Externship Program - Data Analytics

Amazon Top 50 Bestselling Books Analytics using IBM Cognos Analytics

A PROJECT REPORT

Submitted by:

SHENDKAR NARENDRA JANARDHAN

19BME0298

Mentor Name : **Uma Rani**

University : VIT Vellore

Category: Data Analytics

Skills Required: IBM Cognos Analytics

Cognos Link :
https://eu2.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2Fbestsellers%2Bwith%2Bcategories%2BDashboard&action=view&mode=dashboard&subView=model000001812e700a06_00000002

CHAPTER NO.	TITLE	PAGE NO.
1	**INTRODUCTION** **1.1 Overview** **1.2 Purpose**	 **3** **3**
2	**LITERATURE SURVEY** **2.1 Existing Problem** **2.2 Proposed Solution**	 **4** **4**
3	**THEORETICALANALYSIS** **3.1 Block Diagram**	 **5**
4	**FLOWCHART**	**6**
5	**RESULT**	**7**
6	**ADVANTAGES AND DISADVANTAGES**	**16**
7	**APPLICATIONS**	**17**
8	**DASHBOARD**	**18**
9	**CONCLUSION**	**19**

Amazon top 50 bestselling books - Data analysis

Introduction

"Amazon.com, Inc. is a Seattle, Washington-based international technology corporation that specialises in e-commerce, cloud computing, digital streaming, and artificial intelligence. Along with Google, Apple, Microsoft, and Facebook, it is considered one of the Big Five businesses in the United States' information technology industry. The corporation has been dubbed "one of the most powerful economic and cultural forces in the world," as well as "the most valuable brand on the planet.

About

Dataset on Amazon's Top 50 bestselling books from 2009 to 2019. Contains 550 books, data has been categorized into fiction and non-fiction using Goodreads

Purpose

The purpose of this notebook is to perform a data analysis on the dataset and extract insights that could help the business model grow, such as the bestselling genres, authors and a good price/rating relationship. Such insights have a great business value since knowing what is selling a what is not, the company can focus their marketing strategies to improve overall sells: customer might perceive a genre too expensive so they give a low rating or books with high ratings and low sells might not be visible to other customers.

Table of Contents

1. Data loading and data cleaning
2. 'Genre' performance as per 'User Rating'
3. Prices behaviour

4. Prices and User Rating
5. Which books give the best rating for money? 6. Which are the free books with the best rating?

LITERATURE REVIEW

EXISTING PROBLEMS

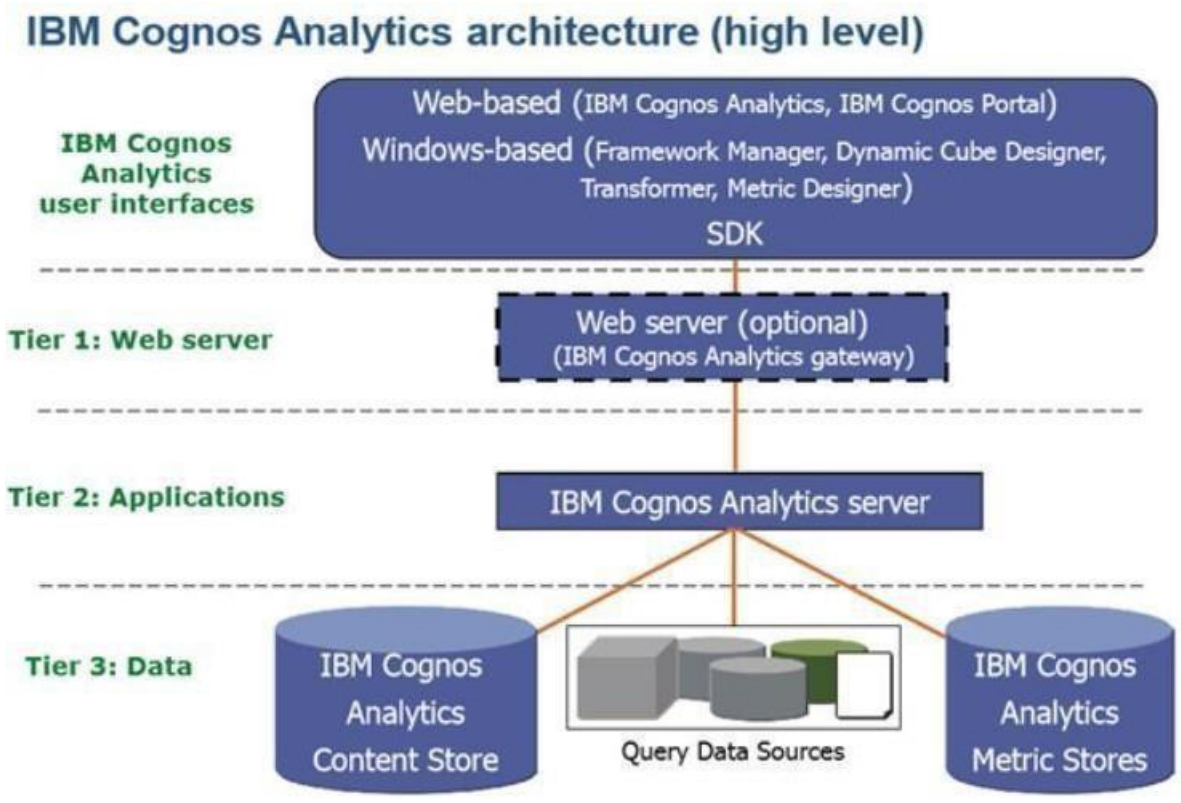
i. If our data analysis reveals strange patterns or our statistical significance is insufficient, we may not have enough data to draw reliable conclusions.
ii. We won't be able to examine the data before developing meaningful strategies unless we undertake data analysis.
iii. It is worthless without context, and we can't turn data into information without it.

PROPOSED SOLUTIONS

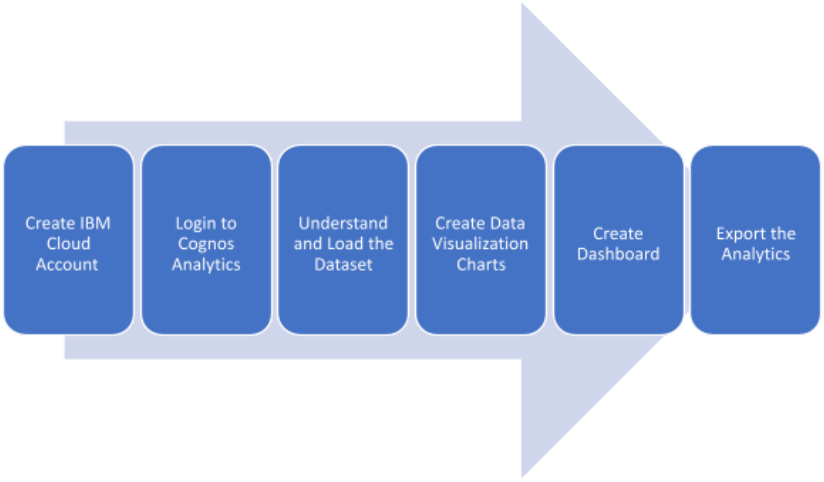
i. To use IBM Cognos to produce various data visualisations.
ii. Using IBM Cognos to create a dashboard
iii. Creating dashboards has the potential to alter our business's success and satisfaction.
iv. Information is pointless unless it can be applied to something.

TECHNICAL ANALYSIS

BLOCK DIAGRAM



FLOWCHART

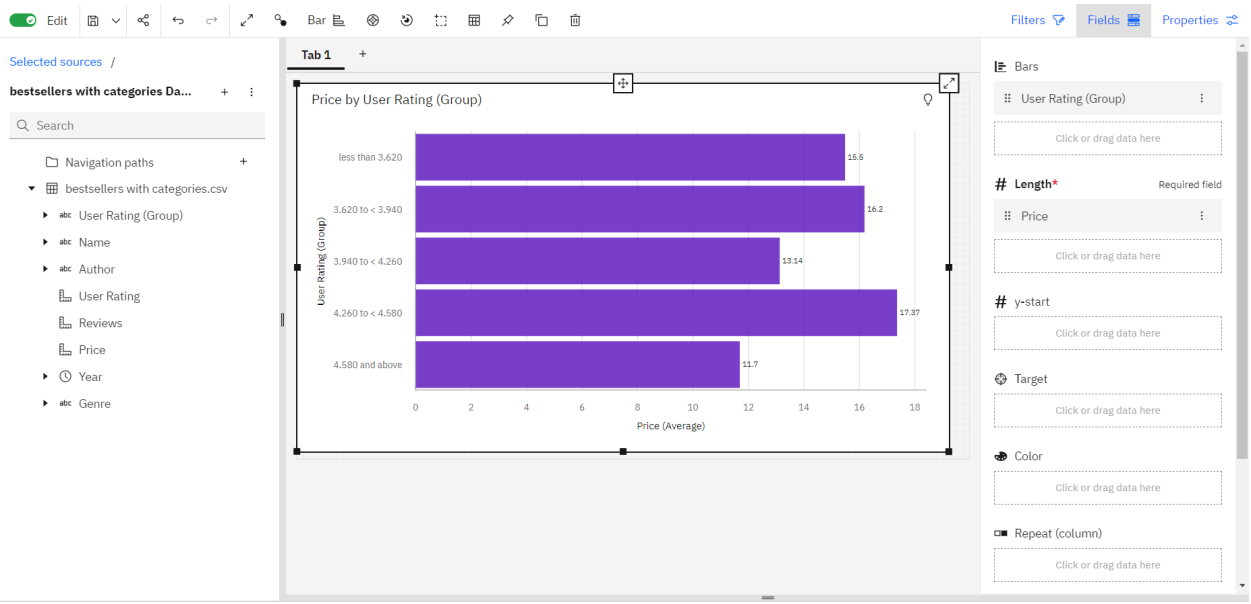

Create IBM Cloud Account
Login to Cognos Analytics
Understand and Load the Dataset
Create Data Visualization Charts
Create Dashboard
Export the Analytics

RESULT

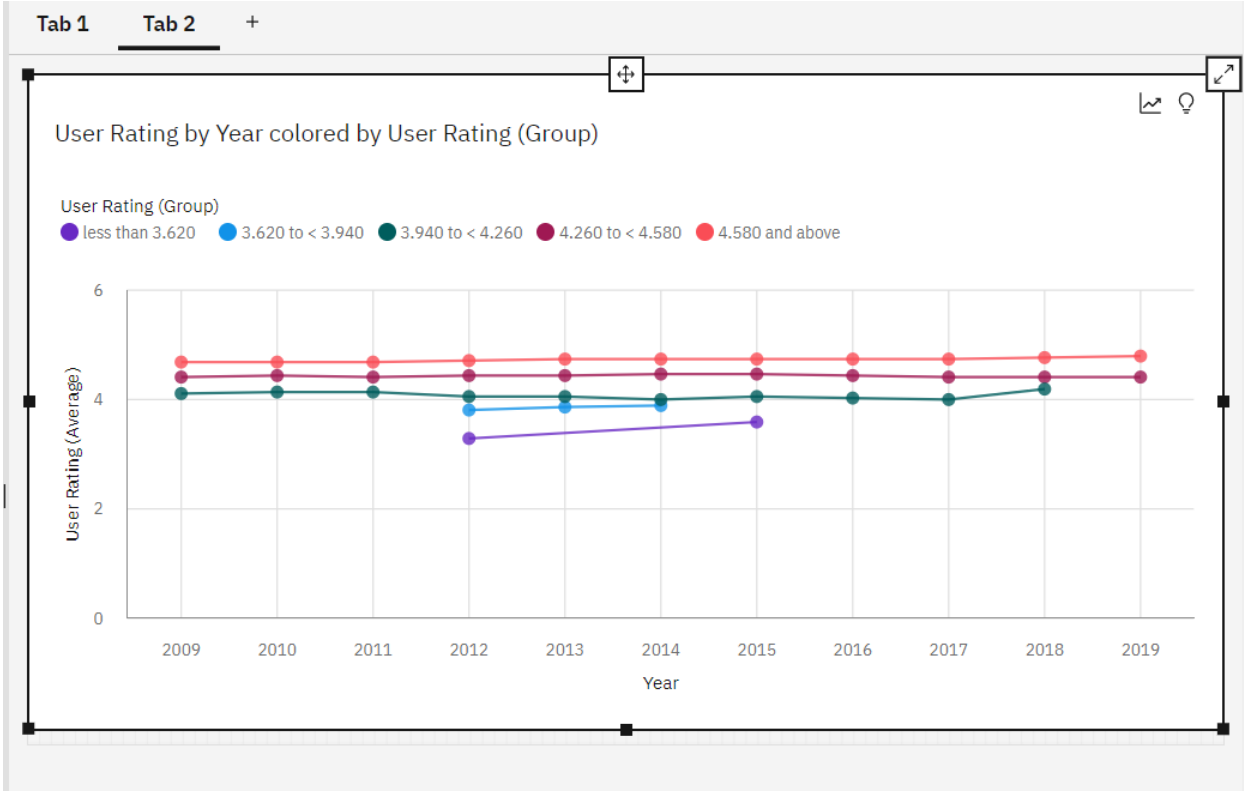
Data Visualization Charts

1 Price by User Rating Groups
2 Price by Year with User Rating Groups Color
3 Line Chart showing User Rating, Price, Year and No. of books.
4 Author-wise No of Books count.
5 User Rating Group wise No of Books.
6 Year-wise No of Books Published.
7 Genre-wise Price Wise No of Books.
8 Year wise Count of Name - Color by Genre.
9 Review Group (X) wise Count of Name -- Waterfall Chart.
10 Review Group (X) wise Count of Name -- Radar Chart.
11 Name Count - Card.
12 Author Count - Card.
13 Line Chart - Showing User Rating Group and Reviews Group wise No of Books with Genre.
14 Column Chart: Reviews Group wise Count of No of Books with Genre.
15 Tree Map: Year wise Count of No of Authors.
16 Hierarchy Bubble: User Group and Review Group wise No of Books by Genre.

1. Price by User Rating Groups

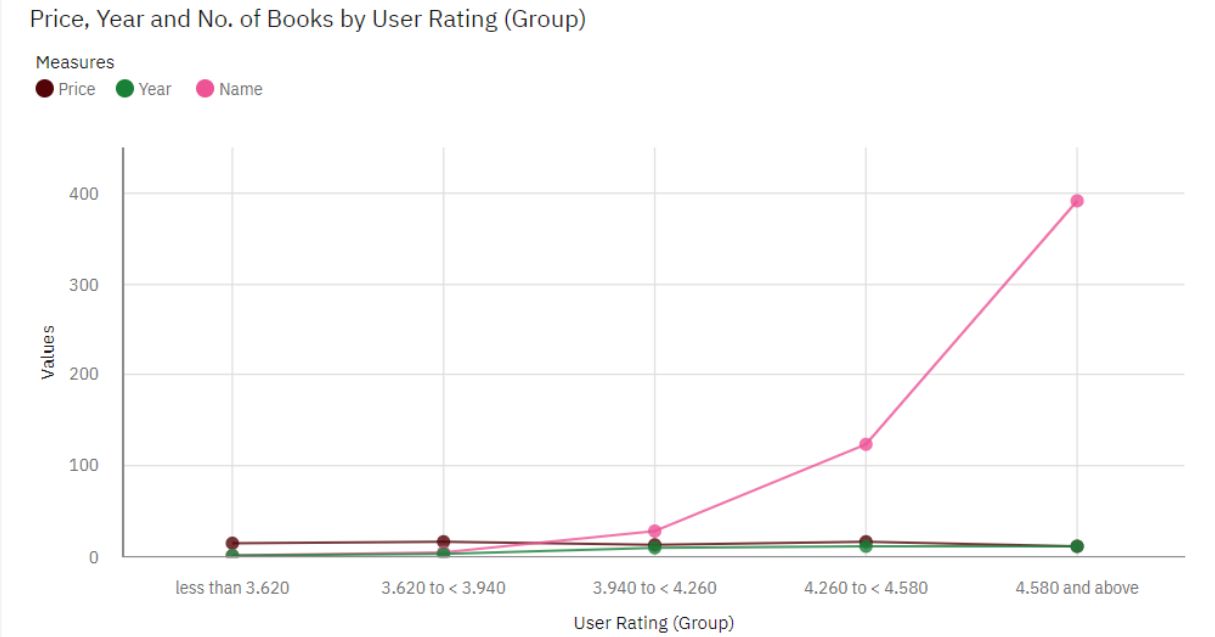


2. Price by Year with User Rating Groups Color



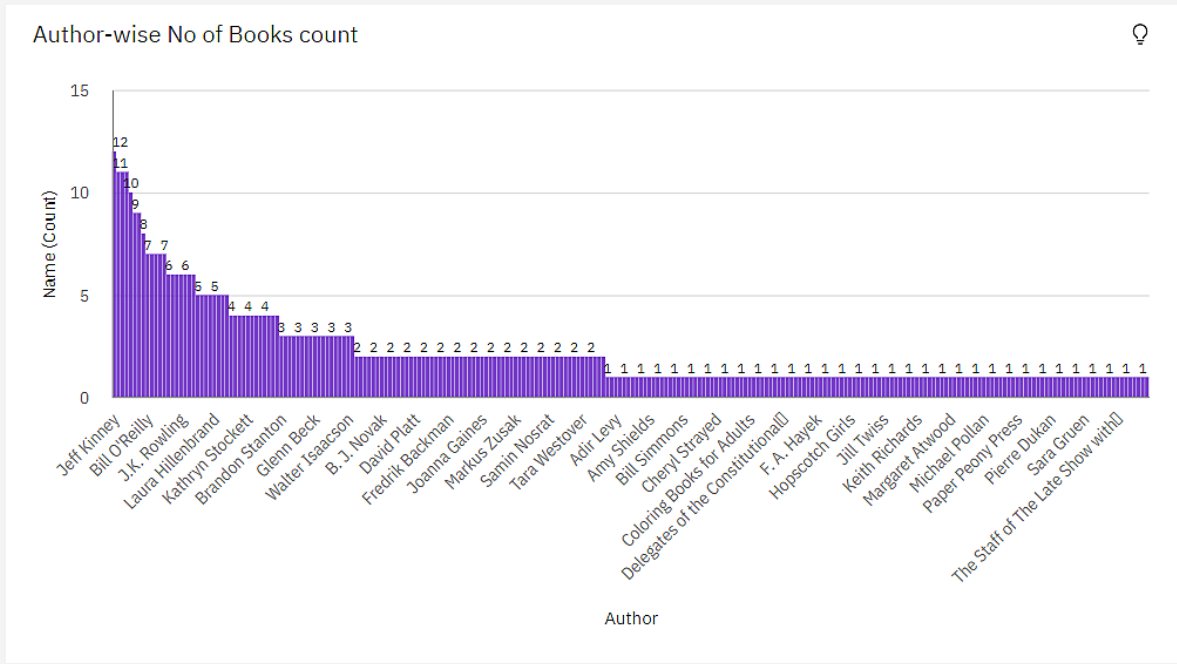
3. Line Chart showing User Rating, Price, Year and No. of books

Tab 1 Tab 2 Tab 3 +

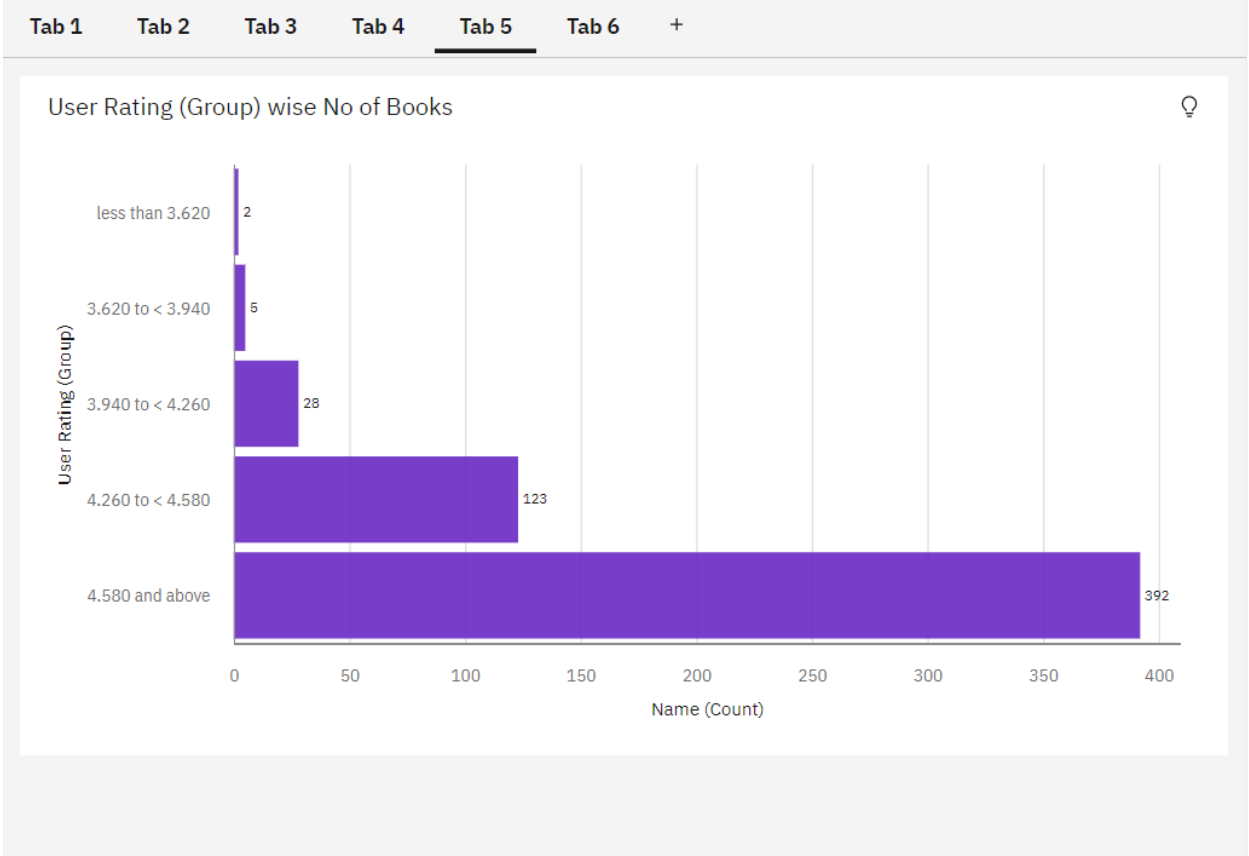


4. Author-wise No of Books count

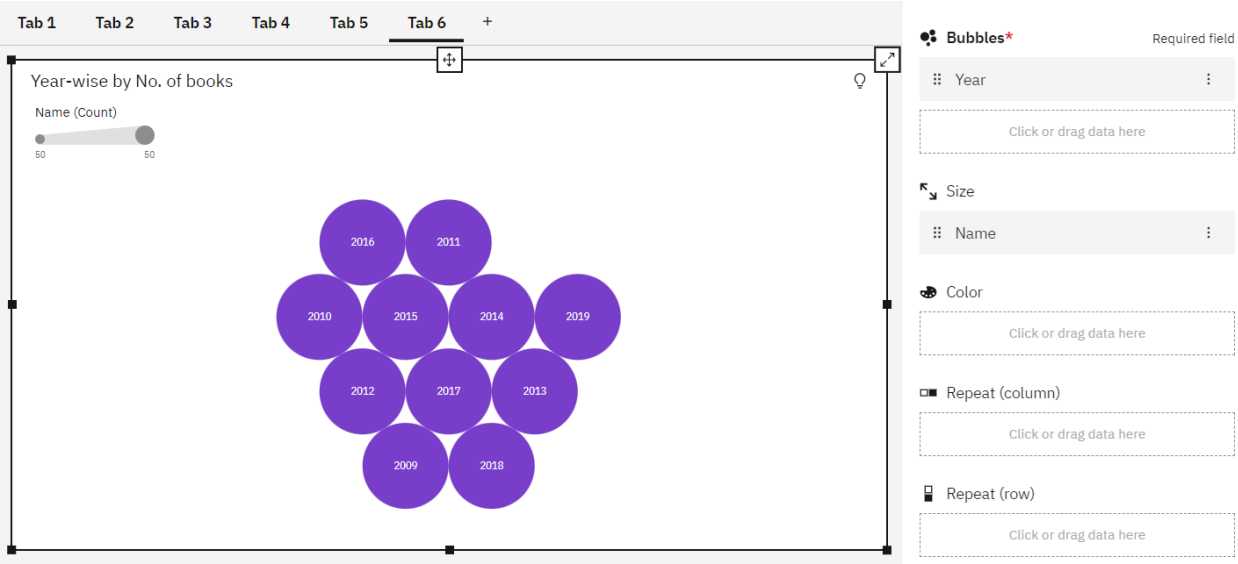
Tab 1 Tab 2 Tab 3 Tab 4 Tab 5 Tab 6 +



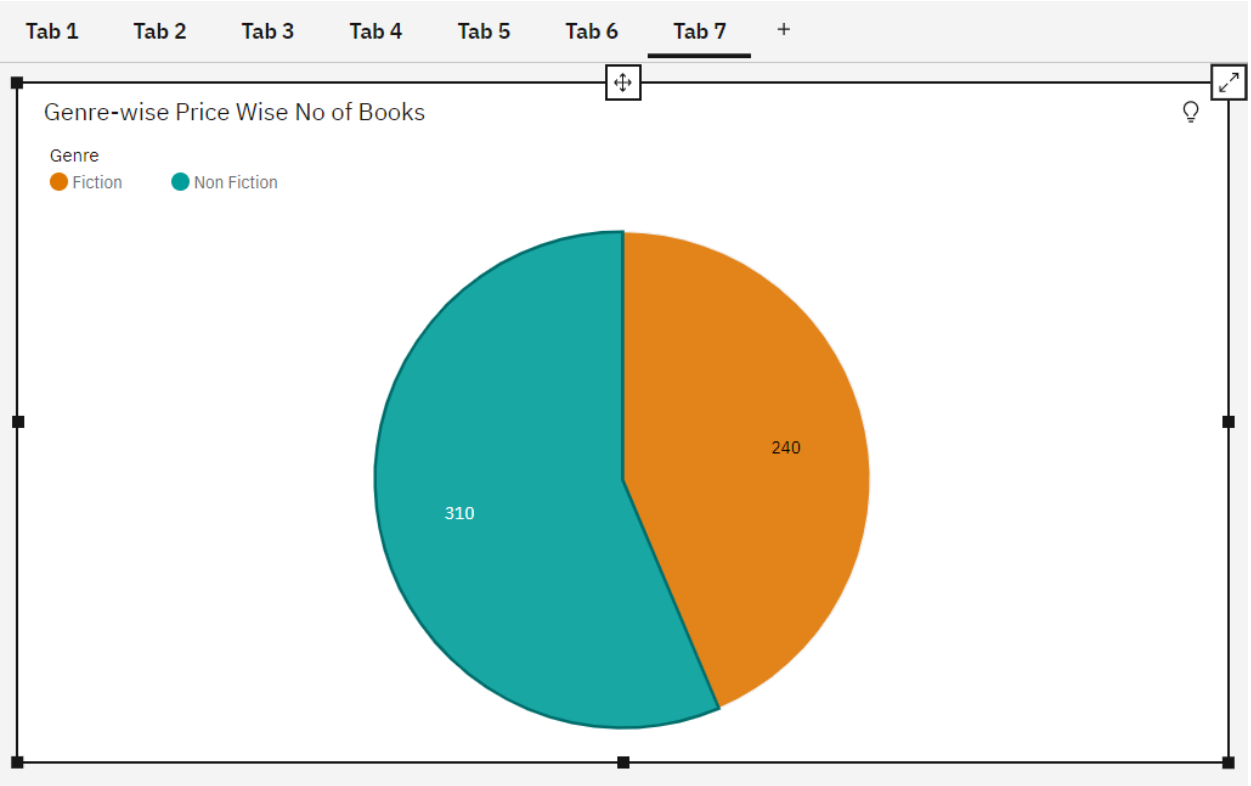
5. User Rating Group wise No of Books.



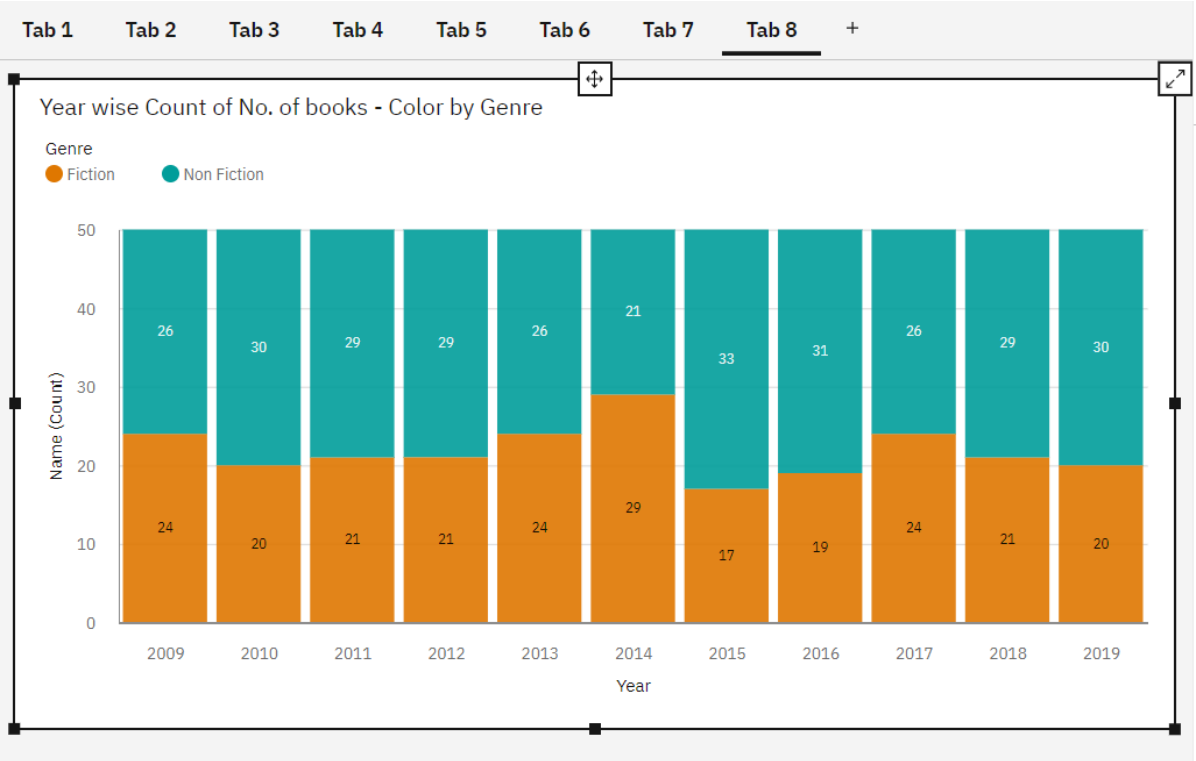
6. Year-wise No of Books Published.



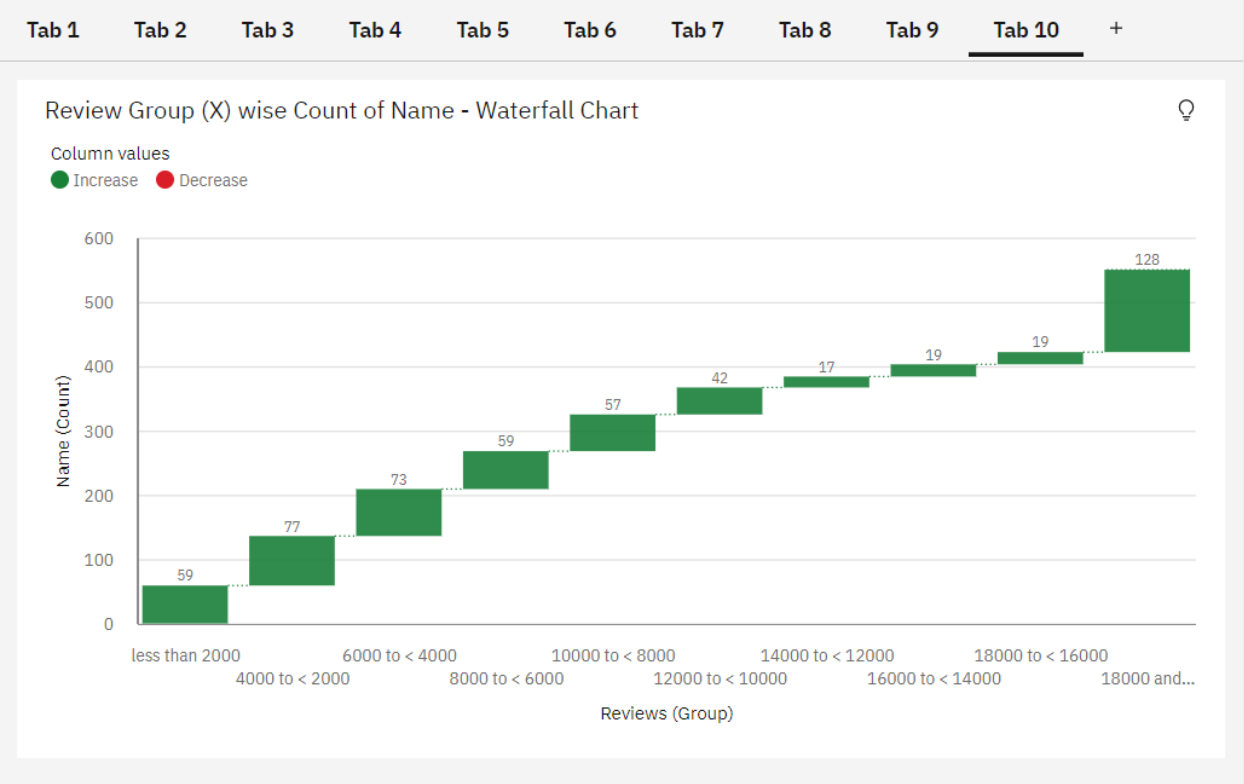
7. Genre-wise Price Wise No of Books.



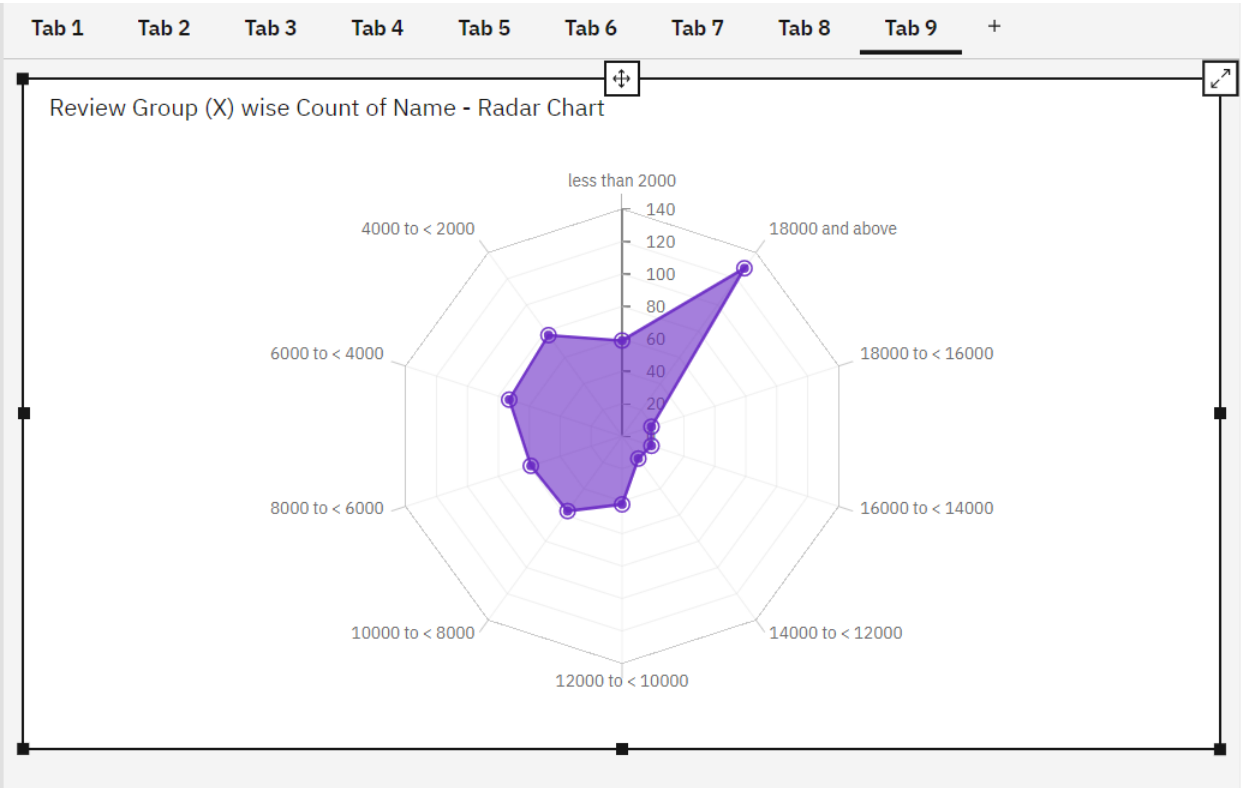
8. Year wise Count of Name - Color by Genre.



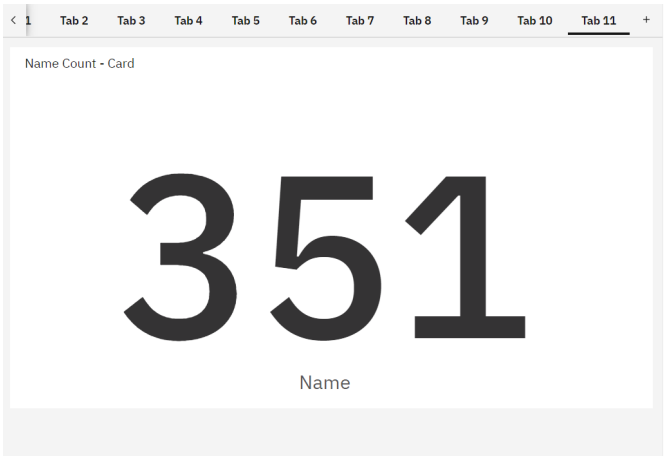
9. Review Group (X) wise Count of Name -- Waterfall Chart



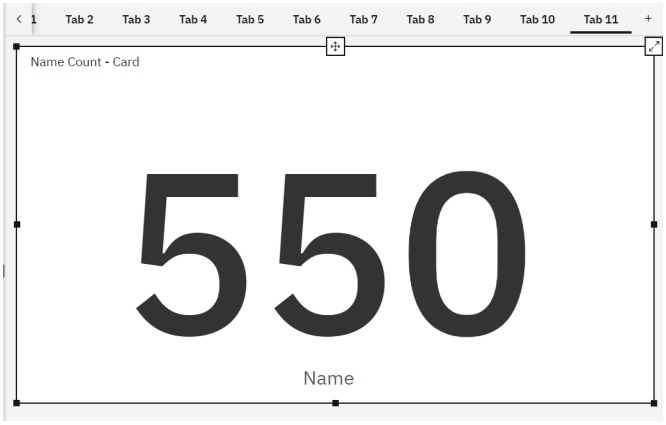
10. Review Group (X) wise Count of Name -- Radar Chart



11. Name Count - Card.



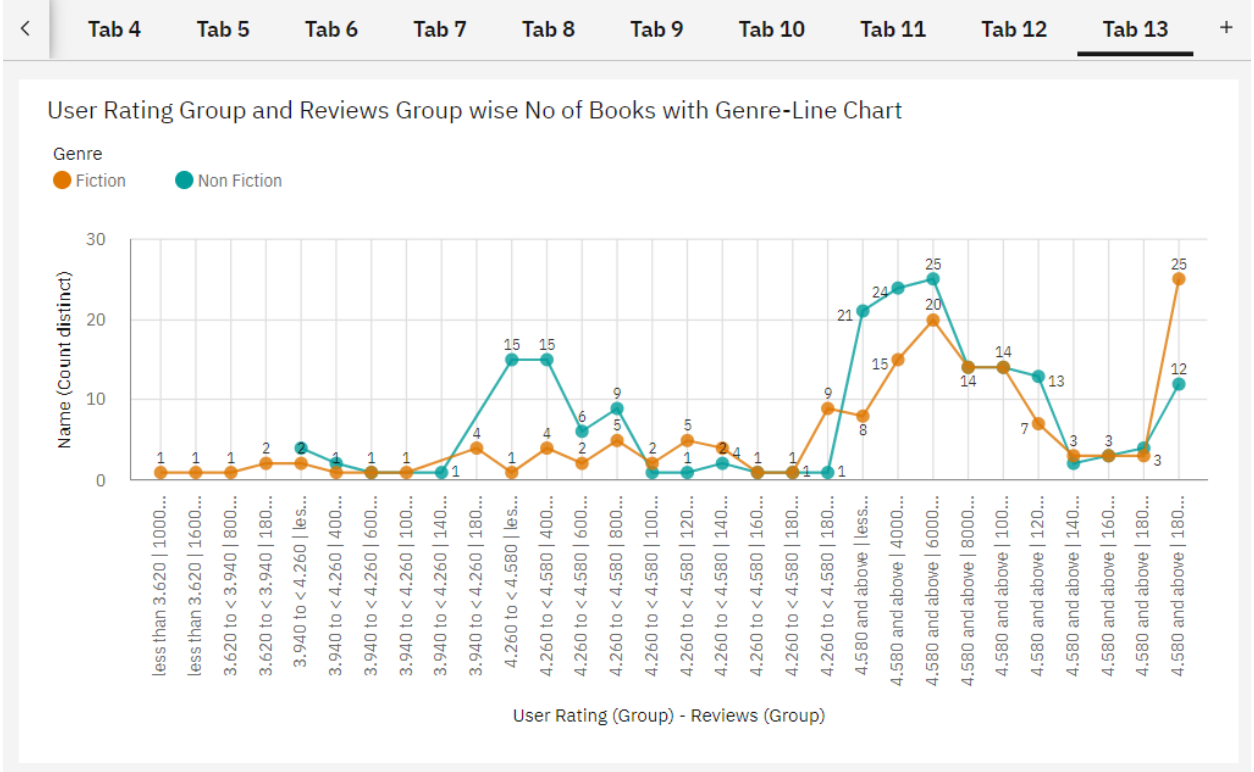
Summarize - Count Distinct



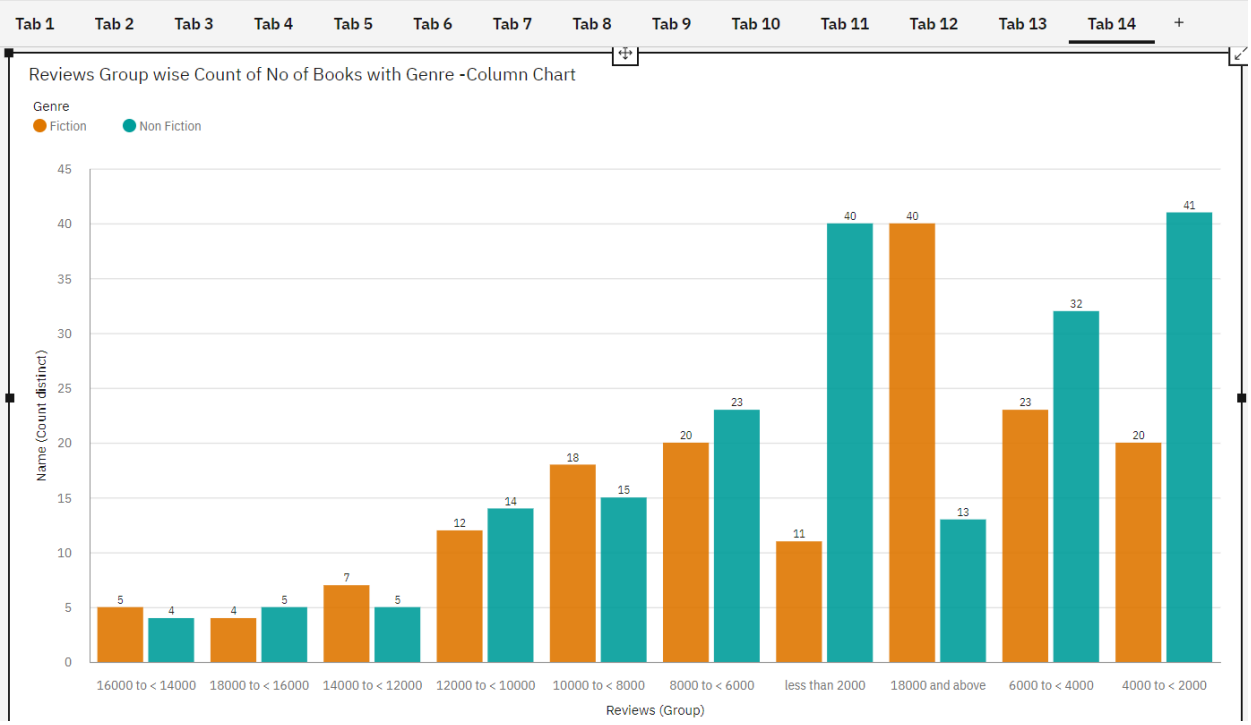
Summarize - Count

12. Author Count - Card.

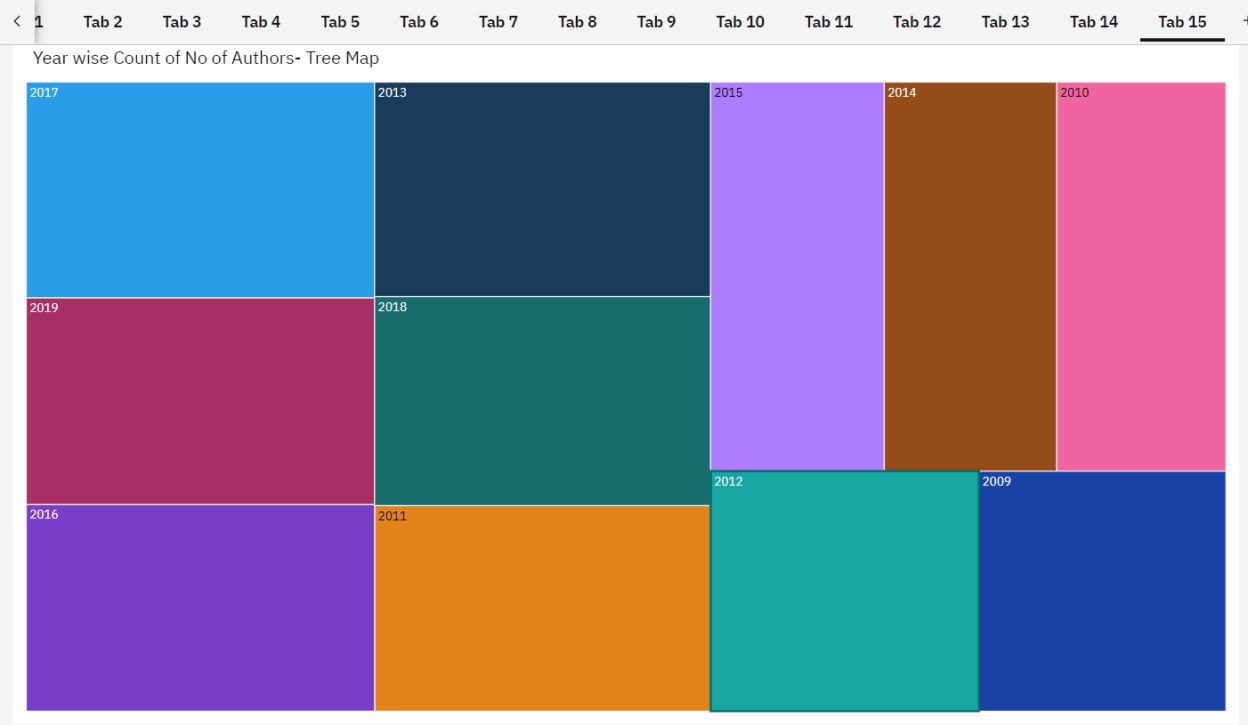
13. Line Chart - Showing User Rating Group and Reviews Group wise No of Books with Genre.



14. Column Chart: Reviews Group wise Count of No of Books with Genre.



15. Tree Map: Year wise Count of No of Authors.



16. Hierarchy Bubble: User Group and Review Group wise No of Books by Genre.

ADVANTAGES AND DISADVANTAGES OF CREATING DASHBOARD

ADVANTAGES

ADVANTAGES

I. Enhanced Visibility: Dashboards give organizations more visibility and make information available whenever they're needed, allowing them to better respond to changing market conditions.
II. Time-saving Efficiency: We no longer waste time generating reports from numerous systems thanks to dashboards. Instead, data is extracted from a source and presented as a simple visual overview.
III. Better Forecasting: With more data knowledge, future demand may be predicted more accurately utilizing historical data. For increased success, businesses can better plan for demand changes by creating measurable targets and deliverables.
IV. Better Decision Making: A dashboard allows firms to review crucial data quickly and thoroughly, whether they're providing reporting and analysis for the entire organization or certain functional areas of the business. Visualized interactivity helps to convey large volumes of data in a simple and understandable manner. Better business decisions can be made with the capacity to quickly discern what the data really means.

DISADVANTAGES

I. Cluttered or flashy design, with users attempting to incorporate too much information without comprehending restrictions or taking into account their individual demands from the wide range of measurables that data analysis provides.
II. The technology used to create dashboards is different from other software solutions already in use in organisations, and it can be difficult to grasp at first. There are no predetermined rules or hierarchy for how dashboard measurements are used in the business. This means that each person can interpret the measurements in their own unique way, resulting in a wide range of data being reported.

APPLICATIONS

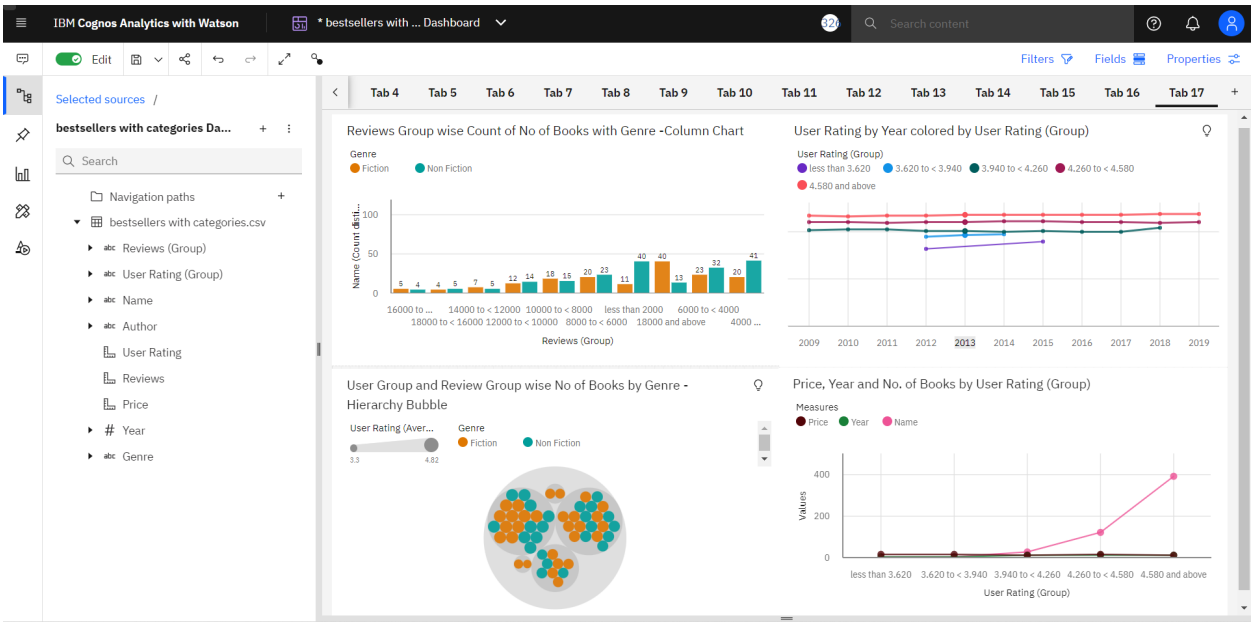
- If you handle complex campaigns, you'll likely wind up with many analytics solutions for each platform, which means you'll have to consult them independently, obstructing your overall picture. Rather, the dashboard shows data from several sources, such as site analytics and social media indicators. It's a lot easier to compare them and see how they develop this way.

- Many critical metrics are plainly shown in a good dashboard, so you don't need to be an analytics specialist to comprehend them. You always have the option of using more specific tools to go deeper into a particular data collection.

- You can create different users so that your complete team can view the same information from wherever if you synchronize your dashboard automatically in the cloud. You can even project the dashboard onto a screen in your office so that everyone can see what's going on in real time.

- A unified dashboard can help you save a lot of time. Dashboards perform all the work for you instead of you having to collect data from various sources and create charts on your own. All you have to do now is spend some time setting up the metrics and deciding how to present them. The reports are generated automatically after that.

DASHBOARD


IBM Cognos Analytics with Watson
* bestsellers with ... Dashboard
Search content
Edit
Filters
Fields
Properties
Selected sources /
bestsellers with categories Da...
Search
Navigation paths
bestsellers with categories.csv
Reviews (Group)
User Rating (Group)
Name
Author
User Rating
Reviews
Price
Year
Genre
Tab 4
Tab 5
Tab 6
Tab 7
Tab 8
Tab 9
Tab 10
Tab 11
Tab 12
Tab 13
Tab 14
Tab 15
Tab 16
Tab 17
Reviews Group wise Count of No of Books with Genre -Column Chart
Genre
Fiction
Non Fiction
Name (Count distri...
Reviews (Group)
User Rating by Year colored by User Rating (Group)
User Rating (Group)
less than 3.620
3.620 to < 3.940
3.940 to < 4.260
4.260 to < 4.580
4.580 and above
User Group and Review Group wise No of Books by Genre - Hierarchy Bubble
User Rating (Aver...
Price, Year and No. of Books by User Rating (Group)
Measures
Price
Year
Name
Values
User Rating (Group)

CONCLUSION

We can understand supplied data using diagrams, graphs, and maps in this way. This knowledge of data enables us to ask the proper questions in order to achieve our objectives through method optimization. We learned how to upload and prepare data with this assignment. We also covered statistical ideas that aided in the computation and presentation of graphs and maps for the dashboard.