**HDSC August ’22 Stage E Project Documentation: Data Curation for data storytelling.**

By Sakshi Jha (Data Storytelling track)

****

*"The legacy of heroes is the memory of a great name and the inheritance of a great example." -* ***Benjamin Disraeli***

**Introduction:**

Each story book is interesting in its own way, and if they are well written the words might as well be live. flipping through an interesting book can feel like searching through a treasure chest. In many ways, I feel like I similar feelings about books as Dylan Thomas. While reading I some time found that words

I could never have dreamt that there were such goings-onin the world between the covers of books,splashing all over the pagesin a million bits and piecesall of which were words, words, words,and each of which were alive foreverin its own delight and glory and oddity and light.

May it be from a friend or AI; a good recommendation is worth it for your weekends, travel or vacation. So, let’s explore the classics on which the algorithm sets our eyes.

**Aim of this project**:

To unravel what type of books were bought frequently, between 2009 and 2019. This project looks to understand the category of books readers enjoy the most and books frequently purchased.

**Data Scraping:** The data was scraped from Amazon book site, using a combination of selenium and BeautifulSoup packages in python.

**Data Curation:** The data obtained from the scraping process was subjected to data curation processes and the data saved in a csv file.

**Dataset description:** the datais displayed in 7 fields including; Name (name of the book), Author (book author), user ratings, review, price, year, genre. This dataset contains information about top selling books of Amazon (550 books) either belonging to fiction or non-fiction.

**Task to be performed:** Juxtapose between the number of books displayed for sale, and the type of books bought frequently by consumers.

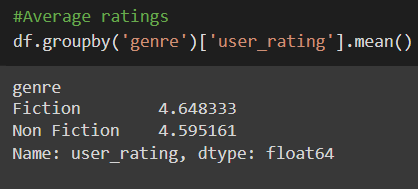
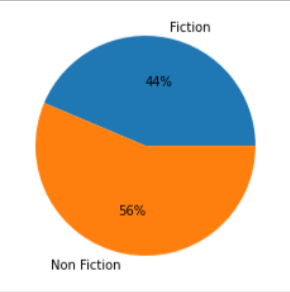
To gain insights in:

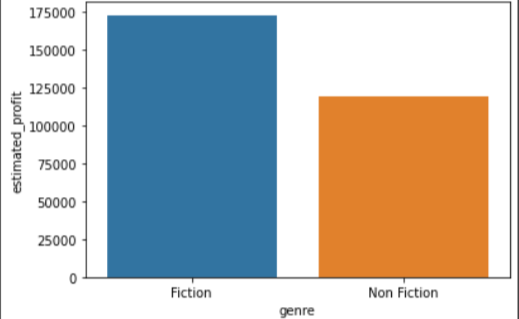
* Average rating of each genre and most sold genre.
* Most popular genre.
* Which genre has the most books in bestseller category.
* Books that have earned the most money in each genre.
* Popularity of each genre and its relationship with time.
* Which year had the most books sold in this category?
* Impact on price of book in respect to time.
* Sales each year.
* Most popular authors.
* The money-making books overall.

**Added feature**:

An extra “Estimated profit” column was added to provide with the briefings for the total earning made by the books.

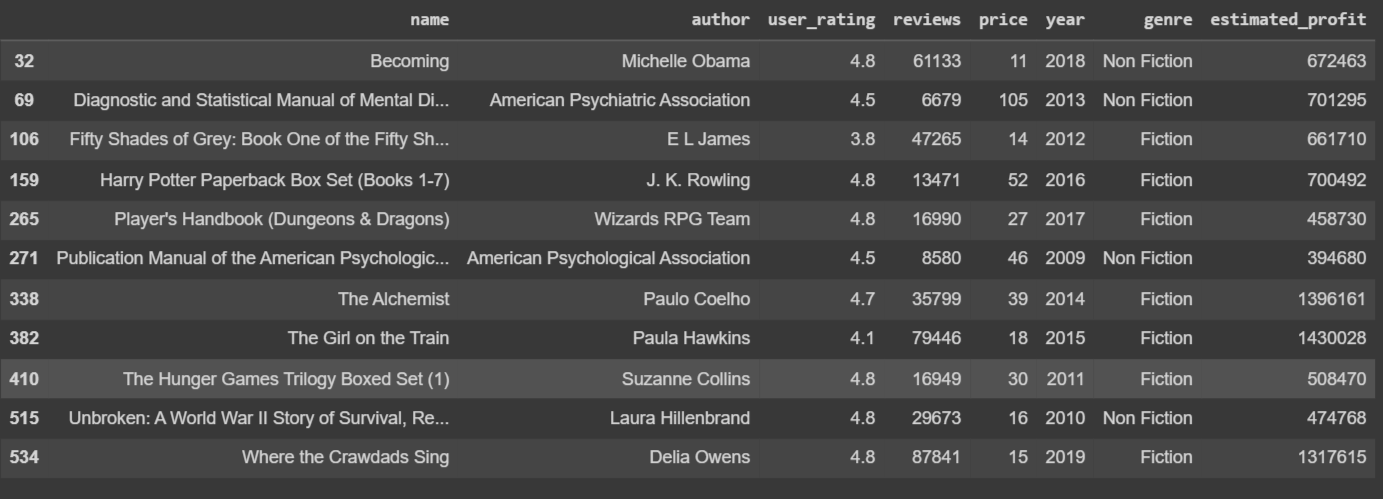
1. **Average rating of each genre and most sold genre.**

* According to EDA, the average rating of fiction books is 4.6 whereas average ratings or non-fiction is 4.5.
* Also, Fiction books experienced a sale of 44 percent while non-fiction of 56 percent. Hence, non-fiction genre is top seller on amazon.

1. **Most popular genre**.

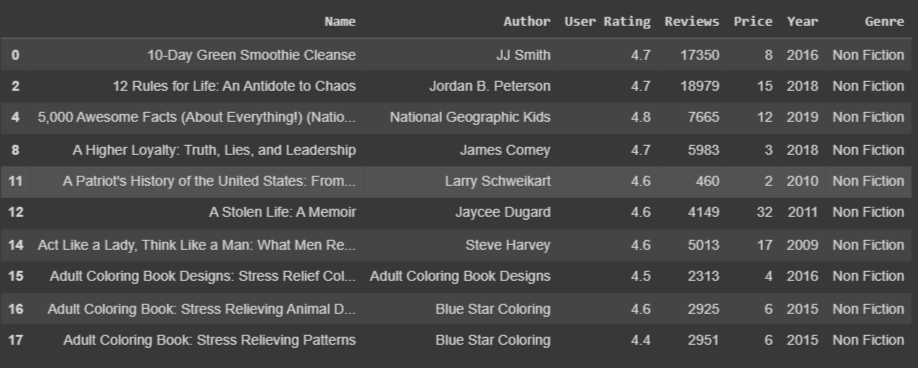
* Although, fiction Books have made maximum profit, making them the more popular genre.

1. **Which genre has the most books in bestseller category.**

* Over 2009 to 2019, the books with highest sales are fiction genre.

1. **Books that have earned the most money in each genre.**

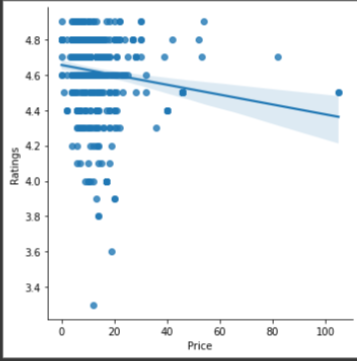
* The above table shows bestsellers in fiction. And below, for nonfiction category.



1. **Which year had the most books sold in this category?**

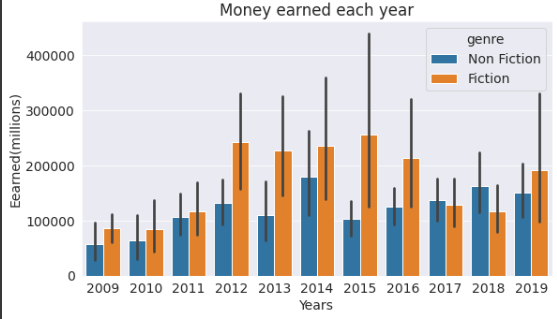
Most sales were observed in 2014.

1. **Impact on price of book in respect to time.**

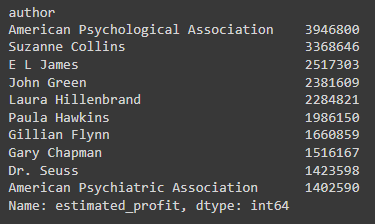
****

This graph shows that there is not a significant relationship between price and ratings but with increasing price, the ratings are falling for both Fiction and Non-fiction. User ratings and prices have inverse relations, most probably.

1. **Sales each year.**

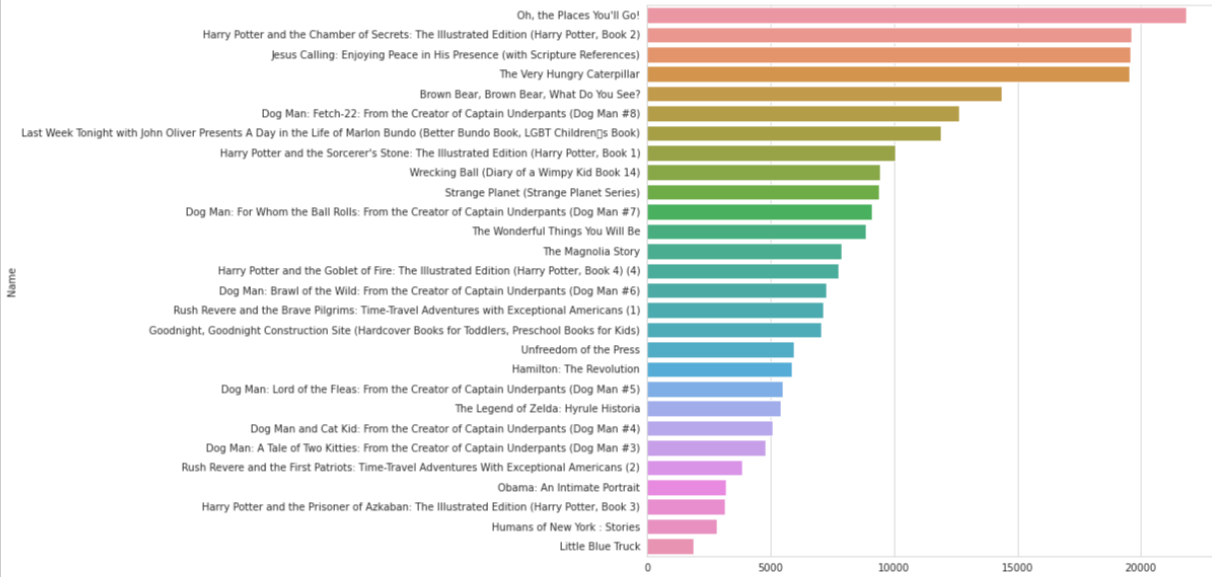
****

The graph shows the earning of these top books, in most cases, are below 0.8 million.

1. **Most popular authors.**

* Above is the list of authors with most sales over the years.

1. **The money-making books overall.**

****

**Conclusion**

The top books are listed under the “money maker books overall”.

These are the top selling books and their author, may you be an avid reader or book seller and collector, investing in these famous books with low paperback cost is profitable for you.

I hope these recommendations are worth your indulgence. Happy reading!