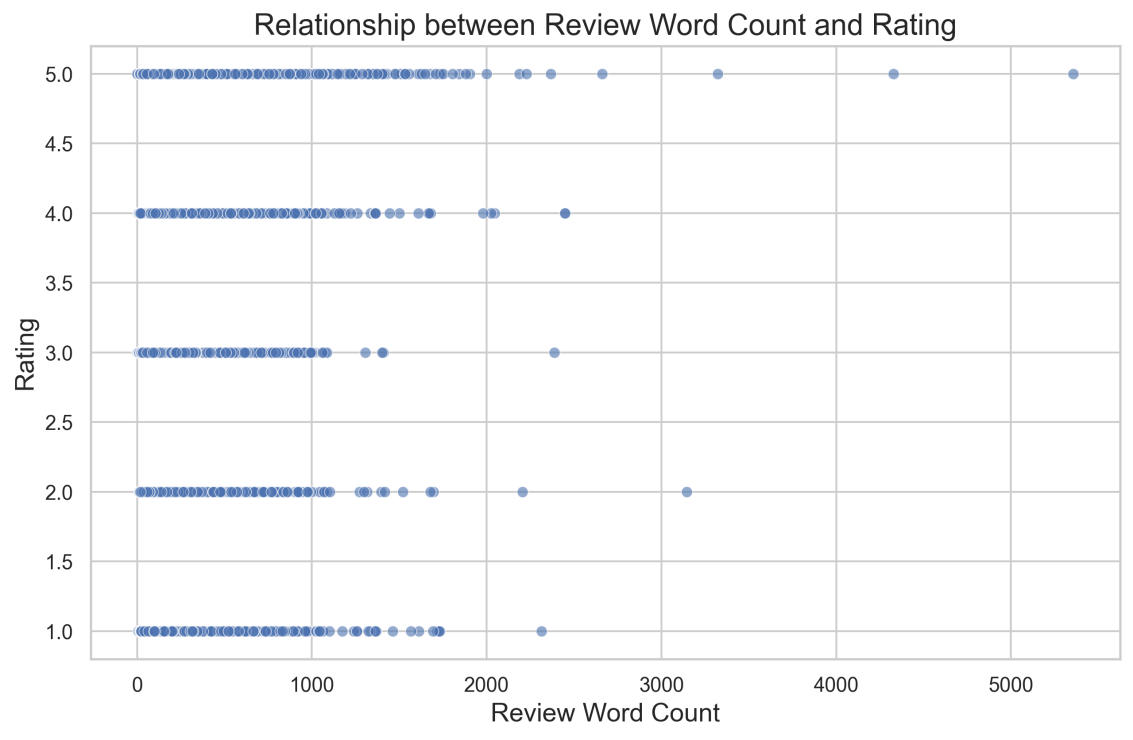
****Amazon Book Reviews****

****Process of Data Crawling and Cleaning****

**Our goal is to gather user review information data on a wide range of books from Amazon, collecting multiple user reviews for each book to make a multi-dimensional analysis. We consider these following factors. The first one is sentiment analysis, we identify whether reviews are positive, negative, or neutral. We can also track rating trends to see how books' popularity changes over time. Besides, reviews offer detailed insights into the strengths and weaknesses of a book, which can help authors and publishers improve their work. Lastly, reviews play a crucial role in determining the reputation and credibility of a book, which helps potential buyers make informed decisions. We collect essential information, including book ID, book title, along with review details like reviewer name, rating, review time, the full review text and so on. This data is stored in a CSV file, which can easily be imported into tools like Python or Excel for further analysis. Once the data is collected, we clean it by removing unnecessary columns, handling missing values, and ensuring data types are consistent. For example, we convert Unix timestamps into readable dates. The cleaned data is saved in a new CSV file, ready for deeper analysis and insights into reader sentiment and trends.**

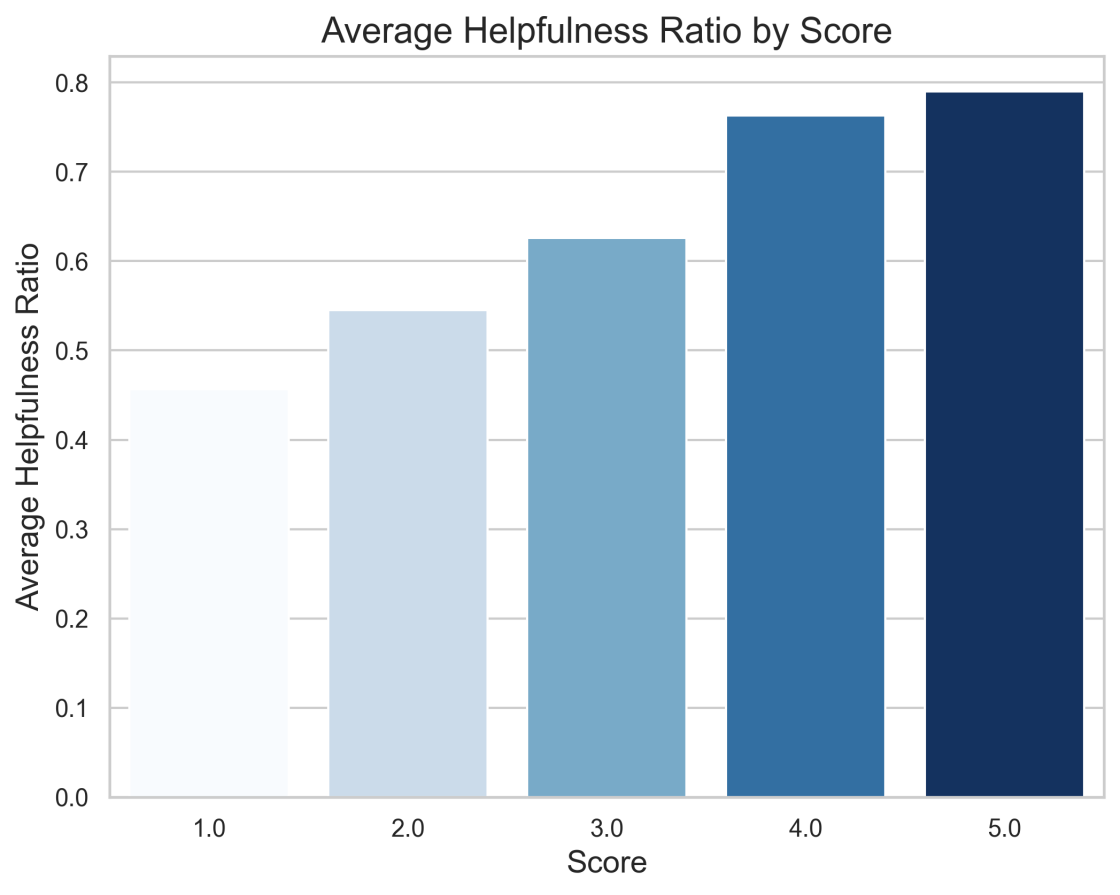
****Data Analyze****

**In all, we collected about 1500 books and related 160000 reviews or so, containing information book title, username, review score, review text and so on. Here are some examples of the analysis we made.**

****

The scatter plot visualizes the relationship between review word count and book ratings. The x-axis represents the number of words in each review, while the y-axis shows ratings from 1 to 5. Points are distributed across the chart, with most reviews being under 1000 words, though a few reach up to 5000. This shows that reviewers share their opinions in different ways, using various review lengths to explain their thoughts and give clear ratings to the books they read.

The data suggests a potential relationship between review length and ratings. Reviews with ratings of 1, 4, and 5 tend to have higher word counts compared to those rated 2 and 3. Notably, only reviews with a rating of 5 exceed 4000 words. This implies that extremely detailed reviews are more likely associated with high satisfaction. To better understand the relationship, additional qualitative analysis such as sentiment evaluation might provide deeper insights into how reviews reflect user opinions.



This bar chart illustrates the average helpfulness ratio of book reviews across different star ratings, ranging from 1 star to 5 stars. The helpfulness ratio measures the proportion of helpful votes to total votes for each review. To ensure data reliability, only reviews with at least 10 total votes were included in the analysis. This visualization highlights the relationship between user star ratings and how helpful their reviews are perceived by other users, which shows differences in perceived helpfulness across positive and negative reviews.

We can see from this figure that reviews with higher star ratings like 4 or 5 stars consistently show a higher average helpfulness ratio. This indicates that users generally find positive reviews more helpful or reliable. In contrast, lower-rated reviews like 1 or 2 stars tend to have lower helpfulness ratios, reflecting that negative feedback may be viewed as less constructive or subjective. These insights emphasize how user perception of review value is strongly tied to rating positivity and sentiment.