

ES-114 Data Narrative 1

Chakradhar Basani, Roll No.:
22110054
Material Science and Engineering
Department
Prof. Shanmuga R, IIT
Gandhinagar
Gandhinagar, India
chakradhar.basani@iitgn.ac.in

Abstract—In this report, we are writing Data narrative for the dataset provided. It gives insight on how python, programming in general can make our lives easy.

Keywords—Python, programming, Data Narrative.

I. OVERVIEW OF THE DATASET

The given data set is named guidebooks-10k. The author of this data set is zygmuntz. This data set contains six million ratings for the 10,000 most popular books. This data set is mainly comprised of 4 csv files.

- ratings.csv: It contains ratings of the books sorted by time
- to_read.csv: It contains the IDs of books marked as "to read" by each user.
- books.csv: The metadata for each book, such as goodreads ID, authors, title, average rating etc, are given in this CSV file. This data is extracted from goodreads XML files.
- book_tags.csv: It has the data about shelves, tags, and genres of the books. Users assign these tags, shelves, and genres.

II. SCIENTIFIC QUESTIONS /HYPOTHESES

1. In which language, most of the books are written?
2. Most of the top 100 books were written in which year?
3. What are the trends in the average ratings of J.K.Rowling books?
4. What number of books that are written by J.K.Rowling are rated above 4.0?
5. Which book was marked "to read" the most number of times?

III. DETAILS OF LIBRARIES AND FUNCTIONS

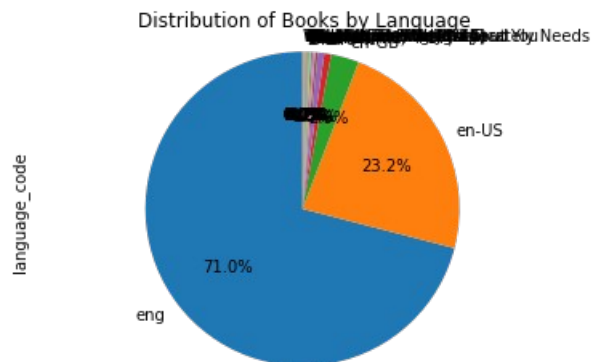
The following libraries and functions are used to answer the above questions.

- Pandas

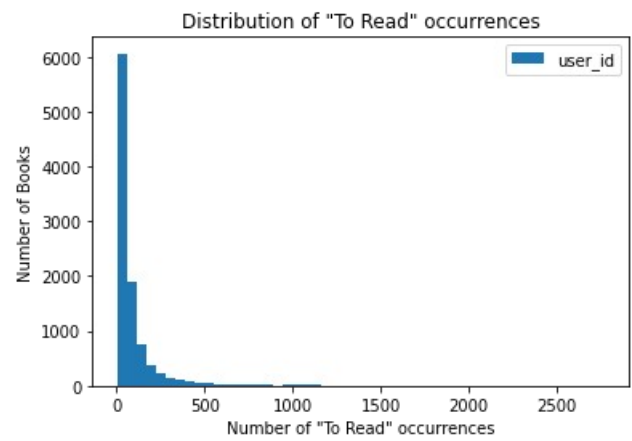
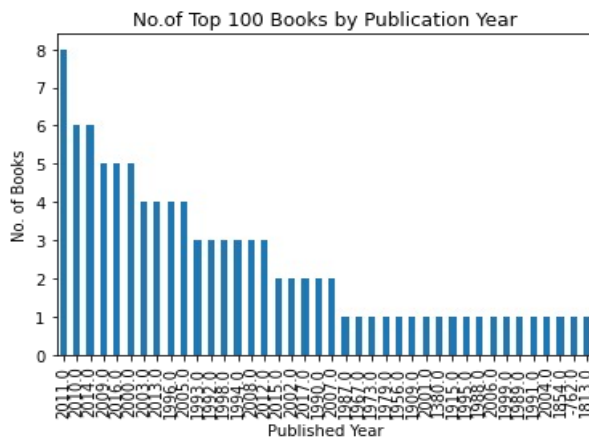
- numpy
- matplotlib
- .read_csv()
- .value_counts()
- .plot()
- .loc[]
- .sort_values
- .sum()

IV. ANSWERS TO THE QUESTIONS

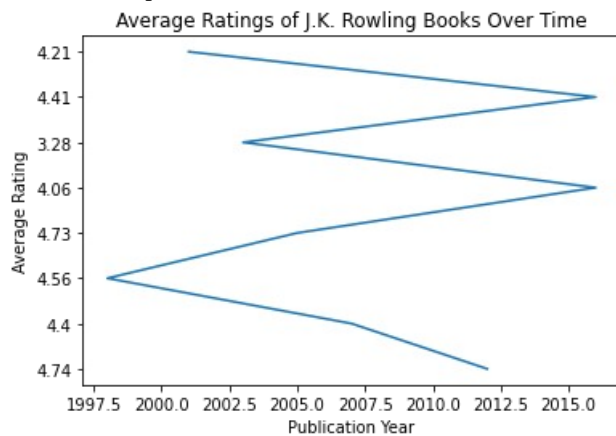
1. Most no. of the books are written in English language



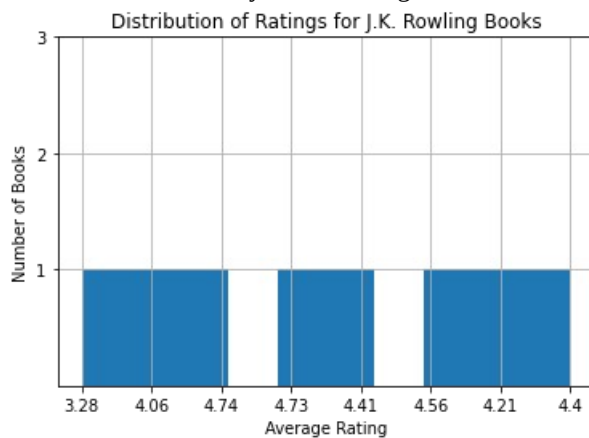
2. Most of the top 100 books were published in year 2011



3. The average ratings of books written by J.K. Rowling didn't follow a particular trend.



4. 7 books written by J.K. Rowling are rated above 4.0



5. The book titled 'The Book Thief' was marked "to read" the number of times. It is marked by 2772 users

SUMMARY OF THE OBSERVATIONS

- I observed that most of the books in top 100 are written in eng language code
- The popular books are mostly published in 20th century
- Average ratings of J.K Rowling's books didn't follow any particular patterns over the years
- It can be JK Rowling is the author because 7 of her books are rated above 4.0
- Most of the books in the given data set were not marked as "to read"

ACKNOWLEDGMENT (Heading 5)

I would like to thank the developeres of python and its libraries for creating these powerful coding language and libraries which made the analysis possible

REFERENCES

- [1] Goodbooks-10k dataset :<https://github.com/zygmuntz/goodbooks-10k>
- [2] Pandas library
- [3] Matplotlib library
- [4] Numpy library