DATA VISUALIZATION PROJECT REPORT

(Project Semester January-May 2024)

GOODREADS BOOK DASHBOARD

Submitted by:

Anshika Nigam

Registration No: 12112925

Program and Section: Bachelor of Technology in Computer Science & Engineering

K21DP

Course Code: INTB233

Under the Guidance of

Maneet Kaur (15709)

Discipline of CSE/IT

Lovely School of Computer Science & Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Anshika Nigam bearing Registration no. 12112925 has completed INTB233 project titled, **"GOODREADS BOOK DASHBOARD"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature:
Name of the Supervisor: Maneet Kaur
Designation of the Supervisor: Assistant Professor
School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab.

Date: 18.04.2024

DECLARATION

I, Anshika Nigam, student of Bachelor of Technology in Computer Science and Engineering 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: 18.04.2024

Signature: Anshika

Registration No: 12112925

Name of the student: Anshika Nigam

ACKNOWLEDGEMENT

I would like to express my profound gratitude to Dr Robin Prakash Mathur of School of Computer Science & Engineering department, and university for their contributions to the completion of my project titled “Goodreads Book Dashboard”.

I would like to express my special thanks to my mentor Maneet Kaur for her time and efforts she provided throughout the year. Your useful advice and suggestions were really helpful to me during the project’s completion. In this aspect, I am eternally grateful to you.

I would like to acknowledge that this project was completed entirely by me and not by someone else.

TABLE OF CONTENT

INTRODUCTION

In today's information age, data plays a crucial role in understanding the world around us. This report utilizes the power of data visualization to unveil fascinating insights into the realm of books. We've meticulously compiled book data and transformed it into an interactive dashboard using Tableau, a leading data visualization software. This dashboard delves into various aspects of books, allowing readers to explore genres, authors, publication trends, and potentially even sales figures, all presented in a visually captivating format.

This report dives into the captivating world of books, visualized through the powerful lens of Tableau. We've harnessed the data's potential to create an interactive dashboard, offering a comprehensive and visually compelling exploration of the literary landscape.

This introduction sets the stage for your report, highlighting the following key points:

The Power of Data Visualization: Briefly mention the importance of data visualization in understanding complex information.

Focus on Books: Indicate the specific data explored in the dashboard, such as genres, authors, publication dates, or sales figures.

Tableau as the Tool: State that Tableau was used to create the dashboard, emphasizing its capabilities for interactive data exploration.

SCOPE OF THE ANALYSIS

Reading Habits:

- Track books read over time
- Analyze genres read most often
- Visualize reading progress by book/author
- Identify trends in reading activity (e.g., seasonal variations)

Book Exploration:

- Discover top-rated books in different categories
- Compare average ratings by genre or publication date
- Find books similar to those a user has enjoyed
- Explore relationships between ratings and reviews

User Engagement:

- Analyze user activity on the Goodreads platform (e.g., reviews, ratings, shelf additions)
- Identify user demographics and reading preferences
- Track social engagement metrics (e.g., friend activity, group discussions)
- Understand user behavior patterns (e.g., most active times, preferred devices)

Additional Considerations:

- The dashboard might incorporate external data sources (e.g., author information, publication details)
- It could allow users to filter and segment data based on their interests
- Interactive features could enable users to drill down for deeper analysis

EXISTING SYSTEM

Public Examples: While there might not be pre-built dashboards for Goodreads, you can find examples built with similar data sets on Tableau Public. Search for dashboards related to "books" or "reading" for inspiration.

Data Sources: You'll need Goodreads data to build your dashboard. Unfortunately, Goodreads doesn't offer an official public API for direct data access. However, you might be able to find datasets related to Goodreads books on platforms like Kaggle.

Remember, building an effective dashboard requires understanding your goals.Once you have a clear objective, you can tailor the visualizations and data points on your dashboard.

Advanced Techniques:

Extracting Data: If you can't find Goodreads datasets, consider unofficial web scraping methods (respecting terms of service) to gather your personal data. However, this can be technical and requires caution.

Data Cleaning and Transformation: Goodreads data might require cleaning and transformation before using it in Tableau. This could involve handling missing values, formatting dates, or creating calculated fields.

Tableau Features:

Tableau Desktop: This is the paid software that allows full development of dashboards. There's a free trial you can explore.

Tableau Public: This is a free platform for sharing and exploring dashboards. While you can't create dashboards directly on Public, you can use it to find inspiration and learn from others' visualizations.

Additional Considerations:

Security: If you use web scraping, prioritize data security. Ensure proper handling of login credentials and avoid techniques that violate Goodreads' terms of service.

Privacy: Using personal Goodreads data requires being mindful of privacy. Only visualize and share data you're comfortable with.

Alternatives to Tableau:

Power BI: Another popular business intelligence tool with similar functionalities to Tableau.

Google Data Studio: A free data visualization tool by Google, with a gentler learning curve than Tableau.

SOURCE OF DATASET

LINK FOR DATASET: https://www.kaggle.com/datasets/jealousleopard/goodreadsbooks

I took this dataset from KAGGLE.

Data Source Documentation: Many data sources, like databases or public datasets, come with their own documentation. This documentation often includes a description of the data, including details about:

Columns: The names and definitions of each data point.

Data Types: The format of the data (e.g., numbers, text, dates).

Units: The units of measurement for quantitative data.

Collection Methodology: How the data was gathered (e.g., survey, sensor readings).

Internal Report Documentation: If the data comes from within your organization, there might be internal documentation that describes the dataset. This could be a separate document, a wiki page, or notes within the data storage itself.

Report Author: Often, the report author themself will include a description of the dataset within the report. This description should summarize key details about the data's origin, structure, and any limitations.

Data Dictionary: In some cases, a data dictionary might be used to describe the dataset. This is a formal document that defines each data point in detail.

Here are some resources to help you find the source of the dataset description:

Check with the Report Author: If you're unsure where the dataset description comes from, the report author is a good first point of contact.

Look for Data Source Information: The report itself might reference the data source and any associated documentation.

Search for Dataset Metadata: If the data comes from a public repository, search for the dataset itself. Often, public datasets have metadata descriptions that detail the data's origin and structure.

ETL PROCESS

In a Tableau project, the ETL (Extract, Transform, Load) process involves preparing and transforming data from its source into a format suitable for analysis and visualization within Tableau. Here's a basic outline of the ETL process in a Tableau project:

Extract (E):

Data Source Connection: Establish connections to the data sources containing the raw data. Tableau supports a wide range of data sources, including databases, spreadsheets, cloud services, and web data connectors.

Data Extraction: Extract the relevant data from the source(s) using Tableau's data connection features. This may involve querying databases, importing files, or connecting to APIs to retrieve data.

Transform (T):

Data Cleaning: Cleanse the extracted data to ensure accuracy and consistency. This may involve handling missing or null values, removing duplicates, correcting errors, and standardizing formats.

Data Integration: Integrate data from multiple sources if necessary. Merge or join datasets to combine related information into a single dataset for analysis.

Data Transformation: Perform transformations on the data to derive new variables, aggregates, or calculated fields. This could include calculations, aggregations, filtering, splitting columns, or pivoting data to reshape it for analysis.

Data Enrichment: Enhance the dataset with additional contextual information or derived metrics to support deeper analysis. This could involve appending external data, applying business rules, or enriching with geospatial information.

Load (L):

Data Modeling: Create a data model within Tableau to organize and structure the transformed data for analysis. Define relationships between tables, hierarchies, and custom groups as needed.

Data Preparation: Prepare the transformed data for visualization within Tableau. This may involve creating calculated fields, sets, groups, parameters, and bins to facilitate analysis and interactivity.

Optimization: Optimize the data model and workbook performance by minimizing the data footprint, optimizing calculations, and leveraging Tableau's data engine for faster querying and visualization.

Data Publishing: Publish the prepared dataset or workbook to Tableau Server or Tableau Online for sharing, collaboration, and interactive analysis with stakeholders.

Throughout the ETL process, it's essential to maintain data integrity, adhere to best practices for data management, and document the steps taken to ensure reproducibility and transparency. Tableau provides a range of tools and features to facilitate each stage of the ETL process, empowering users to prepare and transform data efficiently for insightful analysis and visualization.

ANALYSIS ON DATASET

i. Introduction

Authors: This subheading delves into the world of Goodreads authors. It likely provides basic details like author names, but it might not be an exhaustive list of all Goodreads authors.

Language Code: This subheading focuses on the language a book is written in. The specific language codes might be accompanied by the full language name for better understanding (e.g., "en" for English, "es" for Spanish).

Publication Date: This subheading allows you to explore books based on when they were published. You might be able to filter by specific dates or browse through publication years to see trends over time.

Publisher: This subheading identifies the publishing houses responsible for the books. This section could be helpful to see which publishers dominate Goodreads or discover new publishers based on your interests.

Title: This subheading showcases the titles of the books included in the dataset. It might be the main title of the dashboard, providing an overview of all the books available for exploration. There could potentially be a search function to find specific titles.

1000 Places to See Before You Die: Assuming this is a book title, it might be a singular entry or the first in a list of popular titles. Look for additional context to see if it's a special category or just a single book highlighted.

Text Reviews: This subheading dives into user engagement with the books. The number of text reviews might indicate a book's popularity or how much discussion it has generated on Goodreads.

Number of Pages: This subheading provides information about the book's length. This could be helpful for readers who have a preference for shorter or longer reads.

Ratings Count: This subheading shows how many people have rated a particular book on Goodreads. A high rating count might suggest a book's popularity or widespread recognition.

Average Rating: This subheading goes beyond the total number of ratings and gives you a sense of how the book is generally perceived. The average rating (typically on a scale of 1 to 5) reflects user sentiment towards the book.

Books by Publication Date: This subheading seems to be a dedicated section for browsing books based on their publication year. You might be able to see trends in popularity or discover hidden gems from specific eras.

Top 10 Language Code: This subheading focuses on the most prevalent languages used in the Goodreads book collection. It can reveal interesting patterns about the diversity of languages available or the popularity of certain languages among Goodreads users.

Authors by Total Books: This subheading caters to readers who are interested in following prolific authors. It showcases authors with the most extensive collections on Goodreads, allowing you to discover new authors or explore the works of established ones.

Publishers by Total Books: This subheading sheds light on the publishing houses with the biggest presence on Goodreads. It can be a valuable tool to discover publishers known for specific genres or styles, helping you find books that align with your interests.

Titles by Rating Count: This subheading prioritizes books based on how many ratings they've received. This can be a good way to find books that have sparked a lot of discussion or identify critically acclaimed titles among Goodreads users.

ii. General Description

Tableau Goodreads books: This section displays the program used to create the dashboard (Tableau) and the data source (Goodreads).

Dashboard likely represents the main landing page you're currently viewing.

Layout likely allows you to customize the visual arrangement of the elements on the dashboard.

Authors, Language Code, Publication Date, Publisher, and Title are likely categories used to filter or sort the data.

1000 Places to See Before You Die might be a specific book title, though it's unclear how many other entries are listed.

Text Reviews, Number of Pages, Ratings Count, and Average Rating are likely data points related to the books.

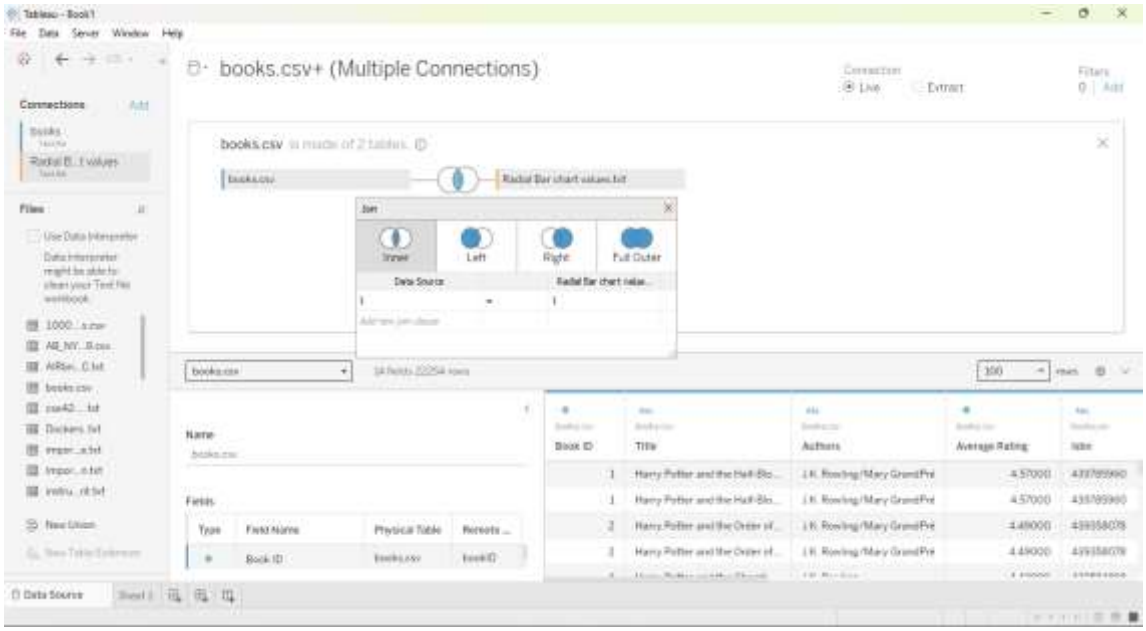
Centre section appears to be a worksheet or table that displays book information. It includes details like author names, publication dates, publishers, book titles, text review counts, number of pages, ratings counts, and average ratings.

Objects section likely represents the visualizations on the dashboard. Here, we see Authors by Total Books and Publishers by Total Books. These seem to be bar charts that show which authors and publishers have the most books in the dataset.

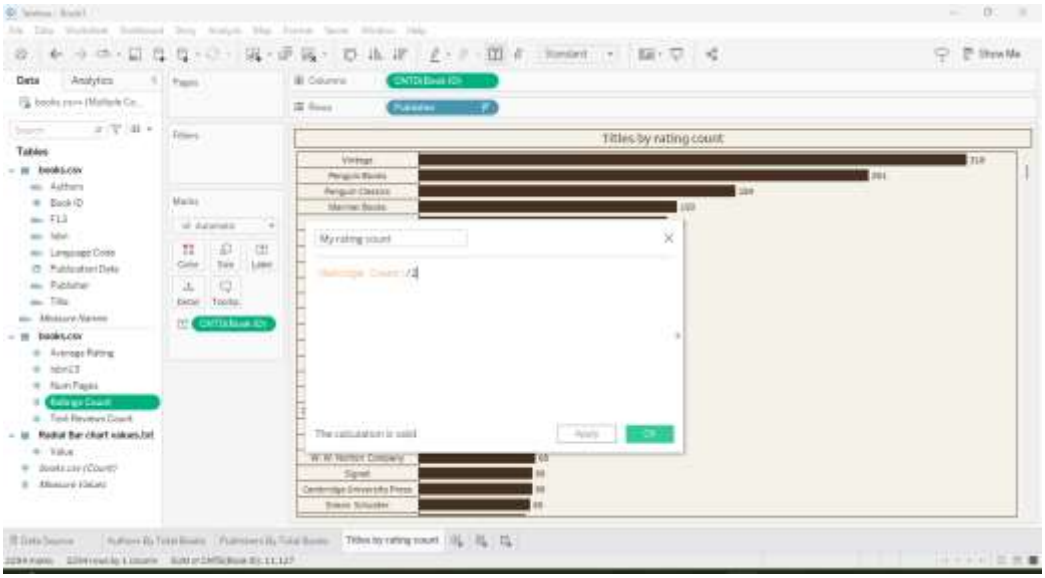
Data Source shows that the data for this dashboard is coming from four sources: Authors by Total Books, Publishers by Total Books, Titles by Rating Count, and Books by Publication Date.

iii. Specific Requirements, functions and formulas

CONNECTED 2 DATASETS

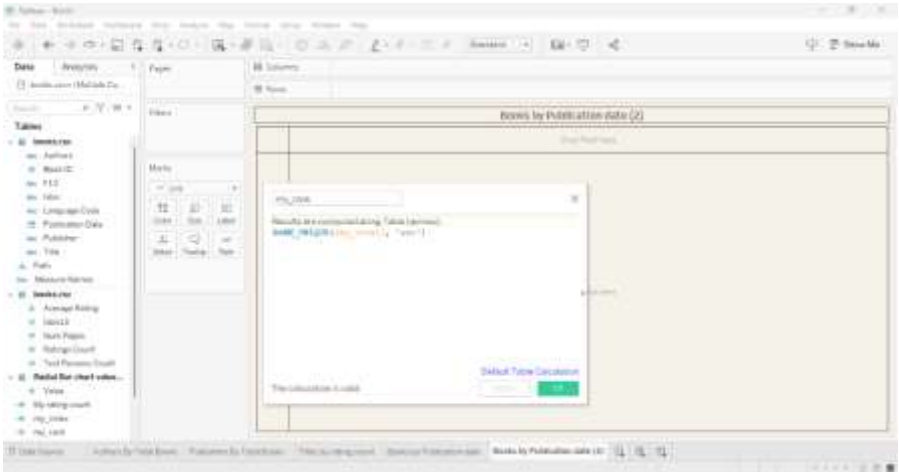


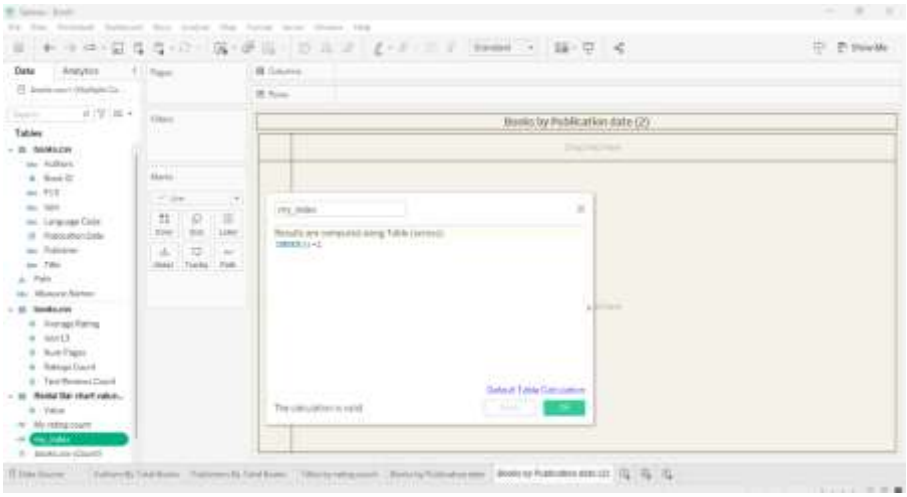
a) Authors By Total Books – Use COUNT DISTINCT(Book ID) with Authors
b) Publishers By Total Books - Use COUNT DISTINCT(Book ID) with Publisher
c) Titles by rating count- Created a calculation My rating count. Use SUM(My rating count) with Title

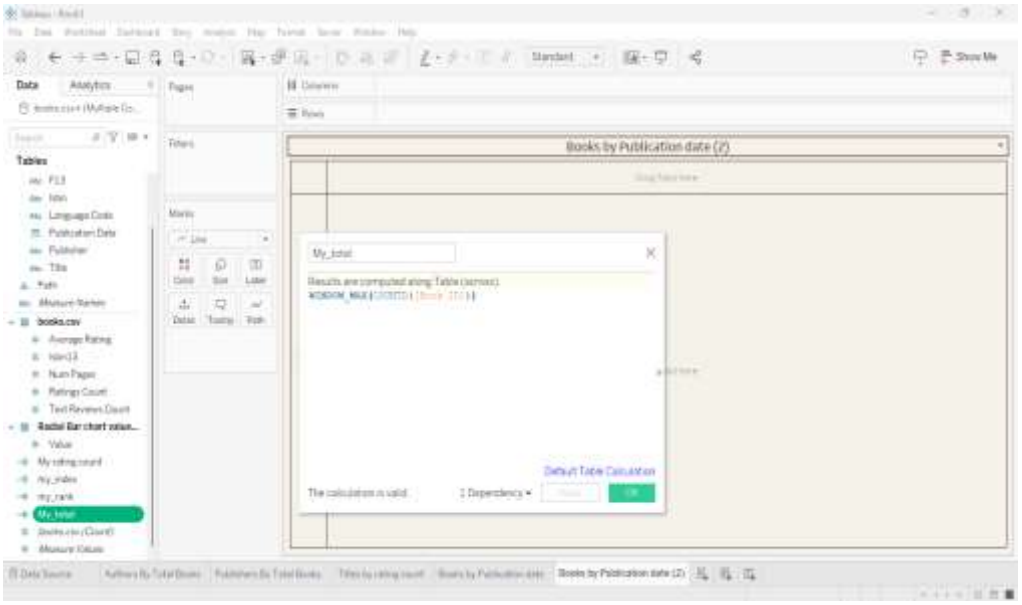


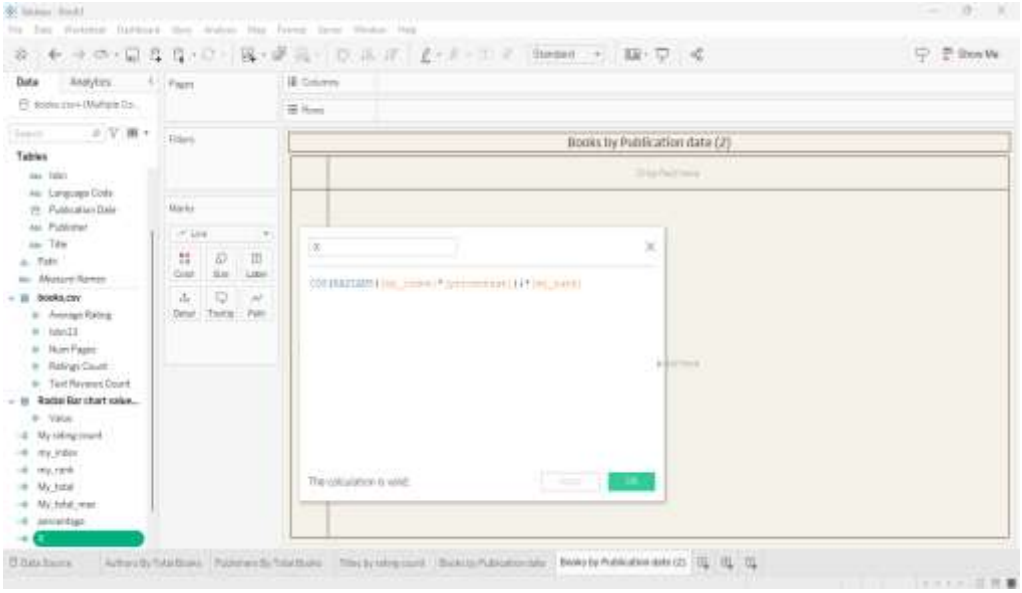
d) Books by Publication date-Use YEAR(Publication Date) with COUNT DISTINCT(Book ID)

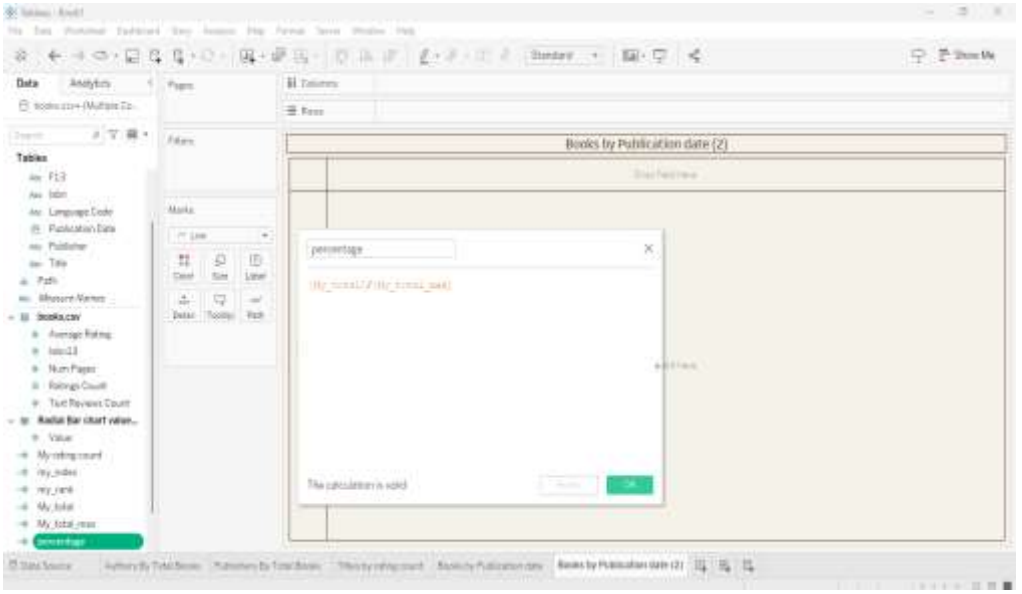
e) Top 10 Language code- Created 3 calculated field X,Y, My_total. Using X, Y and My_total with Language Code and Path.

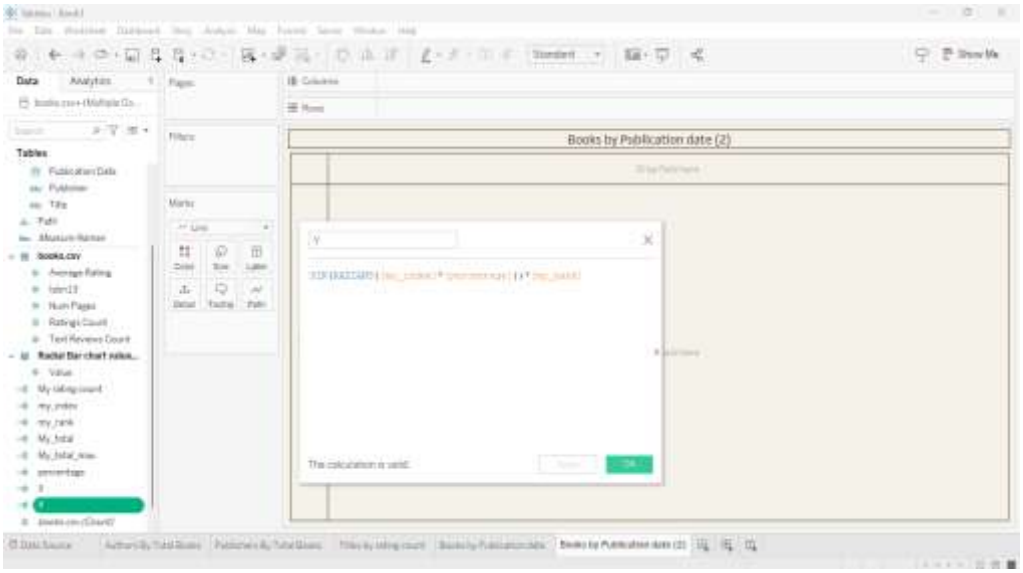




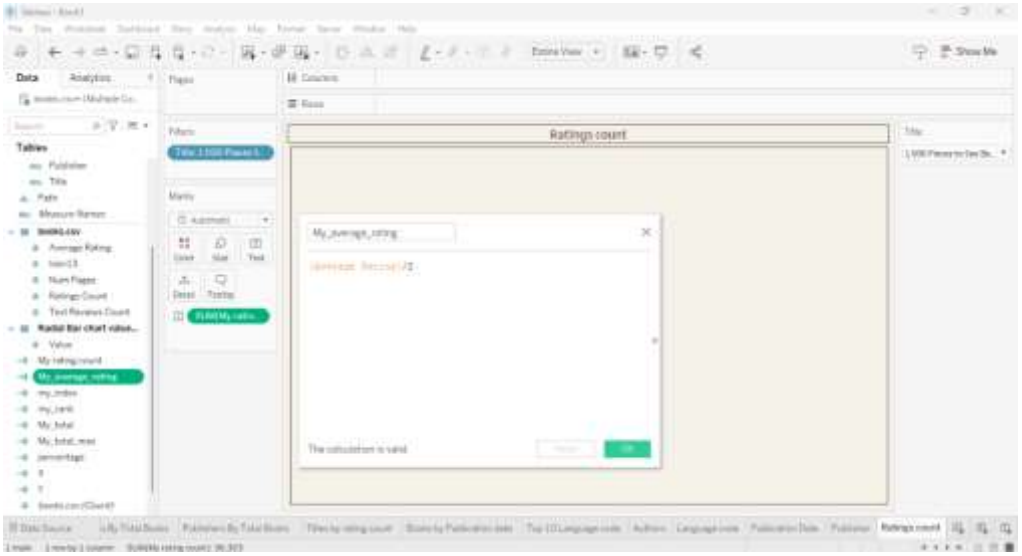

Books by Publication date (2)
My_total
Results are computed along Table (across).
Default Table Calculation
The calculation is valid.
OK

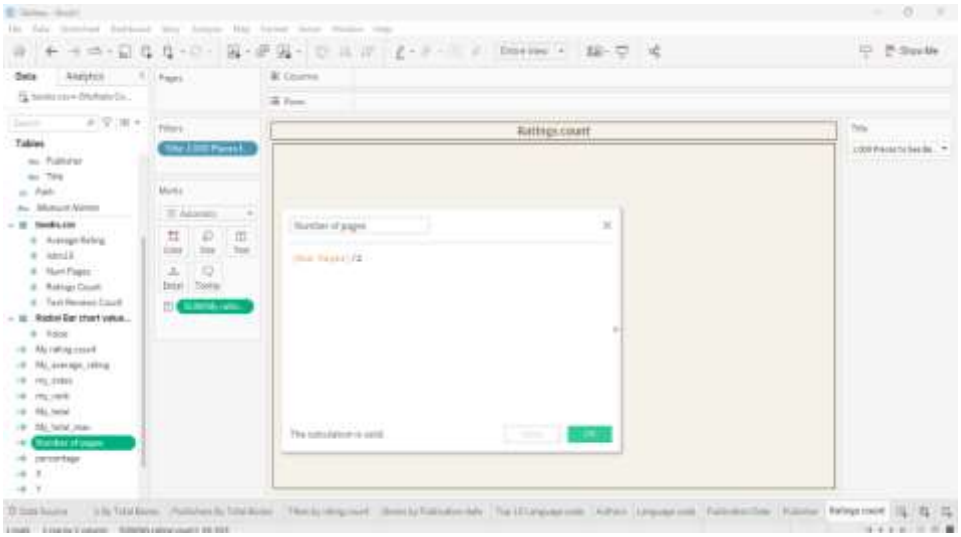

Books by Publication date (2)
The calculation is valid.
OK

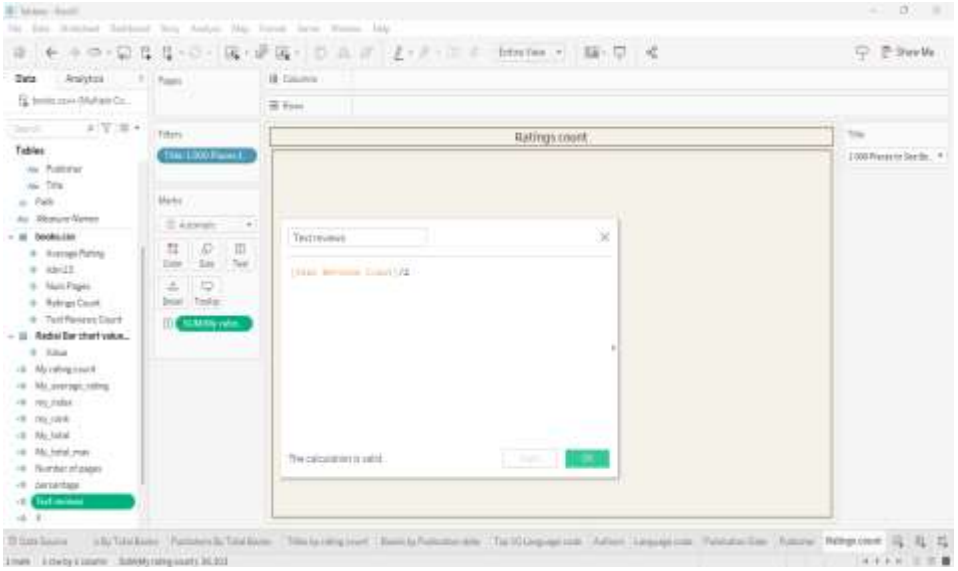

Books by Publication date (2)
percentage
The calculation is valid.
OK



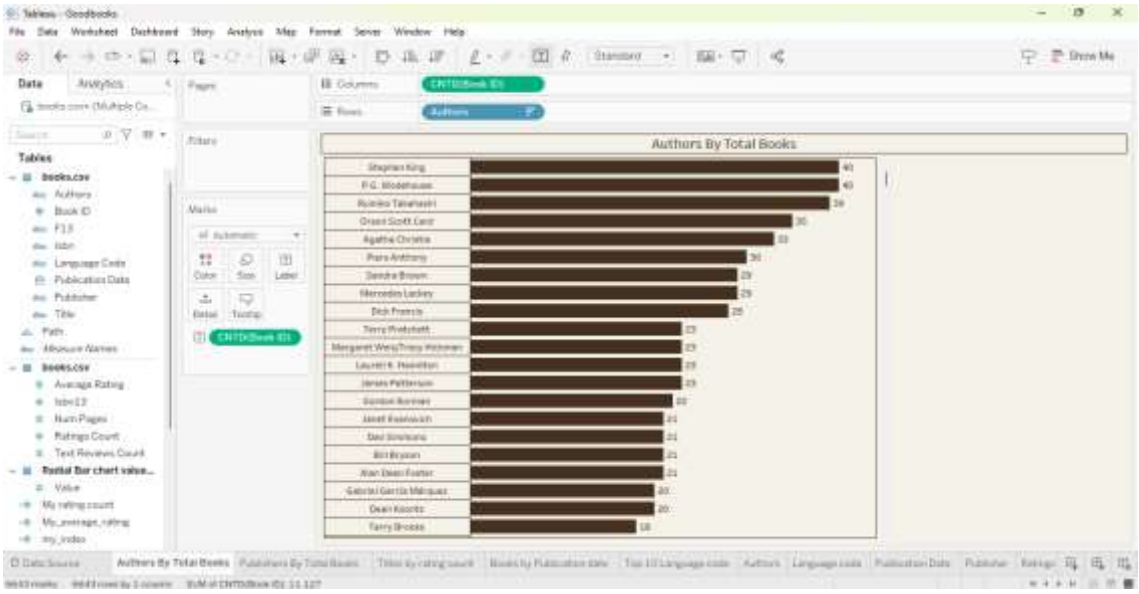
f) Using Text labels created different sheet for small analysis:

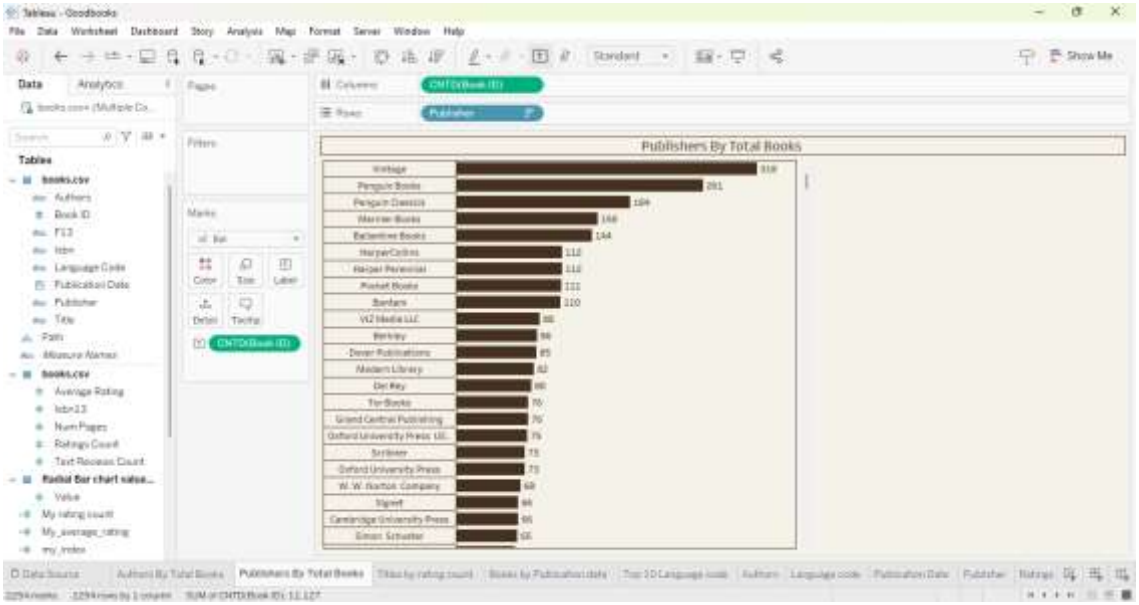


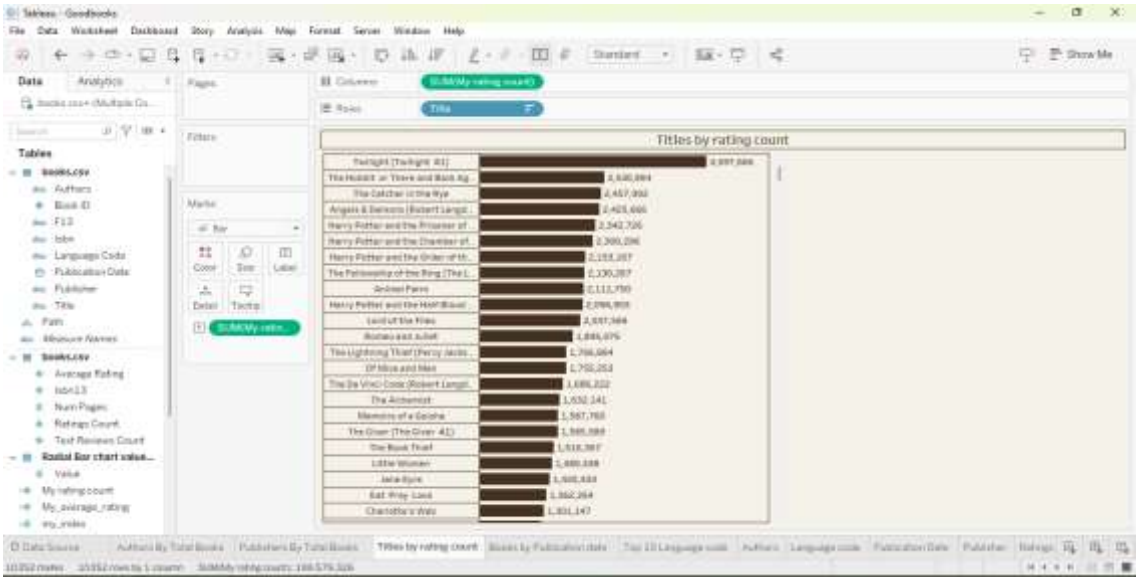

Ratings count
Number of pages
The calculation is valid.
OK

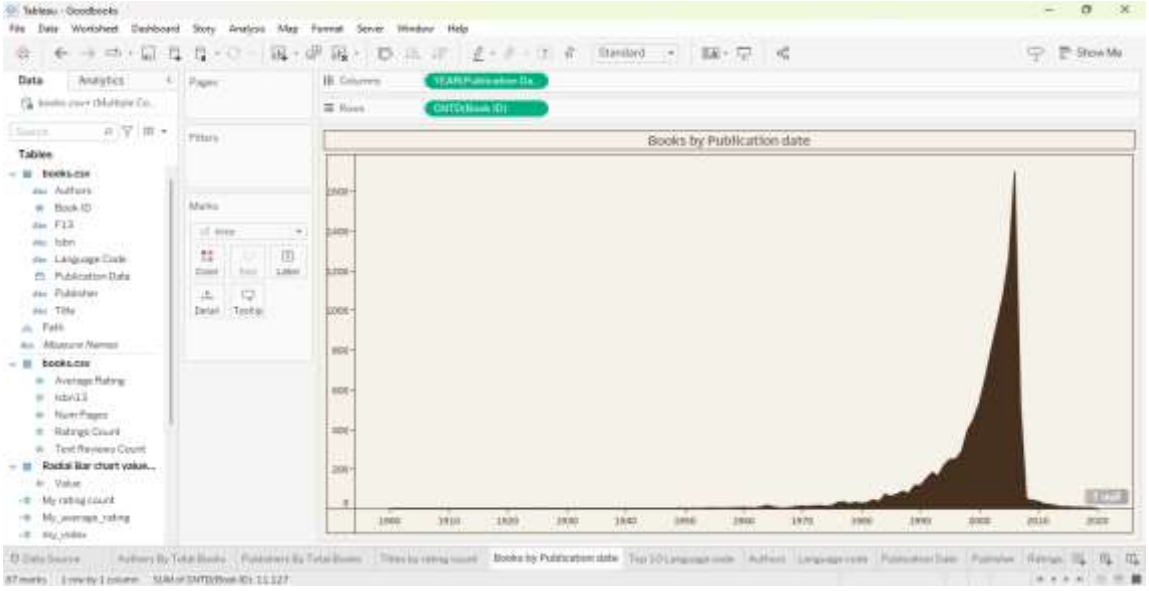

Ratings count
Text reviews
The calculation is valid.
OK

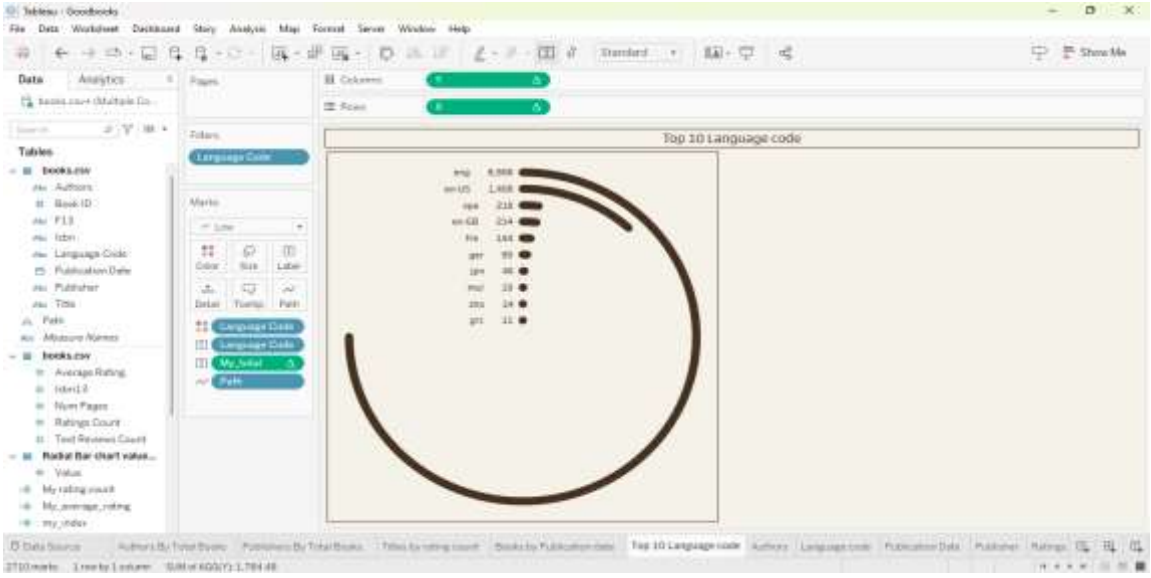
iv. Analysis result

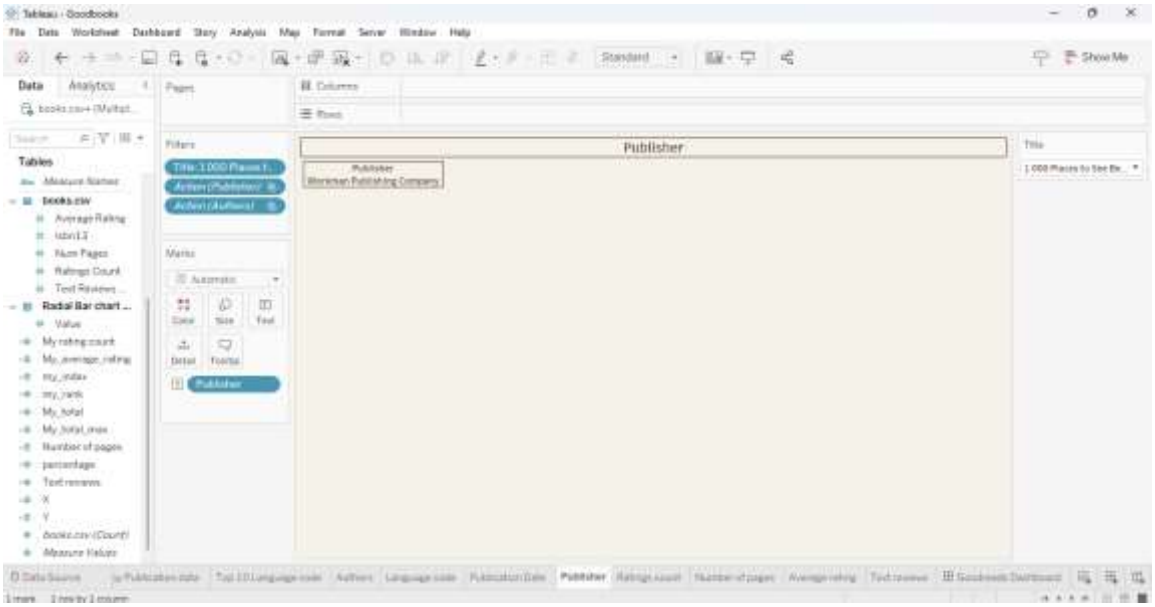


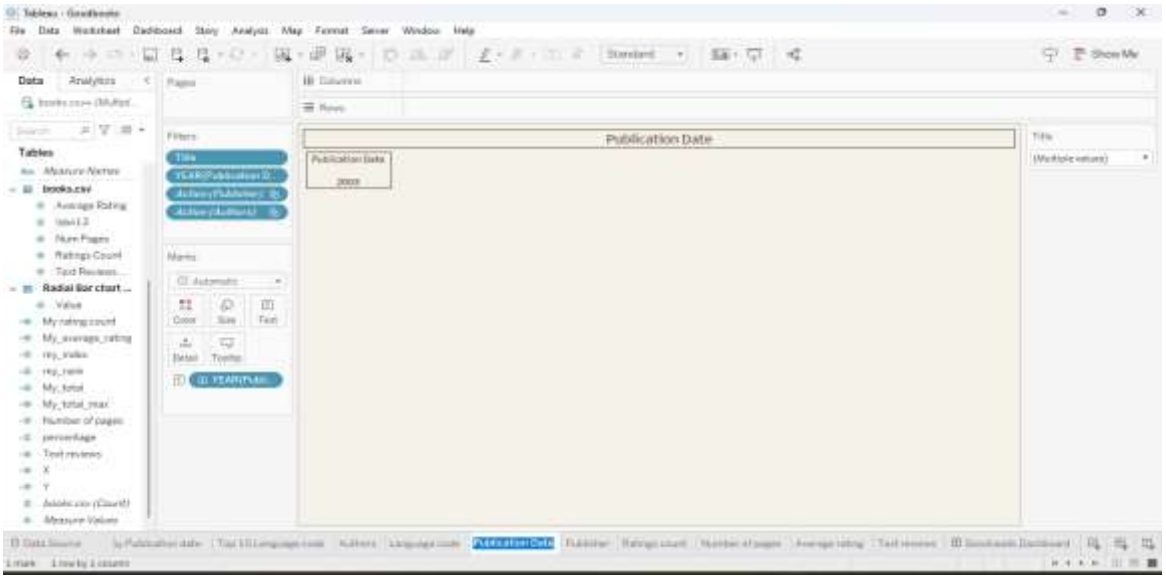


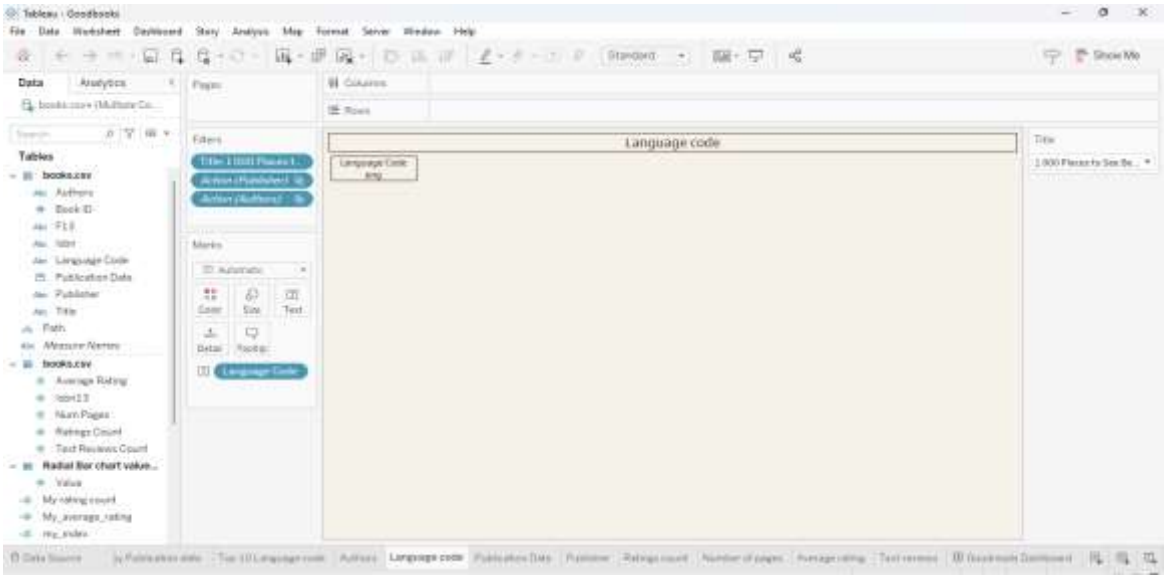

Titles by rating count

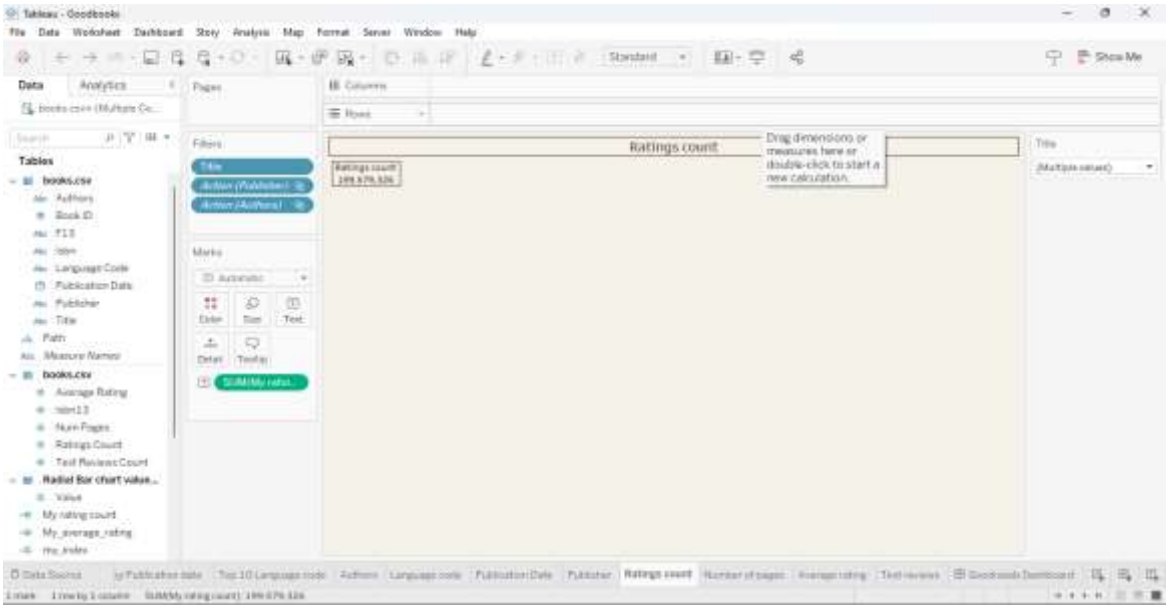

Books by Publication date

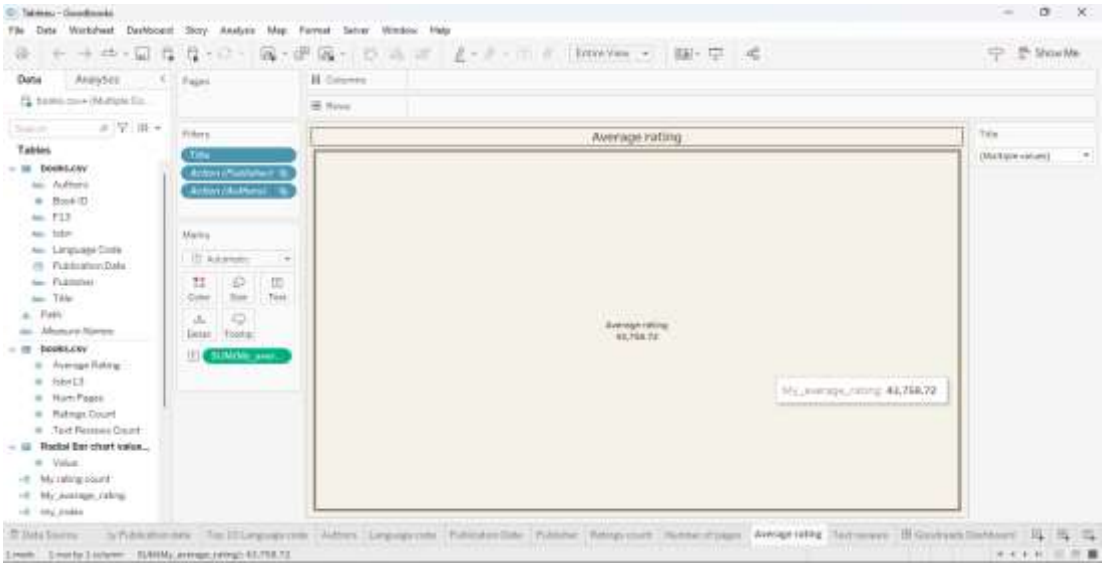

Top 10 Language code

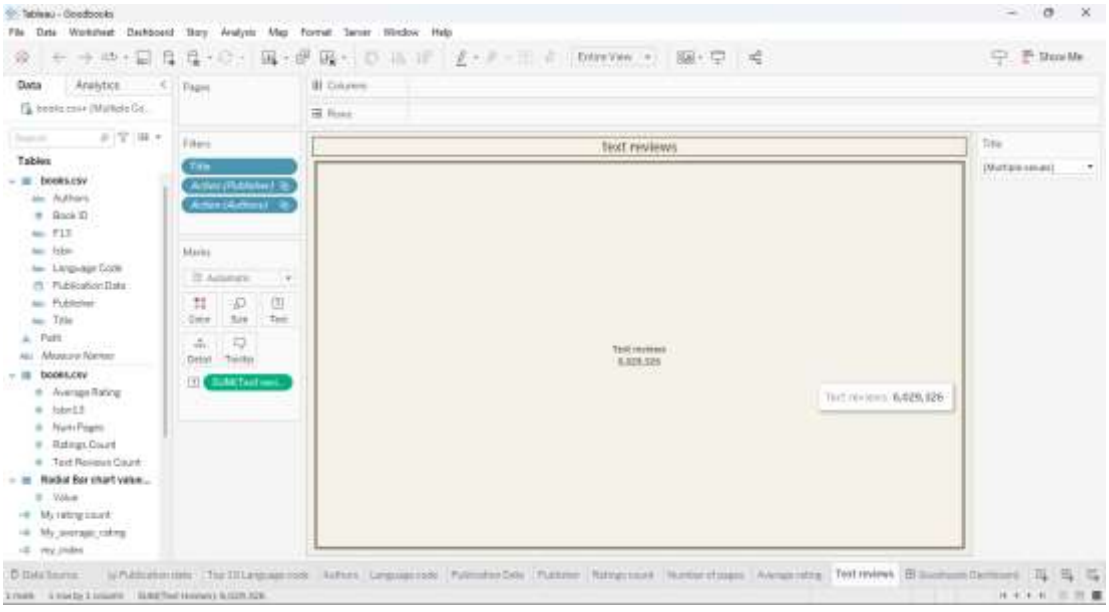

Publisher

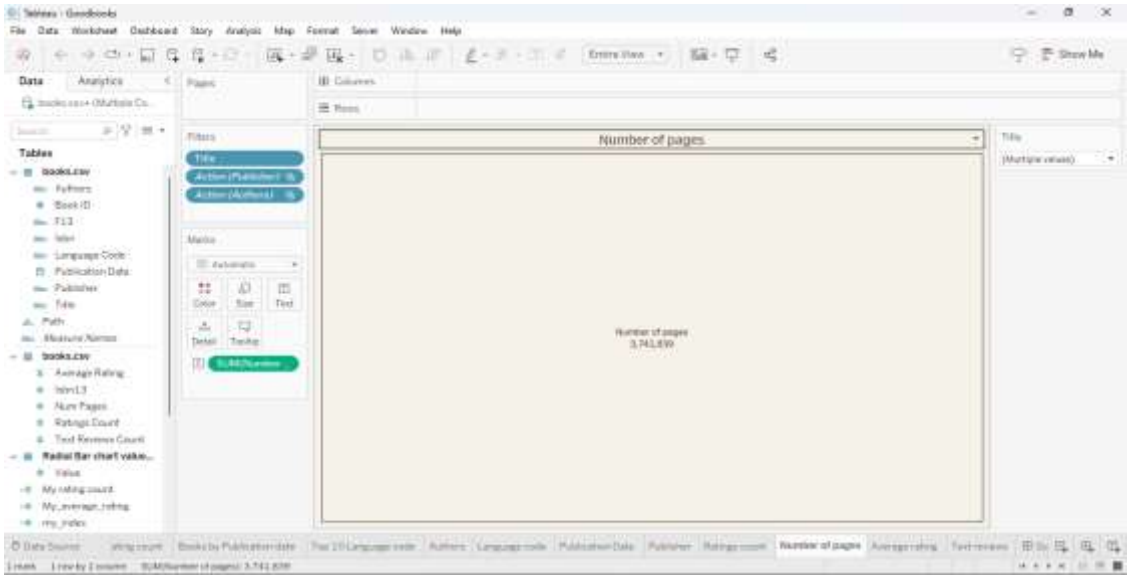

Publication Date

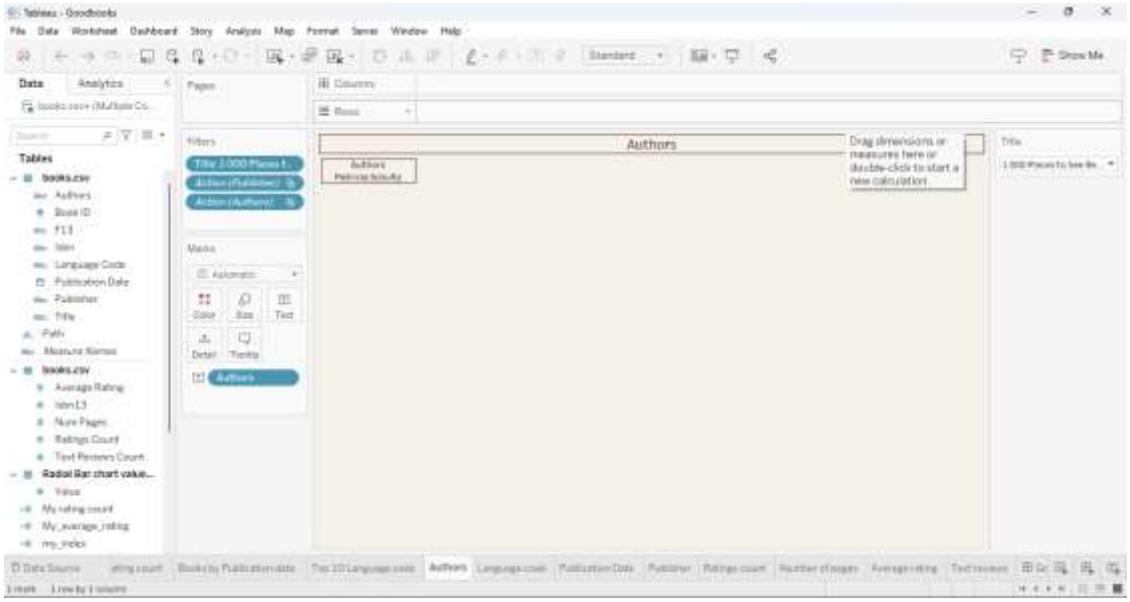

Language code


Ratings count
Drag dimensions or measures here or double-click to start a new calculation.


Average rating


Text reviews

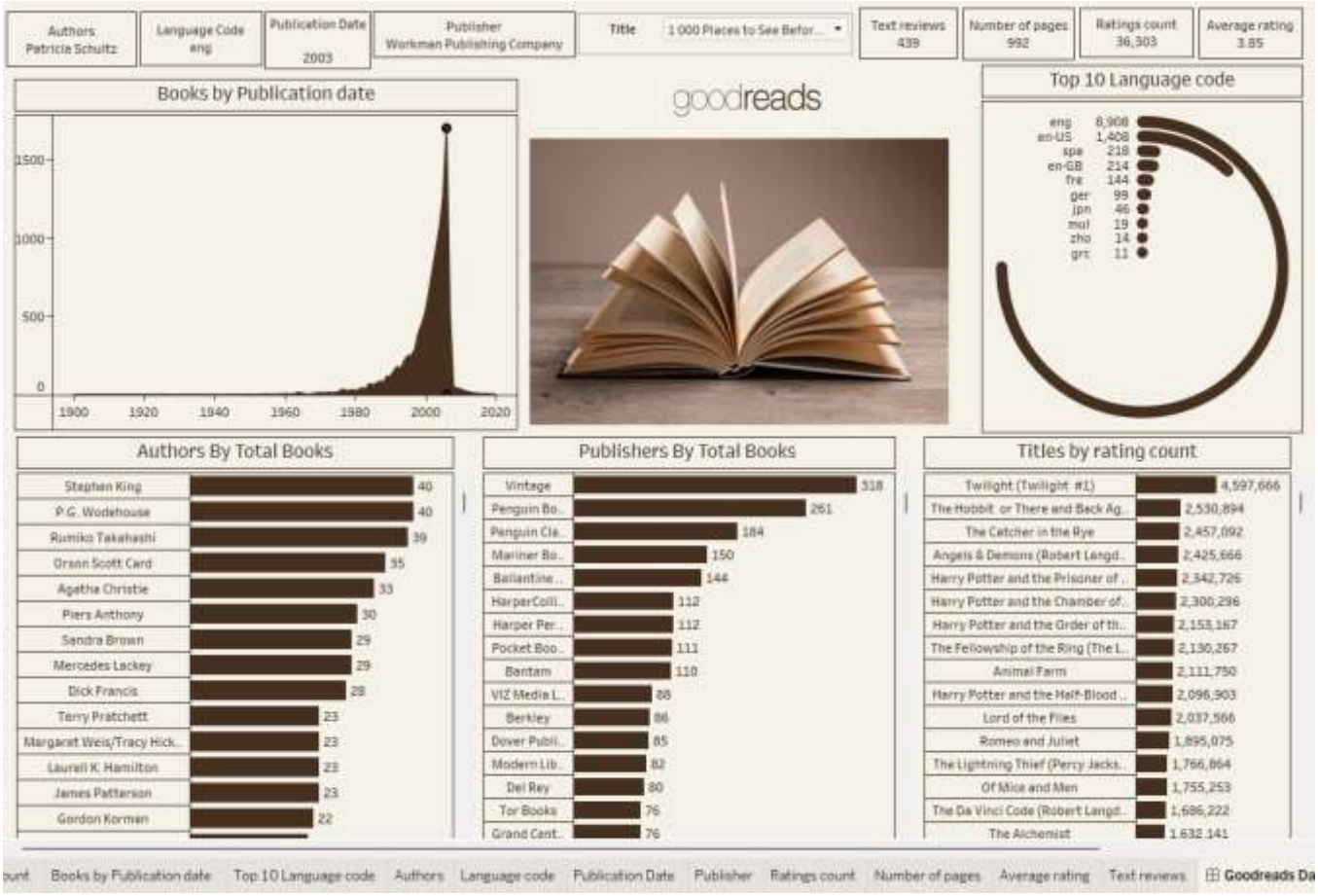

Number of pages
Number of pages
3,741,839


Authors
Drag dimensions or measures here or double-click to start a new calculation

v. Visualization (Dashboard)

Link for Project:

https://drive.google.com/file/d/1WZbgh2Y5EV4ki_QjooaOiGy47VKwykZB/view?usp=sharing



vi. Drawbacks or limitations of existing system

Data Connectivity and Performance:

Data Source Limitations: Tableau excels with clean and structured data. If the underlying data for the Books Dashboard is messy, siloed, or requires complex transformations, it might be cumbersome to prepare for analysis in Tableau.

Large Datasets: While Tableau can handle significant datasets, very large datasets can strain performance. This can lead to slow loading times and sluggish interactions with the dashboard.

Real-time limitations: While Tableau can connect to live data sources, it refreshes data periodically. This may not be ideal if you need real-time insights into book sales or inventory.

Usability and Scalability:

Complexity for Non-Analysts: Tableau dashboards can become complex with many filters and interactions. This can be overwhelming for casual users who just need a quick overview.

Limited Customization: While Tableau offers customization options, it might not be as flexible as coding a dashboard from scratch. This could limit the creation of highly specific or unique visualizations.

Scalability for Many Users: If a large number of users need to access the Books Dashboard concurrently, Tableau Server might be required for proper performance and security. This adds additional cost and complexity.

Advanced Analytics:

Limited Statistical Analysis: Tableau is primarily a data visualization tool. While it offers some statistical functions, it might not be suitable for complex statistical modeling or hypothesis testing.

Machine Learning Integration: Tableau can integrate with some machine learning models, but the level of integration may be limited. This could restrict the ability to use AI-powered insights within the dashboard.

Additional Considerations:

Cost: Depending on the number of users and the deployment method (Tableau Desktop vs Server), there can be associated licensing costs.

Vendor Lock-in: Once a dashboard is built in Tableau, it can be difficult to migrate to another platform.

FUTURE SCOPE

Deeper Data Exploration:

Advanced Drill-Downs: Implement features that allow users to drill down into specific data points to uncover granular details. This could involve nested hierarchies, expandable sections, or context menus.

Trend Analysis: Utilize features like forecasting lines, sparklines, or chart annotations to highlight trends and patterns in book sales, inventory, or other key metrics.

Cohort Analysis: Segment user data by purchase history, demographics, or other relevant factors to understand different reader segments and tailor marketing strategies.

Elevated User Experience:

Interactive Filters: Incorporate interactive filters that update visualizations in real-time, allowing users to explore data from various angles quickly.

Storytelling with Data: Design dashboards that tell a data-driven story. Use annotations, color-coding, and clear visualizations to guide users through insights.

Dashboard Actions: Enable users to trigger specific actions within the dashboard, such as filtering related data tables or exporting subsets of data for further analysis.

Integration and Analytics:

Live Data Connections: Explore the possibility of connecting to real-time data sources for up-to-the-minute insights into book sales, inventory, or social media trends.

Predictive Analytics: If you have historical sales data, consider integrating with Tableau Prep or other tools to create predictive models that forecast future sales trends.

Location Intelligence: If your data includes geographical information, leverage Tableau's mapping capabilities to visualize book sales or reader demographics by location.

Additional Considerations:

Mobile Optimization: With the rise of mobile usage, consider optimizing the dashboard for viewing on tablets and smartphones.

Accessibility Features: Ensure the dashboard is accessible to users with disabilities by incorporating features like proper color contrast, alternative text descriptions for images, and keyboard navigation.

Security and Governance: If the dashboard contains sensitive data, implement proper security measures to restrict access and control user permissions.

Advanced Data Integration:

External Data Sources: Connect the dashboard to external data sources like social media sentiment analysis or weather data to understand how external factors influence book sales.

Text Analysis: If you have access to book reviews or author interviews, explore integrating with natural language processing (NLP) tools to analyze sentiment, topic trends, or emerging genres.

Collaboration and Sharing:

Shared Dashboards: Utilize Tableau Server functionalities to create collaborative dashboards where multiple users can annotate, share insights, and build upon each other's work.

Data Storytelling with Animations: Leverage animation features within Tableau to create dynamic dashboards that guide users through a data narrative, highlighting key trends and insights.

AI and Automation:

Smart Alerts: Set up automated alerts within the dashboard that notify users of anomalies or significant changes in sales figures, inventory levels, or social media trends.

AI-Powered Recommendations: Integrate with machine learning models to recommend books to users based on their purchase history, genre preferences, or similar reader demographics.

REFERENCES

Privacy and Data Access: Goodreads doesn't publicly offer their data for download. You'd likely need to have a business relationship with Goodreads to access their data for building a dashboard.

Focus on User Data: Goodreads is primarily focused on user data like reviews, ratings, and friend connections. While interesting to analyze, this data wouldn't be readily available through Tableau's public data sources.

General Book Sales Data: You can explore public datasets on online bookstores like Amazon or Barnes & Noble's open data platform (https://www.businesswire.com/news/home/20230531005958/en/Barnes-Noble-Education-Announces-the-Sale-of-Its-DSS-Segment-and-Provides-Certain-Preliminary-Fiscal-Year-2023-Results). These datasets might include information on book sales, genres, and author information.

Tableau Resources: The Tableau website offers a wealth of resources for learning Tableau and building dashboards. Here are some starting points:

Tableau Public Gallery: https://www.tableau.com/viz-gallery (Explore dashboards built by the Tableau community for inspiration)

Tableau Blog: https://www.tableau.com/blog (Find articles on best practices, data visualization techniques, and specific use cases)

Books on Tableau: While not specific to Goodreads data, several books offer excellent guidance on building dashboards with Tableau. These books might include examples relevant to the book industry:

Learning Tableau by Joshua Milligan (https://www.amazon.com/Learning-Tableau-2020-visualizations-organization/dp/1800200366)

Tableau Your Data! by Daniel Murray (https://www.amazon.com/Tableau-Your-Data-Analysis-Software/dp/1119001196)

The Big Book of Dashboards by Steve Wexler (https://www.amazon.com/Big-Book-Dashboards-Visualizing-Real-World/dp/1119282713)

BIBLIOGRAPHY

1. Resources on Building Book Dashboards with Tableau:

Articles:

"http://www.markwk.com/reading-data-visualization.html" by Mark Koester: This article details how the author used Goodreads data exported to a CSV file to explore reading trends in Tableau. While it doesn't directly connect to Goodreads, it showcases the analysis of book data within Tableau.

2. Resources on Learning Tableau:

Tableau Learning Path: https://www.tableau.com/learn

Tableau Public Gallery: https://www.tableau.com/viz-gallery

Tableau Blog: https://www.tableau.com/blog

These resources will equip you with the skills to build dashboards in Tableau, which can then be applied to analyse book data from alternative sources.

3. Books on Tableau:

Learning Tableau by Joshua Milligan (https://www.amazon.com/Learning-Tableau-2020-visualizations-organization/dp/1800200366)

Tableau Your Data! by Daniel Murray (https://www.amazon.com/Tableau-Your-Data-Analysis-Software/dp/1119001196)

The Big Book of Dashboards by Steve Wexler (https://www.amazon.com/Big-Book-Dashboards-Visualizing-Real-World/dp/1119282713) These books provide a comprehensive foundation for building dashboards in Tableau, applicable to various data sets, including potentially book data.

4. Alternative Book Data Sources:

Amazon Open Data: While not directly from Goodreads, Amazon offers open data on book sales and related information. You can explore this data for building a book dashboard in Tableau: https://www.businesswire.com/

Remember: While Goodreads data might not be readily available, these resources can equip you with the knowledge and tools to build a book dashboard using alternative data sources and Tableau.

5. Data Analysis of Books and Reviews:

Books:

Beautiful Data by Colin Ware (https://www.amazon.com/Beautiful-Data-Stories-Elegant-Solutions/dp/0596157118) : This book delves into the fundamentals of data visualization, offering valuable insights into presenting book data effectively.

R for Everyone by Jared P. Lander (https://www.amazon.com/Everyone-Advanced-Analytics-Graphics-Addison-Wesley/dp/0321888030) : While not specific to Tableau, this book explores data analysis with R, a powerful tool you could potentially use to analyze Goodreads data (if it were accessible) or similar book review datasets.

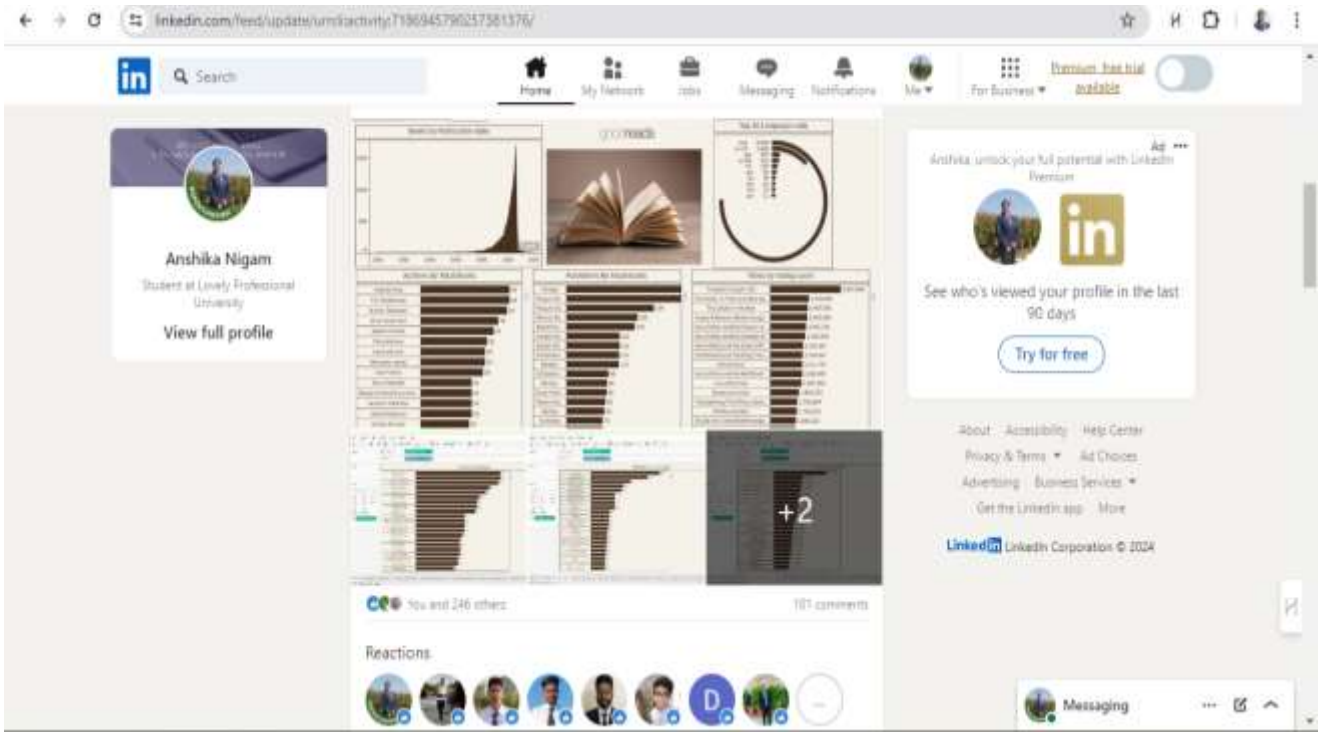
Mining the Social Web by Matthew A. Russell (https://www.amazon.com/Mining-Social-Web-Facebook-LinkedIn/dp/1449367615) : This book explores techniques for analyzing social media data, which could be relevant if you consider user reviews from Goodreads or other platforms as a data source (assuming ethical considerations are addressed).

Articles:

"https://arxiv.org/abs/2311.11250" by S. P. Mukherjee and P. K. Pal: This research paper explores sentiment analysis of book reviews, showcasing techniques that could be applied to analyze Goodreads data (if accessible).

"https://ashima96.medium.com/building-a-book-recommendation-system-a98c58a4f1bb" by Sangeetha Anandakrishnan: This article discusses recommender systems for books, a concept you could explore when building a book dashboard using user data (with appropriate privacy considerations).

SCREENSHOT OF LINKEDIN POST



Achieved 250+ and 100+ comments for my work and got appreciated by the IT community.

Link to Post: https://www.linkedin.com/posts/anshikanigam19_datavisualization-goodreads-books-activity-7186945790257381376-JD7O?utm_source=share&utm_medium=member_desktop

Till Now More to Go…Thank You.