

# **DATA VISUALIZATION PROJECT REPORT**

(Project Semester January-June 2023)



## **GOODREADS DASHBOARD**

Submitted by

**Ranjith SK**

Registration No: 12007594

Programme and Section: Data Science, KM027

Course Code: INT 233

Under the Guidance of

**Tanima Thakur - 23532**

**Discipline of CSE/IT**

**School of Computer Science Engineering**

**Lovely Professional University, Phagwara**

## **DECLARATION**

I, Ranjith SK, student of B. Tech CSE (Hons) – Data Science, under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 10.04.2023

A handwritten signature in black ink, appearing to read 'Ranjith SK', with a stylized flourish at the end.

Signature

Registration No. 12007594

Name of the student: Ranjith S K

## **ACKNOWLEDGEMENT**

I would like to express my respect and gratitude to the teachers and staffs at Lovely Professional University for providing this opportunity to do this program and project regarding Data Visualization in Tableau.

I would also like to extend my gratitude to Tanim Thakur (Instructor for the course - INT 233) who helped by giving guidance and support all through the classes and also in completing the project.

The success and final outcome of this project required a lot of guidance and assistance from my instructor and my friends, and I am extremely fortunate to have got this along the completion of my project work. Whatever I have done is only due to such guidance and assistance and I would not forget to thank them.

## **TABLE OF CONTENT**

### 1. Introduction

### 2. Scope of the Analysis

- Data Source
- Data Cleaning and Preparation
- Visualization Design
- Interactive Elements

### 3. Objectives

### 4. Existing System

- Limitations of Existing System

### 5. Analysis on dataset (for each analysis)

- General Description
- Visualization
- Summary

### 6. List of Analysis with Results

### 7. Future Scope

### 8. References

## INTRODUCTION

Data visualization tools have gained popularity over the years as organizations and businesses look for more approachable and user-friendly ways to gain insights from their data. As a reader, you may find it challenging to keep track of the books you have read and the insights you gained from them. I made a Tableau dashboard visualization of my Goodreads records to learn more about the books I have read and wish to read. Goodreads, an online community for book enthusiasts, offer its users the platform to review and rate books, monitor their reading progress, and discover new books for their next read.

My Goodreads data was the subject of exploratory data analysis for this project, which additionally aimed to produce a Tableau dashboard that presented the results in a clear and approachable way. Here, we can examine my book reading habits and requirements on the dashboard, including the genres I enjoy reading, the format I prefer to read, and the novels I desire to read.

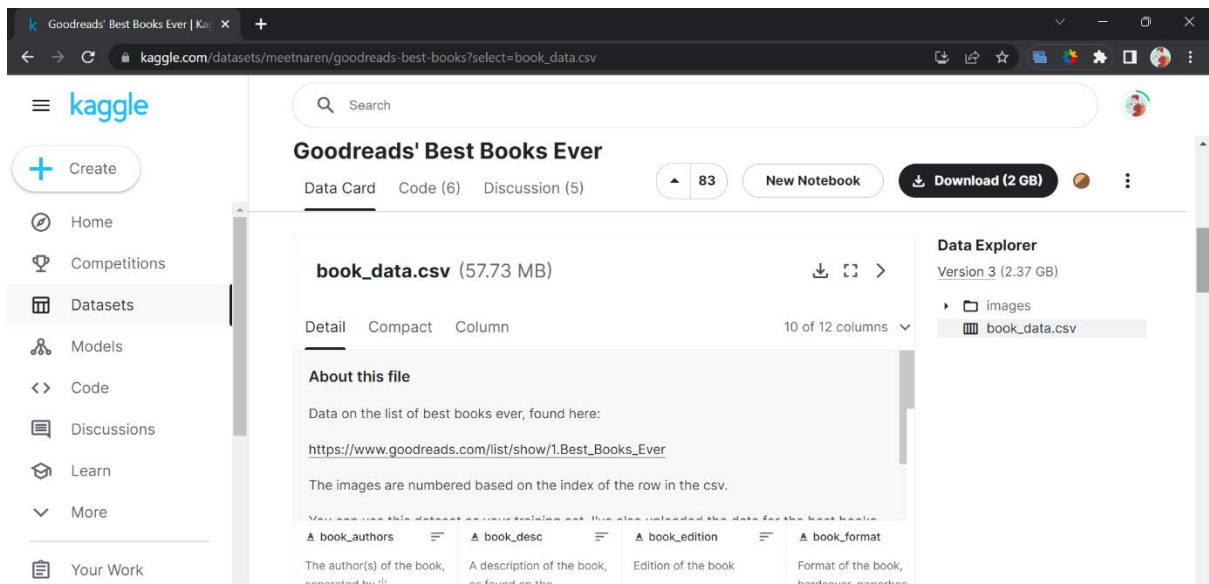
This project report gives you an overview of the scope of analysis used for the visualization project using my Goodreads data in a Tableau dashboard. The insights gained from this project can be used to track my reading choices, discover new books, and analyze my reading behaviour. We can quickly explore and analyze my Goodreads data using this Tableau dashboard to learn more about the books I read and want to read.

## SCOPE OF THE ANALYSIS

### Data Source:

The data source which I have used in this project is derived from an existing data set from Kaggle. The data source is on Kaggle by the name Goodreads' Best Books Ever. I have used the term derived here, because from the original data set, I have only included Top 50 books which I have read and want to read to this project.

If in future, I read more books I would like to update the data, so that I can analyze the same using this dashboard project.



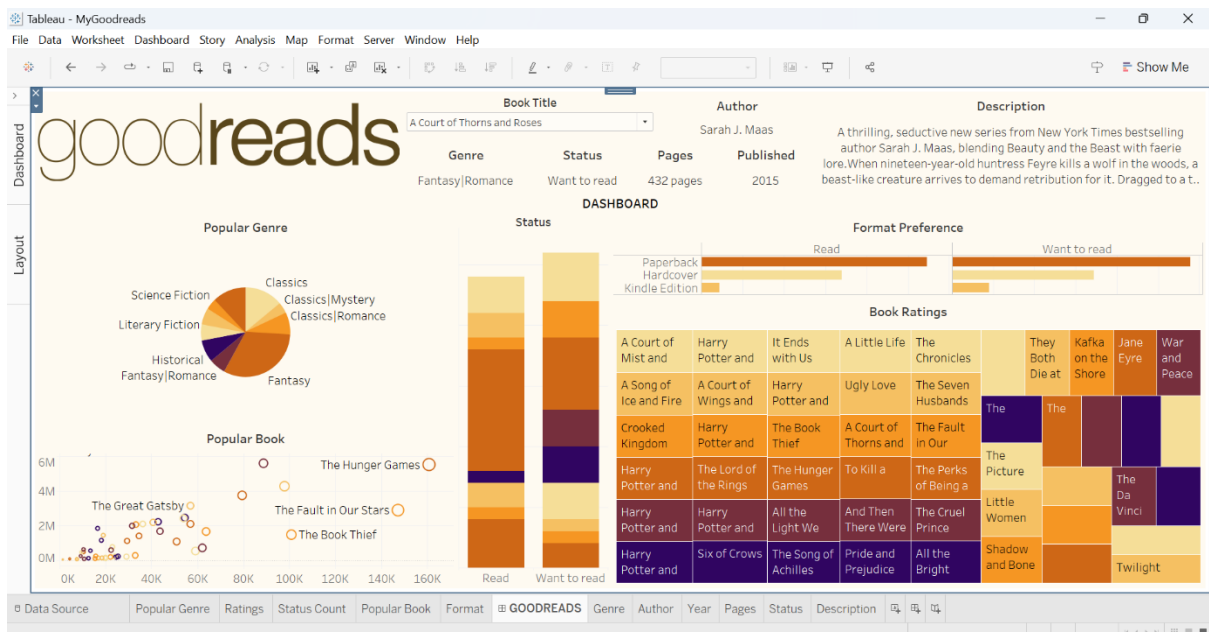
### Data Cleaning and Preparation:

To prepare the data for my need, from the original source, I have cleaned and processed the same. To explain further, I have removed unwanted and irrelevant data from the original which I don't need for my project. I have formatted and included some extra content like the status of the book and published year of the books in the data set using Microsoft Excel.

	A	B	C	D	E	F	G	H	I	J	K	L
1	book_authors	book_desc	book_edition	book_forr	Published	book_isbn	book_pag	book_rating	book_rati	book_revi	status	book_title
2	Colleen Hoover	Sometimes it is the one who lov	Bloomsbury	Kindle Edi	2016		367 pages	4.41	133163	19877	Read	It Ends with Us
3	Sarah J. Maas	A thrilling, seductive new series	US Edition	Kindle Edi	2015		432 pages	4.27	280710	32852	Want to r	A Court of Thorns and Roses
4	Madeline Miller	Greece in the age of heroes. Pa	Bloomsbury Mo	Paperback	2011	9.78E+12	368 pages	4.31	92247	12688	Read	The Song of Achilles
5	Jane Austen	Clever, rich - and single - Emma	Penguin Classics	Paperback	1815	9.78E+12	474 pages	3.99	545845	13313	Want to r	Emma
6	Jane Austen	When Elizabeth Bennet first me	Penguin Classics	Paperback	1813	9.78E+12	432 pages	4.25	2454879	54359	Read	Pride and Prejudice
7	Emily Bronte	Set on the stormy moors of nor	Wordsworth Cla	Mass Mar	1847	9.78E+12	245 pages	3.84	1121618	29417	Want to r	Wuthering Heights
8	Sarah J. Maas	Feyre survived Amarantha's clu	US Edition	Kindle Edi	2016		626 pages	4.69	196527	29188	Want to r	A Court of Mist and Fury
9	Sarah J. Maas	Looming war threatens all Feyr	US Edition	Paperback	2015	9.78E+12	699 pages	4.51	135356	18450	Want to r	A Court of Wings and Ruin
10	J.K. Rowling	Harry Potter's third year at Hog	First Edition	Mass Mar	1999	9.78E+12	435 pages	4.55	2217128	42873	Read	Harry Potter and the Prisoner of Azkaban
11	J.K. Rowling	Harry Potter is a wizard. He is	First Edition	Hardcover	1998	9.78E+12	251 pages	4.4	2163470	40720	Read	Harry Potter and the Chamber of Secrets
12	J.K. Rowling	Harry Potter is midway through	First Scholastic T	Paperback	2000	9.78E+12	734 pages	4.54	2077766	36237	Read	Harry Potter and the Goblet of Fire
13	J.K. Rowling	Harry has been burdened with	First Edition (UK	Hardcover	2007	9.78E+12	607 pages	4.62	2069089	56797	Read	Harry Potter and the Deathly Hallows
14	J.K. Rowling	Harry Potter's life is miserable.	First Edition (UK	Hardcover	1998	9.78E+12	320 pages	4.46	5588580	88816	Read	Harry Potter and the Sorcerer's Stone
15	J.K. Rowling	When Harry Potter and the Hal	First Edition (UK	Paperback	2005	9.78E+12	652 pages	4.55	1991117	31620	Read	Harry Potter and the Half-Blood Prince
16	John Tiffany/Jack Thor	Based on an original new story	Special Rehears	Hardcover	2016	9.78E+12	343 pages	3.7	515658	59194	Read	Harry Potter and the Cursed Child - Parts One and Two
17	J.K. Rowling	There is a door at the end of a	: US Edition	Paperback	2003	9.78E+12	870 pages	4.48	2041594	33264	Read	Harry Potter and the Order of the Phoenix
18	Colleen Hoover	An unforgettable love story tha	Bloomsbury	Paperback	2014	9.78E+12	322 pages	4.28	134601	16708	Want to r	Ugly Love
19	Charlotte Bronte	Fiery love, shocking twists of fa	Penguin Classics	Paperback	1847	9.78E+12	507 pages	4.11	1381404	34670	Want to r	Jane Eyre
20	William Shakespeare	In Romeo and Juliet, Shakespea	New Folger editi	Mass Mar	1597	9.78E+12	283 pages	3.74	1818262	16442	Read	Romeo and Juliet
21	Stephen Chbosky	The critically acclaimed debut n	US / CAN	Paperback	1999	9.78E+12	213 pages	4.2	1063711	50846	Read	The Perks of Being a Wallflower

## Visualization Design:

I have used the same background idea as the original Goodreads website to make it look good visually. I have used appropriate chart types, colours, and layouts so that viewers can check my dashboard in ease. My dashboard is also user-friendly and easy to navigate making it more effective.



### Interactive Elements:

I have added some interactive elements to my dashboard to allows users to view the books I have read and want to read in detail. I have used drop down functionality so that we can view the details of the books. I will be explaining the same in detail in the data analysis part of the report.

### **OBJECTIVES**

1. To gain insights into my reading habits.
2. To track my reading progress.
3. To improve my reading experience.
4. To check the popularity of a book.
5. To learn about my book format preference.

To gain insights on  
my reading habits.

To learn about my  
book format  
preference.

To track my  
reading  
progress.

To check the  
popularity of a  
book.

To improve my  
reading  
experience.



#### To gain insights into my reading habits:

One of main objectives of this project is to analyze my Goodreads data to gain insights on my reading habits. This dashboard will provide information on the genres that I read, the ratings of the book which I have read, and the books I want to read, among other factors. This will help me in understanding my reading preferences and it will help make choices about the books I read in the future.

#### To track my reading progress:

Another objective of this project is to use my Goodreads data to track my progress towards reading goals that I have set for myself. For example, This dashboard will provide information on how much books I have read in a specific genre and the books I want to read in a specific genre.

#### To improve my reading experience:

By analyzing the Goodreads data, the dashboard will help me to understand my reading preferences in detail by checking the ratings of the books I have read and will help me make well informed choices about the books I want to read in the future.

#### To check the popularity of a book:

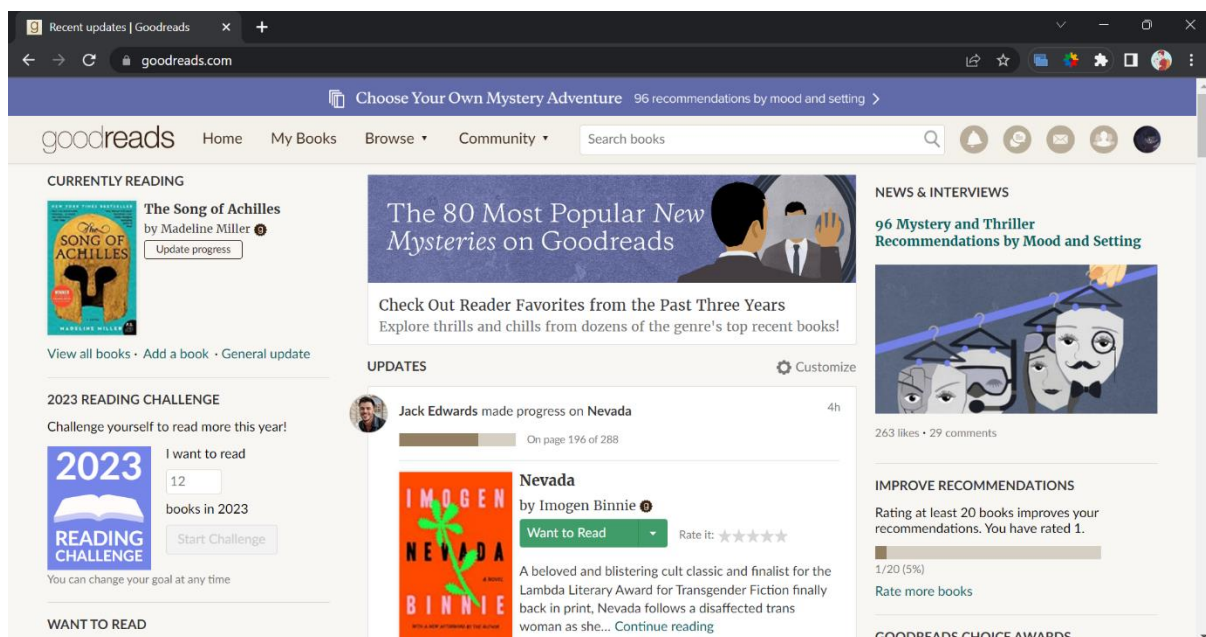
The dashboard can provide insights of the popular book I have read or want to read based on the rating count and the review count of a book, which will be helpful to me in finding more books of popular choice.

#### To learn about my book format preference:

The dashboard can show which format: like paperback, hardcover, or kindle edition; I have preferred to read of all the books I have read and want to read, which will help me in choosing my next book in the format I want to read.

## EXISTING SYSTEM

Goodreads allows users to create profiles, add friends and join book clubs. Users can add books to their virtual bookshelves by searching for books by genre, author, or keyword. They can also create reading lists, comment on books, rate them, and participate in discussions. Personalized recommendations are another service Goodreads provides based on users' reading preferences and history.



## LIMITATIONS OF EXISTING SYSTEM:

- Despite its popularity and usefulness, Goodreads have some limitations. One of the biggest issues is the site's design, which is difficult to navigate, particularly for new users. The site's search function is also limited, making it difficult to find specific books and authors.
- There have been some concerns raised about Goodreads' privacy policies and data collection practices. Some users have reported that the site collects more than necessary data and it doesn't provide or be transparent about how the data is being used.
- Finally, some users have expressed difficulties with the site's review system is susceptible to review manipulation and review bombing. This can make it difficult for the users to determine the quality and relevance of reviews on the site.

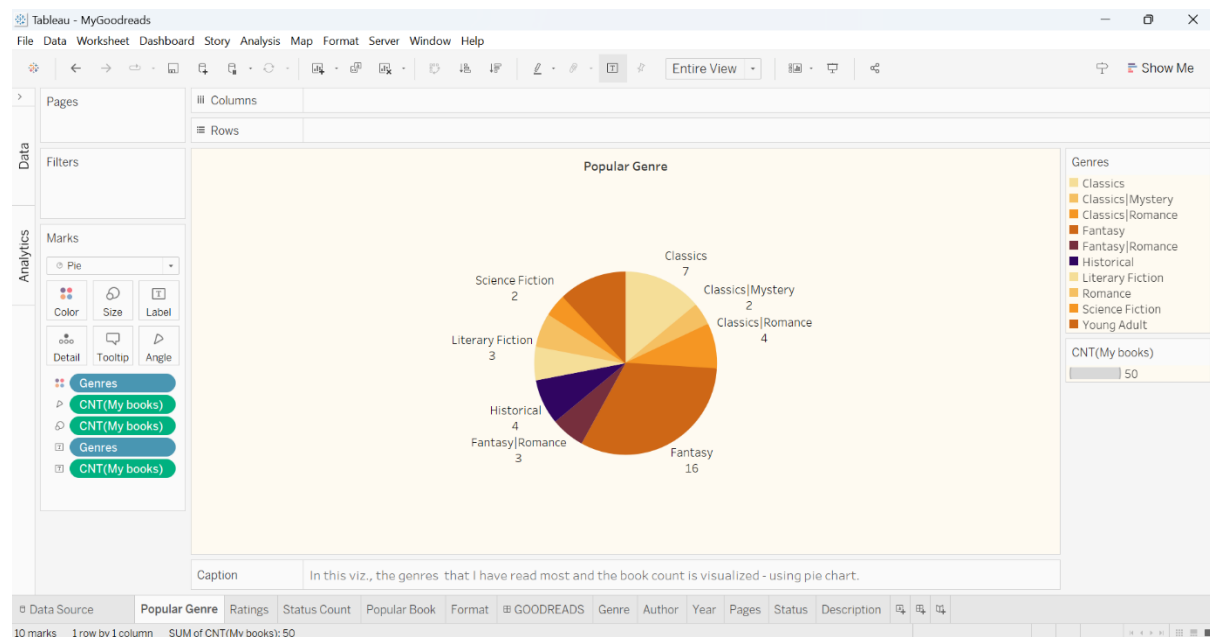
# ANALYSIS ON DATASET

## 1. Genre Analysis

### Analysis:

- ✓ Using my Goodreads data, I have identified the popular genre preference that I've read.
- ✓ I have visualized the number of books read from each genre using Pie Chart.
- ✓ From the visualization, we can analyze that I have mostly read from the 'Fantasy' genre.
- ✓ And we can further analyze that I have least read from the 'Science Fiction' genre.

### Visualization:



### Summary:

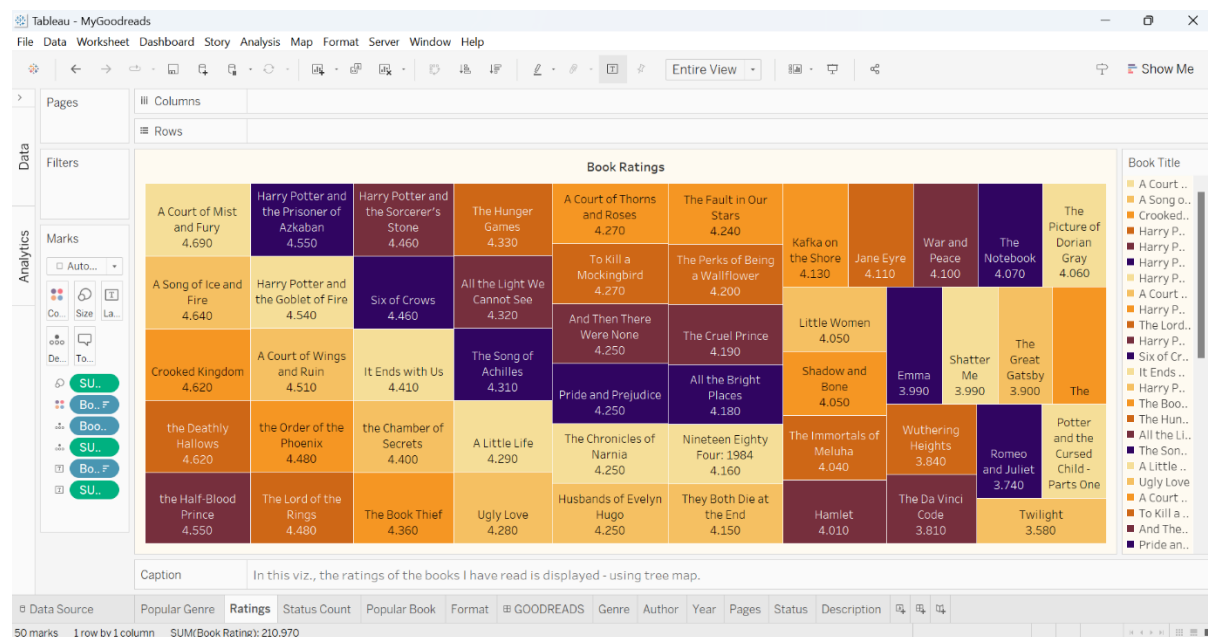
In summary, genre analysis can be a great way to better understand your reading preferences through your personal Goodreads data. By using Tableau to visualize your genres, you can identify which ones you enjoy and the ones that may not be as interesting to you, leading you to make more informed decisions for future book choices.

## 2. Rating Analysis

### Analysis:

- ✓ Using my Goodreads data, I have observed the popular book by the average rating given to a book.
- ✓ I have visualized the most popular book by rating using Tree Map. And I have also coloured the same based on the book titles.
- ✓ From this visualization, we can analyze that 'A Court of Mist and Fury' book has the most rating among other books.
- ✓ And we can further analyze that the 'Twilight' book has the least most rating among all the other books.

### Visualization:



### Summary:

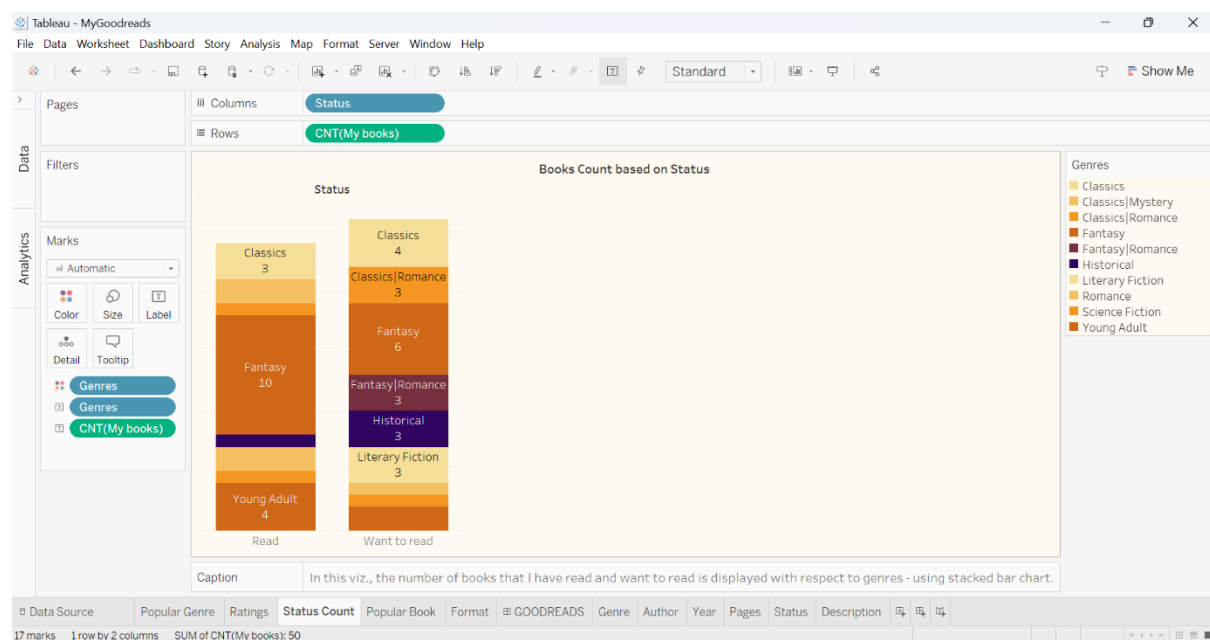
Rating analysis can help you identify which elements of a book you enjoy and identify the factors that affect your ratings. By utilizing Tableau to visualize your data, you can gain valuable insights into how you read books based on ratings, and how to make better-informed decisions for future book choices.

### 3. Status Analysis

#### Analysis:

- ✓ Using my Goodreads data, I have analyzed my overall status of the books I've covered.
- ✓ I have visualized the status of the books by categorizing them into 'Read' and 'Want to Read' with Genre as the base using Stacked Bar Chart.
- ✓ From this visualization, we can analyze that I have books that I 'Want to Read' the most in count than the books I have 'Read'.
- ✓ We can further analyze that I have categorized the chart using colours based on Genre.

#### Visualization:



#### Summary:

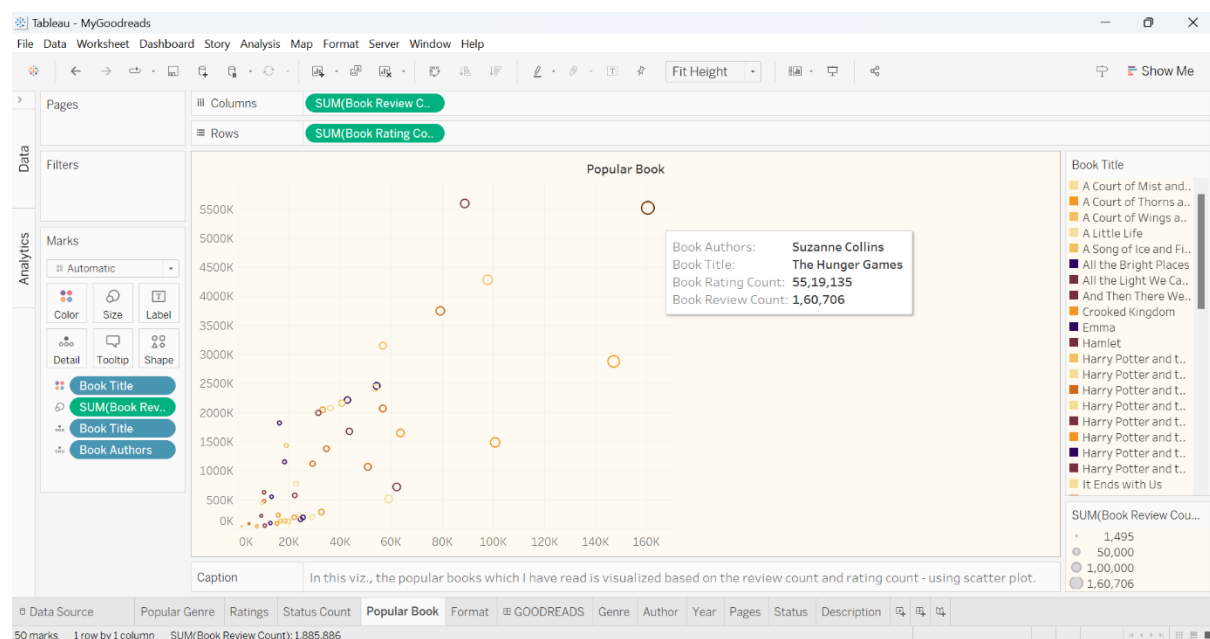
Analysis of status can help you identify which types of books you enjoy and how often you complete what you set out to read. You can use Tableau to visualize different statistical aspects of your statuses, gain valuable insight into your reading preferences, and make informed decisions for future book choices.

## 4. Book Popularity Analysis

### Analysis:

- ✓ Using my Goodreads data, I have analyzed the overall popularity of books based on rating count and review count.
- ✓ I have visualized this analysis, by plotting books along the axes of review count and rating count and examined the popular book using Scatter Plot.
- ✓ From this visualization, we can analyze that the 'Hunger Games' is the most popular book based on the rating and review count.
- ✓ Further we can analyze that 'A Song of Ice and Fire' is the least popular book based on the rating and review count.

### Visualization:



### Summary:

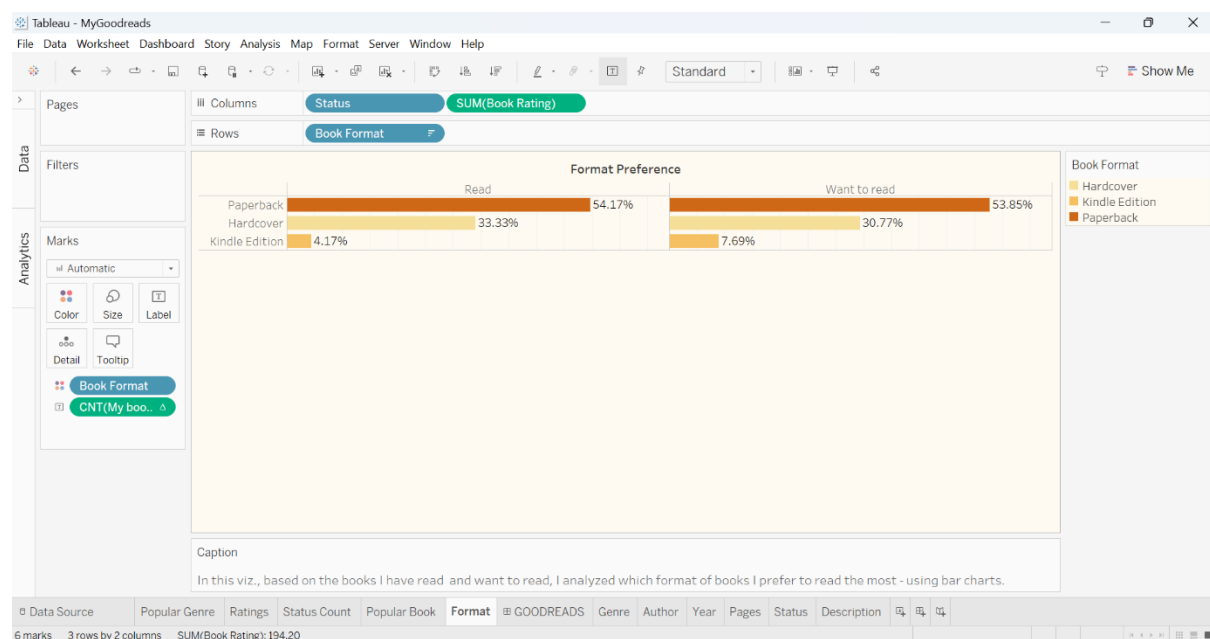
Popular book analysis can help you identify what books and categories are worth reading and which writers are worth following. By using Tableau to visualize this insightful analysis, you can get an idea of where the majority of people's interests lie and improve your chances of gaining popularity as either an author or book reviewer.

## 5. Format Preference Analysis

### Analysis:

- ✓ Using my Goodreads data, I have examined the overall format preference for my books and categorized them based on the Status.
- ✓ I have visualized the data by comparing the number of books I have read or want to read in each format using horizontal Bar Chart.
- ✓ From this visualization, we can analyze that the I have preferred 'Paperback' format the most.
- ✓ Further we can analyze that I have preferred the 'Kindle Edition' format less compared to the other two formats.

### Visualization:



### Summary:

Format preference analysis can be useful for both authors and readers who want to decide which format is best for their target audience or which format they want to read. By using Tableau to visualize this insightful analysis, you can gain valuable insight into which format is best suited for your readership.

## LIST OF ANALYSIS WITH RESULTS

### 1) Genre Analysis

#### Result:

Based on my preference for fantasy over science fiction, the pie chart might look something like this:

Genre	Book Read/Want to Read
Fantasy	32%
Science Fiction	4%

This pie chart indicates that out of all the books I have read or want to read, 32% are fantasy, 4% are science fiction, and the remaining 64% includes all other genres. This visualization suggests that I have a strong preference for fantasy books over science fiction.

We can use this information to gain insights into my personal reading habits and preferences, and to inform future book selections.

### 2) Rating Analysis

#### Result:

Based on my visualization using a heat map, 'The Court of Mist And Fury' has a much higher average rating compared to 'Twilight'. The results of the rating analysis can be visualized using the following heat map:

Book Title	Rating
TCOMAF	4.690
Twilight	3.580

The visualization shows that TCOMAF received an average rating of 4.69, while Twilight received an average rating of only 3.58. This indicates that TCOMAF is more popular than Twilight, at least based on the ratings.



### 3) Status Analysis

#### Result:

Based on my visualization using a stacked bar chart, it appears that I have read more books than I still want to read. The results of the status analysis can be visualized using the following chart:

Books I have Read	Books I Want to Read
24	26

The chart shows that out of the total number of books that I have either read or want to read, I have read 24 books, while I still want to read 26 books. The chart is also broken down by genre, with the majority of the books I have read and want to read falling under the fantasy genre.

This analysis can be used to inform future reading choices or to gain insights into my reading habits. For example, if I want to read more books in other genres or if I want to catch up on books that I have been meaning to read for a long time, I can use this information to prioritize your reading list.

### 4) Book Popularity Analysis

#### Result:

The Scatter Plot chart shows that among the books I analyzed, "The Hunger Games" has the highest number of ratings (5,519,135) and reviews (160,706), which makes it the most popular book. On the other hand, "A Song of Ice and Fire" has the lowest number of ratings (42,029) and reviews (1,495), making it the least popular book.

This analysis can be used to understand which books are the most popular among readers, based on their ratings and reviews. Books with higher ratings and more reviews may be more likely to attract new readers or generate interest in a particular genre or author.

## 5) Format Preference Analysis

### Result:

The chart shows that among the formats I analyzed (Paperback, Hardcover, and Kindle edition), Paperback books have the highest number of preferences, with the average total of 55%. Hardcover books are the second most preferred format, with the average of 35%, while Kindle editions have the lowest preference, with only average below 10% .

This analysis can be used to understand which book format I prefer to read the most among all the other formats.

## FUTURE SCOPE

The project that I have created can be expanded and improved in a number of ways, and can be used to generate further insights into readers' preferences and behaviour. Some potential areas for future scope include:

- 📊 Adding more data: You can add more data to the visualization, such as the authors other work, the books you would read in the near future, and the price of each book. This will help readers make more informed decisions about which books to read and will also provide additional insights into trends and patterns in the book market.
- 📊 Adding filtering and sorting options: You can add filtering and sorting options to the dashboard, allowing readers to sort or filter the data based on specific criteria, such as author name, genre, language, or publication date. This will help readers find books that are most relevant to their interests.
- 📊 Adding more charts and graphs: You can add additional charts and graphs to the dashboard, such as bubble charts, line charts, and heat maps, to visualize data in different ways and provide more detailed insights into readers' preferences.

## REFERENCES

Original Dataset:

[https://www.kaggle.com/datasets/meetnaren/goodreads-best-books?select=book\\_data.csv](https://www.kaggle.com/datasets/meetnaren/goodreads-best-books?select=book_data.csv)

Used Dataset:

<https://github.com/Ranjith-13/Data-Visualization/blob/main/My%20Goodreads%20Dashboard/My%20Goodreads.xlsx>

Goodreads:

<https://www.goodreads.com/>