Amazon Bestselling Books from 2015-2025 EDA

*Marta Gil Antunano*

**Amazon Bestsellers (2015-2025)**

**Overview:**

This project begins to analyze the trends for the most popular books, according to Amazon sales, from the year 2015 to 2025, which can be found here: [https://www.amazon.com/gp/bestsellers/{year}/books/ref=zg\_bsar\_cal\_ye](https://www.amazon.com/gp/bestsellers/%7byear%7d/books/ref=zg_bsar_cal_ye)

This was selected due to the global reach for amazon books and its percentage of the book market across the U.S. (estimated 71% of total readers).

**Objective:**

The goal of this exploratory analysis is to begin to discern trends in the bestseller list that could aid us into deciding whether to invest resources for a new publishing house.

1. **Data Collection**

* Datawas obtained by Webscrapping using headless browser, obtaining the 60 most popular books for each year.
* Data was cleaned and organized in a dataset with 12 columns: Title, Author, Theme, Rating, Price, Published year, Total page count, Number of reviews, Amazon year as a bestseller, Categories, genres and Publisher.

1. **Univariant Analysis**

* We reviewed the relationship between Geres of fiction and Non Fiction, skewing more towards popularity in Non Fiction:

A green and blue squares with white text

AI-generated content may be incorrect.

* Next, we reviewed for any outliers for books in price and in rating, finding some higher priced books (above $25) and books below the 4.6 mean.
* Due to small amount of entries (655) we cannot infer any further conclusions from the univariant analysis.

1. **Multivariant Analysis**

* Looking at the most popular books over each year vs. the cluster of top 15 titles over the decade, there was a dissonance:

**A screenshot of a computer

AI-generated content may be incorrect.**

**A graph of different colored squares

AI-generated content may be incorrect.**

* There appears to be a difference in between the most popular books each year ( usually Fiction) vs the top 15 best selling books across the decade (Non Fiction).
* We can see this also in our top most popular categories and genres too: A graph of a bar graph

  AI-generated content may be incorrect.A graph of different colored rectangular shapes

  AI-generated content may be incorrect.
* Which leads to the question of: who are our most popular authors, where, and what percentage of the bestselling do they take?A chart with different colored squares

  AI-generated content may be incorrect.
* The most popular authors seem to be overwhelmingly in the Fiction category, what percentage is it of the total?
* Pie chart of popular categories based on book titles:

A colorful pie chart with text

AI-generated content may be incorrect.

* They are only 25% of the total bestselling authors who hold 50% of the most popular book titles, which happen to be in Fiction:A colorful pie chart with white text

  AI-generated content may be incorrect.
* This makes Fiction behave like a clique, few hold the success of the bestsellers category.
* For the Publishers, it appears this holds true as well: A graph of a bar

  AI-generated content may be incorrect.

A blue squares with a red and blue center

AI-generated content may be incorrect.

* Top 3 publishers hold the most frequent bestsellers in Fiction.
* However, the most popular Genre is still Non fiction, as seen for the last decade.
* We began to investigate if this could be a profitable venture, since the market wasn’t so in control of a couple of authors. When we analyzed the average price for each book:

A graph of a number of colored bars

AI-generated content may be incorrect.

* We can see that Non Fiction books sold by average, at 1.5 times a higher price than Fiction. We will need to investigate what forms was it sold in ( kindle vs hardcover), the cost of developing and manufacturing the book, but it could be a profitable path in the publishing sphere.

1. **Takeaways**

* Fiction Bestsellers market appear to be controlled by few authors, speciallyover the last 5 years: Collen Hoover, Sara J. Maas, Delila Owens etc.
* However the Non Fiction market, by aggregate is still more popular every year and overall (**52%** vs **48**%) and there doesn´t appear to be only a few authors controlling the large share of the bestsellers market.
* The average price of the Non Fiction book bought was up to **1.5** times the price of a Fiction one. Worth exploring what forms is it usually consumed in (hardcover vs kindle) and cost of manufacturing.
* For further analysis, it could be interesting to obtain a larger amount of book titles that were in the bestseller (~200), most common form of purchase, to further tease out trends in the Non Fiction sphere as well as consumer direction.