Do We Judge Children's Books by Their Covers?*

Exploring Title, Exclamation Marks, and Reviewer Bias

David Qi

December 3, 2024

This study examines the impact of first impressions on children's book ratings, focusing on how attributes such title, page count, and review count influence ratings. Using data from Goodreads and both simple linear regression and fine-tuned transformer models, we found that book titles alone explained 7.94% of the variation in ratings, indicating the significance of first impressions. This study provide support for previous studies in concluding that children may favor the theme of excitement and celebration, which are useful in building recommendation system and selecting optimal titles.

Table of contents

*Code and data are available at: https://https://github.com/UTDQi/children_book_review

1 Introduction

The rise of user-generated content has transformed how consumers interact with media, especially when it come to book ratings. These ratings serve as an indirect reflection of a book's quality and readers' perceptions, serving as one of the few pieces of information that help potential readers to form initial impressions. As they play a crucial role in guiding decisions,

consequently, it has become standard practice to use book ratings as a predictor variable in recommendation systems, serving as a key indicator of how much a reader enjoys a book.

However, children's books, as a unique genre, present distinct dynamics in the realm of user-generated ratings and recommendation systems. Unlike adult literature, children rarely contribute directly to ratings or reviews. Instead, parents, educators, and caregivers typically provide feedback. Which have made is difficult to create recommendation systems (Milton et al. 2020). And since many reviewer are not direct reader of the book, we hypothesize their rating may subject to greater bias from first impression.

The estimand of this study is the level of reader satisfaction with a book, as reflected by its average rating. Specifically, this paper aims to determine the extent to which first impressions influence overall satisfaction.

This paper made use of simple linear regression models and fine-tuned pretrained transformer models, by analyzing children's book data from Goodreads(Wan and McAuley 2018; Wan et al. 2019), to uncover how much such first impression will influence the rating of a book. The study provides guidance for author and publishers to create attractive titles, identifies traits in children's book rating, which may be useful for designing recommendation systems. Where these recommendation systems, when it comes to children, may be difficult to create due to lack of user generated data, and will likely benefit from understanding provided by this study.

As a result, we found that one may explain 7.94% variation in the data by title alone. This fact along with the significant F statistics of linear regression model, have reflected the small but clearly present influence of first impression on book ratings. We have develop theories, by looking at our model output and relating to relevant literature, that as a sign of excitement and celebration, exclamation marks can provide a positive first impression and improve rating. We also note the influence of number of pages and review to rating ratio on average rating, and considered possible causation.

The R scripts in this paper are written in R(R Core Team 2023), The table in our study are generated by knitr (Xie 2014), and graph with ggplot2(Wickham 2016). Data input and output were done using arrow(Richardson et al. 2024).Data manipulation made use of tidyverse(Wickham et al. 2019). The data download, cleaning and transformer model made use of python(Van Rossum and Drake 2009) with supporting packages pytorch(Paszke et al. 2019), Scikit-learn(Pedregosa et al. 2011), transformers(Wolf et al. 2020), tqdm(Costa-Luis and Contributors 2023), numpy(Harris et al. 2020), pandas(McKinney 2010), requests(Reitz 2023).

The rest of the paper is structured as follows: in Data section we will discuss the various aspects of the data we use, how they were gathered and the variable we shall use. In the Model section we will discuss details of our models, why it is justified and the coefficient summaries. In the Results section, we present the predictions of the model. In the Discussion section we will discuss the interpretation to the predictions, and what we may learn from the coefficients for the model, we will also look as possible weakness and future improvement for the model, data collection, and analysis.

2 Data

We will be using the children's book data from goodreads.com, collected in late 2017, originally for the propose of academic research conducted by Wan and McAuley (2018) and Wan et al. (2019). We would also thank Alexander (2023)'s book for providing information on this dataset.

2.1 Overview

The children's book dataset contains 124,082 books. After removing books that lacks the requires information, or that have incorrect entries easily identifiable(such as having June 31 as date), we have selected 65,118 books for analysis. The justification and risks of such selection will be discussed in Discussion.

2.2 Data Measurement

Goodreads is an American book recommendation and social networking platform. Its primary purpose is to allow users to catalog their books, keep track of their reading progress, share reviews, and engage in discussions about literature. The included books are mainly in English, with some books in other languages available. Any registered user on Goodreads may give rating to any book, the rating comes in integers from 1 to 5, with no half-star option. Along with a rating, user also may write text reviews for the book, which may be seen by other users. Descriptions of a book were updated by volunteer and staffs that maintains the website, while the rating contents are user generated.

One issue this may raise is as each user has different understand of rating, the aggregate effect may lower the effectiveness of rating's role as a estimand for satisfaction.

The data collection was done through a web-scarping on contents available for any user(Wan and McAuley 2018; Wan et al. 2019).

2.3 Varibles

For this study, we intend to use average rating as an estimand, and select predictor variables that has influence one one's first impression. Below is a brief overview of these variables.

2.3.1 Average Rating

Figure 1 displays the distribution of average book ratings('average_rating'), which appears to approximate a normal distribution centered around 3.88. The data suggests a tendency for people to give higher ratings. An average book will not receive a 3, but will have average rating close to 3.88 instead.

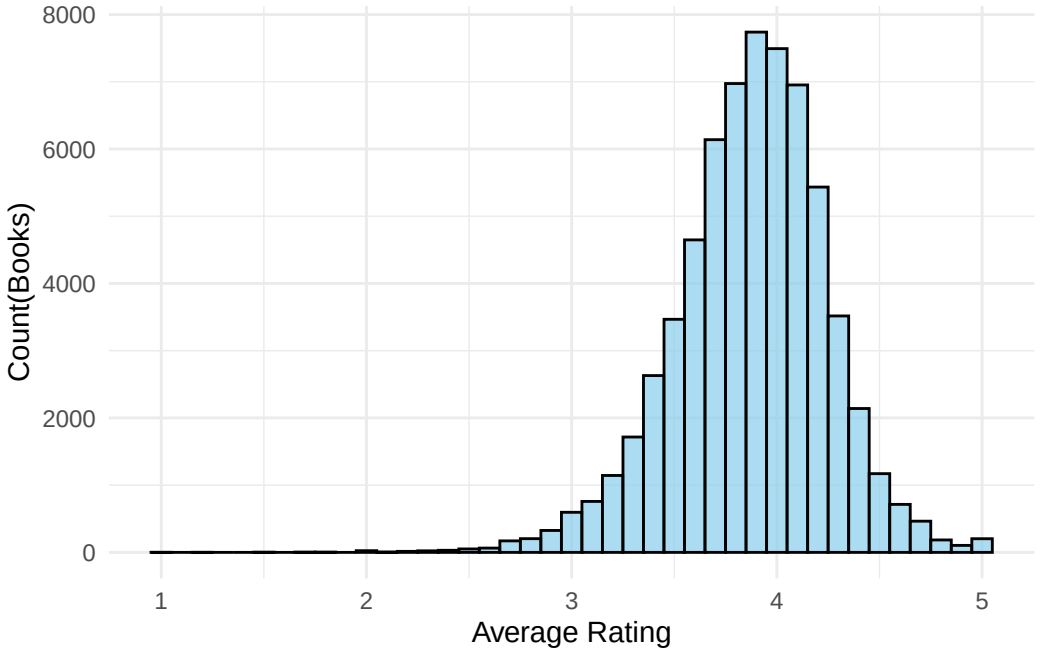

Figure 1: Distribution of average rating

2.3.2 Publication Day

Publication day(`publication_day`) is the day in the month the book is published, it is used to capture monthly trends. Such pattern do exists, as Figure 2 clearly shows that there is a great increase in book published one the first day of each month, and a slight rise the fifteenth day. We would expect the difference in level of competition between different days in a month can influence people's perception toward a book.

2.3.3 Publication Day In The Year

Publication Day In The Year(`publication_day_in_year`) is the day in the year the book is published, it is used to capture yearly trends. Similar to previous part, Figure 3 is a bar chart

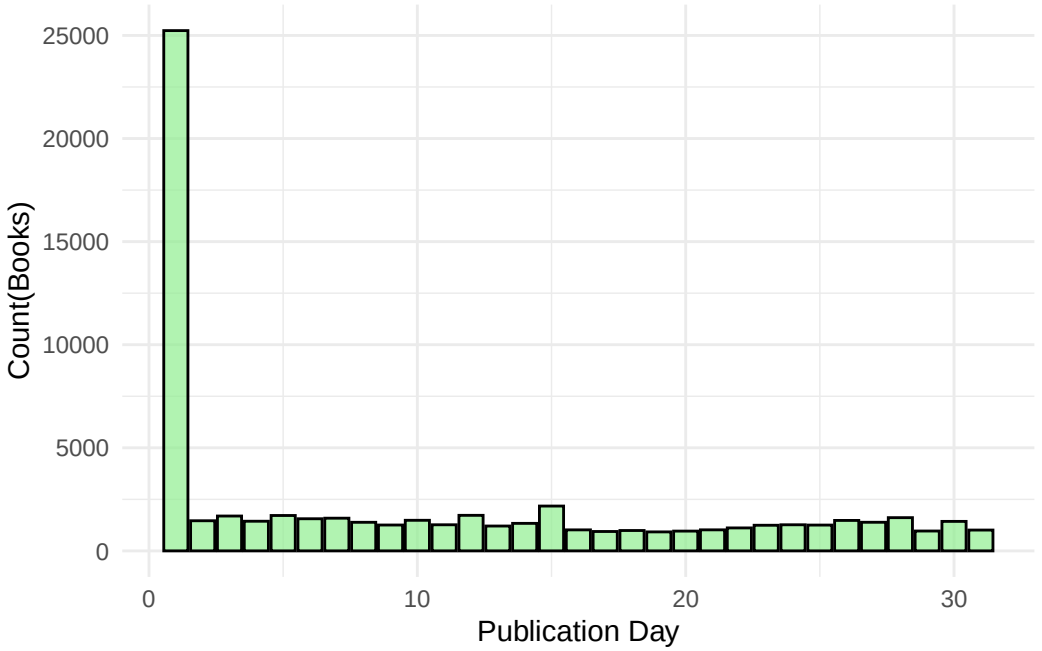

Figure 2: Distribution of Publication Days

that shows the yearly pattern on publication of books. We Observe a decline in book sales in June and December, and peaks in January and Sepetember.

2.3.4 Publication Year

Publication Year is the year the book is published, people may treat old books differently from new books in the way they give reviews. Also as Figure 4 shows, there is an exponential increase in the number of books people can choose form. The drop in the end of the graph is because that many newly published books need time for then to gain their first rating, and the book published in 2018 haven't been published at the time of data collection and naturally don't have reviews.

2.3.5 Number of Text Reviews

Number of text reviews (`text_reviews_count`) is the number of written text reviews for a book. The number of text reviews is visible to new readers, and provides them with a first impression for the popularity of the book. Figure 5 shows the distribution of books with different number of text reviews, note that this is not a full plot as there are very few books that received far more than 1000 reviews, the most being 31,536 reviews. We see that of most book getting very few review and a few book have very high reviews.

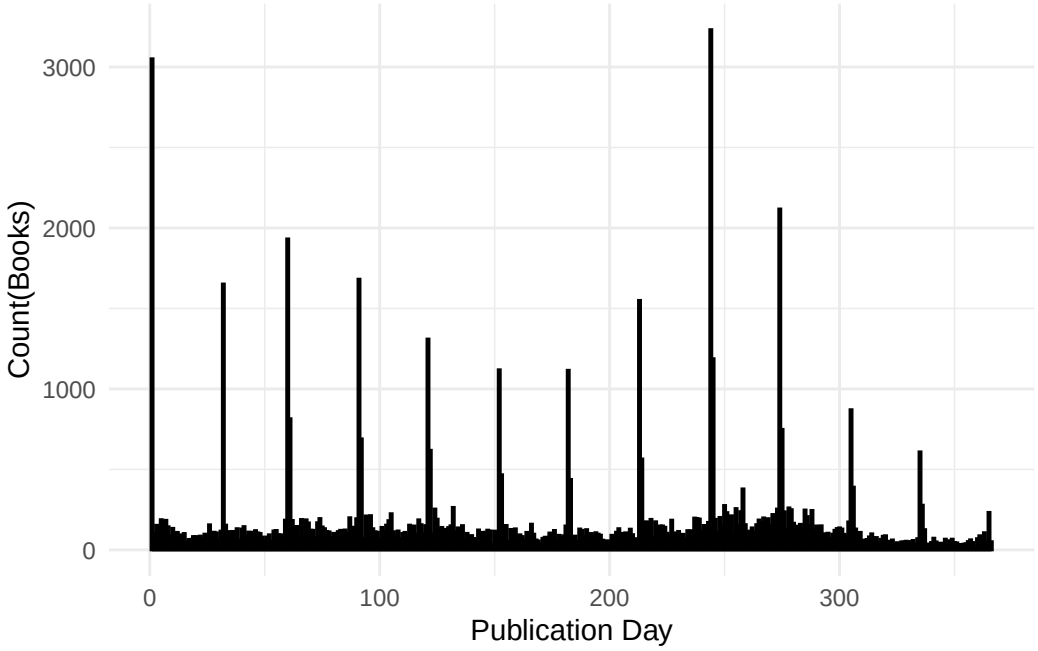

Figure 3: Distribution of Publication Days(Yearly)

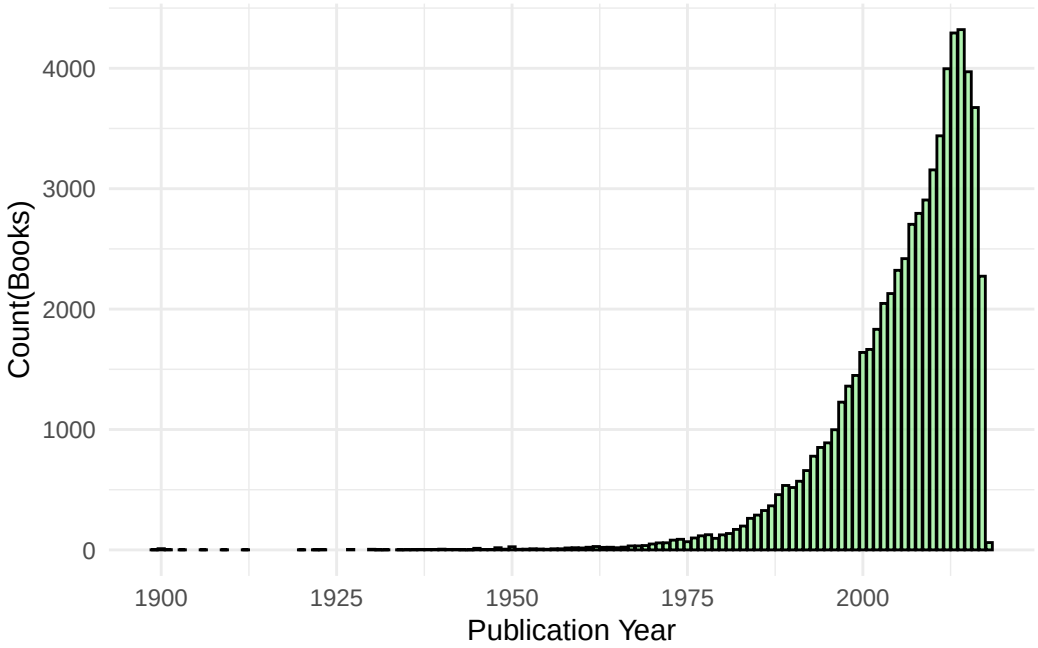

Figure 4: Distribution of Publication Year

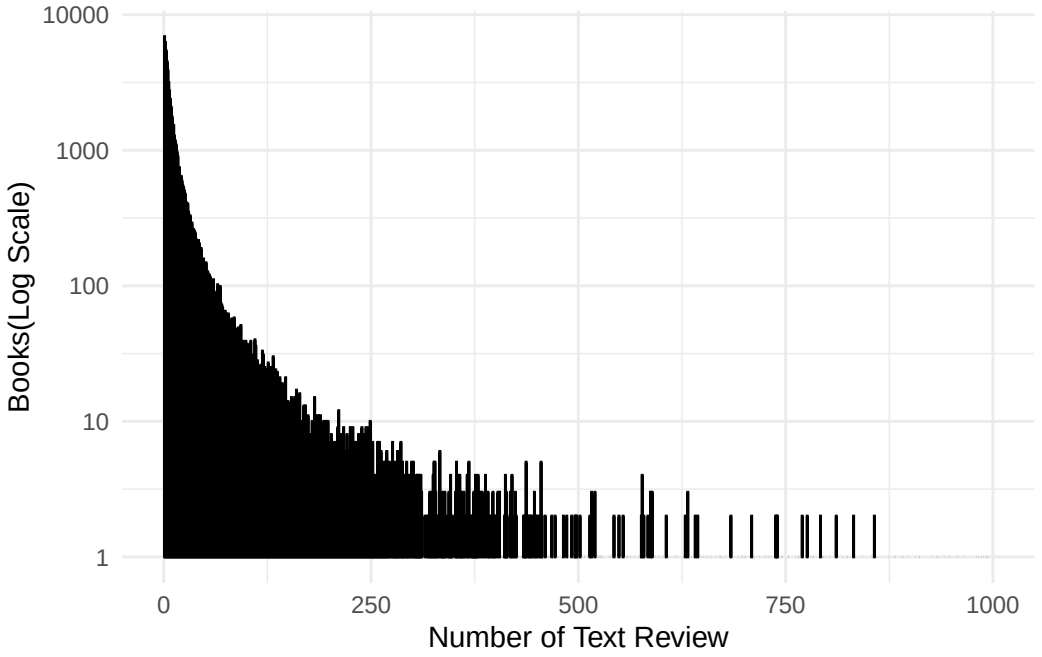

Figure 5: Distribution of Number of Text Review

2.3.6 Text Reviews To Rating Ratio

Text Reviews To Rating Ratio (`text_reviews_ratio`) is the ratio between number of text review and number of ratings. It serves as a measurement of reader engagement and the atmosphere of the reader community. Which can have substantial impact on one's impression of a book. Figure 6 shows the distribution of books with different number of this ratio, showing readers have different level of engagement for different books. Note this variable displays a weak correlation(correlation = 0.2112446) with year of publication, including this variable in the model will minimize the effect of year of publication in the linear regression model. However we have included both variables in consideration for interpretiblity.

2.3.7 Title

Title (`title`) is the most visible trait on the front cover of a book. The wording of the title will directly shape one's first impression of the book, and directly determines if one have interest in the book.

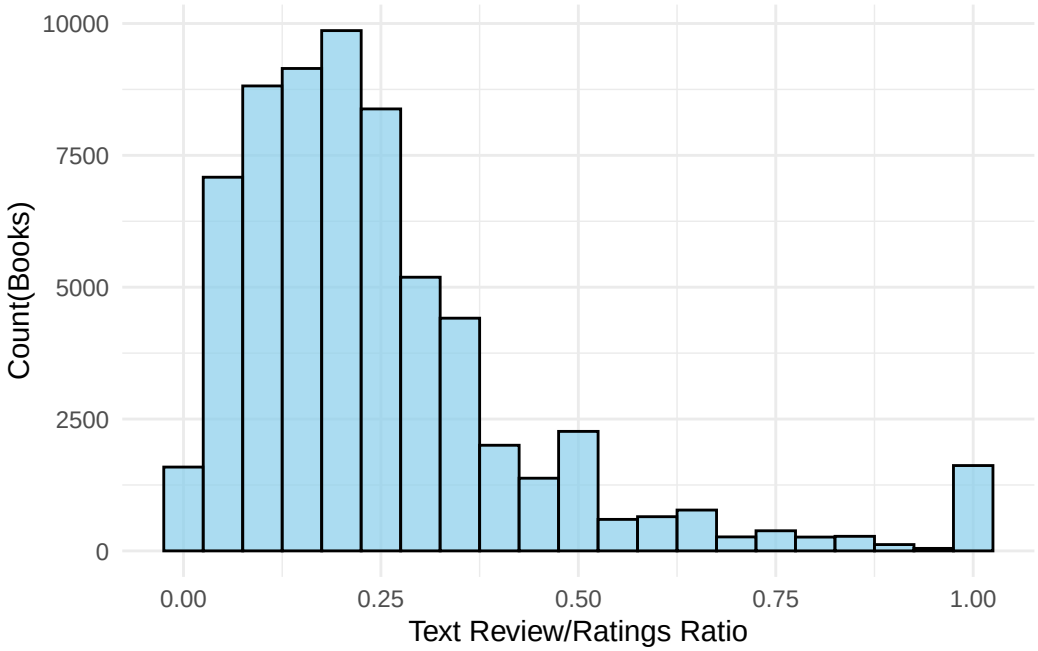

Figure 6: Dirtribution of the Ratio of Text Review

2.3.8 Number of pages

Number of Pages (`num_pages`) is the sometimes more notable than the title of the book, one often realize how thick a book is without knowing the title. If we want to evaluate the impact of first impression on book reviews, number of pages is an important factor to take into consideration. Figure 7 shows that most books in the dataset have a low number of pages, as the number of pages increases, the frequency of books with higher page counts drops off significantly. Agreeing with our impression of children;s book.

2.4 Unused Varibles

There are certain variables that could be drawn from the dataset, which may fit the topic of this paper, but will not be used for modeling. In this section we justify why they are not selected.

2.4.1 Publication Month

The effect of publication Month have been fully covered by days in a year. While the latter is more accurate.

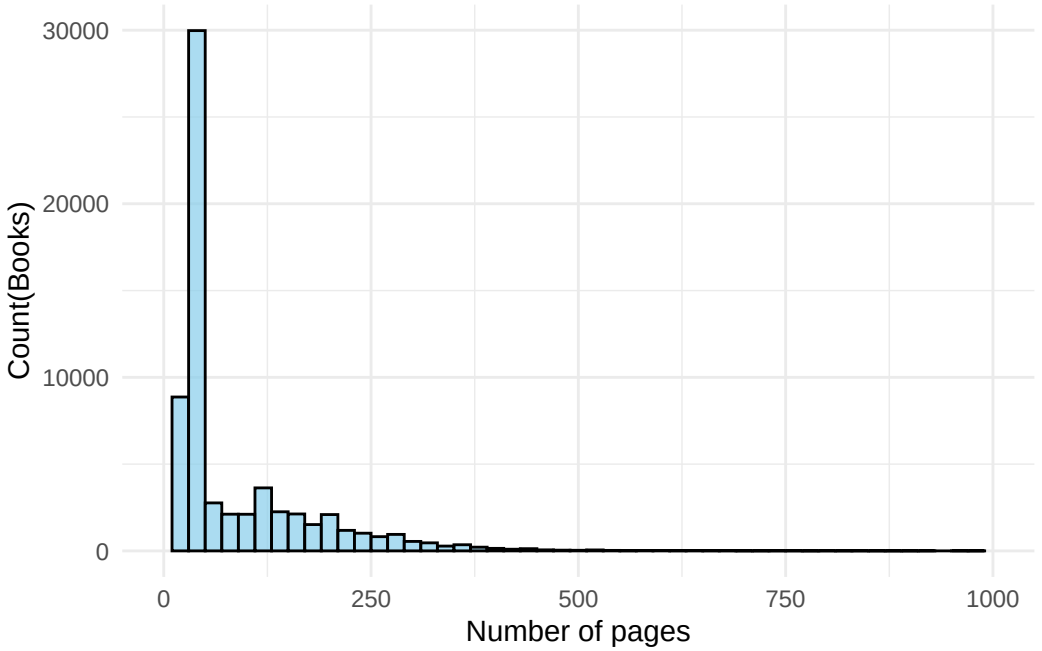

Figure 7: Distribution of number of pages

2.4.2 Rating Count

The total number of ratings is an important indicator of book popularity. However including it does not improve the model by much, and it can be captured by text review to rating ratio, along with number of text reviews. So we did not inculde the varible for simplicity.

3 Model

3.1 Linear Regression Model

We will first attempt to use linear regression model to predict the average rating of a book.

3.1.1 Model set-up

For each book, we assumed they are being sampled similarly.
Let the rating of a book be denoted by R. Then we will use linear regression model:

$$R = \beta_0 + \beta_1 P_{day} + \beta_2 P_{day\_in\_year} + \beta_3 Pages + \beta_4 Reviews + \beta_5 Review\_ratio + \beta_6 P_{year} + \epsilon$$

Where ϵ is an independent, normally distributed error term.
$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ are regression coefficients.
P_{day} is the day in the month.
$P_{day\_in\_year}$ is the number of day in the year.
$Pages$ is the number of pages.
$Reviews$ is the number of text reviews.
$Review\_ratio$ is the ratio of text reviews.
P_{year} is the year of publication.
We will predict the average rating to be:

$$\hat{R} = \beta_0 + \beta_1 P_{day} + \beta_2 P_{day\_in\_year} + \beta_3 Pages + \beta_4 Reviews + \beta_5 Review\_ratio + \beta_6 P_{year}$$

We fit this model using standard linear regression: finding the set of regression coefficients that minimized mean squared error. The model diagnostics are in Appendix 2

3.2 Transformer Model

To analysis how title may affect the rating of a book, we use a transformer model to process the title. To implement this model, we used python(Van Rossum and Drake 2009) with supporting packages pytorch(Paszke et al. 2019), Scikit-learn(Pedregosa et al. 2011), transformers(Wolf et al. 2020), tqdm(Costa-Luis and Contributors 2023), numpy(Harris et al. 2020), and pandas(McKinney 2010).

Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. 2019) is a language model built on the Transformer encoder architecture, which has been pretrained on text sequence from Wikipedia and BookCorpus (Zhu et al. 2015).

We will fine tune the BERT model to predict average rating from the book title. A book title is a string input, which first passes through BERT pretrained tokenizer to be made into a sequence of tokens. For example the title "Bradford Street Buddies : Springtime Blossoms" will become:

[CLS], bradford, street, buddies, : , spring, -time, blossoms, [SEP], [PAD], [PAD]

Where [CLS] is special token for "class", [SEP] is special token for "separator", and [PAD] is special token for "padding". The pretrained BERT model accept this tokenized input and outputs a token-wise embeddings tensor that encodes meaning and information related to each token as it learned from pretrained data. As BERT have been pretrained on massive data, the embedding it produces has real connections with Semantic and Syntactic meaning of the token. While the embedding related to the [CLS] token carries information regarding the overall meaning of the title(Devlin et al. 2019; Zhang et al. 2019).

For our specific use, we need to fine-tune the model. Which is preformed by adding a fully connected layer after the output related to [CLS], then train the resulting model with this new

layer added, Figure 8 from original paper of Devlin et al. (2019) illustrated this process when the task is classification. In this case we are replacing the classification fully connected layer by a regression layer. Fine tuning BERT for regression tasks is also a common practice, with examples demonstrated by people such as Zhang et al. (2019) and Kang et al. (2022).

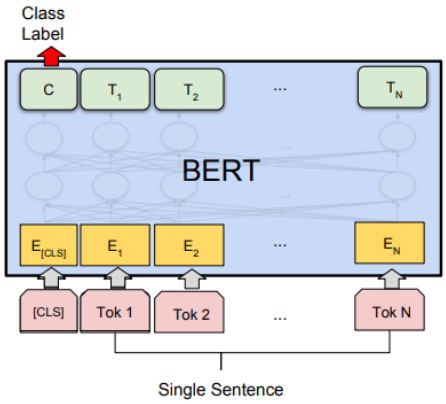

Figure 8: BERT fine tuning process

We use a 70%,15%,15% train,validation,test separation. We train the model on training data with mean squared error loss on validation set to measure model performance. We then adjust the learning rate, number of epoch trained, and batch size until we find a optimal hyperparameter that produces the best model on validation data. We then choose this model for further discussion.

4 Results

4.1 Linear Regression Model

Table 1 are the model coefficients fitted. The model have adjusted R-squared 0.038, showing the model have limited predictive power. However, the F-statistic(426 on 6 and 65111 Degrees of Freedom) and Pr(>|t|) provide concrete evidence that these first impression variables that we draw out have influence on average rating. Discussion of each variable is in the Discussion section.

Table 1: Linear Regression Model summary

	Estimate	Standard Error	T value	P value
(Intercept)	3.5800000	0.29700000	12.100	0.00
Publication Day	0.0013900	0.00014500	9.600	0.00
Publication Year	0.0000909	0.00014800	0.613	0.54

Table 1: Linear Regression Model summary

	Estimate	Standard Error	T value	P value
Number of Pages	0.0006160	0.00001350	45.800	0.00
Number of Text Reviews	0.0000737	0.00000738	9.990	0.00
Text Review Ratio	0.0801000	0.00737000	10.900	0.00
Publication Day in The Year	0.0001260	0.00001420	8.880	0.00

4.1.1 Note on Publication Year

This variable's effect is not statistically significant, which requires justification for its inclusion in the model. The rationale lies in its role as an underlying factor influencing the Text Reviews Ratio, a key variable associated with higher ratings. There is a clear trend: as the Publication Year approaches the present, the Text Reviews Ratio tends to increase on average, indicating higher levels of reader engagement. This suggests that more recent publications garner greater textual interaction, which, in turn, may positively impact overall ratings.

4.2 BERT regression Model prediction

After some testing, we have decided on a final model.Table 2 shows the final hyperparemeter settings. The model have resulted a mean squared error loss of 0.1271 and R squared of 0.0794 on test dataset, with mean squared error of 0.1219 on validation dataset. This means that the model explains 7.94% of the variation in average ratings based solely on the book title, suggesting the influence of first impressions on how books are rated.

We have also tested with modified test data, by changing periods into exclamation marks, or removing every "the" from the title. The model suggests that by replacing every period with exclamation marks, the average rating of the test set would increase by 0.026, and by removing every "the", the rating drops by 0.012.

Table 2: hyperparameter setting

Hyperparameter	Value
MAX_LEN	30
BATCH_SIZE	16
EPOCHS	2
LEARNING_RATE	0.000013

5 Discussion

5.1 Interpretation of Model

5.1.1 Linear Regression

The linear regression model constructed using numerical data have provide many interesting findings around each variables. And have pointed out directions for future research: using experiment and surveys to verify the causation we have proposed. However there are many limitations to linear regression model, the underlying assumptions need to be fulfilled, it have difficulties processing text data such as titles, which is why we proposed using BERT model for a more thorough analysis of this data.

Below we discuss the result for some particular interesting variables.

5.1.1.1 Publication Day

Although the change per day is small, it is highly significant. This suggests that books published later in the calendar year (closer to December) might receive slightly higher ratings on average. The relationship is positive, but the effect size is quite small. This possibly reflect the high competition on the first day of each month have damaged rating for book published on that day.

5.1.1.2 Number of Pages

For each additional page in a book, the average rating increases by 0.000616. Books with more pages tend to have slightly higher average ratings. This suggests that readers might associate longer books with higher quality, that they may tend to believe the book is more in depth, more rigorous and is of higher value. We may also link this to the psychological phenomenon of cognitive dissonance, where some people may tend persuade themselves to believe the book is good to revolve the discomfort of purchasing something that is worthless (Festinger and Carlsmith 1959).

5.1.1.3 Text Reviews Number

There is a positive relationship between the number of text reviews and the average rating, meaning that books with more reviews are likely to have higher ratings. Which may be explained by that social proof or engagement through reviews influences ratings positively.

5.1.1.4 Text Reviews Ratio

For each 1% increase in the ratio of text reviews to total reviews, the average rating increases by 0.0801. This is one of the most influential variables in the model. Books that have a higher ratio of text reviews and more detailed feedback from readers tend to receive significantly higher ratings. Indicating that detailed text-based reviews are more strongly associated with higher ratings than numerical ratings. We could also explain association by the positive impression created through higher engagement.

5.1.1.5 Publication Day in the Year

Like the publication day variable, the publication day in the year also shows a positive relationship with the average rating. This means that books published later in the calendar year may be slightly more favorably rated, although the effect is again small.

5.1.2 BERT Model

The fine tuned BERT model have provided understanding on our starting question: how much do people judge a book by its cover? Our answer is that by title alone, we can explain 7.94% of the variance. While the model can be used to provide guidance in creating individual titles. It may also be experimented to see the overall effect of having one type of title over another.

For example, our experiment of replacing periods with exclamation marks have aligned with the result of Milton et al. (2020), who find that "young children are inclined towards joy, favoring books including terms like celebration, excitement, gratitude, cheerful, and smile". Which are hinted by the use of exclamation marks. This has reinforced our previous understanding of the subject, and further verifies the predictions of our model. While the lowered rating by removing "the" may come from the resulting grammatical error in some books. There are other similar experiments one can do with this model to further develop knowledge on how wording of the title affect reader perception. One example would be replacing masculine pronouns into feminine or neutral pronouns, and observe the effect it has on average rating. And alternative testing preformed by removing the exclamation marks and replacing them with period could be conducted.

5.2 Weaknesses and Next Steps

There may be several important limitations in our study and data which restrict the effectiveness of our result, we will discuss three of them below and also provide possible direction for future studies to address these limitations.

5.2.1 Correlation Versus Causation

In the paper we have observed various Correlations, however, they may not demonstrate a causation. For example, we attempts to explain variation in score by text reviews ratio, however there is a important potential confounder variable, which we are not investigating since the nature of research is to find relation of first impression to ratings. This confounder variable is the quality of the book, a high quality book will naturally receive positive reviews, and may potentially attract people to comment more on the book. Where in the case of number of pages, a very important variable that we do not have hold of is formatting, larger fonts may improve reader experience while resulting in pages. A more comprehensive study that, likely in form of a survey may be applied to uncover causation, and whether our explanation are accurate.

5.2.2 Missing Data

There are certain important facts about the book that highly link to first impression that we can't acquire form this dataset.Including fonts, page size and paper quality.

Also around half of of the data point are removed due to missing or incorrect data. As there may be systematic trends to what data are missing(e.g. data for popular books are less likely to be wrong since more people concerns about it), removing these data may have systematically introduced bias into our data.

5.2.3 Sources of Bias.

There may be other sources of data not from the data cleaning process. An an American website, Goodread is not representative for books in other languages than English. Not only are the books of these languages, the readers from other cultures may not be well represented. Which increase this bias further. Also there may be economical and technological source of bias. Books that are more accessible or affordable to readers in the U.S. may dominate the data. Titles from smaller publishers or independent authors, particularly those outside of major Western markets, may not receive the same level of visibility or reviews. While since the data is form an online platform, it only allow data collection on people who have the money and skill to access such platforms. Which will exclude many of the young children and elderly. As well as views from people in poverty. Further research could be targeted specifically toward these groups. In fact, as Milton et al. (2020) points out, there are many existing, "state of the art" recommendation algorithm designed for adults. However, similar algorithms experience difficulties to be implemented on children, as there are not similar amount of data available. This lack of data have increased the importance further research through traditional methods such as surveys. See Appendix 1 for a discussion on how such survey could be conducted.

A Appendix 1: Ideal methodology and survey

In the discussion section, we touched on using surveys to validate and further develop our existing findings. In this section, we present description to an idealized survey such that one may use for further research into the topic.

A.1 Objective

We want to validate and measure the actual effects of three of our main findings:

1. Books expressing themes such as celebration, excitement, gratitude, cheerful, and smile are preferred. While exclamation in the title hints this.
2. Higher number of pages create positive attitude toward the book.
3. High level of community engagement create positive attitude toward the book.

A.2 Sampling Approach

Keep in mind that this survey aims to be targets to collect more data one a particularly hard-to-reach group: children. While we also aims at identifying the part of rating pattern in children that are due to adults. Quota sampling is quite suitable for this circumstance, which enables us to place more effort on the groups we want to focus on(Chen, Felt, and Henry 2018).

A.3 Stratified Quotas

We will use the age groups used by Milton et al. (2020) to test the alignment of results, which are under 5, 6 to 8, and 9 to 11.The quota should be evenly spread geographically in the United States, and equal in age group. For a total of 500 participants, we will recruit 100 participants from each age group, evenly spread geographically. And the rest 100 is left to parents and teachers. The sample will be offline, since we do not expect to reach our target population online.

A.4 Incentives

We will offer the following incentives to complete the survey:

1. A letter explaining the propose, theme and length and possible questions of the survey.
2. A $20 reward for completing the survey questionnaire.

3. Selection of any book or book series with market price less than 30$. Which will be mailed within 30 days of completing the survey questionnaire.

These incentives matches the outline given by Stantcheva (2023). And may incentivize the participants to provide more accurate replies.

A.5 Survey Structure

The survey will begin with an introduction explaining its purpose, scope, and confidentiality terms. The participant may continue if they and their guardian fully understands and agrees to the terms and conditions.

The survey questions will be grouped logically, for children it will covering the topics below:

- Demographics: Collecting background information on participant.
- Reading practices: Amount of reading done by participant lately, which genres and theme is the participant in favor of. Question regarding book length will be in this section.
- Understanding of Rating: Should book rating reflect experiences apart from reading, such as community engagement?
- Rating of Sample Titles: Several random titles are provided for the participants to rate their first impression. Questions will randomly include real book titles, and book titles with period replace by exclamation marks.

For adults, the "reading practices" section will be replaces by a section collecting the participant's experience with children's book. And the Understanding of Rating will be from a adult perspective, asking questions such as "Will you write reviews based on your chlidren's fondness of the book?"

A.6 Ethical Considerations

As we are interviewing children, ensuring the ethical integrity of the survey is crucial.The following considerations are taken into account:

1. **Informed Consent**:
    - Both parents/guardians and children must provide informed consent before participating.
    - Consent forms will clearly explain the purpose, procedures, risks, benefits, and data confidentiality.
2. **Confidentiality and Data Protection**:
    - Personal information collected will be anonymized and securely stored.

- Data will be used solely for research purposes and will not be shared with third parties.

3. **Minimizing Harm and Stress**:
    - Survey questions will be age-appropriate, avoiding sensitive or distressing topics.
    - Guardians will be encouraged to assist younger participants if needed to reduce any potential stress.
4. **Review by an Ethics Board**:
    - The survey design will be submitted to an Institutional Review Board (IRB) or equivalent ethics committee for approval to ensure adherence to ethical standards.
5. **Child-Centric Language and Format**:
    - The survey will be designed using simple, clear language suitable for children.
    - Visual aids, such as illustrations or emojis, may be used to engage younger participants and enhance understanding.

By addressing these ethical considerations, the survey ensures the protection and well-being of participants while maintaining the integrity and reliability of the research findings.

A sample survey can be found at: Link.Thank Stantcheva (2023) and Chen, Felt, and Henry (2018)'s paper in providing valuable guidance in creating survey.

B Appendix 2: Model diagnostics

Figure 9 are the model diagnostics for the model. Q-Q plot is generally acceptable. There are certain influential points in the data set, however, these points does not occur incorrectly and result of some very popular books, so we are not removing them from the data set. There are some potential violation of linear regression assumptions visible from the residue vs. fitted plot, however we will accept this model for now since we have better model available and this only serves as a help in interpreting the numerical varibles.

B.1 Selection

Our model selection is based on AIC(Akaike information criterion) and interpretation of data.

The best model in terms of AIC in selecting was "average_rating ~ publication_day + publication_month + num_pages + text_reviews_count + text_reviews_ratio + ratings_count + publication_day_in_year", the reason for selecting 'publication_year' and un-selecting 'publication_month' and 'ratings_count' have been discussed in the paper.

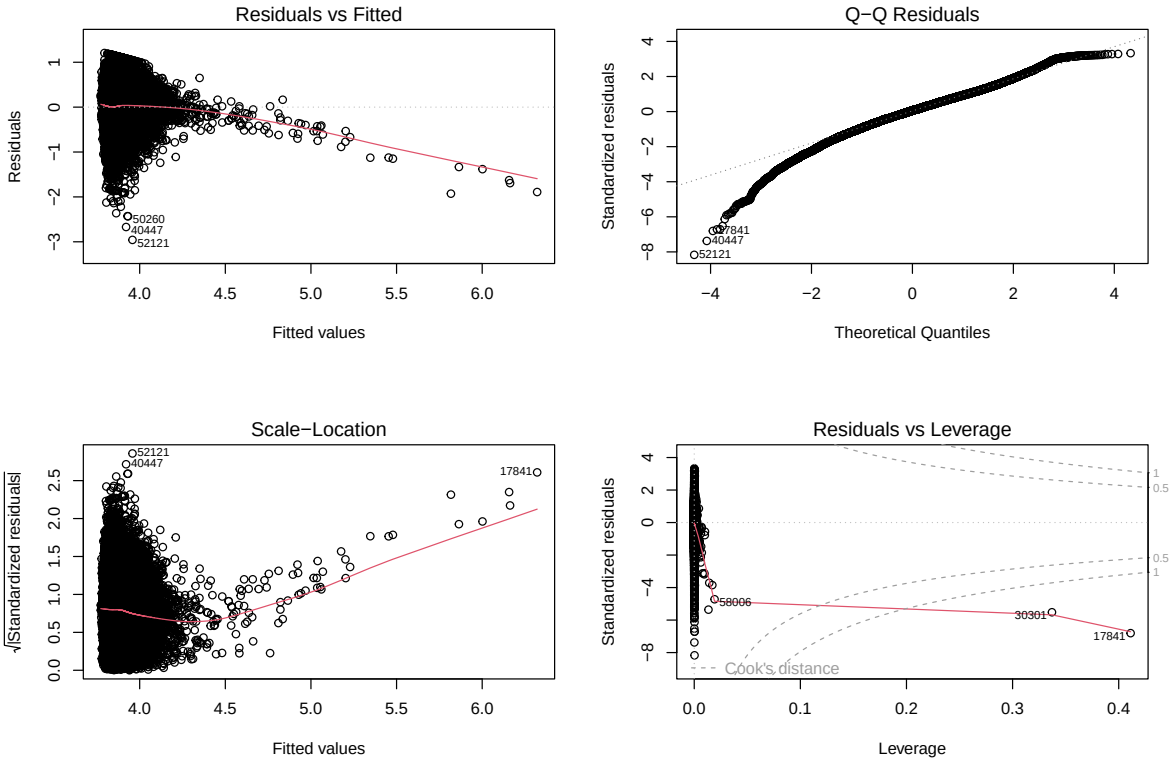

Figure 9: Model diagnostics for Linear regression model

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Chen, Heng, Marie-Hélène Felt, and Christopher Henry. 2018. "2017 Methods-of-Payment Survey: Sample Calibration and Variance Estimation." Bank of Canada. https://doi.org/10.34989/tr-114.

Costa-Luis, Casper da, and Contributors. 2023. "Tqdm: A Fast, Extensible Progress Bar for Python and CLI." https://tqdm.github.io/.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–86. Minneapolis, Minnesota: Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423.

Festinger, L., and J. M. Carlsmith. 1959. "Cognitive Consequences of Forced Compliance." *The Journal of Abnormal and Social Psychology* 58 (2): 203.

Harris, Charles R, K Jarrod Millman, Stéfan J van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, et al. 2020. "Array Programming with NumPy." *Nature* 585 (7825): 357–62.

Kang, Hyeunseok, Sungwoo Goo, Hyunjung Lee, Jung-woo Chae, Hwi-yeol Yun, and Sangkeun Jung. 2022. "Fine-Tuning of BERT Model to Accurately Predict Drug–Target Interactions." *Pharmaceutics* 14 (8). https://doi.org/10.3390/pharmaceutics14081710.

McKinney, Wes. 2010. "Data Structures for Statistical Computing in Python." *Proceedings of the 9th Python in Science Conference* 445 (1): 51–56.

Milton, Ashlee, Levesson Batista, Garrett Allen, Siqi Gao, Yiu-Kai D Ng, and Maria Soledad Pera. 2020. "'Don't Judge a Book by Its Cover': Exploring Book Traits Children Favor." In *Proceedings of the 14th ACM Conference on Recommender Systems*, 669–74. RecSys '20. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3383313.3418490.

Paszke, Adam, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, et al. 2019. "PyTorch: An Imperative Style, High-Performance Deep Learning Library." In *Advances in Neural Information Processing Systems*. Vol. 32.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–30.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Reitz, Kenneth. 2023. "Requests: HTTP for Humans." https://docs.python-requests.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/packag

e=arrow.

Stantcheva, Stefanie. 2023. "How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible." *Annual Review of Economics* 15 (1): 205–34. https://doi.org/10.1146/annurev-economics-091622-010157.

Van Rossum, Guido, and Fred L. Drake. 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Wan, Mengting, and Julian J. McAuley. 2018. "Item Recommendation on Monotonic Behavior Chains." In *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018*, edited by Sole Pera, Michael D. Ekstrand, Xavier Amatriain, and John O'Donovan, 86–94. ACM. https://doi.org/10.1145/3240323.3240369.

Wan, Mengting, Rishabh Misra, Ndapa Nakashole, and Julian J. McAuley. 2019. "Fine-Grained Spoiler Detection from Large-Scale Review Corpora." In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, edited by Anna Korhonen, David R. Traum, and Lluís Màrquez, 2605–10. Association for Computational Linguistics. https://doi.org/10.18653/V1/P19-1248.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wolf, Thomas, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, et al. 2020. "Transformers: State-of-the-Art Natural Language Processing." In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. Online: Association for Computational Linguistics. https://aclanthology.org/2020.emnlp-demos.6.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.

Zhang, Aston, Zachary C. Lipton, Mu Li, and Alexander J. Smola. 2019. *Dive into Deep Learning*. Self-published. https://d2l.ai.

Zhu, Yukun, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. "Aligning Books and Movies: Towards Story-Like Visual Explanations by Watching Movies and Reading Books." In *Proceedings of the IEEE International Conference on Computer Vision*, 19–27. IEEE.