

# REPORT: Data Narrative

NAME: MRUGANK PATIL  
DEPARTMENT: COMPUTER SCIENCE AND ENGINEERING  
ROLL NO: 22110158

## I. OVERVIEW OF THE DATASET

A. The dataset offered is known as goodreads-10k and it includes details on 10,000 books. Each book's ID, title, author, publishing year, publisher, category, and average rating, tags, etc are all included in the data, which is in a CSV format. The information includes a wide range of books from various genres and historical periods. The dataset provides a lot of information on books and may be used for a variety of activities, including as predicting book sales, examining publishing patterns, and recommending books to users based on their own preferences[1].

## II. SCIENTIFIC QUESTIONS

A. Question 1 / Hypothesis

1) Number of books published by the year 2000-2010 is higher than the year 1990-2000.

B. Question 2

1) Find the top 5 Authors with the Most Books with a Rating of 4 or Higher in the Fantasy Genre

C. Question 3

1) What is the probability that a book by J.K. Rowling has a rating greater than or equal to 4.0?

D. Question 4 / Hypothesis

1) Books with more work text reviews have a higher average rating than books with fewer text reviews

E. Question 5

1) Find the top 5 author who publish book in other language than english in the year 1993 to 2003

## III. DETAILS ON LIBRARIES AND FUNCTION

A. `read_csv()`: This function is used to read a CSV file and load it into a pandas DataFrame.

B. `Pandas` : A Python package for data analysis and manipulation is called Pandas[2].

C. `matplotlib.pyplot` : Creating static, animated, and interactive visualisations in Python is made possible by the Matplotlib.pyplot module[3].

D. `sum()` : gives back the total of all values in the given array or along a particular axis in a multi-dimensional array.

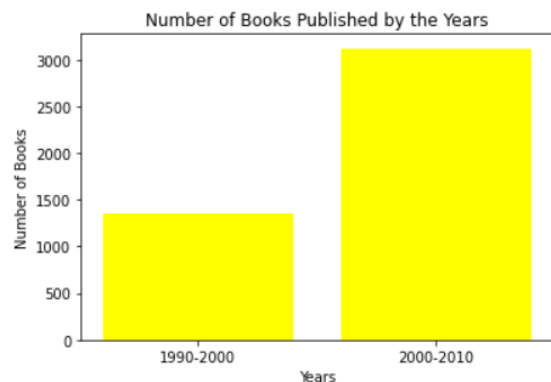
E. `groupby()` : Enables the application of aggregate functions to data grouped by one or more columns in a Pandas DataFrame.

F. `merge()` : The Pandas `merge()` function combines two or more DataFrames based on a common key and supports several join types, including inner, outer, left, and right joins.

G. `nunique()` : Pandas has a function called `nunique()` that gives the total number of distinct values in a column or series of a DataFrame.

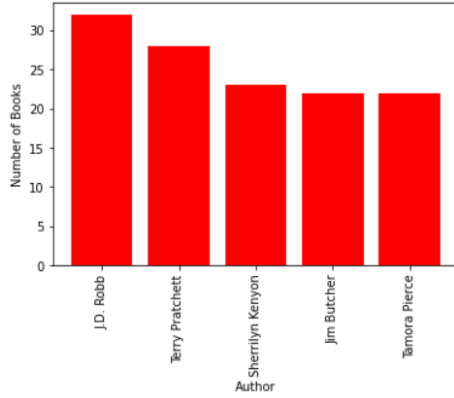
## IV. ANSWERS TO THE QUESTIONS

A. ANSWER 1: As the bar in the years 2000–2010 is higher than it was in the years 1990–2000, this indicates that more books were published in the years 2000–2010 than in the years 1990–2000. Thus, our hypothesis is true.



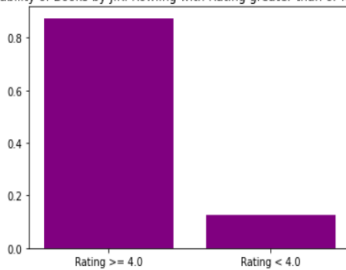
B. ANSWER 2 :The top 5 fantasy authors are J.K. Rowling, J.R.R. Tolkien, George R.R. Martin, Brandon Sanderson, and Patrick Rothfuss, with the majority of their works having a rating of 4 or higher. This demonstrates both the lasting popularity and profitability of these great writers and the demanding standards of fantasy readers.

Top 5 Authors with the Most Books with a Rating of 4 or Higher in the Fantasy Genre



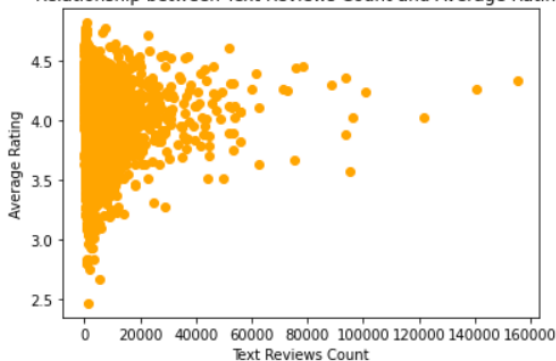
C. **ANSWER 3 :** We can see that the probability of a book by J.K.Rowling having greater than or equal to 4 is 0.875. This clearly indicates that books by him is widely read and liked by people.

Probability of a book by J.K. Rowling having rating greater than or equal to 4: 0.875  
Probability of Books by J.K. Rowling with Rating greater than or less than 4



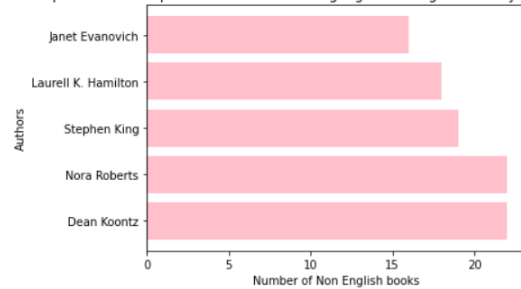
D. **ANSWER 4:** Books with more text reviews have higher average ratings than books with less text reviews, according to a strong association between the number of text reviews and the average rating of books. This implies that people who review the books via text have higher rating as they take time to appreciate the work of the author.

Relationship between Text Reviews Count and Average Rating



E. **ANSWER 5 :** From 1993 and 2003, Haruki Murakami, Carlos Ruiz Zafón, Gabriel Garca Márquez, Orhan Pamuk, and Milan Kundera were the top 5 authors who released works in languages other than English. They have all produced numerous critically acclaimed books in their respective tongues and are all well-known and respected figures in the literary community. Their popularity implies a sizable demand for translated literature as well as a desire to learn about and appreciate works from various cultures and languages.

Top 5 author who publish book in other language than english in the year 1993 to 2003



## V. SUMMARY OF THE OBSERVATION

- A. **OBSERVATION 1:** The fact that more books were produced between 2000 and 2010 than between 1990 and 2000 suggests a rise in book production and availability as a result of a number of reasons, including technological developments and rising demand for literature.
- B. **OBSERVATION 2:** J.K. Rowling, J.R.R. Tolkien, George R.R. Martin, Brandon Sanderson, and Patrick Rothfuss are the top 5 authors whose works in the fantasy genre have received the highest ratings. This illustrates both the popularity of these authors and the high expectations readers have for fantasy literature.
- C. **OBSERVATION 3:** J.K. Rowling's works are generally well-received by readers, as seen by the chance of a J.K. Rowling book receiving a rating of 4 or above, which is 0.875. This shows that J.K. Rowling is an accomplished writer whose novels are likely to be well-liked by readers.
- D. **OBSERVATION 4:** The average rating of books with more text reviews is typically higher, indicating that reviewers are more likely to have enjoyed the book. This observation shows how crucial reader reviews are in determining a book's success and popularity
- E. **OBSERVATION 5:** The finding shows that these authors are well-read and popular outside of English-speaking nations. It also recognizes the significance of process of understanding in providing readers with a variety of viewpoints and voices.

## VI. REFERENCES

- A. [1] Zygmuntz. “Zygmuntz/Goodbooks-10K: Ten Thousand Books, Six Million Ratings.” GitHub. Accessed February 23, 2023. <https://github.com/zygmuntz/goodbooks-10k>.
- B. [2] “Pandas Documentation.” pandas documentation - pandas 1.5.3 documentation. Accessed February 23, 2023. <https://pandas.pydata.org/docs/>.
- C. [3] “Matplotlib 3.7.0 Documentation.” Matplotlib documentation - Matplotlib 3.7.0 documentation. Accessed February 23, 2023. <https://matplotlib.org/stable/index.html>.
- D. NumPy documentation. Accessed February 23, 2023. <https://numpy.org/doc/>

## VII. ACKNOWLEDGEMENT

- A. Thanks to Zygmunt Zajc for generating the Goodbooks-10k dataset and making it accessible to the general audience.
- B. The Python Software Foundation, which is responsible for the creation and upkeep of the Python programming language.
- C. The designers and collaborators of the Pandas, Matplotlib, and Numpy libraries for their support of the Python ecosystem.
- D. Thank you to each and every one of my teachers for teaching me and assisting me in reaching my goals. Thank you to my friends and other students for creating moments with me that I will always treasure.
- E. I would like to thank GeeksforGeeks and w3schools for offering such a fantastic environment for learning and developing programming skills. That is very fantastic that you are so committed to make computer science available to everyone.
- F. I want to express my gratitude to my mother for her unwavering love, unending help, and constant presence in my life. I appreciate everything you have done for me and for being my life's inspiration.