What can we get from "Goodreads" Indonesian Language Books Data?

Result Highlight by Renyta Dangin

What did we want to achieve?

This analysis aims to find insight from the dataset regarding the list of Indonesian language books on the "goodreads" website.
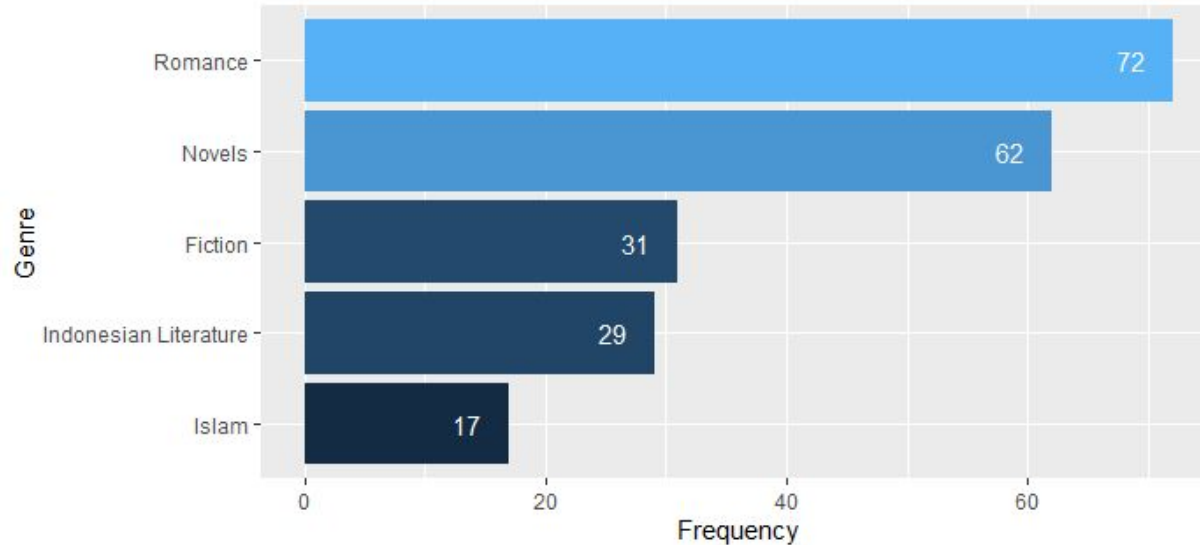
About The Book

About The Author

About The Publisher

About Everything Else

Summary

First Analysis is to find <mark>The top 5 genres for Indonesian literature</mark>. **Finding the majority of each book's genres is the goal and the main genres of the novels are the ones that were counted.**

Data of top 5 most popular genres that have been published
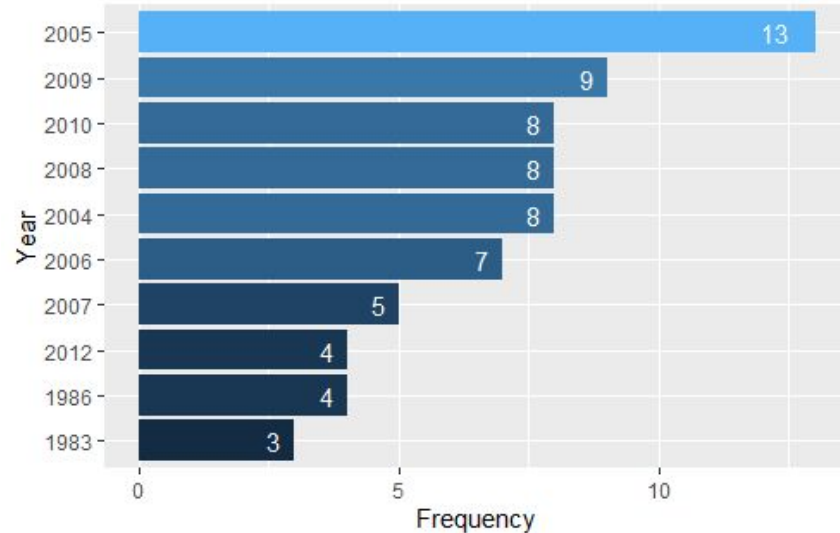Goodreads Indonesian Books 2020

- The most popular genre that have been published is romance.
- There was a significant difference of frequency between the top 5 in each genre.
- This analysis does not justify the preferences of genres for the reader.

The 2nd analysis looks at <mark>the relationship between the number of books released each year.</mark> where the first figure displays the ten years with the most books produced and the second figure displays the trendline of the number of books released between 1940 and 2000.
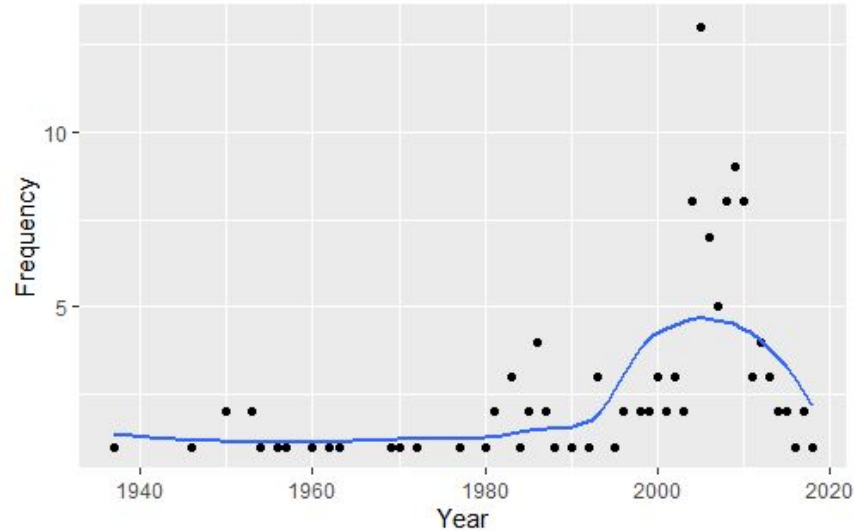

Data of top 10 years with the most published books
Goodreads Indonesian Books 2020

Data on the published book trend line over time
Goodreads Indonesian Books 2020

- 2005 is the year with the most published books, with 13 books published.
- The majority of books were released less than 5
- Each of the top ten years has a significant gap in terms of books frequency
- Start from year 2000, the number of books published were increased. .

The **third analysis is to find the top 10 books according to ratings.** Books should be taken into consideration depending on their rating because most readers rank their books based on the CAWPILE method (Character, Atmosphere, Writing, Plot, Intrigue, Logic, and Enjoyment).

Data of top 10 Indonesian books based on Rating
Goodreads Indonesian Books 2020


```
        rating
Min.    :3.280
1st Qu. :3.840
Median  :4.010
Mean    :3.997
3rd Qu. :4.200
Max.    :4.550
```

- Ranking is calculated using books that have a minimum number of rating equal to the average number of rating of the book which is 2222.
- There isn't a book that meets the requirement and has a perfect rating.
- The average rating of books is 3.99 ~ 4.00

About The Book

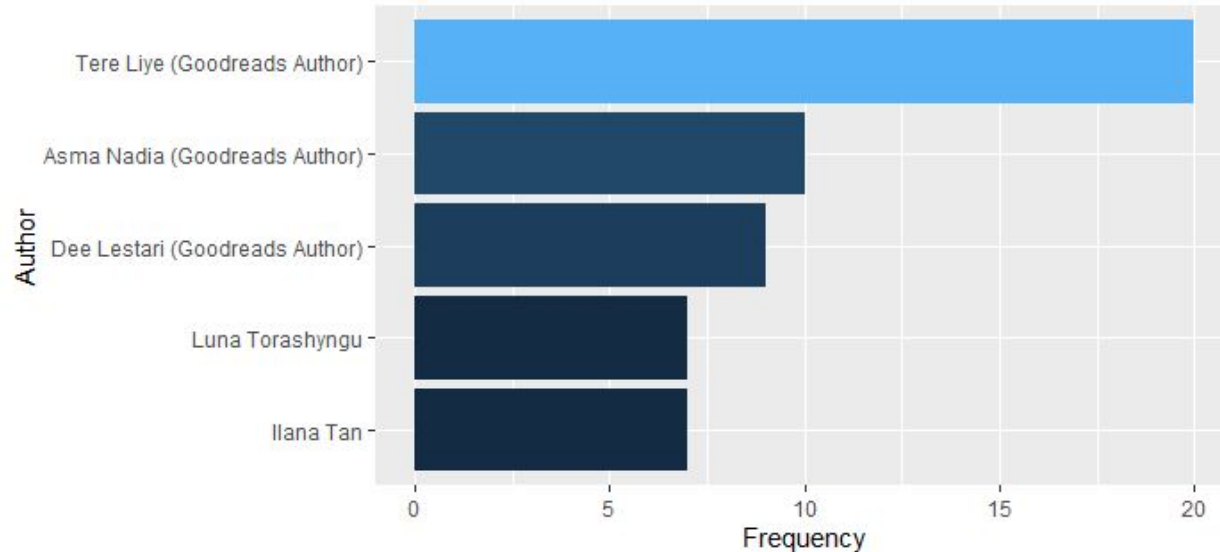**About The Author**

About The Publisher

About Everything Else

Summary

Based on how frequently they write books, the 4th study concentrates on active authors. An <mark>understanding of how each author performed</mark> is sought after by this analysis. However, it has no impact on how each book does.

Data of top 5 Indonesian author with active book release
Goodreads Indonesian Books 2020
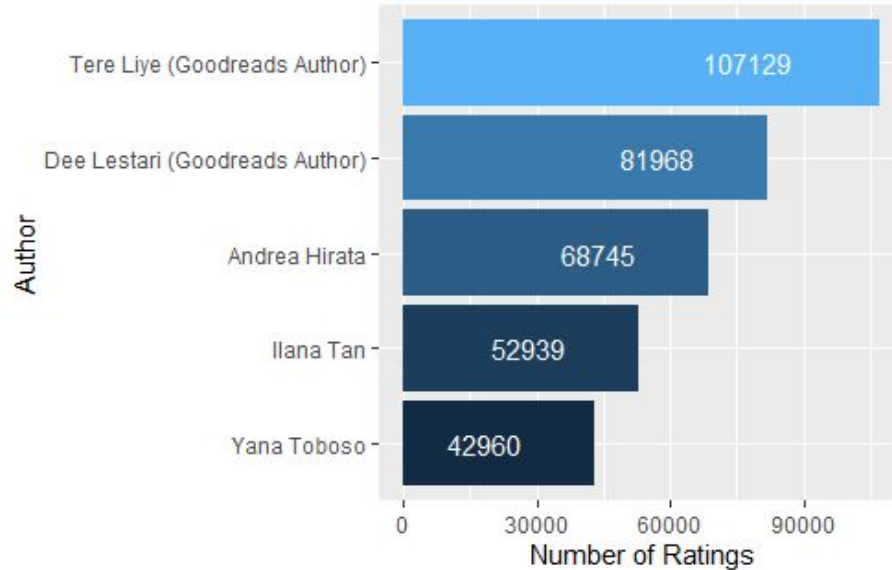


```
                      n
Min.      :  1.000
1st Qu.:  1.000
Median :  1.000
Mean      :  1.702
3rd Qu.:  1.750
Max.      : 20.000
```

- In terms of the quantity of books published, authors in positions 1 and 2 have a big disparity.
- The rest of the author has a close resemblance in terms of the number of books that have been published.
- The average of books from the author is 1,7 ~ 2 books.

The 5th analysis will identify <mark>the top five authors with the highest rating and number of rating.</mark> This analysis is an attempt to find the best author's based on those 2 criteria.

Data of top 5 most rated author based on the number of rating
Goodreads Indonesian Books 2020

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	1.00	94.25	546.00	3781.89	2343.00	107129.00

there is an author who only received 1 number of rating, which is not adequate, particularly if the author received a 5-star rating for a total of only 1 rating.That's why it's important to include the total number of ratings

- The gap between each author in the chart is equal.
- The average number of ratings received by each author is 3781.89 ~ 3782.

About The Book

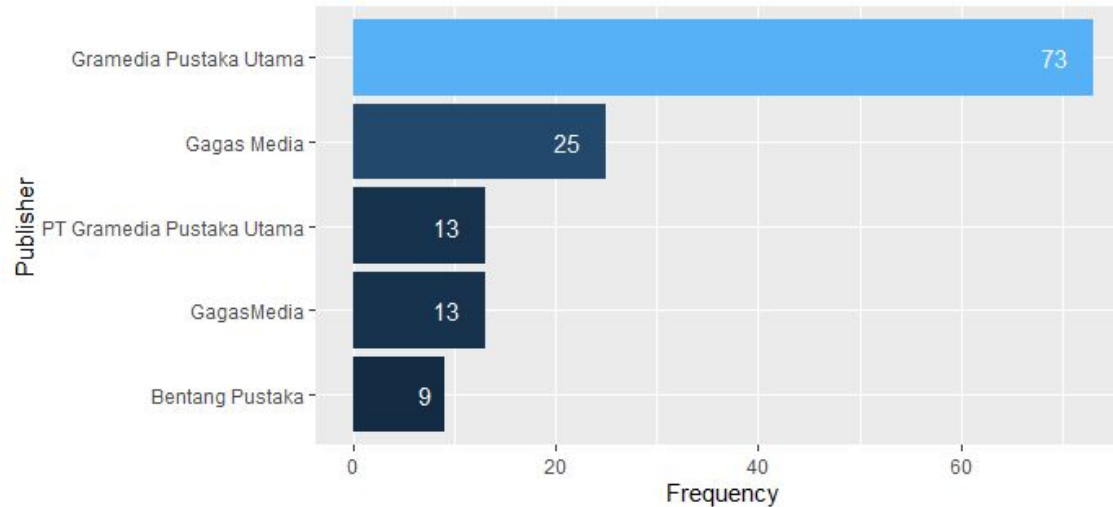About The Author

About The Publisher

About Everything Else

Summary

The 6th analysis is to <mark>determine the most active publisher based on the frequency with which they were published.</mark>

Data of top 5 most published publisher
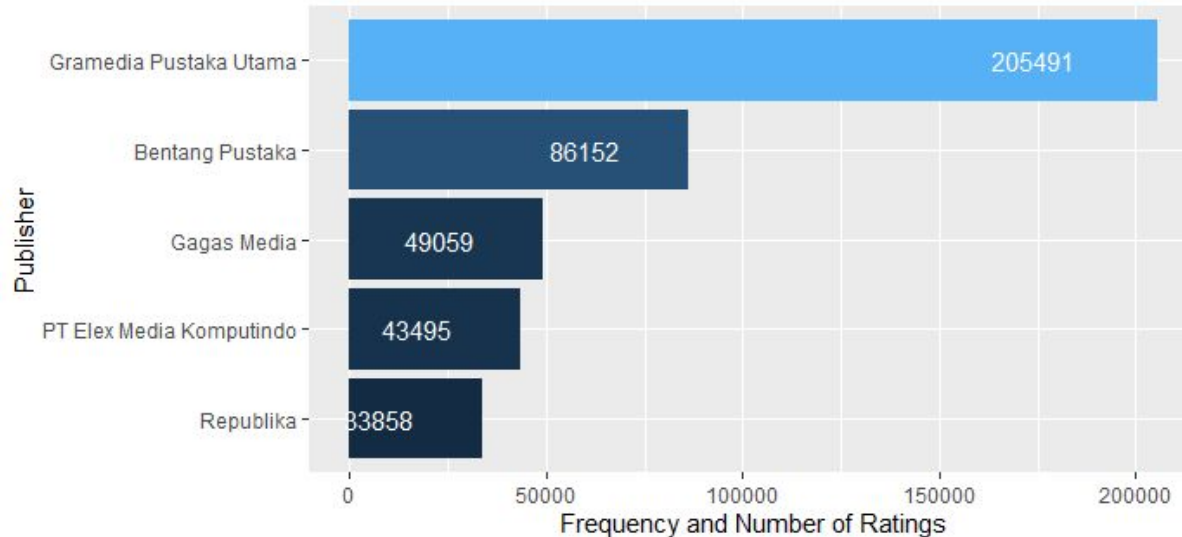Goodreads Indonesian Books 2020


```
      Freq
Min.     : 1.000
1st Qu.: 1.000
Median : 1.000
Mean     : 2.355
3rd Qu.: 2.000
Max.     :73.000
```

- There is a significant gap in terms of the frequency in each publisher.
- Gramedia Pustaka Utama is the most published publisher for 73 books been published in range of year 1940 - 2018.
- Most of the publishers had an average of 2 published books.

Based on the average of released books, this 7th analysis seeks to identify the top publisher with at least two books published (average book been published from each authors) and more than 5296 ratings (based on the average ratings of publisher).

Data of top 5 best published publisher
Goodreads Indonesian Books 2020


```
      numRatings
Min.    :       1.0
1st Qu.:     117.5
Median :     577.0
Mean    :   5295.5
3rd Qu.:   3878.5
Max.    :205491.0
```

- There is a significant gap between each set of data, especially the top one.
- It can be concluded that Gramedia Pustaka Utama is the best-published publisher based on Goodreads data.

About The Book

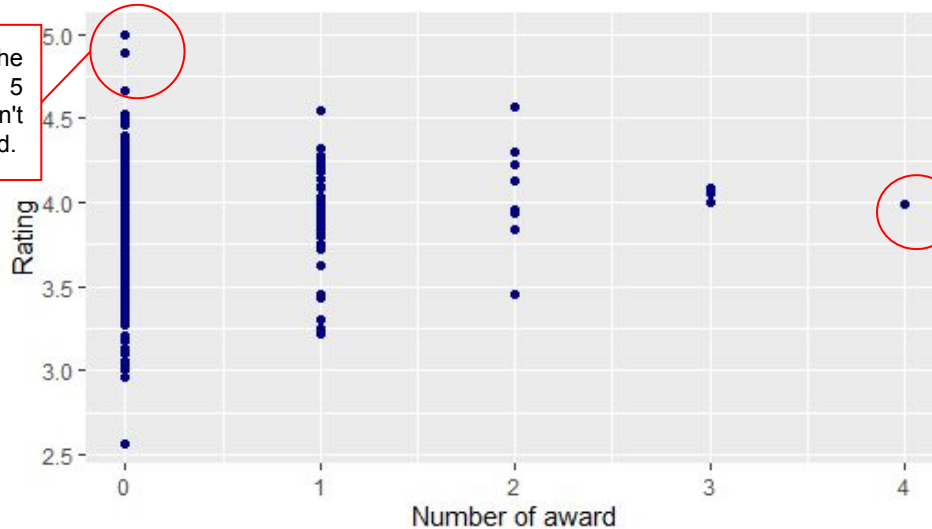About The Author

About The Publisher

About Everything Else

Summary

"Do most awarded books have the best rating books?"
This 8th analysis attempts to find the correlation between the number of awards and the rating of each book.

Data of correlation between the number of awards given for
each book and the rating

Goodreads Indonesian Books 2020

the majority of the books that had a 5 star rating didn't have a single award.

Instead, the book that received four awards only had four stars.


```
booklist.numaward
Min.      :0.0000
1st Qu.:0.0000
Median :0.0000
Mean      :0.1914
3rd Qu.:0.0000
Max.      :4.0000
```
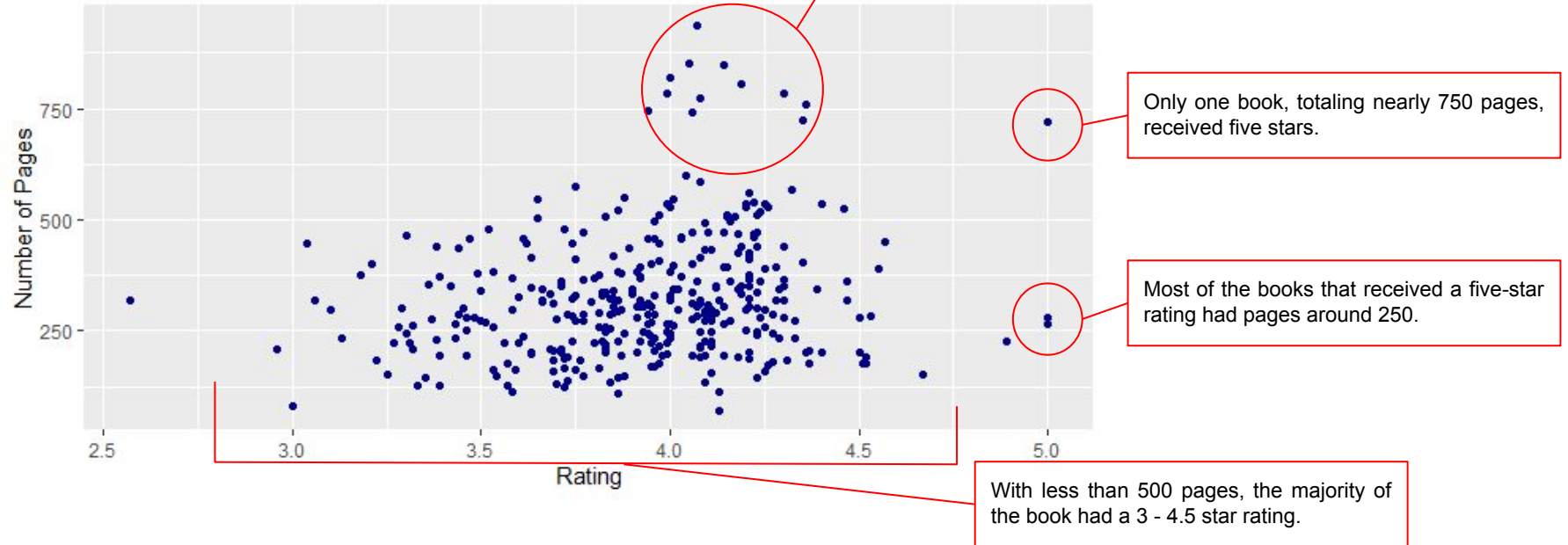
It is supported by the fact that the average number of awards won by each book is only 0.19 ~ 0 awards.

Do pages affected the rating of books?
This 6th analysis attempts to find the correlation between the number of book pages and the rating.

Data of correlation between the number of pages and the rating
Goodreads Indonesian Books 2020

Some of the books with up to 750 pages only received 4 stars.

Only one book, totaling nearly 750 pages, received five stars.

Most of the books that received a five-star rating had pages around 250.

With less than 500 pages, the majority of the book had a 3 - 4.5 star rating.

The conclusion is that the number of pages has no effect on the rating of each book.

In Summary

1. The author's most well-known genre for publications is romance.

2. Between 1940 and 2020, the year with the most books published is 2005.

3. From the data of the top 10 books according to ratings, it is known that "Black Butler, Vol. 3" is the highest-rated book on the Goodreads website with 4.52 ratings. (conditioning applied).

4. Tere Liye has released 20 novels, making her the author with the most active book releases.

5. According to the quantity of ratings from the Goodreads website, Tere Liye has the most amount of ratings of any author.

6. In terms of the quantity of books published, Gramedia Pustaka Utama is the most active publisher.

7. According to the quantity of books released and the quantity of ratings, Gramedia Pustaka Utama is the best publisher.

8. It is discovered that the number of awards a book has won does not affect its rating.

9. It is also discovered that the books' ratings are unaffected by their page count.