Analysis and Recommendation of Good Books

Jagadeesan D

Asst Professor, Dept., of Business Administration (Logistics), Hindusthan College of Arts and Science, Coimbatore, TN, IND.

Dr. Sathya D★

Asst Professor II, Dept., of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, TN, IND.

Pavithra N

UG Student, Dept., of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, TN, IND.

ARTICLE INFO

ABSTRACT

An exploratory examination of the idea of information is used to evaluate the application of conceptual analysis in data science. It is important to understand the associations between many aspects of a book, such as the book's rating, the author's trend, and the number of languages in which the book is published. More than a hundred thousand people have given their evaluations, indicating that these are well-liked novels. A second reason why this study is significant is because it allows us to understand what kinds of books today's readers are most motivated by, as we've always assumed that only magical personal novels carry this type of power.s Correspondences are shown to exist between the concepts of books rating, author's trend, language, title and thereby pointing in promising directions for further research. Thus, conceptual analysis will be used to upgrade our understanding of archival concept.

1.0 INTRODUCTION

The Primary objective of this analysis is to examine the Book data set for distribution, outlier and anomalies to direct testing of specific hypothesis. Also aims in understanding the dataset visually through graphical representation. Scope of our project is to get a fair idea about books, Also, choosing the right book at right time to inculcate the knowledge from books from this analysis. Using a combination of proprietary algorithms and billions of data sets, Goodreads' recommendation engine can effectively predict what books readers will be interested in next (Chung, 2011). The algorithm checks to see how frequent they appear on the same bookrack, and they are enjoyed by the same population. The engines also differentiate the people of who are interested in the similar kind of books with others who have different taste.

★ *Corresponding author's e-mail:* sathya.d.cse@kct.ac.in (Dr. Sathya D)

In Markwick (2019), it is a website which recommends and gives rating about books. After reading the book the user can give rating and can see what others thought about the book. It also gives the crowd ranking about the book. There are two ratings one is my rating and other one is average rating the plotting against these two points gives the correlation between the two. A positive correlation indicates that my rating accepts with the crowd rating if not then the negative correlation. It also provides the rating of the book by considering the page length too.

As many people feel bored and uneasy while reading some books: which they didn't like or the books they didn't need at that time. This is most common things, and this may reduce their book addiction. Also, this habit of reading books develops one's personality, knowledge, decision making, etc.

So, good book data analysis helps people to analyse, customize and pick the right books at right time. This helps people in all age groups (irrespective of age).

2.0 PROPOSED SYSTEM AND RESULTS

2.1 Exploratory Data Analysis

An essential part of any Data Analysis or Data Science project is exploratory data analysis (EDA). Discovering patterns, anomalies, and hypotheses (outliers) in our knowledge of the dataset is the goal of EDA.

Using EDA, the dataset's numerical data may be summarised statistically, and different graphs can be generated to aid in data interpretation.

Steps followed in EDA

1) Importing libraries
2) Reading data
3) Descriptive statistics
4) Missing value imputation
5) Graphical Representation

2.2 Dataset Description

Our group analysed a Goodreads Dataset, collected by the Goodreads API. This data set contains detailed information about individual books. This dataset was obtained via the Kaggle site, from user Soumik. There are 10,352 unique titles and 11,126 unique ISBNs with a variety of attributes and data types. The twelve columns contain strings, integers, date Time, and characters (Ralli, 2020; Jingchenliu, 2020).

Information contained in the dataset includes:

- *Title:* The title released by the publisher.
- *Authors:* Identifying information about the book's writers.
- *Average rating:* The overall average rating for the book was obtained.
- *ISBN:* The International Standard Book Number (ISBN) is also a unique number that may be used to identify a specific book.
- *ISBN-13:* Instead of the conventional 11-digit ISBN, the book will have a 13-digit ISBN to identify it.
- *Language code:* The major language used to explain the content of the book is described below. For instance, Eng., indicates that the book was written in English.

- *Number of pages:* Its total number of pages in the book.
- *Ratings count:* The total number of times the book has been rated by readers.
- *Text reviews count:* The book got a lot of good reviews in writing.
- *Publication date:* The first publication date of the book.

The various data analysis and the implemented results are shown as screen shots. The pair plot and Normalization are shown in the Figure 1 and Figure 2. The correlation matrix and bar plot are shown in Figure 3 and Figure 4.
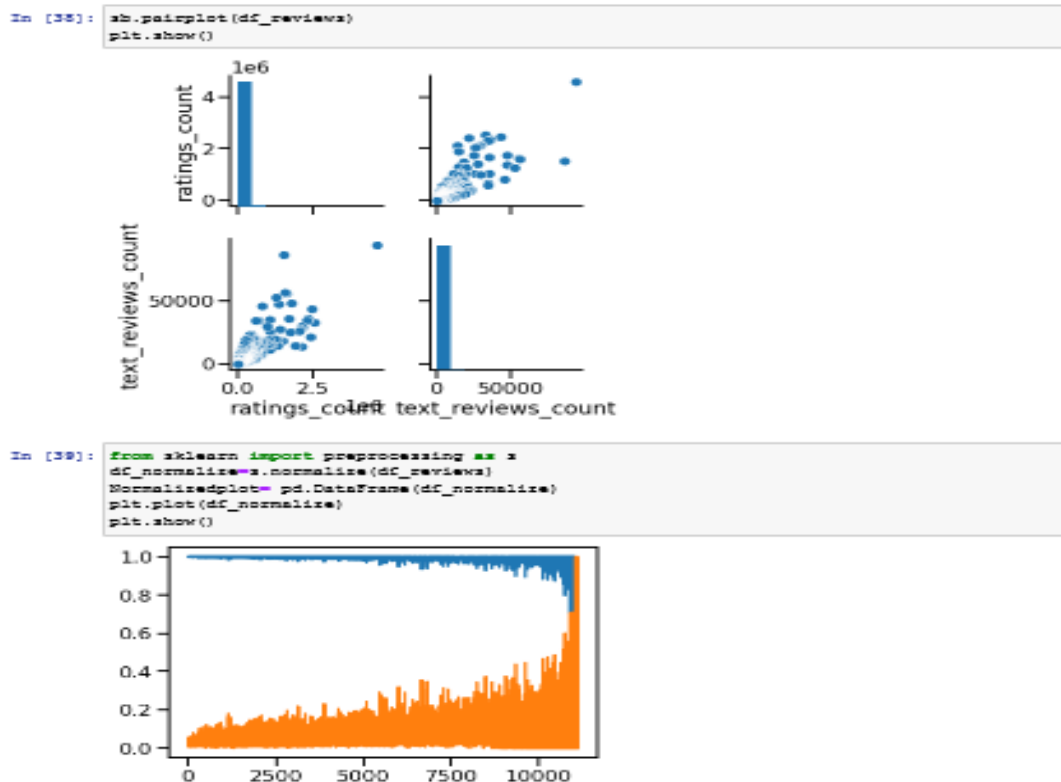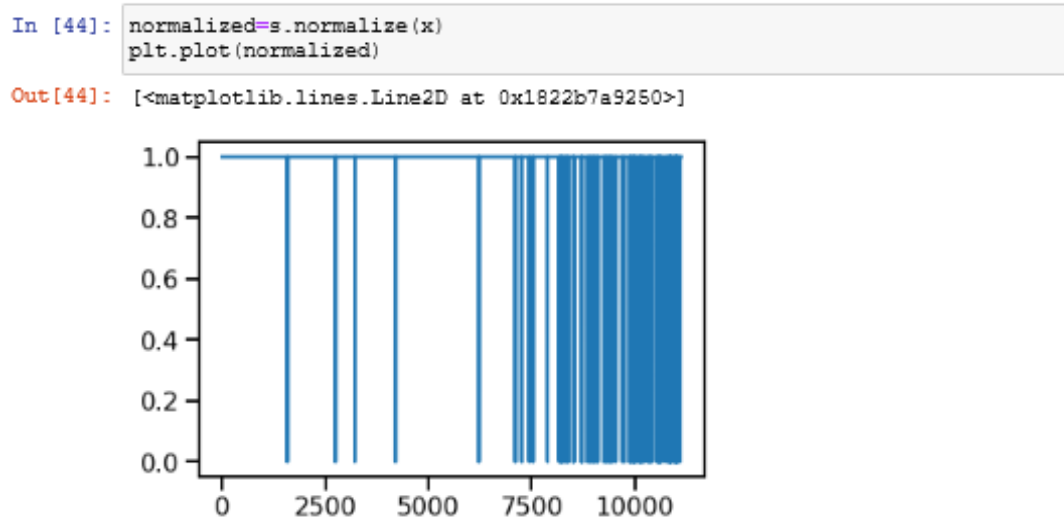

Figure 1 - Pairplot

Figure 2 - Normalization

```
In [46]:  from scipy import stats
          from mlxtend.preprocessing import minmax_scaling
```

```
In [47]:  fig, axes = plt.subplots(figsize=(10, 5))
          sb.heatmap(data=df.corr(), annot=True, linewidths=1.5, fmt=".1f", ax=axes)
          plt.show()
```
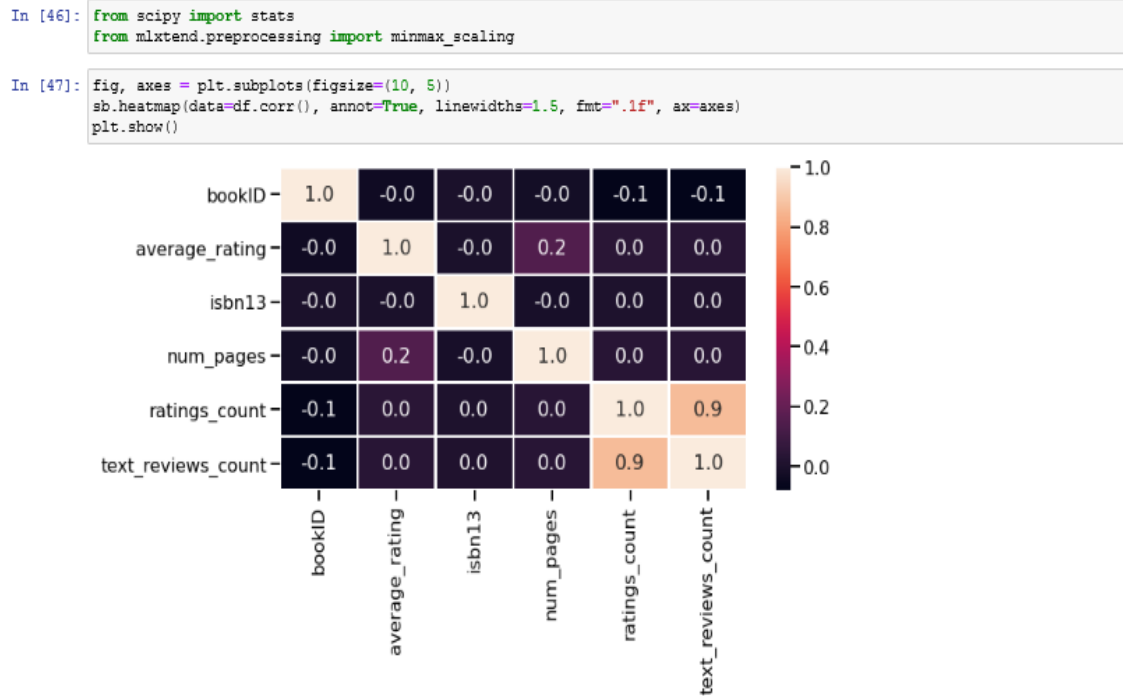

Figure 3 – Correlation Matrix and Heatmap

```
In [55]:  # visualise the above comparison result
          pred.plot(kind='bar', figsize=(13, 7))
```

```
Out[55]:  <matplotlib.axes._subplots.AxesSubplot at 0x1822a5689d0>
```
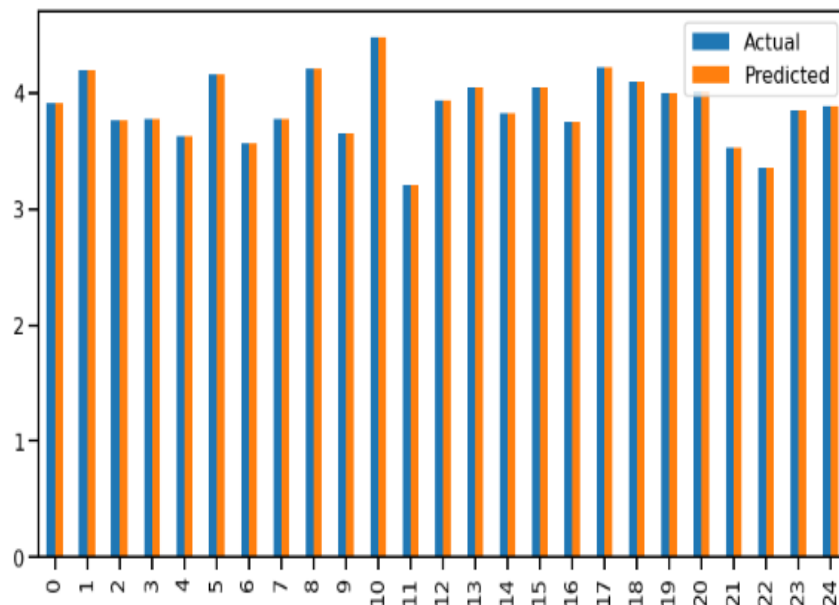

Figure 4 – Bar Plot

CONCLUSION

The scale, scope, complexity, and tempo of good books are likely to increase in the upcoming years. People would not consume a lot more time in choosing the books based on their need with our good books. It will be people's favourite choice.

REFERENCES

Chung, K., (2011) "Announcing Goodreads Personalized Recommendations", [Online] available at https://www.goodreads.com/blog/show/303-announcing-goodreads-personalized-recommendations?page=4.

Cicke, G., (2020) "EDA with Python - Goodreads Data", [Online] available at https://www.kaggle.com/code/gokhancicek/eda-with-python-goodreads-data/notebook.

Jingchenliu (2020) "GoodReads Books Data Analysis" [Online] available at https://www.kaggle.com/code/liujchen/goodreads-books-data-analysis/notebook.

Markwick, D., (2019, Apr) "Goodreads Analysis", [Online] available at https://dm13450.github.io/2019/04/03/Goodreads.html.

Ralli, S., (2020) "Goodreads: Analysis and Recommending Books", [Online] available at https://www.kaggle.com/code/hoshi7/goodreads-analysis-and-recommending-books/notebook.

Sareen, S., (2020, Jan) "Breaking Down Goodreads Dataset using Python", [Online] available at https://towardsdatascience.com/breaking-down-goodreads-dataset-using-python-388e9b9d6352?gi=5e58def3625.

Singh, A., (2018, Aug) "A Practical Introduction to K-Nearest Neighbors Algorithm for Regression (with Python code)", [Online] available at https://www.analyticsvidhya.com/blog/2018/08/k-nearest-neighbor-introduction-regression-python/.

Soumik (2020) "Goodreads-Books", [Online] available at https://www.kaggle.com/jealousleopard/goodreadsbooks.