

# Book Reviews Exploratory Dataset Analysis

## 1. Relationship between Price of the Book and Its Ratings

- Dropped unnecessary columns and removed null values to ensure data quality

### Observations:

#### 1. Scatter Plot Analysis:

- Most low-priced books receive high ratings.
- Mid-priced books show a more varied distribution of ratings, with some achieving very high ratings.
- Some books are significantly costlier, potentially skewing the data.

#### 2. Statistical Analysis:

- **Negative Correlation:** A correlation coefficient (r-value) of -0.4769 suggests a moderate negative relationship between book prices and ratings. Higher-priced books generally have lower ratings, but this is not a strong rule.
- **Significance:** A p-value of 1.0247e-06 indicates that the negative correlation is statistically significant, ruling out random chance.
- **Standard Error:** A standard error of 0.002652 shows that the slope estimate is precise, indicating consistency in the negative relationship across the data.

### Conclusion:

While there is a statistically significant negative correlation between book prices and ratings, it is moderate in strength. Other factors like content quality, genre, and author popularity also play significant roles in determining ratings.



## 2. Genres with the Highest Ratings

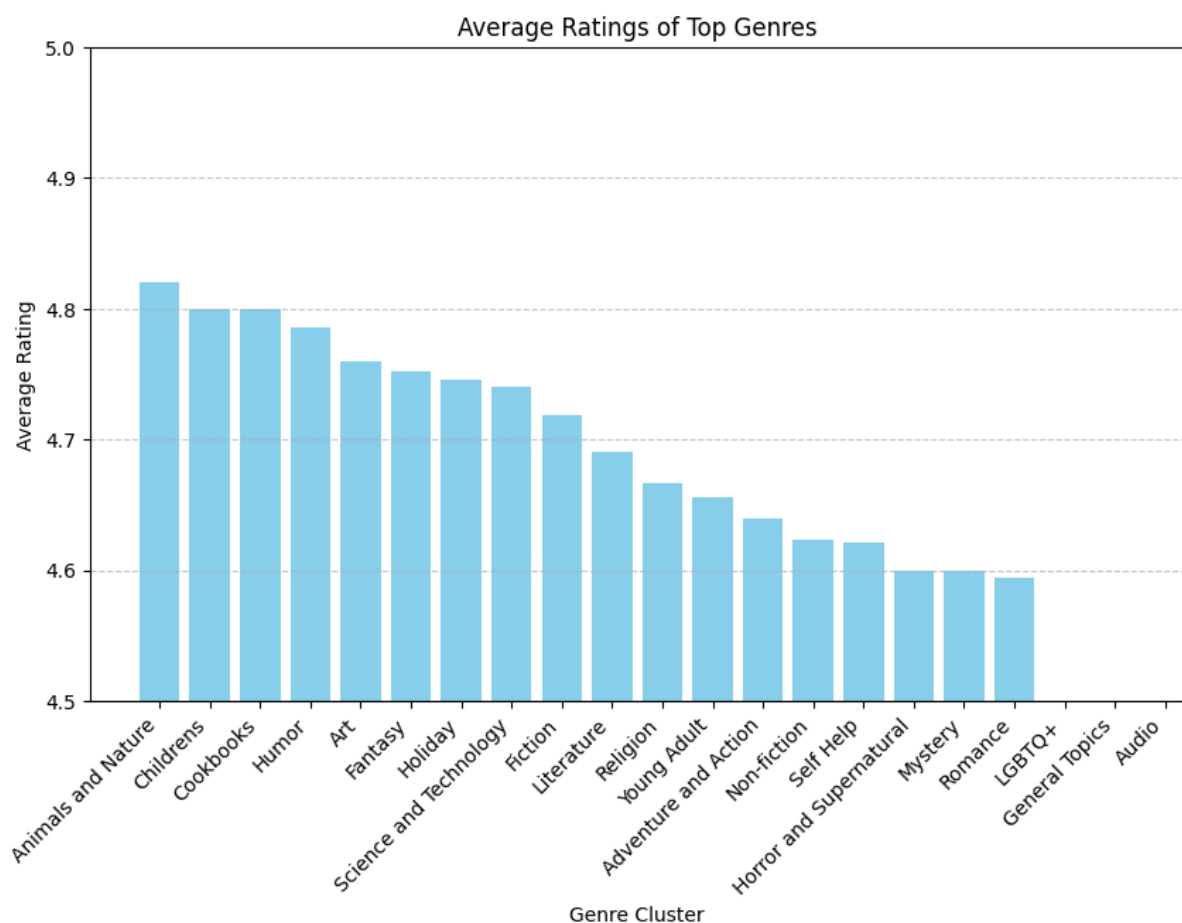
- Separated genres which were comma separated
- Clustered genres to get a better understanding

### Observations:

- **Top-rated Genres:**
  1. Animals and Nature: Average rating 4.82 (9 books)
  2. Children's: Average rating 4.8 (39 books)
  3. Cookbooks: Average rating 4.8 (2 books)
  4. Humor: Average rating 4.7857 (7 books)
  5. Art: Average rating 4.76 (3 books)

### Conclusion:

The genres "Animals and Nature" and "Children's books" have the highest ratings, indicating their popularity among readers.



### 3. Most Popular and Highest-Rated Authors

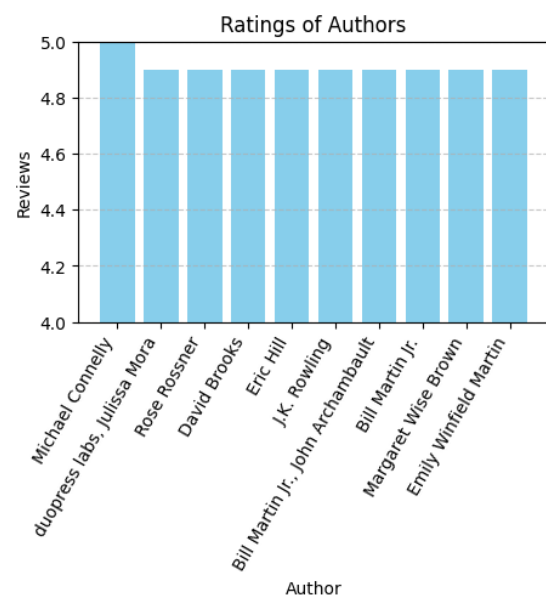
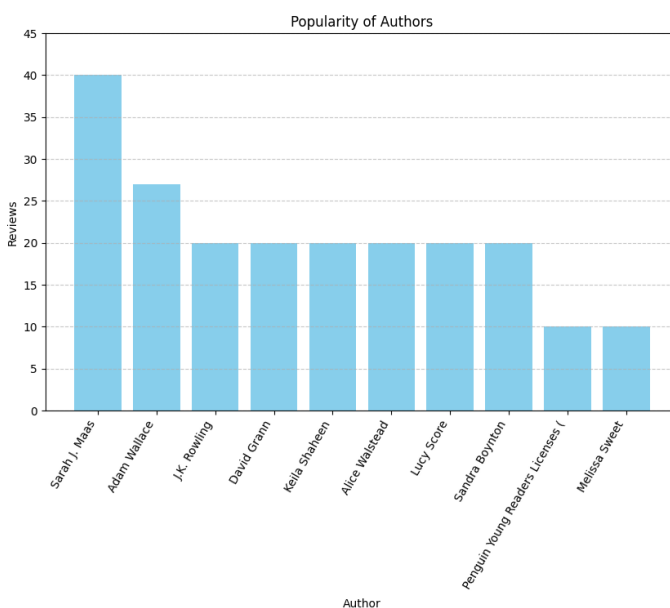
- Measured popularity of the author by the number of reviews they are getting
- Rating of an author is the average rating for the books written by them

#### Observations:

- **Popularity (Number of Reviews):**
  1. Sarah J. Maas: 40 reviews
  2. Adam Wallace: 27 reviews
  3. J.K. Rowling: 20 reviews
  4. David Grann: 20 reviews
  5. Keila Shaheen: 20 reviews
- **Highest Ratings:**
  1. Michael Connelly: 5.0
  2. duopress labs, Julissa Mora: 4.9
  3. Rose Rossner: 4.9
  4. David Brooks: 4.9
  5. Eric Hill: 4.9

#### Conclusion:

Sarah J. Maas is the most popular author based on the number of reviews, while Michael Connelly is the highest-rated author.



#### 4. Change in Book Prices Over Time

- Considered Median for Analysis as mean can be biased due to less number of books

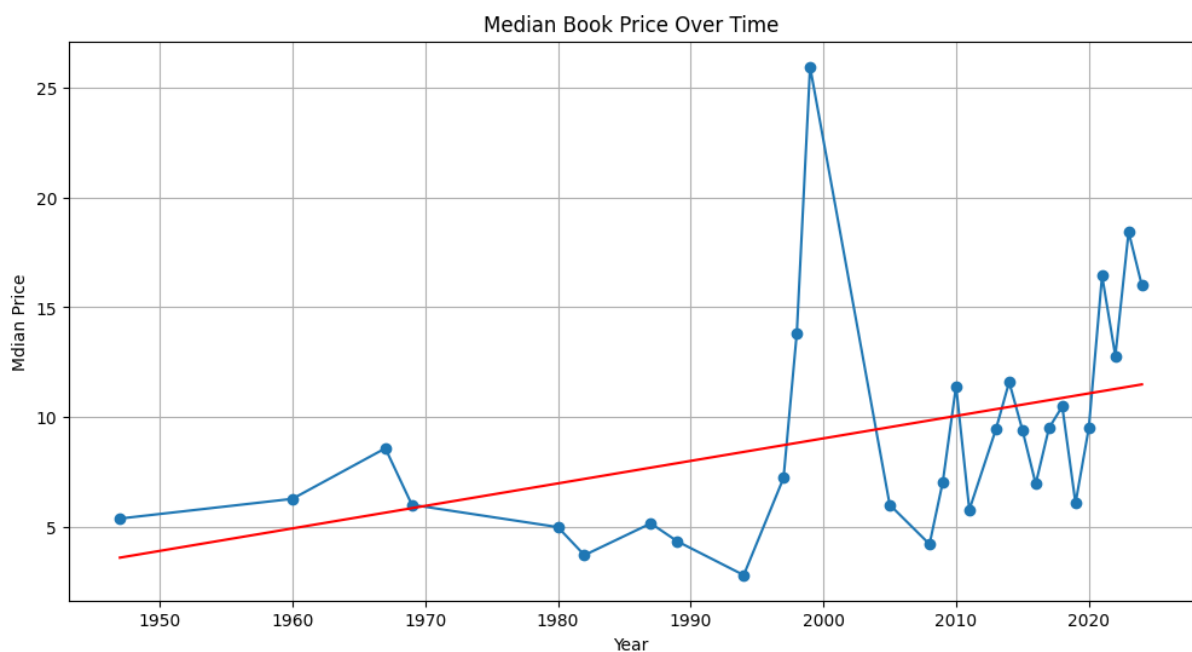
##### Observations:

##### 1. Median Price Analysis:

- Prices have generally risen over time.
- Data before 2014 may have a sample bias due to fewer books, so insights post-2014 are more reliable.
- From 2014 to 2020, median prices fluctuated around \$10.
- After 2020, book prices have increased significantly.

##### Conclusion:

Books have become much costlier in the last three years compared to earlier periods.



## 5. Other Insights

### Observations:

- **Trend in Ratings Over Time:**
  - There is a clear declining trend in book ratings over time, indicating that books in the past were rated higher than those published more recently.
- **Distribution of Ratings:**
  - Ratings are mostly clustered between 4.6 and 4.9, with 4.8 being the most common rating.

### Conclusion:

There is a declining trend in book ratings over time, and most books receive ratings between 4.6 and 4.9.

