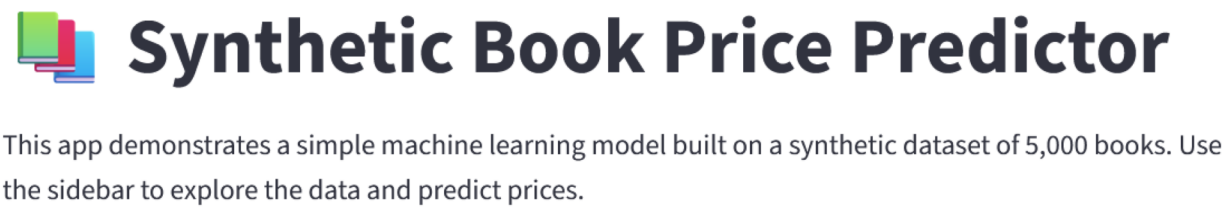
Insights (examples):

• Fairly uniform prices across categories (simulated).

• A safe playground to test preprocessing and feature pipelines.

4. Application & Deployment

A simple web app exposes data exploration and prediction.



Header of the demo application.

Key points for production:

- Data lineage and versioning.
- Feature validation and model evaluation (CV, metrics).
- Post-deployment monitoring (drift, performance).
- CI/CD for both the app and the model.

Next steps:

- Enrich the dataset and try additional algorithms.
- Add model explainability (SHAP, PDP) to the UI.
- Integrate data testing/validation (e.g., Great Expectations).