A Song of Reviews and Q&As

Investigating the Role of Q&A Sections on Review Sentiment on Goodreads

MASTER THESIS MARKETING ANALYTICS

SPRING 2023

Name:	Jeroen Maagdenberg
SNR:	2060436
Supervisor:	Shrabastee Banerjee
Co-reader:	Henk C.A. Roest

Executive Summary

User-generated content is becoming more and more important in the buying process of consumers. However, the recently adopted Q&A section within online platforms has not yet been investigated to a great extent. This study focused on examining the role of the presence of a Q&A section on the average review sentiment on Goodreads.com, a platform where users can receive book recommendations and engage in discussions with authors and other users. Next to this main effect, two potential moderators had been identified: the number of likes to the featured question and the number of answers to the featured question.

The study used data from Goodreads and included data on questions, reviews, and related information. Fixed effect models and a difference-in-difference analysis were conducted with the goal to investigate the impact of Q&A on review sentiment. In the difference-in-difference analysis, reviews on Goodreads were considered as the treatment group while Amazon reviews were part of the control group.

Although a positive relationship between having a Q&A section and the average sentiment of reviews was hypothesized, the fixed effect models showed a negative relationship. The moderating effects, however, showed no statistically significant influences. Though, it is important to note that various assumptions of fixed effect regressions were violated. Therefore, it is advised that these results are considered as suggestions rather than instructions. The results of the difference-in-difference analysis did not find a significant influence on the main effect. However, it did find that the sentiment of reviews drops over time. It also implied that reviews on Goodreads are generally less positive compared to the reviews on Amazon. Finally, it also suggested that an increase in question likes led to a more positive sentiment on Goodreads reviews.

Based on the findings, it is not recommended for managers to directly pursue Q&A sections. Given the generally lower sentiment on Goodreads, is suggested that managers first investigate the differences in audience. For instance, it may be the case that users on Goodreads are more critical. Next, managers can replicate the study with other sentiment lexicons as they might reveal different results. Finally, managers are advised to gather question data on the actual first question rather than the featured question.

Preface

The process of writing my thesis has been profound and fulfilling. I've done a lot of research during this time, improved my analytical thinking, and learned more about user-generated content. I am extremely grateful for the advice and assistance I received from my supervisor, Shrabastee Banerjee, whose knowledge and mentoring helped to shape my thesis. In addition, I am very thankful to my friends and family for their full support during both my master’s.

As my time as a student comes to an end, I am optimistic and eager to see what the future holds. Without a doubt, the knowledge and abilities I've acquired throughout both my master's degrees in Marketing Management and Marketing Analytics will serve as a strong foundation for my future endeavors. I'm excited to apply the knowledge I have acquired and have a positive influence on the world. I look forward to the journey that lies ahead with gratitude and enthusiasm, eager to start new endeavors and never stop learning.

Table of Contents

1. Introduction

Over the past decade, online sales have skyrocketed. During the recent Covid-19 period, e-commerce in the U.S. experienced a growth from $571.2 billion in 2019 to $815.4 billion in 2020 (Census Bureau, 2022). When purchasing online, consumers are not able to physically touch or consume products, meaning that they must rely on information that is available online. Among this information is content that comes from other users and consumers, so-called user-generated content. User-generated content is richer in opinions, more up-to-date, and more persuasive than information provided by vendors, which makes it more attention-deserving (Lu, Ye, & Law, 2014). People consult websites with user-generated content to acquire relevant information before purchasing. In general, their motivation is to reduce the risk of making the wrong choice (Cheng & Ho, 2015). On these websites, the user-generated content that they usually encounter are ratings and reviews. However, online retailers and brands have started to integrate community-driven Q&A sections. Considering the recent rise in popularity, little is known about its impact (Banerjee et al., 2021). Therefore, it is crucial to study how the integration of a Q&A section influences the review sentiment.

The influence of a Q&A section on the average review sentiment is worth studying because 40% of customers do not buy electronics without reading reviews pre-purchase (Castelli et al., 2017). When consumers are given more accurate information about a book, it is believed that they are better able to predict whether they will enjoy it or not. When a book does not satisfy the demands of a consumer, it is more likely that it is returned. However, if a book fits their preferences, consumers are more likely to write a more positive review, which influences other customers' intention to buy. It is crucial to minimalize the likelihood of returning as returning books costs retailers about 66% of the original retail price, regardless of the state (Park et al., 2014).

The world's largest website for readers and book recommendations is Goodreads, a community website for book enthusiasts (Goodreads, n.d.; Scuderi, n.d.). On Goodreads, users can discover new books but also become more informed on a specific book. On each product page, users can rate, write reviews, post quotes, join discussions and collaborate on the Q&A section. While these are all forms of user-generated content, users use these functionalities for different purposes. On Goodreads, users can use the Q&A section to become more informed about the attributes or characteristics of a book. This need for information can occur pre-purchase, but also while and after reading a book. In the pre-purchase stage, consumers are engaged and looking for validation since they want to lower the risk that the book is not fit for them (Horton, 2022). In addition, users can ask about certain events or characters that they have read about in the book. However, not every book page has a Q&A section on Goodreads. In this study, it is assumed that a consumer rates, reviews, and answers questions on a book post-purchase as their opinion and knowledge are based on their experience.

When a Q&A section is present, it can differ significantly in size and thus it may inform users on different levels. Some books have only a couple of questions answered, while other books have more than 100 (e.g., ‘A Game of Thrones’ by George R.R. Martin with more than 150 answered questions) (Goodreads, n.d.). The questions may also have different characteristics such as the number of likes or the number of answers to a specific question. A better idea about a book is expected to lead to higher satisfaction. A satisfied customer often spreads word of mouth that is positive (Anderson, Fornell, & Lehmann, 1994). A common form of word of mouth online is a review. According to a study by Lackermair, Kailer, and Kanmaz (2013), 75% of consumers said that the level of quality of reviews significantly affects their choice to buy a product from an online retailer. On Goodreads, questions with a higher number of likes are usually placed higher on the question page. The highest-placed question tends to have the most likes and answers. From here on, it will be referred to as the featured question.

Various studies have been conducted on the impact of user-generated content. Ahn et al. (2016) investigated the motivation for people to write user-generated content. Another study looked at the effect of ratings and reviews on profits of new products (Fainmesser et al., 2021), while Reich and Maglio (2020) explored the impact of mentioning a prior purchase mistake in reviews. The implementation of a Q&A section, a relatively new form of user-generated content, has been investigated significantly less compared to reviews and ratings. A Q&A section enables consumers to ask questions about a product or a specific feature with the goal to mitigate fit uncertainty. This study relates to the study by Banerjee et al. (2021) who found that the integration of Q&As resulted in a decrease in the fraction of reviews with a negative sentiment. These negative reviews are often related to the mismatch between the consumer and the product. It is also known that in an online environment, consumers must rely on mechanisms that express the credibility and reliability of reviews and other forms. On many platforms, this is expressed by voting. Though, Sipos et al. (2014) argue that voting on an answer is not only based on the quality of an answer but also on other users' agreement with the answer.

The current research is mainly focused on either the impact of reviews or Q&As on product rating, such as the findings by Park et al. (2014). In addition, it is mainly focused on Amazon and other retailers while Goodreads is a book review and recommendation website that functions as a community space for book-enthusiasts. Furthermore, the findings of Banerjee et al. (2021) are based on a dataset that consists of questions with a single answer and questions with answers provided by the platform rather than the customer. They pointed out that almost every question receives multiple answers on various platforms. Having multiple answers to a single question has benefits such as diverse perspectives. Next to that, if multiple users provide similar answers, a reader may believe that the information is reliable. However, it could also increase ambiguity. As Branco et al. (2016) pointed out, purchase behavior is affected by the amount of information that is shown. This is supported by Khern-am-nuai et al. (2022) who found that the number of answered questions is a driver of sales. Therefore, as a second contribution, this research will explore the influence of having multiple answers to a question on the beforementioned relationship. This research will contribute to the existing literature in a third manner, by investigating the effect of the number of likes to a question. An indication of the reviews' quality may be seen in the number of likes where a high number usually indicates that a review is more helpful. It has been found that these helpful reviews have a stronger impact on sales (Dhanasobhon et al., 2007; Cheng & Ho, 2015; Mousavi et al., 2021). This is similar to the beforementioned argument of Sipos et al (2014), but the effect of likes on questions is not yet explored.

In the following chapter, a review of existing literature can be found. This is followed by the conceptual framework and an overview of the data. This research uses existing datasets of panel data concerning books and reviews on Goodreads and Amazon. First, the data will be explored by the use of descriptive statistics. Then, the effect of a Q&A section will be estimated by the use of fixed effects regression models. For analyzing the moderation effects, interaction terms will be added to the regression models. This is followed up by a Difference in Difference analysis to see whether the sentiment of reviews on Goodreads differs from the sentiment on Amazon. Here, the reviews on Goodreads will be used as the treatment group, and the reviews on Amazon as a control group. Since Goodreads offers a purchase link to various web shops such as Amazon, spillover effects are expected to exist. However, this will be countered by considering the parallel trend assumption: The average result for the treated and comparison groups would have changed concurrently in the absence of the treatment (Marcus & Sant'Anna, 2021). Hereafter, the results are reported. Finally, a discussion and conclusion on the findings are presented.

2. Literature review

User-generated content can take many different forms. The most prominent forms are ratings and reviews, and a newer form is the Q&A section. These forms of content are often referred to as electronic word-of-mouth. In order to function, a direct commercial relationship is not required for user-generated content (Ramirez, 2020). The impact of user-generated content on consumer behavior has been extensively covered in the literature. For example, Chevalier and Mayzlin (2006) found that reviews at Amazon significantly affect the sales of books. Dhanashobhon et al. (2007) confirm the findings of Chevalier and Mayzlin's (2006) study and add that helpful reviews are related to higher sales. Furthermore, Zhu and Zhang (2010) showed that certain product characteristics amplify the influence of reviews on sales in the video game industry. More recent studies have focused on the characteristics of reviews, such as the writer of the review.

A first literature stream aims at the role of managers, where much research is focused on how the actions of marketers affect the behavior of people as users of a platform but also as purchasers. Albuquerque et al. (2012) explored the relationship between marketing activities (such as price promotions) and user-generated content. They examined that price promotions have a strong influence on purchase behavior, while more than 50% of HP's MagCloud purchases are generated by the referrals of content creators. Similarly, Scholz et al. (2013) explored the effects of marketer- and user-generated across three stages: awareness, interest, and purchase. Their results indicate that awareness is positively related to marketer-generated content while neutral user-generated content has a positive effect on the conversion rate. Also, Colicev et al. (2019) explored the effects of marketer- and user-generated content across the different stages of the marketing funnel, and across services and durable and non-durable goods. They concluded that user-generated content has a stronger informative effect, but weaker persuasive effect than content produced by firms. In addition, Colicev et al. (2019) found that the positive relationships were larger for goods than for services. Furthermore, Chevalier et al. (2018) argue that consumers write reviews with the goal to not only impact other consumers but also managers. Their results indicate that managerial responses encourage customers to write reviews more critically. The role of industry critics is related to this. Critics are very influential in the movie industry, and Basuroy et al. (2003) investigated the effects of reviews by film critics on the box office success of movies. They found that critics can influence and predict the performance of a movie and that negative reviews hurt performance more than positive reviews help. They also believe that these effects are present in other industries such as books, music, and financial markets. Reviews typically act as a one-way communication channel between the reviewer or critic and the reader. In contrast, Q&As offer an interactive, informative, and conversational perspective as they allow for asking and answering questions. As a result, Q&As may offer a distinct point of view compared to reviews (Mousavi et al., 2021).

Next to the actions of managers on user-generated content, a second stream is concerned with the influence of other users. For instance, Naylor et al. (2011) state that reviews written by a user who is considered similar to the reader are more persuasive than those written by dissimilar users. However, these findings contradict the results of the study by Colicev et al. (2019). Next, many platforms allow users to respond to reviews written by other users, similar to managerial responses to consumer reviews. Esmark Jones et al. (2018) show that by allowing consumers to respond to negative consumer reviews, companies can effectively manage product satisfaction. This is in line with the congruity theory (Osgood & Tannenbaum, 1955). According to this theory, consumers feel under pressure to make sense of messages that are contradicting their own. When a reader receives a positive response to their negative review from a different user or a company, a tension is created which causes the reader to make a conscious effort to realign their opinion and change their attitude (Osgood et al., 1957). Furthermore, Bondi (2019) showed that reviews on Goodreads can affect how books are discovered, and, subsequently, change consumer behavior by enabling users to discover lesser-known products that suit their preferences. These findings highlight the importance of investigating the role of user-driven Q&A sections on Goodreads. Q&A sections offer an interactive perspective that can be shaped by the behavior of other users, making it important to explore how they impact the way users perceive and review books on the platform.

In contrast to the extensive research on reviews, the role of a Q&A section on other forms of user-generated content is less investigated. Much of the current literature on Q&A concerns question-answering communities such as Quora. Ravi et al. (2014) explored what defines a high-quality question and how question quality influenced user behavior. Also, the relationship between reputation and question difficulty and answer quality was examined (Lappas, Dellarocas, & Derakhshani, 2016). Rath and Shah (2016) have shown that community Q&A pages should increase moderation in order to reduce the number of unanswered questions. Banerjee et al. (2021) addressed the influences of a Q&A section on consumer reviews in an e-commerce setting. They concluded that by answering questions, consumers can reduce the number of negative reviews for the product categories Technology and Home & Garden. In these product categories, a fairly clear distinction can be made between the quality and the fit of a product. In contrast, this study will explore the impact of Q&A technology on reviews in the book industry, a product category that has hedonic and utilitarian purposes (Dhar & Wertenbroch, 2000), and where quality and fit are more closely related (Banerjee et al., 2021). In addition, their study was based on data that comes from a U.K.-based big-box retailer. The data includes sales data, reviews, and Q&As posted on its website. Goodreads does not sell books, but only offers links to bookstores such as Amazon, Barnes & Noble, and Walmart. Therefore, the influence of Q&A on the review sentiment may be different.

Consumers who shop online are exposed to two different types of uncertainty: product quality uncertainty and product fit uncertainty (Banerjee et al., 2021). In an offline shop, consumers can touch products and create expectations about the quality of a product and its performance (Sun et al., 2022). However, this is not the case in an online shopping environment (Kim & Krishnan, 2015). Online shoppers rely on the information provided by retailers. According to Dimoka et al. (2012), product quality uncertainty is the result of the inability of a retailer to describe a product's characteristics, quality, and performance. By allowing users to ask and respond to questions about the performance and quality of a book, the addition of a Q&A section may reduce this uncertainty. Banerjee et al. (2021) argue that reviews are very useful as they allow consumers to read about the quality based on the experiences of others. Product fit uncertainty, on the other hand, describes the extent to which a product fits the needs of a consumer. In this case, a product may be returned to the retailer, regardless of the quality (Banerjee et al., 2021). It is one of the main causes of the high rates of online product returns (Wang et al., 2021). Several studies have been conducted on these two types of uncertainty. Weather et al. (2007) argue that consumers should have control over information when they look for products with predominantly search attributes. This means that allowing consumers to look at all information (including Q&A and reviews) for as long and whenever they want, leads to a reduction in quality uncertainty. The quality of a service can be uncertain even after purchase (Ju Choi & Kim, 1996). Further, Yu et al. (2023) concluded that for a movie with low-fit uncertainty, its search attributes (such as information on tickets and release date) should be shared through social media instead of traditional advertising.

3. Conceptual framework

This study focuses on the impact of a Q&A section on the average sentiment of reviews on Goodreads. In addition to these variables, two moderators have been identified that may influence the relationship between the independent variable and the dependent variable. The first moderator is the number of answers on the most liked question, while the second is the number of likes on that specific question. From here on, the most liked question is referred to as the featured question. In this chapter, a conceptual framework is outlined that shows the relationships between the variables in this study.

The Q&A section on Goodreads enables users to ask and answer questions related to books and it serves as a platform for knowledge sharing and community building. By doing so, users may feel more informed about the books they read, which may contribute to a more positive reading experience and indirectly to a more positive review sentiment. It may be the case that users have asked questions about a book but have not received an answer. Therefore, the Q&A section is present when at least one question has been answered. As explained before, a Q&A section allows a consumer to acquire relevant information about a book before but also after purchase (Horton, 2022). Furthermore, allowing consumers to respond to negative consumer reviews, lead to a more positive product satisfaction as found by Esmark Jones et al. (2018). While this is the case for reviews, it is assumed that a similar relationship is present for Q&A technology. Therefore, the following is hypothesized:
H1: *The presence of a Q&A section has a positive effect on the average sentiment of reviews.*

Consumers who participate in a community, such as Goodreads, are known to interact with other members and share their opinions and experiences. Lee, Han, and Suh (2014) found that such interactions tend to motivate customers to make their first purchase of a good or service. By frequently discussing product-related issues and seeking out product knowledge, users who have read books can inform other users and have a significant impact on their buying decisions. However, studies on the impact of information load on consumer behavior show inconsistent results. Based on the limited capacity model, information overload can occur when consumers are provided with too much information at once (Lee & Lee, 2004). Lee and Lee (2004) and Sicilia and Ruiz (2010) found an inverted U-shaped relationship between information load and consumer behavior. This indicates that the behavior is positively influenced as the information load increases, but beyond a certain point, more information will have a negative effect. However, Huang et al. (2013) showed that more information leads to increased perceived trustworthiness from consumers toward online retailers. This is supported by Mousavi et al. (2021) who found that the presence of a Q&A section has a positive effect on the average rating of products, and that this effect is greater as the number of answers to a single question increases. This leads to the hypothesis:

H2: *The effect of the presence of a Q&A section on the average sentiment of reviews increases (decreases) as the number of answers on the featured question goes up (down).*

When users browse through the Q&A section, they may pay more attention to questions that have a significant number of likes. When a question received many likes, it has been deemed as a relevant or important question by other Goodreads users. Therefore, it increases the visibility of the question, and it will appear as the featured question and at the top of all questions. These featured questions are above the fold, wherefore they are given most of the attention (Mahfooz & Nadeem, 2019). Since these questions tend to have more attention, it is also likely that they receive more interaction, such as more answers. Wu et al. (2018) found that consumers' community-participation increases both the likelihood of post-purchase reviews and the likelihood that these reviews are positive. It is therefore expected that the following effect occurs:

H3: *The effect of the presence of a Q&A section on the average sentiment of reviews increases (decreases) as the number of likes on the featured question goes up (down).*

Based on the three hypotheses mentioned above, the following conceptual framework is constructed:

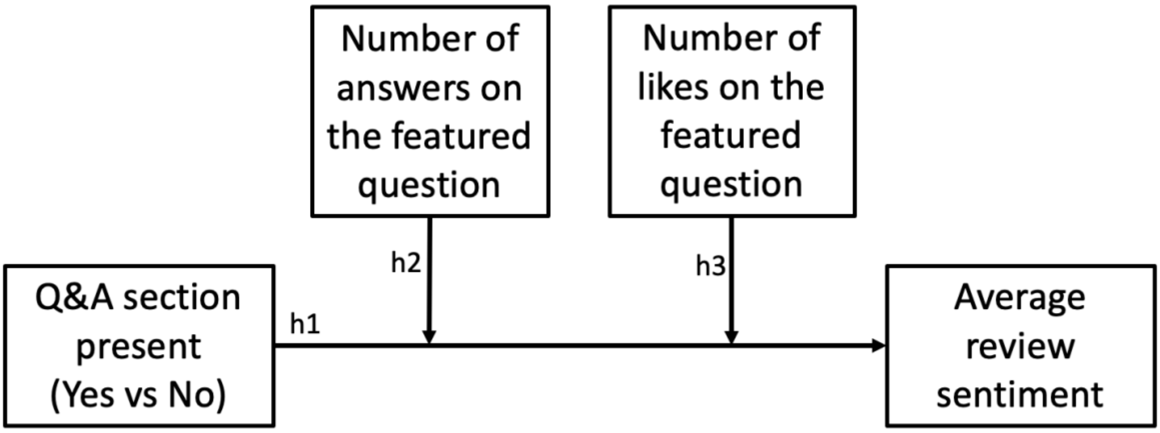


Figure 1: Conceptual framework

4. Data

This research used a large collection of panel data concerning books on Goodreads and Amazon to estimate the model. The data consisted of information on the books, but also on related information that is available on Goodreads and Amazon. The data was obtained from an external source and provided to the researcher by supervisor S. Banerjee. The dataset was provided to assist the progress of the analysis and to answer the research question and supporting hypotheses. The process of data preparation is explained in the following section. RStudio was used to transform the raw data into the final dataset and to analyze the data. All source code files for data preparation and analyses can be found on a GitHub repository. The link to this repository can be found in Appendix A.

4.1. Goodreads

4.1.1. Raw data overview

In order to have a complete dataset on Goodreads data, three datasets with data from Goodreads had to be merged. The first dataset, from here on called 'subset_questions', consisted of 132,626 observations. Each observation includes various elements regarding product information, questions, and answers per book. In this dataset, there were 100,034 unique books. The number of likes on a question and the number of answers to a specific question were also present for each observation. Furthermore, this dataset had a column that specifies the number of days between the date of scraping and the date of the question until one year.

To specify the exact date from one year onwards, data had been scraped from the website Internet Archive ("web.archive.org"), a digital archive of the internet. This second dataset, from here on called 'web_archive', is similar to 'subset_questions', however, the information on books is less detailed, and thus the number of likes on a question and the number of answers to a specific question were not available. Though it provided the exact date of the featured question. This dataset consists of 12,139 books that are all present in 'subset_questions'.

The third dataset focused on reviews posted on Goodreads. The reviews are only from books that are present in 'subset_questions'. This dataset consists of 1,138,076 observations that include the ID of a book and of a review, the review text itself, and the date of the review.

4.1.2. Data preparation

Since this study is focused on Q&A, the observations of 'subset_questions' without at least one question were omitted. This resulted in 31,328 observations across 14,174 unique book IDs. As mentioned before, it is only possible to specify the exact date of a question until one year after it was posted on Goodreads. Therefore, this dataset was filtered on only observations that contain "days ago" or "months ago" in the date column. The strings with "month ago" were converted to "days ago" and the exact date was extracted.

This was followed by transforming the 'web_archive' dataset. In this dataset, the column 'Url_timestamp' contained timestamps that were extracted from the URLs that the Web Archive uses. Similar to the steps with the previous dataset, the question timestamp had been deducted from the 'Url_timestamp' to extract the exact date of the question. Then these questions are added to the subset_question dataset. As mentioned before, the observations scraped from Internet Archive did not have Likes and Number of Answers. However, the subset_question dataset does. A temporary data frame was created with all book IDs and their corresponding Likes and Number of Answers. These were then added to the complete list of questions.

Next, the dataset on Goodreads reviews had to be prepared for the analyses. First, all observations with whitespaces for reviews were omitted. This was followed by removing links, HTML codes, special characters, and numbers from the reviews. Next, the string of the review date was transformed into a readable format. Then, the reviews were merged with the questions into one single dataframe. Lastly, dummies were introduced to indicate whether a review was posted before or after the featured question was posted.

Finally, the sentiment of the reviews had to be computed and added to the merged file. The cleaned reviews were used for this. Sentiment analysis is the process of identifying and quantifying subjective information from text, such as attitudes (Rocklage, He, Rucker, & Nordgren, 2023). This technology is increasingly being used to measure sentiment in user-generated content (Baek, Ahn, & Choi, 2012). A common approach is to use a lexicon, a dictionary where each word is connected with a score (Bessa, 2022). The tidytext package by Silge & Robinson (2016) offers several popular lexicons. Each lexicon has been validated via crowdsourcing and has been tested on data, such as Twitter data and reviews on restaurant movies. The tidytext package is an easy and effective tool for handling large datasets and it allows data manipulation with other popular tidy packages such as dplyr and tidyverse. Therefore, it has been considered as a suitable option for Goodreads reviews. This study uses the AFINN lexicon, where each word has been assigned with a numerical score from -5 to 5, indicating a negative or positive sentiment for a particular word. The AFINN lexicon is simple to use and has a large vocabulary. The overall sentiment score of a review is calculated by summing the scoring of all words in a review. As expected, an overall score that is positive indicates a positive sentiment, and a negative overall score indicates a negative sentiment. The output of AFINN is a single score that represents the overall sentiment of the text (Silge & Robinson, 2016).

To measure the sentiment of a review, the review text had to be transformed into a tidy text format. This format is a ‘table with one-token-per-row’, where a token is a unit of text, such as a word. The AFINN lexicon is based on unigrams, which means that it only works with tokens that are a single word. By using the unnest_tokens() function, each review was separated into tokens of single words. To aid the sentiment measurement, common stop words such as “the”, “to” and “and” had to be removed as these words are generally not laden with sentiment (Silge & Robinson, 2016). The Goodreads tidy text dataframe consisted of 48,788,900 words, and the top 5 most frequent words in Goodreads reviews can be found in Appendix B.1. The tidy text dataframe was then used for sentiment measurement and its output is added to the file that contains both Goodreads reviews and questions. Finally, a new variable called ‘Year_Month’ is created that contains the year and month of the review.

4.1.3. Estimation Sample

The final dataset is a merged dataset that combines book information, questions, answers, and reviews found on book pages on Goodreads' website. Due to the cleaning process, various observations were omitted. This led to 1,132,755 observations (reviews) across 14,123 books. A random selection of 3 positive, 3 negative and 3 neutral reviews with their AFINN score can be found in Appendix B.2. As seen in Appendix B.3, the number of reviews rose steadily over the past decade to a total of 1,132,755. This again shows why it is crucial to understand how managers can assert influence on the sentiment of reviews. Table 1 includes some descriptive statistics on the Goodreads reviews.

	Mean	**SD**	**Max**	**Min**
Likes	1.7751	17.0440	1504	0
Number of Answers	1.5562	5.2014	185	0
AFINN score at review level	2.6965	11.5136	7,260	-287

Table 1: Descriptive statistics on Goodreads reviews

An interesting observation is the positive AFINN score of 7260 of a single review compared to the mean of 2.6965. Figure 3 suggests that the review with a score of 7,260 is an outlier and that all other reviews have a score lower than 1,000. As it is believed that an outlier could disrupt the regressions, the observation was omitted by filtering out reviews with a AFINN score larger than 1,000 (Mahajan, Sharma, & Wind, 1984). Furthermore, Figure 4 shows that the average sentiment was very fluctuating during beginning phase. Once the number of reviews increased (as seen in Appendix B.3), the sentiment became more stable and experienced a small upward trend.

Figure 3: Scatter Plot of AFINN Scores on Goodreads

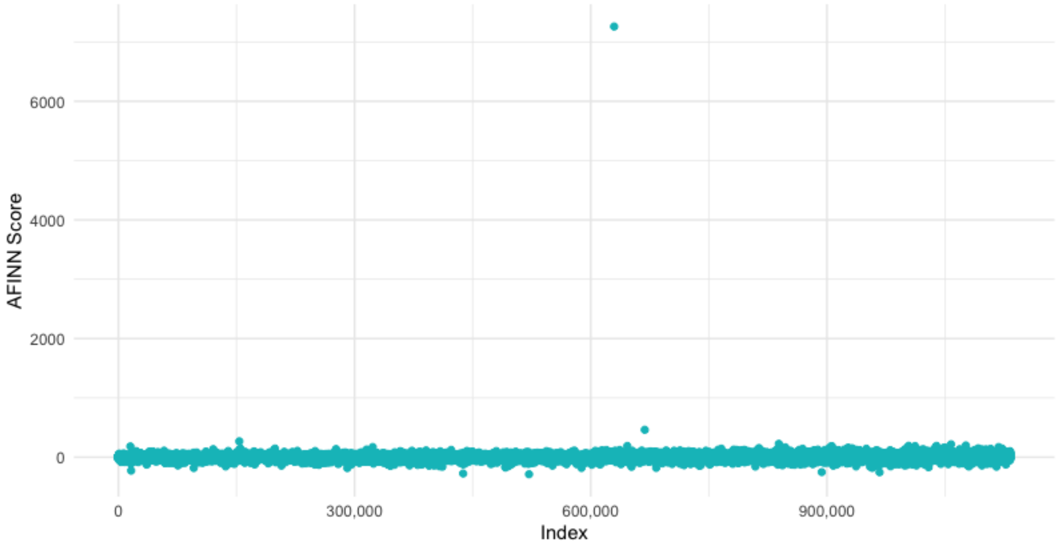
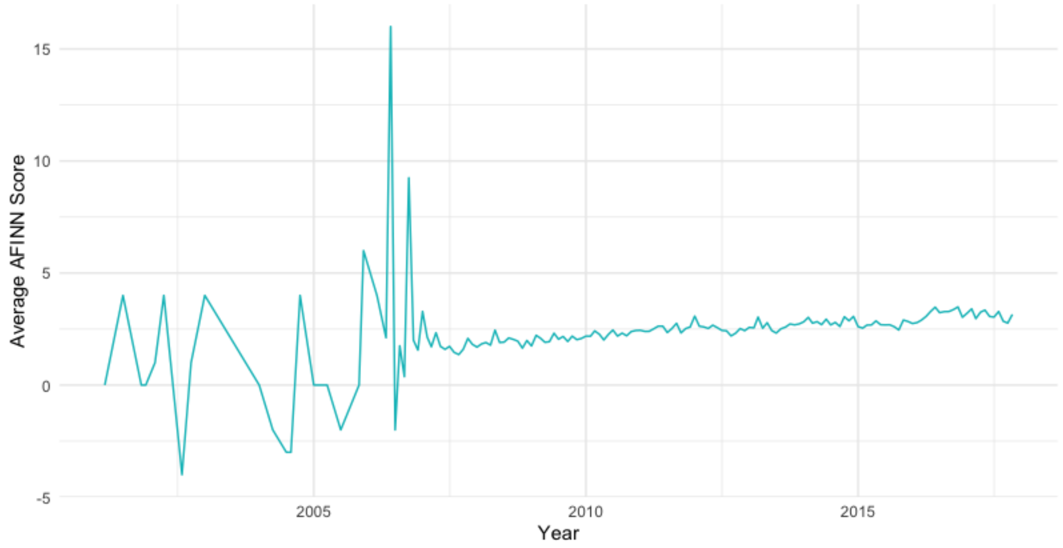


Figure 4: Average AFINN Score on Goodreads



With the exclusion of the outlier, 1,132,754 reviews across 14,123 books remained for estimations. Therefore, the dataframe is referred to as the estimation sample. Table 2 shows the updated descriptive statistics based on the books and reviews that were taken into consideration for this research. As seen in the Max column, the review with the most positive sentiment score now has a score of 462 instead of 7,260.

	Mean	**SD**	**Max**	**Min**
Likes	1.7751	17.0440	1504	0
Number of Answers	1.5562	5.2014	185	0
AFINN score at review level	2.6901	9.2772	462	-287

Table 2: Updated descriptive statistics on Goodreads reviews without the outlier

Since both the date of posting a review and the date of posting the featured question were known, it was possible to determine which reviews were written before and after. As seen in Table 3, the average sentiment of the reviews written after the date of posting the featured question was slightly lower (2.5310 versus 2.7420).

	Before	**After**
Number of reviews	854,301	278,453
AFINN mean	2.7420	2.5310
AFINN SD	9.2995	9.2068

Table 3: Comparison number of reviews and sentiment score between Before and After

4.2. Amazon data

4.2.1. Raw data overview

As it is plausible that review sentiment on Goodreads might be affected by unobservable influences, a comparative analysis with Amazon data was implemented to control for these influences. This had been done by using a difference-in-differences analysis. To start this analysis, the reviews from Amazon were matched with the books that were present in the Goodreads dataset. To match the reviews with their corresponding book ID, two datasets were used. The first dataset contained the review text while the second contained the book ID. Both datasets had an ASIN (Amazon Standard Identification Number) on which they were to be matched (Rogers, n.d.). The dataset that contained the review text consisted of 2,649,037 observations.

4.2.2. Data preparation

When a large dataset is used and has to undergo cleaning and measurement tasks, it is likely that it is confronted with computational limitations and time constraints. With the sizable dataset that had been used in this study, the cleaning and sentiment measurement tasks were very computational. In this case, it would mean that each word in the 2,649,037 reviews had to be separated or unnested. Hence, a subsample of approximately 50% was used to facilitate a smoother preparation process. To avoid potential biases arising from varying numbers of reviews per book, a more refined sampling approach was used. Instead of simply selecting a random sample of 50%, an extra step was included. By using the random_frac() function in R, a fraction of 50% of the Goodreads books is chosen randomly. This subset did not include any biases and is believed to represent the whole population. Next, the Amazon review dataset was filtered based on the book IDs in this subset, resulting in only reviews corresponding to those books. By using this approach, a book in the subset will have all its reviews, and thus disproportionate review counts are avoided, which increases validity and reliability.

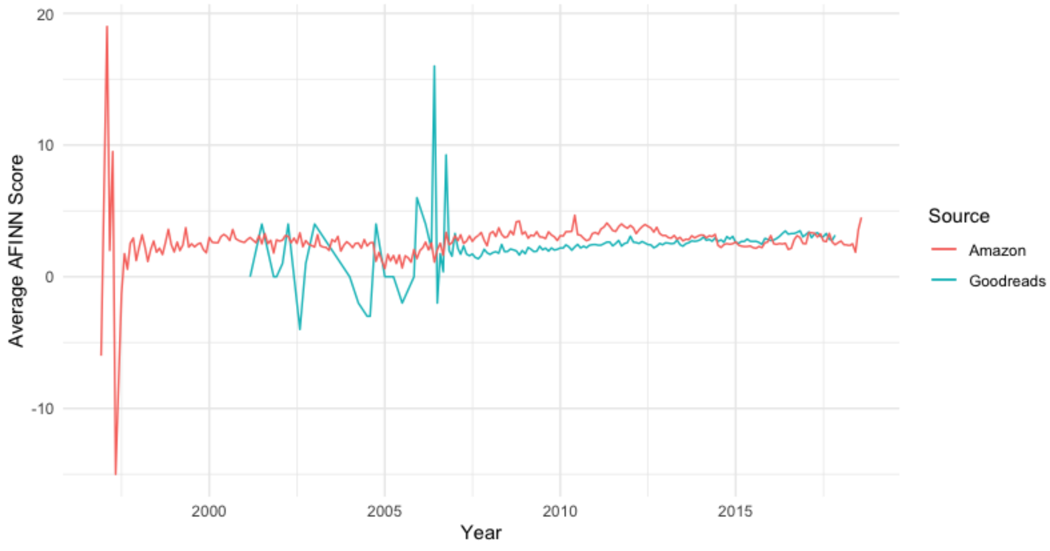
As mentioned before, there were 14,174 unique book IDs in the Goodreads dataset. Using the overlap file, 13,773 books were matched and thus 6,886 books remained in the subset. Adding each review to these books resulted in a total of 1,256,647 reviews. Similar to the Goodreads reviews, these reviews were cleaned and unnested, and the sentiment scores were measured for 38,018,485 words. The five most frequent words can be found in Appendix B.4. Finally, the year and month of the review date are stored in 'Year_Month'.

4.2.3. Estimation Sample

The Amazon reviews dataset used for the difference-in-difference analysis contained 1,256,647 reviews and their relevant information on books such as dates, likes, and number of answers. Reviews on both platforms experienced a significant growth over the years (Appendix B.5). Similar to the Goodreads data, a scatter plot of the AFINN score of Amazon reviews can be seen in Figure 6. The scores were distributed closer to each other and did not show any significant outliers. Therefore, no observations were removed in this case. Similar to the average sentiment score of Goodreads, the reviews on Amazon also experienced a more stable slope once the number of reviews on the platform increased (as seen in Figure 7).

Figure 6: Scatter Plot of AFINN Scores on Amazon

Figure 7: Comparison of Average AFINN Score on Goodreads vs Amazon



As mentioned above, the Amazon reviews are only for a subset of books. Therefore, the descriptive statistics below were adjusted to this. The means of all three variables are similar to the means of the Goodreads reviews. However, the largest number of Likes is significantly smaller (365 versus 1,504). Though, it is expected that the 1,504 is not an outlier but rather deemed as a very good question.

	Mean	**SD**	**Max**	**Min**
Likes	1.6020	10.1265	365	0
Number of Answers	1.5200	5.0682	185	0
AFINN score at review level	2.7513	7.3748	344	-459

Table 4: Descriptive statistics for Amazon reviews within the subset of books

5. Model

5.1. Regressions

As described in chapter 3, it is expected that there is a positive relationship between the presence of a Q&A section and the average review sentiment. To gain an idea about the relationship, an independent two-sample t-test could be employed. However, it is hypothesized that there are also moderating effects in play. This means that an independent two-sample t-test is not sufficient. Though, it can provide an indication. Its results and implications can be found in Appendix C. In order to analyze the effect of Q&A on sentiment of book reviews and how the moderators affect this relation, various regressions have been developed. A simple OLS regression was used to predict the main effect of including a Q&A section on the overall review sentiment. In this case, the dependent variable, Y_{ijt}, stands for the sentiment score of a consumers' review i of a particular book j at a particular day t. The dummy variable, $Post_{ij}$, is the independent variable and indicates whether a review was posted after or before the question was posted. After is represented by a 1 and before by a 0. The model is represented by the equation below:

$$Y_{ijt} = \beta_0 + \beta_1 \times Post_{ij} + \varepsilon_{ij} \quad (1)$$

The intercept, β_0, stands for the average sentiment score of book reviews before any treatment. β_1 is the coefficient that indicates the effect that happens due to addition of a Q&A, and ε_{ij} is the error term. As mentioned before, this model represents only the main effect. In the following models, additional variables, such as moderators and interactions, are included. First, the moderators and their interaction with the main effect will be added to the model:

$$Y_{ijt} = \beta_0 + \beta_1 \times Post_{ij} + \beta_2 \times Likes_j + \beta_3 \times Number\ of\ Answers_j + \beta_4 \times \left(Post_{ij} \times Likes_j\right) + \beta_5 \times \left(Post_{ij} \times Number\ of\ Answers_j\right) + \varepsilon_{ij} \quad (2)$$

The coefficient β_2 indicates the effect on the sentiment when the featured question receives one more like, ceteris paribus. Similarly, β_3 shows the impact when the featured question receives one more answer. β_4 and β_5 are the interaction effects of the respective variable with the main effect. Next to the moderating effects, it is possible that there exists a three-way interaction between the moderators and the main effect. To predict this, the following model is constructed:

$$Y_{ijt} = \beta_0 + \beta_1 \times Post_{ij} + \beta_2 \times Likes_j + \beta_3 \times Number\ of\ Answers_j + \beta_4 \times \left(Post_{ij} \times Likes_j\right) + \beta_5 \times \left(Post_{ij} \times Number\ of\ Answers_j\right) + \beta_6 \times \left(Post_{ij} \times Likes_j \times Number\ of\ Answers_j\right) + \varepsilon_{ij} \quad (3)$$

where β_6 is the coefficient for the joint interaction of the moderators on the main effect. It is worth nothing that this three-way interaction is not in line with the conceptual model. As panel data is used, it is likely that unobserved variables are exerting influence in these models. To capture these effects, it is recommended to add product fixed effects and time fixed effects to the models (Cunningham, 2021). In such a fixed effects model, the model is based on fixed effects specific to an observation. Therefore, the intercept β_0 is replaced with the α_i parameter. Using the previous three models as baseline, the following three models include both types of fixed effects:

$$Y_{ijt} = \alpha_i + \beta_1 \times Post_{ij} + \gamma_j + \delta_{ij} + \varepsilon_{ij} \quad (1)$$

$$Y_{ijt} = \alpha_i + \beta_1 \times Post_{ij} + \beta_2 \times Likes_j + \beta_3 \times Number\ of\ Answers_j + \beta_4 \times (Post_{ij} \times Likes_j) + \beta_5 \times (Post_{ij} \times Number\ of\ Answers_j) + \gamma_j + \delta_{ij} + \varepsilon_{ij} \quad (2)$$

$$Y_{ijt} = \alpha_i + \beta_1 \times Post_{ij} + \beta_2 \times Likes_j + \beta_3 \times Number\ of\ Answers_j + \beta_4 \times (Post_{ij} \times Likes_j) + \beta_5 \times (Post_{ij} \times Number\ of\ Answers_j) + \beta_6 \times (Post_{ij} \times Likes_j \times Number\ of\ Answers_j) + \gamma_j + \delta_{ij} + \varepsilon_{ij} \quad (3)$$

In the models above, γ_j captures the book-specific fixed effects and δ_{ij} the time fixed effects. Y_{ijt} and the betas still represent the same effects as mentioned before. Finally, covers the residual error term that is not captured by the other elements in the models.

To ensure that the aforementioned regressions are reliable, the models should meet four crucial assumptions. The first assumption is justified when the error term is uncorrelated across all observation. This is the case when ε_{ij} has an expected mean of 0. Secondly, each variable has to be independently drawn and identically distributed from the same distribution. The third assumption is met when there are no large outliers. Finally, it is not allowed to have perfect multicollinearity between two or more variables (Hanck et al., 2023). As there is a possibility that homoskedasticity is present, the regressions include clustered standard errors based on book ID.

5.2. Difference-in-difference model

It is possible that the fixed effect model described above may not fully capture the relationship. Therefore, a difference-in-difference analysis has been conducted. A difference-in-difference analysis allows comparison between the outcome of a treatment group and a control group, before and after the occurrence of the treatment. By taking the difference in the periods before and after the treatment for both groups, an estimation can be made of the average treatment effect on the treated group (ATT) (Goodman-Bacon, 2021). Developing the model is done in two steps: first, a simple comparison of the sentiment between before and after, followed by a comparison between Goodreads and Amazon. The equation on the effect of the treatment by comparing before and after is stated as followed:

$$ATT = E\left(Y_{j1} - Y_{j0} \,\middle|\, D = 1\right)$$

Where ATT stands for the average treatment effect on the treated. Y_{j1} represents the sentiment score for a book j after the featured question was posted, while Y_{j0} stands for the score of a book j before the Q&A section was present. Finally, $D = 1$ indicates that a book j is among those after the featured question was posted. Next, the comparison between Goodreads and Amazon, denoted by Z, is introduced:

$$E(Y_{i1} - Y_{i0} \mid D = 1, X) = E(Z_{i1} - Z_{i0} \mid D = 0, X)$$

In this equation, the post-treatment differences in review sentiment of a book on Goodreads is compared to the post-treatment differences of the same book's sentiment on Amazon. Since Amazon is the control group, its review sentiment is considered as a baseline. This is also referred to as the parallel-trends assumption. It posits that when both the treatment and control group had not received treatment, they would have experienced similar differences over time (Proserpio & Zervas, 2016). Now that the difference-in-difference equation is formulated, it will be transformed to a regression that also includes the aforementioned fixed effects. As previously mentioned, the three-way interaction of 'Post x Likes x Number of Answers' is not aligned with the conceptual framework. Since the difference-in-difference analysis also adds the variable 'Treated', this interaction becomes a four-way interaction. Thus, the difference-in-difference regressions without and with four-way interaction are presented below:

$$Y_{ijt} = \alpha_i + \beta_1 \times Post_{ij} + \beta_2 \times Treated_{ij} + \beta_3 \times \left(Post_{ij} \times Treated_{ij}\right) + \beta_4 \times \left(Post_{ij} \times Treated_{ij} \times Likes_j\right) + \beta_5 \times \left(Post_{ij} \times Treated_{ij} \times Number\ of\ Answers_j\right) + \gamma_j + \delta_{ij} + \varepsilon_{ij}$$

$$Y_{ijt} = \alpha_i + \beta_1 \times Post_{ij} + \beta_2 \times Treated_{ij} + \beta_3 \times (Post_{ij} \times Treated_{ij}) + \beta_4 \times (Post_{ij} \times Treated_{ij} \times Likes_j) + \beta_5 \times (Post_{ij} \times Treated_{ij} \times Number\ of\ Answers_j) + \beta_6 \times (Post_{ij} \times Treated_{ij} \times Likes_j \times Number\ of\ Answers_j) + \gamma_j + \delta_{ij} + \varepsilon_{ij}$$

Where $Post_{ij}$ denotes whether the review was written before ($Post_{ij}$ = 0) or after ($Post_{ij}$ = 1) the featured question was posted on Goodreads. When a review was posted on Goodreads, the variable $Treated_{ij}$ is equal to 1. For a review on Amazon, $Treated_{ij}$ equals 0. The purpose of the difference-in-difference analysis is to capture the differential effect of having a Q&A. In these regressions, this effect is captured by the interaction $Post_{ij} \times Treated_{ij}$ and its corresponding coefficient, β_3. A positive and significant coefficient implies that having a Q&A section has a positive effect on a book's average review sentiment. In that case, it would suggest that, on average, books on Goodreads experiences a larger increase in review sentiment compared to books on Amazon during the post-treatment period.

6. Results

6.1. Fixed effects regression

The data described in Chapter 4 is used to estimate the three proposed models from the chapter above. The models include fixed effects, for unit-specific fixed effects the Book ID is used and for the time-fixed effect the Year_Month variable is used. The results are described below.

As shown in Table 5, 1,132,754 reviews on Goodreads are used for estimating the three models. The moderators Likes and Number of Answers are removed due to collinearity and have no results. Model 1 demonstrates the main effect of the presence of a Q&A section on the average review sentiment. The variable 'Post' is significant with a p-value smaller than 0.001 and has a coefficient of -0.702. This indicates that the average review sentiment is, on average, 0.702 less positive after a book receives a question compared to the average review sentiment before a question is posted. Model 1 shows that the independent variable explains 13% of the variance in review sentiment (Adjusted R^2 = 0.12954). In the second model, both the moderators and their interaction with the variable 'Post' are added. Similar to the first model, the 'Post' variable is significant and has a coefficient of -0.754. Though, it is estimated that both moderators are statistically insignificant. Therefore, no conclusion can be made about their influence. Model 2 also explains 13% of the variance (0.12956). Lastly, Model 3 extends on the previous model by including a three-way interaction. Again, the variable 'Post' is significant and shows a negative relationship (-0.826), and 13% of the variance is explained by the independent variable (0.12958). Though, the interaction of Post and Likes is only significant when a significance level of $p < 0.05$ is used.

	Model 1	**Model 2**	**Model 3**
Post	-.702***	-.754***	-.826***
	(.056)	(.058)	(.066)
Likes			
Number of Answers			
Post x Likes		.002	.007 .
		(.002)	(.004)
Post x Number of Answers		.001	.006
		(.003)	(.004)
Post x Likes x Number of Answers			-.000
			(.000)
Observations	1,132,754	1,132,754	1,132,754
Adjusted R^2	.12954	.12956	.12958
Within R^2	.00039	.00041	.00044

*Table 5: Outcome fixed effect regressions of Model 1 through Model 3. Standard errors (clustered by book ID) are shown in parentheses. '***' $p < 0.001$ & '.' $p < 0.05$*

The second model is used to test the four regression assumptions, despite Model 3 having a slightly larger adjusted R^2. The second model is based on existing literature and the conceptual model described in Chapter 3. The first assumption is tested by using the Breusch-Pagan test (see Appendix D.1). The test statistic has a value of 228.75 with a corresponding p-value of 0.000. As a result, the assumption is violated since the model could contain heteroscedasticity. To test whether residuals are independent, the Breusch-Godfrey test has been used (Appendix D.2). Again here, the p-value is 0.000 which implies that there might be a correlation. The third assumption, normality, is also violated as the Jarque-Bera test shows a very large X-squared

(Appendix D.3). This is confirmed by the slope shown in the normal Q-Q Plot in Appendix D.4. The only assumption that is not violated is multicollinearity. This assumption is tested by using a correlation matrix which can be found in Appendix D.5.

6.2. Difference in difference

As previously mentioned, the models estimate that the independent variable, the presence of a Q&A section, and the dependent variable, the average review sentiment, only have a statistically significant, negative relationship. Since the relationship was hypothesized as positive, it is interesting to see how this effect is compared to a group that has not been exposed to the presence of a Q&A section. Here, the reviews on Amazon come into play and will act as a control group. In case the average sentiment on both Goodreads and Amazon experience a similar trend, it is highly likely that there were other factors at play. When the opposite is discovered, it can be argued that having a Q&A section does have a negative effect on the average review sentiment. The difference-in-difference analysis is conducted using both regressions with and without the four-way interaction. However, based on the previous analysis, it has been determined that Model 2 is the suited model. Therefore, the next section is focused on the difference-in-difference analysis based on Model 2. Though, the output with the four-way introduction can be found in Appendix D.6.

According to the difference-in-difference output in Table 6, both the 'Post' and 'Treated' variables have statistically significant effects on the average sentiment score of reviews. The coefficient estimate for the 'Post' variable is -0.417 which suggests that there is a negative relationship between posting a review after the featured question was posted and the average score of review sentiment. In other words, a review that was posted after the featured question has, on average, a 0.417 lower sentiment compared to a review posted before that question. A negative relationship is consistent with the results of the fixed effects model. The coefficient for 'Treated' (-0.543) indicates that it is likely that reviews on Goodreads have a lower sentiment score than those on Amazon, which was used as a control group.

	Review sentiment
Post	-.417***
	(.070)
Treated	-.543***
	(.045)
Post x Treated	-.021
	(.087)
Post x Treated x Likes	.005***
	(.001)
Post x Treated x Number of Answers	-.000
	(.002)
Observations	1,816,642
Adjusted R^2	.12916
Within R^2	.00084

*Table 6: Output Difference-in-Difference based on Model 2. Standard errors (clustered by book ID) are shown in parentheses. '***' $p < 0.001$*

The combined effect of these two influences was found to be insignificant. This means that no conclusions can be made about this interaction term. However, the interaction term 'Post x Treated x Likes' is significant with an estimate of 0.005. On the other hand, the interaction 'Post x Treated x Number of Answers' was found to be statistically insignificant. Hence, no conclusions can be made about this interaction term. Finally, the adjusted R^2 of 0.12916 indicates that approximately 13% of the variance in the model is explained by the independent variable which is similar to the fixed effects regressions.

As seen in Appendix D.6, the interaction term of 'Post x Treated x Likes' becomes insignificant when the four-way interaction is included in the difference-in-difference. The outcomes of the other variables and interactions remain fairly similar to the difference-in-difference analysis based on Model 2.

7. Discussion

The goal of this study was to investigate the relationship between the presence of a Q&A section and the average review sentiment on Goodreads. In order to express the presence of the Q&A section, a dummy variable called 'Post' was created. This dummy identified whether a review was posted before or after the featured question was posed. To gain an indicative idea, an independent two-sample t-test was conducted. It found that the relationship was significant and implied that the presence of a Q&A section influences the average sentiment of reviews on Goodreads.

Three fixed effect regressions were conducted to find out what role the presence of a Q&A section plays and what affects it. The model of the first regression (Model 1) found a negative effect of the presence of a Q&A section and suggested that the sentiment score of a review is, on average, 0.702 lower when it is written after the featured question is posed. In Model 2 the moderators Number of Answers and Likes and their interaction with the presence of a Q&A section were introduced. However, it did not appear that the moderators have a statistically significant influence. Model 3 also added a three-way introduction between the independent variable and both moderators. Though, only the main effect was found to be significant at a significance level of $p < 0.001$. All three models estimated that the presence of a Q&A section on Goodreads had a negative influence on the average sentiment of reviews. This finding contradicts with hypothesis H1, which hypothesized a positive relationship. A possible explanation may be that a Q&A section at first increases the expectations of a user, but it increases the disappointment even more when expectations are not met. The variables estimated in Model 2 followed the conceptual model based on literature. Therefore, this model was tested on the four assumptions for fixed effect models. Three out of the four assumptions were violated, except for the multicollinearity assumption. This implies that there is a possibility that the outcome is biased and hence inaccurate.

The difference-in-difference analysis based on Model 2 showed a significant, yet negative effect for the variable 'Post' (-0.417), meaning that the average sentiment of reviews goes down over time. The difference-in-difference analysis also revealed that the 'Treated' variable was significant. This suggests that reviews on Goodreads are generally less positive than those on Amazon. However, the output showed that the interaction between these two variables does not significantly affect review sentiment. In other words, it indicated that there is no differential impact of the presence of a Q&A section on Goodreads. Therefore, no conclusions could be made about this relationship and no support can be given to hypothesis H1. Also, the interaction term 'Post x Treated x Number of Answers' was insignificant, and thus nothing can be concluded about hypothesis H2 either. It might be the case that this moderator has an inverted U-shape effect as found by Lee and Lee (2004) and Sicilia and Ruiz (2010). Since the difference-in-difference analysis is a linear regression, it is unable to determine an inverted U-shape, which could explain its insignificance. On the other hand, the interaction of the moderator 'Likes' with 'Post' and 'Treated' showed a significant and positive effect. This positive coefficient suggests that the sentiment of a review posted on Goodreads after a featured question was posted tends to increase with a score of 0.005 for each additional like the featured question gets. This is consistent with hypothesis H3. Below, an overview of the hypotheses can be found.

	Hypothesis	**Result**
H1	*The presence of a Q&A section has a positive effect on the average sentiment of reviews.*	Not supported
H2	*The effect of the presence of a Q&A section on the average sentiment of reviews increases (decreases) as the number of answers on the featured question goes up (down).*	Not supported
H3	*The effect of the presence of a Q&A section on the average sentiment of reviews increases (decreases) as the number of likes on the featured question goes up (down).*	Supported

Table 7: Overview of hypotheses and their outcomes

8. Conclusion

This study examined the influence of having a Q&A section present on the average sentiment of reviews on Goodreads, a popular online platform for book enthusiasts to discover and discuss books. The dependent variable was the average sentiment score of reviews, measured by using the AFINN sentiment lexicon. The presence of a Q&A section was investigated as the independent variable and was denoted as the variable 'Post'. This variable indicated whether a review was written after or before the featured question was posted. Next to this main effect, this study investigated the potential moderating effects of the Number of Answers and Likes to the featured question.

Based on relevant academic literature, fixed effect models were constructed that controlled for book- and time-specific fixed effects. While each model showed similar results, Model 2 was selected due to its match with the conceptual framework. The findings of Model 2 indicated that there was a significant influence of the presence of a Q&A section on the review sentiment on Goodreads.com, however, this influence was estimated to be negative. This finding contradicts with the main hypothesis H1. Moreover, no significant influences by the moderating effects were found. To gain further knowledge on the impact of the Q&A section, a difference-in-difference analysis was conducted to see measure whether the presence of a Q&A section influenced the average review sentiment or whether other factors played a role. Here, Goodreads was considered as the treatment group while reviews on Amazon were part of the control group. The analysis revealed that the average sentiment decreases over time and that the reviews on Amazon were generally more positive compared to Goodreads. However, the analysis was not able to identify an effect of having a Q&A section present on Goodreads. Nonetheless, it implied that the sentiment of a Goodreads review increased with each additional like on the featured question when it is written after when compared to on Amazon.

The results of this study have implications for both managers and academics. From a managerial perspective, understanding the role of a Q&A section on the sentiment of book reviews is important for integrating user-generated content on community-driven platforms such as Goodreads. The results from the fixed effects regression suggested that there is a negative influence of introducing a Q&A section on the average sentiment of a review while the difference-in-difference analysis did not find any significant influence. Therefore, it is likely that any attempts by publishers and authors through Q&A interactions have no effect or a negative effect. The difference-in-difference analysis also found that reviews were generally more positive on Amazon. Goodreads managers should investigate the potential reasons that contribute to the lower average sentiment and explore strategies to increase it. A possible explanation is that the audience of both platforms is different, and that those different audiences review their reading experience with different criteria. It is crucial to note that three of the four fixed effects model assumptions were violated. Hence, the results should be treated as indicative. The results also contribute on an academic level. It fills an academic gap by focusing on the role of Q&A sections, which was researched less extensively compared to other forms of user-generated content. Also, the findings contradict with the hypotheses based on existing literature. Next, it is also relevant to the academic field since it investigates the impact of a Q&A function on a community-driven platform where primarily users pose questions and answers. This is unlike the typical focus of managerial responses to user questions in online shopping environments. As a final contribution to existing literature, a difference-in-difference analysis was conducted using Goodreads as the treatment group and Amazon as the control group.

Although various relevant insights were provided, it is important to address several limitations. Firstly, the data only included questions that were listed as the 'featured question', the question with the highest interactions by Goodreads users. Therefore, there is a possibility that there are older questions that were not taken into consideration and that reviews might be categorized as 'Before' while in reality they were written after the first question was posed. It is important to note that only 24.58% of the 1.1 million Goodreads reviews were classified as 'after'. For the Amazon reviews used in the difference-in-difference analysis, this proportion was 33.88%. However, it is likely that these percentages are higher in reality. Secondly, this study used the AFINN sentiment lexicon for the measurement of review sentiment scores. The AFINN lexicon is a so-called 'unigram' lexicon, which means it measures words independently rather than in a combination with the words before or after. An example is the review "*This was the tenth time that I have read this book, and every time i go through it I discover other pieces of it that I enjoy. Great book and still kind of sad that the series had to finish.*". Although this review is fairly positive, when measuring the words independently, the sum of AFINN scores equaled 0, indicating a neutral sentiment. Given the size of the Goodreads reviews and Amazon review datasets, it is not feasible to manually check every review. A third limitation is the use of a subset of Amazon reviews due to technical constraints. 50% of all books with questions were selected, and for each of these, all their Amazon reviews were included. While this approach ensured complete data per selected book, it is possible that relevant information was omitted.

Based on these limitations, suggestions can be made for future research. First, the data only contained the featured question per book that had the most interaction. The independent variable, the presence of Q&A, was represented by a dummy variable that indicated whether a review was written before or after the featured question was posted. To improve the validity of the findings, it is suggested to replicate the current study with the actual first question instead.

Next, future research should explore alternative sentiment lexicons. The current study used the AFINN lexicon, which measures sentiment based on unigrams. However, it would be interesting to see if the outcomes change when a lexicon is used that captures the contextual sentiment. Thirdly, it is suggested that the research is replicated by using all relevant Amazon reviews rather than the subset based on 50% of the books. In addition to addressing these limitations, other suggestions can be made. The current study investigated the role of the Q&A section, a form of user-generated content, on the book platform, Goodreads. The literature on community-driven Q&A is limited and thus it could be beneficial to investigate beyond the book domain. For instance, one could compare the influences of Q&A sections on platforms across different product types, such as hedonic vs utilitarian products, to see how product characteristics influence the role of Q&A on review sentiment. Furthermore, future studies could investigate the moderating effects of the demographics of users. It might be the case that there is a different audience on Goodreads wherefore the average sentiment is lower than on Amazon as the difference-in-difference analysis showed. In addition, it is worth noting that the Goodreads data included reviews that were written before the launch of Goodreads in January 2007. Therefore, managers should pay more attention to the data cleaning and preparation processes to exclude reviews with dates that predate the Goodreads launch. To continue on this, Figure 7 also shows very fluctuating averages around the launches of Goodreads and Amazon. Thus, for future research, it is advised to use reviews from the periods shortly after launch. Finally, three out of the four assumptions for fixed effect regressions were violated. In addition, it is possible that H2 is not supported because the number of answers leads to an inverted U-shaped relationship rather than a linear relationship. Therefore, it is suggested to explore alternative analysis methods.

9. Works Cited

Ahn, D.-Y., Duan, J. A., & Mela, C. F. (2016). Managing User-Generated Content: A Dynamic Rational Expectations Equilibrium Approach. *Marketing Science, 35*(2), 284-303.

Albuquerque, P., Pavlidis, P., Chatow, U., Chen, K.-Y., Jamal, Z., Koh, K.-W., & Fitzhugh, A. (2012). Evaluating Promotional Activities in an Online Two-Sided Market of User-Generated Content. *Marketing Science, 31*(3), 406-432.

Anderson, E. W., Fornell, C., & Lehmann, D. R. (1994, July). Customer Satisfaction, Market Share, and Profitability: Findings from Sweden. *Journal of Marketing, 58*(3), 53-66.

Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of Online Consumer Reviews: Readers' Objectives and Review Cues. *International Journal of Electronic Commerce, 17*(2), 99-126.

Banerjee, S., Dellarocas, C., & Zervas, G. (2021). Interacting User-Generated Content Technologies: How Questions and Answers Affect Consumer Reviews. *Journal of Marketing Research, 58*(4), 742-761.

Basuroy, S., Chatterjee, S., & Ravid, S. A. (2003). How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets. *Journal of Marketing, 67*, 103-117.

Bessa, A. (2022, March 17). *Lexicon-Based Sentiment Analysis: A Tutorial*. Retrieved from KNIME: https://www.knime.com/blog/lexicon-based-sentiment-analysis

Bondi, T. (2019). *Alone, Together: Product Discovery through Consumer Ratings*. Retrieved from NET Institute Working Paper No. 19-09: https://ssrn.com/abstract=3468433

Branco, F., Sun, M., & Villas-Boas, J. M. (2016). Too Much Information? Information Provision and Search Costs. *Marketing Science, 35*(4), 605-618.

Castelli, M., Manzoni, L., Vanneschi, L., & Popovič, A. (2017, October 30). An expert system for extracting knowledge from customers’ reviews: The case of Amazon.com, Inc. *Expert Systems with Applications, 84*, 117-126.

Census Bureau. (2022, January 13). *Annual Retail Trade Survey: 2020.* Retrieved from Census Bureau: https://www.census.gov/data/tables/2020/econ/arts/annual-report.html

Cheng, Y.-H., & Ho, H.-Y. (2015, April). Social influence's impact on reader perceptions of online reviews. *Journal of Business Research, 68*(4), 883-887.

Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research, 43*(3), 345-354.

Chevalier, J. A., Dover, Y., & Mayzlin, D. (2018). Channels of Impact: User Reviews When Quality Is Dynamic and Managers Respond. *Marketing Science, 37*(5), 688-709.

Colicev, A., Kumar, A., & O'Connor, P. (2019). Modeling the relationship between firm and user generated content and the stages of the marketing funnel. *International Journal of Research in Marketing, 36*(1), 100-116.

Cunningham, S. (2021). *Causal Inference: The Mixtape.* Yale University Press.

Dhanasobhon, S., Chen, P.-Y., & Smith, M. (2007). An Analysis of the Differential Impact of Reviews and Reviewers at Amazon.com. *ICIS 2007 Proceedings*.

Dhar, R., & Wertenbroch, K. (2000). Consumer Choice Between Hedonic and Utilitarian Goods. *Journal of Marketing Research, 37*(1), 60-71.

Dimoka, A., Hong, Y., & Pavlou, P. A. (2012). On Product Uncertainty in Online Markets: Theory and Evidence. *MIS Quarterly, 36*(2), 395-426.

Esmark Jones, C. L., Stevens, J. L., Breazeale, M., & Spaid, B. I. (2018). Tell it like it is: The effects of differing responses to negative online reviews. *Psychology & Marketing, 35*(12), 891-901.

Fainmesser, I. P., Lauga, D. O., & Ofek, E. (2021, November). Ratings, Reviews, and the Marketing of New Products. *Management Science, 67*(11), 7023-7045.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics, 225*(2), 254-277.

Goodreads. (n.d.). *A Game Of Thrones - Reader Q&A*. Retrieved from Goodreads: https://www.goodreads.com/book/13496/questions

Goodreads. (n.d.). *About Goodreads*. Retrieved from Goodreads: https://www.goodreads.com/about/us

Hanck, C., Arnold, M., Gerber, A., & Schmelzer, M. (2023). *Introduction to Econometrics with R.* Essen, Germany.

Hayes, A. (2023, April 5). *T-Test: What It Is With Multiple Formulas and When To Use Them*. Retrieved from Investopedia: https://www.investopedia.com/terms/t/t-test.asp

Horton, J. (2022, October 3). *4 reasons to implement a question and answer platform on your site*. Retrieved from Bazaarvoice: https://www.bazaarvoice.com/blog/boost-sales-with-a-question-and-answer-platform/

Huang, M., Zhu, H., & Zhou, X. (2013). The effects of information provision and interactivity on e-tailer websites. *Online Information Review, 37*(6), 927-945.

Ju Choi, C., & Kim, J.-B. (1996). Reputation, learning and quality uncertainty. *Journal of Consumer Marketing, 13*(5), 47-55.

Khern-am-nuai, W., Ghasemkhani, H., Qiao, D., & Kannan, K. (2022, August 28). *The Impact of Online Q&As on Product Sales: The Case of Amazon Answer.* Retrieved from https://ssrn.com/abstract=2794149

Kim, Y., & Krishnan, R. (2015). On Product-Level Uncertainty and Online Purchase Behavior: An Empirical Analysis. *Management Science, 61*(10), 2449-2467.

Lackermair, G., Kailer, D., & Kanmaz, K. (2013). Importance of Online Product Reviews from a Consumer's Perspective. *Advances in Economics and Business, 1*(1), 1-5.

Lappas, T., Dellarocas, C. N., & Derakhshani, N. (2016, February 16). *Reputation and Contribution in Online Question-Answering Communities.* Retrieved from SSRN: https://dx.doi.org/10.2139/ssrn.2918913

Lee, B.-K., & Lee, W.-N. (2004). The Effect of Information Overload on Consumer Choice Quality in an On-Line Environment. *Psychology & Marketing, 21*(3), 159-183.

Lee, H., Han, J., & Suh, Y. (2014, May-June). Gift or threat? An examination of voice of the customer: The case of MyStarbucksIdea.com. *Electronic Commerce Research and Applications, 13*(3), 205-219.

Lu, Q., Ye, Q., & Law, R. (2014). Moderating effects of product heterogeneity between online word-of-mouth and hotel sales. *Journal of Electronic Commerce Research, 15*(1), 1-12.

Mahajan, V., Sharma, S., & Wind, Y. (1984). Parameter Estimation in Marketing Models in the Presence of Influential Response Data: Robust Regression and Applications. *Journal of Marketing Research, 21*(3), 268–277.

Mahfooz, H., & Nadeem, S. (2019). Game of Eyeballs: What Should Be Above and Below the Fold of an E-Commerce Website: A Biometric Study. *International Journal of Experiential Learning & Case Studies, 4*(2), 193-206.

Marcus, M., & Sant'Anna, P. H. (2021, March). The Role of Parallel Trends in Event Study Settings: An Application to Environmental Economics. *Journal of the Association of Environmental and Resource Economists, 8*(2).

Mousavi, R., Hazarika, B., Chen, K., & Razi, M. (2021). The Effect of Online Q&As and Product Reviews on Product Performance Metrics: Amazon.com as a Case Study. *Journal of Information & Knowledge Management, 20*(1).

Naylor, R. W., Lamberton, C. P., & Norton, D. A. (2011). Seeing Ourselves in Others: Reviewer Ambiguity, Egocentric Anchoring, and Persuasion. *Journal of Marketing Research, 48*(3), 617-631.

Osgood, C. E., & Tannenbaum, P. H. (1955). The principle of congruity in the prediction of attitude change. *Psychological Review, 62*(1), 42-55.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The Measurement of Meaning.* University of Illinois Press.

Park, D.-H., Lee, J., & Han, I. (2014, December 8). The Effect of On-Line Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement. *International Journal of Electronic Commerce*, 125-148.

Proserpio, D., & Zervas, G. (2016, November 7). Online Reputation Management: Estimating the Impact of Management Responses on Consumer Reviews. *Forthcoming, Marketing Science*.

Ramirez, D. (2020, March 26). *User-generated content vs eWord-of-Mouth (UGC vs eWOM)*. Retrieved from TINT: https://www.tintup.com/blog/user-generated-content-vs-eword-of-mouth-ugc-vs-ewom/

Rath, M., & Shah, C. (2016). Deconstructing the failure: Analyzing the unanswered questions within educational Q&A. *Computer Science, 53*(1), 1-6.

Ravi, S., Pang, B., Rastogi, V., & Kumar, R. (2014). Great Question! Question Quality in Community Q&A. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*, pp. 426-435.

Reich, T., & Maglio, S. J. (2020). Featuring Mistakes: The Persuasive Impact of Purchase Mistakes in Online Reviews. *Journal of Marketing*, *84*(1), 1-145.

Rocklage, M. D., He, S., Rucker, D. D., & Nordgren, L. F. (2023). Beyond Sentiment: The Value and Measurement of Consumer Certainty in Language. *Journal of Marketing Research*.

Rogers, K. (n.d.). *What is Amazon ASIN number & how to get it?* . Retrieved from DataFeedWatch: https://www.datafeedwatch.com/blog/amazon-asin-number-what-is-it-and-how-do-you-get-it

Scholz, M., Dorner, V., Landherr, A., & Probst, F. (2013). Awareness, Interest, and Purchase: The Effects of User- and Marketer-Generated Content on Purchase Decision Processes. *International Conference on Information Systems.*

Scuderi, R. (n.d.). *10 Best Book Recommendation Sites You Need To Know*. Retrieved from Lifehack: https://www.lifehack.org/articles/technology/10-best-book-recommendation-sites-you-need-know.html

Sicilia, M., & Ruiz, S. (2010). The effects of the amount of information on cognitive responses in online purchasing tasks. *Electronic Commerce Research and Applications*, *9*, 183-191.

Silge, J., & Robinson, D. (2016). tidytext: Text Mining and Analysis Using Tidy Data Principles in R. *Journal of Open Source Software*.

Sipos, R., Ghosh, A., & Joachims, T. (2014). Was This Review Helpful to You? It Depends! Context and Voting Patterns in Online Content. *Proceedings of the 23rd International Conference on World Wide Web*, (pp. 337-348).

Statistics Solutions. (n.d.). *Mann-Whitney U Test*. Retrieved from Statistics Solutions: https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/mann-whitney-u-test/

Sun, C., Fang, Y., Kong, M., Chen, X., & Liu, Y. (2022). Influence of augmented reality product display on consumers' product attitudes: A product uncertainty reduction perspective. *Journal of Retailing and Consumer Services, 64*.

Wang, Y., Ramachandran, V., & Sheng, O. R. (2021). Do Fit Opinions Matter? The Impact of Fit Context on Online Product Returns. *Information Systems Research, 32*(1), 268-289.

Weathers, D., Sharma, S., & Wood, S. L. (2007). Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods. *Journal of Retailing*, *83*(4), 393-401.

Wu, J., Fan, S., & Zhao, J. L. (2018, March). Community engagement and online word of mouth: An empirical investigation. *Information & Management*, *55*(2), 258-270.

Yu, Y., Qiu, L., Chen, H., & Yen, B. (2023). Movie fit uncertainty and interplay between traditional advertising and social media marketing. *Marketing Letters*.

Zhu, F., & Zhang, X. (. (2010). Impact of Online Consumer Reviews on Sales:The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing, 74*(2), 133-148.

10. Appendices

Appendix A: GitHub Link

https://github.com/jeroenmaagdenberg/Goodreads-QandA-ReviewSentiment

Appendix B: Data exploration

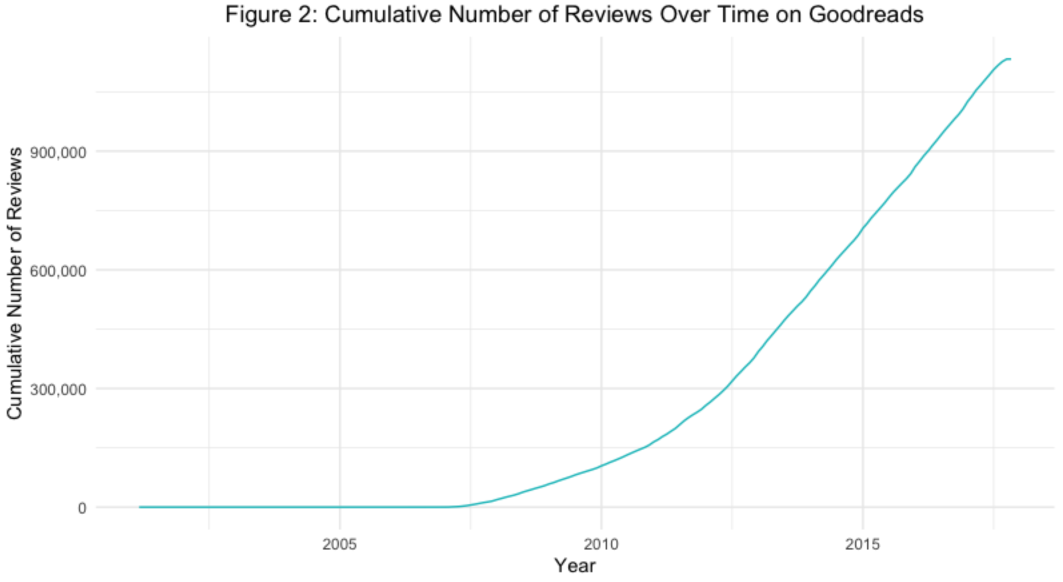
Appendix B.1: Top 5 most frequently used words in reviews on Goodreads.com

	word	n
1.	book	1,394,726
2.	read	694,499
3.	story	579,341
4.	love	330,293
5.	characters	287,221

Appendix B.2: Example of cleaned reviews with accompanying AFINN scores

Title	Review Text	AFINN Score
Ghost Summer	a favorite new to me writer This selection of short stories is an excellent introduction to her work and all the stories make me yearn for more from her characters and her world	5
Golden	Compelling story complete with mystery The choices we make	0
Hexwood	I have to read this one twice backtoback because the ending is absolutely stunning Truly didnt see that twist coming Its a scifi after all	4
Ready Player One	stars	0
Songs of a Dead Dreamer and Grimscribe	Ligotti writes stories you have to read one at a time then perhaps reread to really understand and feel all the nuances twists and turns One of my new favorites	2
The 7 Habits of Highly Effective People: Powerful Lessons in Personal Change	One of the most important books Ive ever read	0
Flora & Ulysses: The Illuminated Adventures	DONE And I absolutely hated it The plot was boring the characters uninterested and stupid and the pictures broke my flow of reading They gave spoilers to what was on the next page I hated it	-12
The Island of Dr. Libris	Overall decent book fast starting but a Disappointing ending	-2
Stones from the River	Its not bad writing I just dont like the story being told	-3

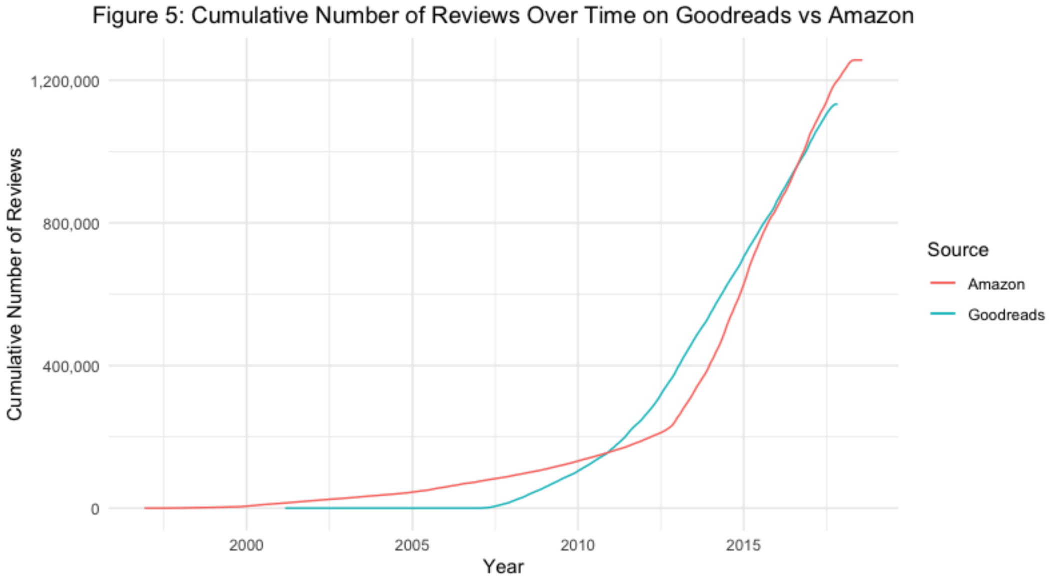
Appendix B.3: Cumulative Number of Reviews Over Time on Goodreads



Appendix B.4: Top 5 most frequently used words in reviews on Amazon.com

	word	n
1.	book	1,448,645
2.	read	674,927
3.	story	438,633
4.	love	242,544
5.	life	234,389

Appendix B.5: Cumulative Number of Reviews Over Time on Goodreads vs Amazon



Appendix C: Excerpt Welch's Two Sample t-test

For each review is known whether it has been posted before or after the featured question. As seen in Table 3, the sample sizes and variances are not equal. Taking this into consideration, the Welch's Two Sample t-test is used as it adjusts for these inequalities (Hayes, 2023). The T-statistic is 10.041 which implies that the mean sentiment score of the non-treated group is higher than the treated. In addition, the p-value of this test close to 0. Therefore, it is assumed that there is a significant difference between the sentiment scores of the groups. In order to confirm this assumption, the outcome is tested on normality. Both the Shapiro-Wilk and the Anderson-Darling normality tests provide a p-value close to 0 and thus the assumption of normality is violated. In the case that an assumption of a two-sample t-test is not met, a non-parametric test can be used. One of such tests is the Wilcoxon rank sum test (Statistics Solutions, n.d.). The test resulted in an estimate and a p-value that are both close to 0. This means that there is a significant difference between the treated and non-treated groups, but the effect is only minimal.

Appendix D: Results models

Appendix D.1: Output Breusch-Pagan test

```
	studentized Breusch-Pagan test

data:  AFINN_score ~ post + Likes + Number_of_Answers
BP = 228.75, df = 3, p-value < 2.2e-16
```

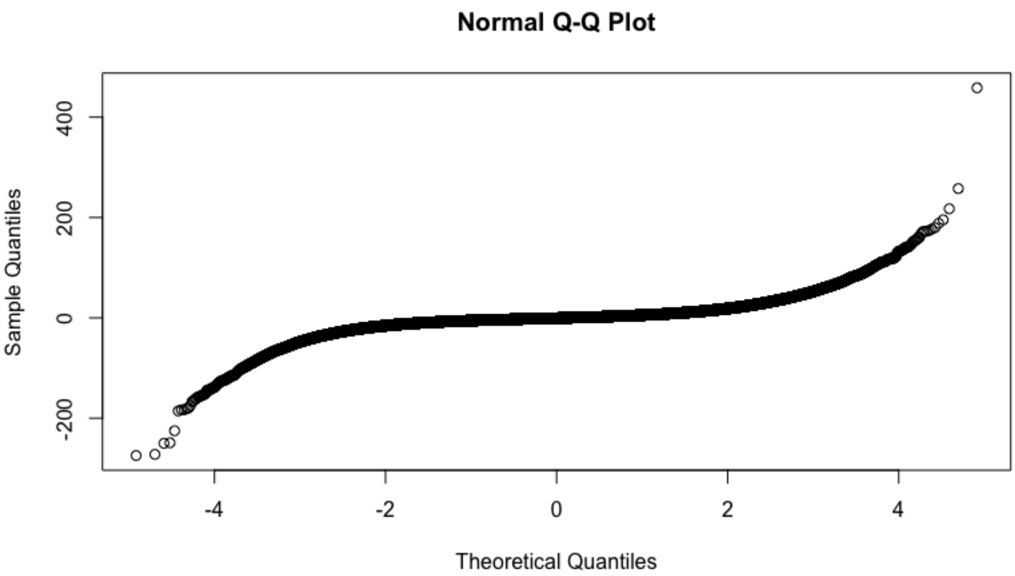
Appendix D.2: Output Breusch-Godfrey test

```
	Breusch-Godfrey test for serial correlation of order up to 3

data:  AFINN_score ~ post + Likes + Number_of_Answers
LM test = 42686, df = 3, p-value < 2.2e-16
```

Appendix D.3: Output Jarque Bera test

```
	Jarque Bera Test

data:  resid(model3)
X-squared = 60517227, df = 2, p-value < 2.2e-16
```

Appendix D.4: Normal Q-Q Plot



Appendix D.5: Correlation matrix for multicollinearity

	Post	**Likes**	**Number of Answers**
Post	1.00000000	0.08325816	0.04573200
Likes	0.08325816	1.00000000	0.62561763
Number of Answers	0.04573200	0.62561763	1.00000000

Appendix D.6: Output Difference-in-Difference with four-way interaction

```
OLS estimation, Dep. Var.: AFINN_score
Observations: 1,816,642
Fixed-effects: Book_Id: 6,860,  Year_Month: 259
Standard-errors: Clustered (Book_Id)
                                      Estimate Std. Error    t value   Pr(>|t|)
post                                 -0.416743   0.070204  -5.936175 3.0598e-09 ***
treated                              -0.542743   0.045391 -11.957091  < 2.2e-16 ***
post:treated                         -0.040779   0.094912  -0.429656 6.6746e-01
post:treated:Likes                    0.006425   0.006068   1.058938 2.8967e-01
post:treated:Number_of_Answers        0.001062   0.004238   0.250499 8.0221e-01
post:treated:Likes:Number_of_Answers -0.000024   0.000084  -0.289901 7.7190e-01
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 7.4643     Adj. R2: 0.129163
               Within R2: 8.379e-4
```