Analyzing Amazon's Bestselling Books for Self-Publishing Insights

Soumyalatha Bojaraj
soumyalathabojaraj@gmail.com

We use tech to connect human potential and opportunity with dignity & humility

PROJECT OBJECTIVE

GOAL:

What kind of books sell well on Amazon's self publishing platform?

KEY QUESTIONS:

- What genres sell best on amazon?
- What is the average page count for a bestselling book?
- What is the optimal price range?
- How are the user ratings and reviews distributed across the successful genres?
- Do high user rating or higher number of reviews have an impact on the price?

What did I do?

The Process in simple words

**Fetched
Data from
Kaggle.com**

**Cleaned
the data**

**Gained
insights**

**Loaded
into
Python**

**Explored
with charts
(EDA)**

**Documented
findings &
conclusions**

Presentation

DATA PRE-PROCESSING

INITIAL OBSERVATIONS

- Shape of data: **4868** rows and **10** columns

- Missing values in 2 columns: **Brand** and **Number of pages**.

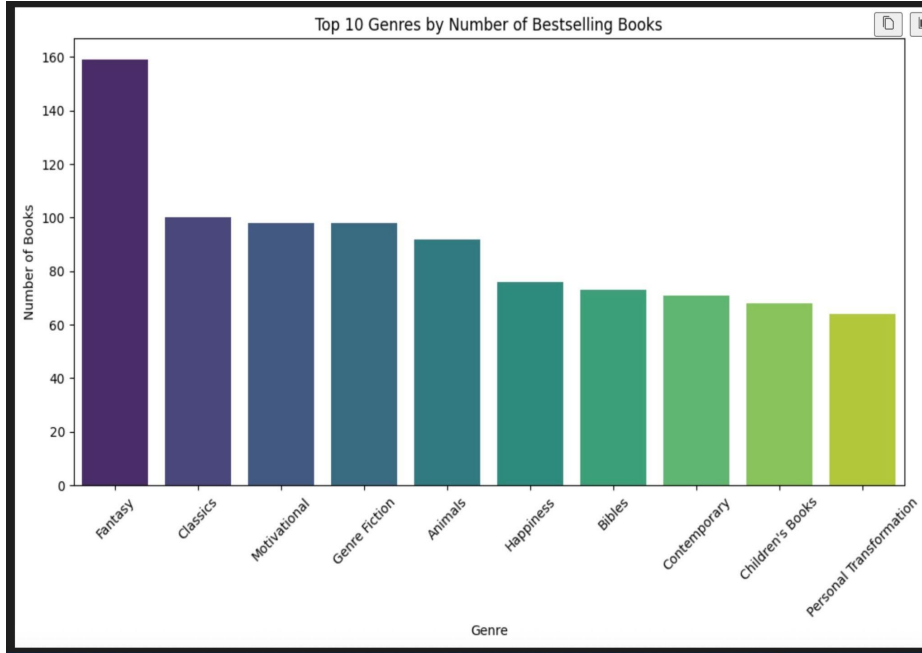- Outliers in the column **Number of pages**.

DATA CLEANING

- Dropped the less necessary column: **Brand**

- Dropped the rows with missing **number of pages**

- Removed outliers
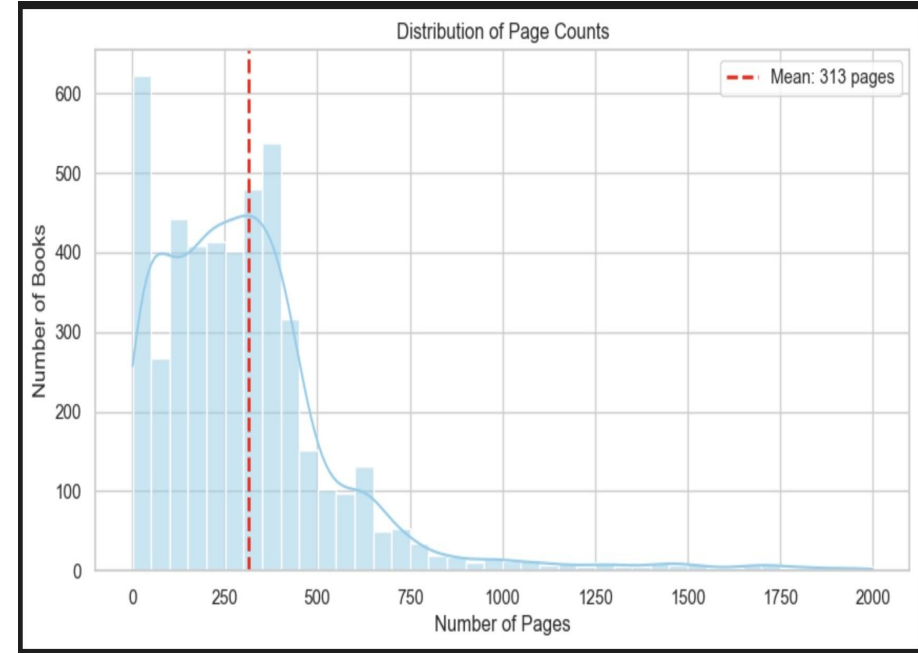
- Renamed column '**Reviews**' to '**Ratings**'

CLEANED DATA

- Shape of data: **4710** rows and **9** columns

- The dataset includes information about: **Title, Rank, Ratings, Review count, Price, Genre, Manufacturer, Author** and **Number of pages**.
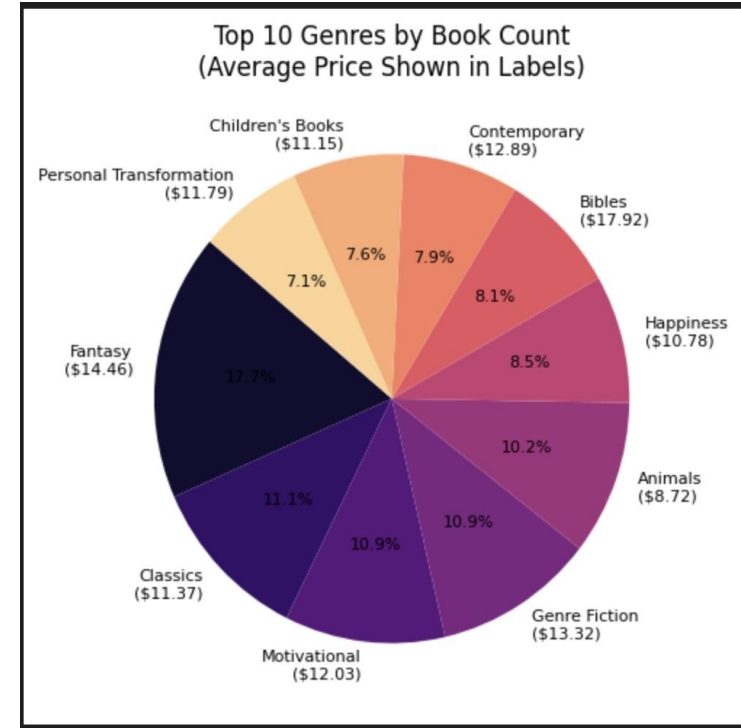
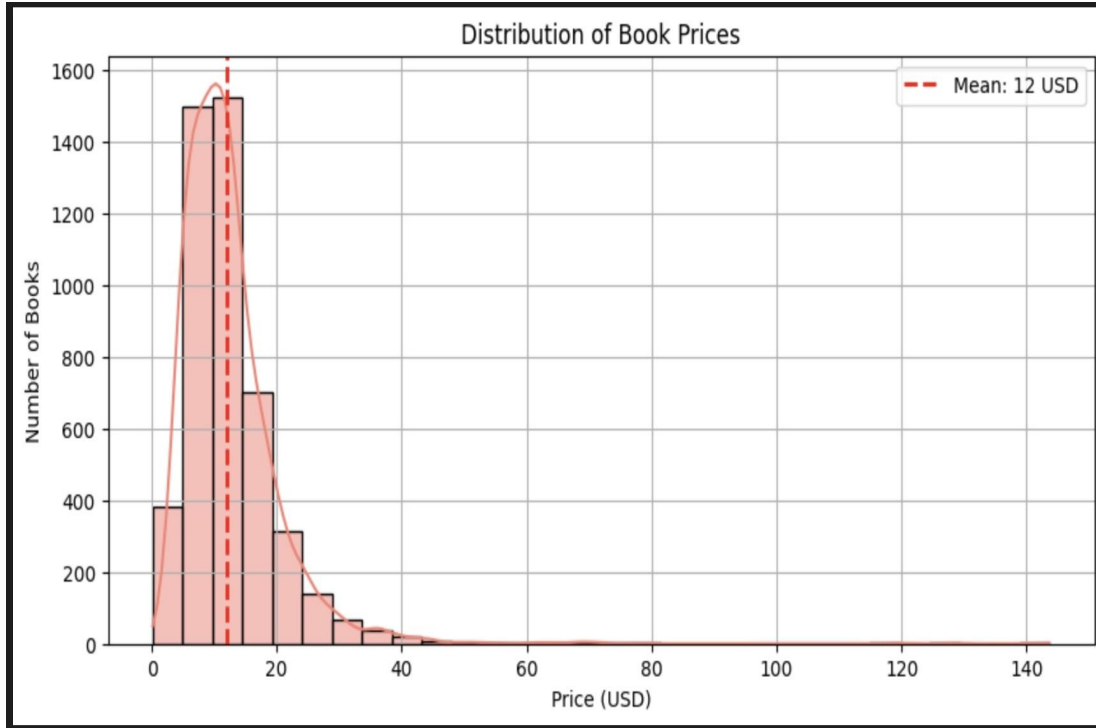# Exploring Data (EDA) with Visualizations
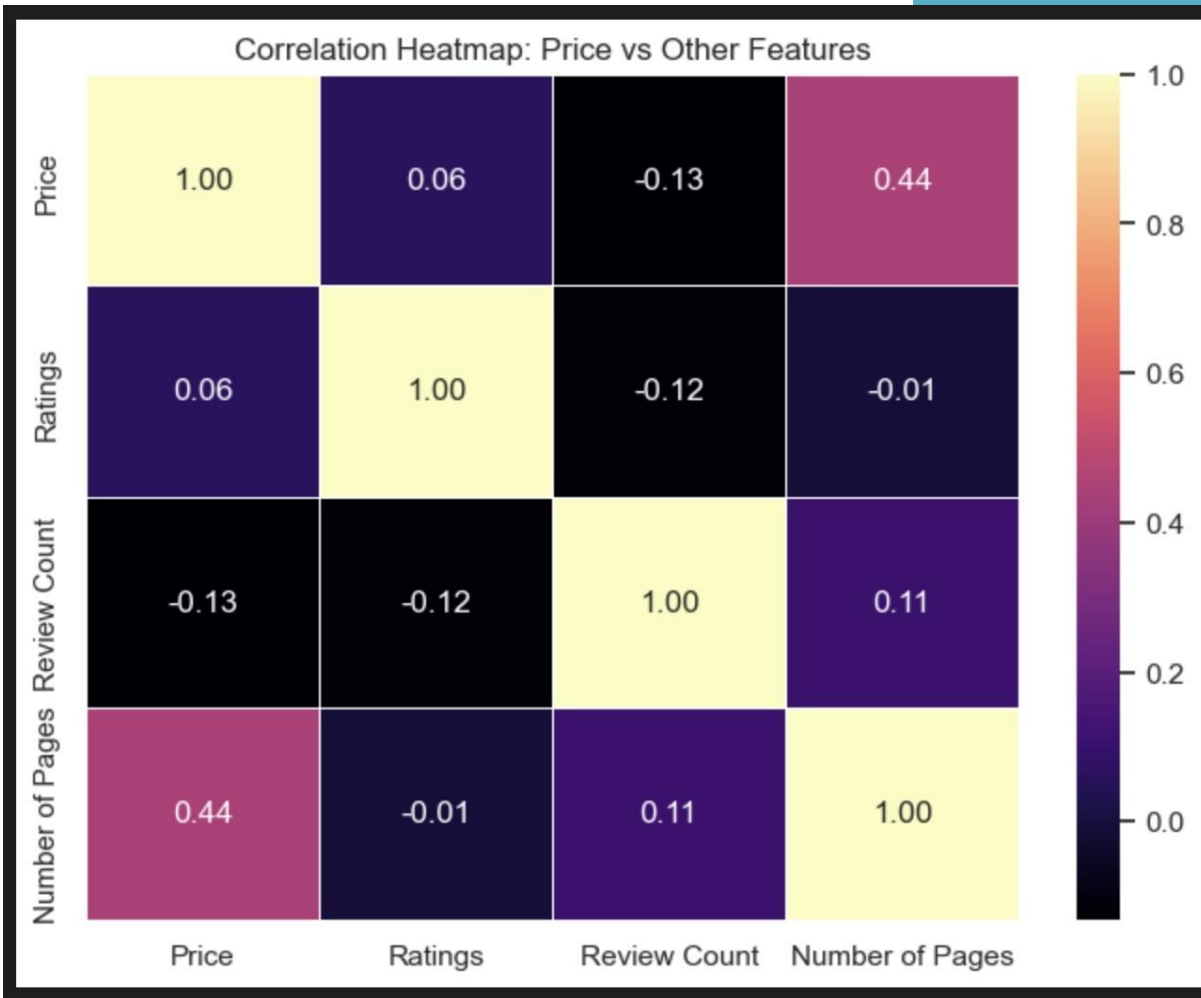
Top 10 Genres by Number of Bestselling Books

Distribution of Page Counts

Genres

Number of pages

Average price analysis

Distribution of Book Prices

Top 10 Genres by Book Count
(Average Price Shown in Labels)

Correlation Heatmap: Price vs Other Features

Relationship of Price, Ratings, Review Count and Number of Pages

- Price seems to increase a little when number of pages increase, but it is only moderate.

- The Ratings and Review count seem to have weak relation with Price

KEY INSIGHTS

- What genres sell best on Amazon? ➡ **Fantasy, Classics, and Motivational content**

- What is the average page count for a bestselling book? ➡ **313 pages.**

- What is the optimal price range? ➡ **$10 and $15**

- How are user ratings and the number of reviews distributed among the most successful genres? ➡ **Higher ratings(～4.8) and more reviews.**

- Do high ratings or review counts impact the price of a book? ➡ **No strong correlation (High-rated or highly-reviewed books exist at all price points.)**

ROADBLOCKS AND RETHINKING

Finding suitable dataset was tough.

- I decided on the topic first.

- Narrowed down available datasets.

- Picked the most recent dataset (2024).

Narrowing down the right questions to ask took time

- Wrote down all questions I could think.

- Prioritized what could be more relevant.

- Chose top 5 questions to analyse.

Coding part for visualizations was challenging.

- Searched for pieces of code to create visualisations.
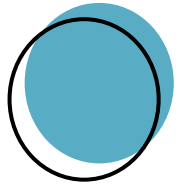
- Reworked on what exactly I wanted to include

LIMITS and LOOSE ENDS

Limitations in the Dataset

- Publishing type is unknown — no information to distinguish between self-published and traditionally published books.

- Book format details are missing — hardcover/paperback, dimensions of a book or printing specifications- impacts price of a book

- Publication year is not included — analysing trends over time was not possible.
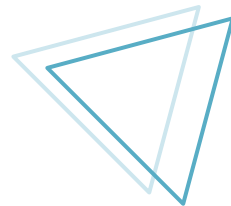
WHAT COULD BE EXPLORED?

Future

★ Author column can be explored to see if Author popularity has impact on sales.

★ Manufacturers can be looked for pricing insights/comparison.

Publishing dreams?

Try these data-backed recommendations

- **Focus on genres like Fantasy, Classics or Motivational content.**

- **Keep the pagecount between 300-320.**

- **Set price range between $10 to $15.**

- **For honest feedback- Consider including book inserts or reach out to reader communities.**

Thank You!