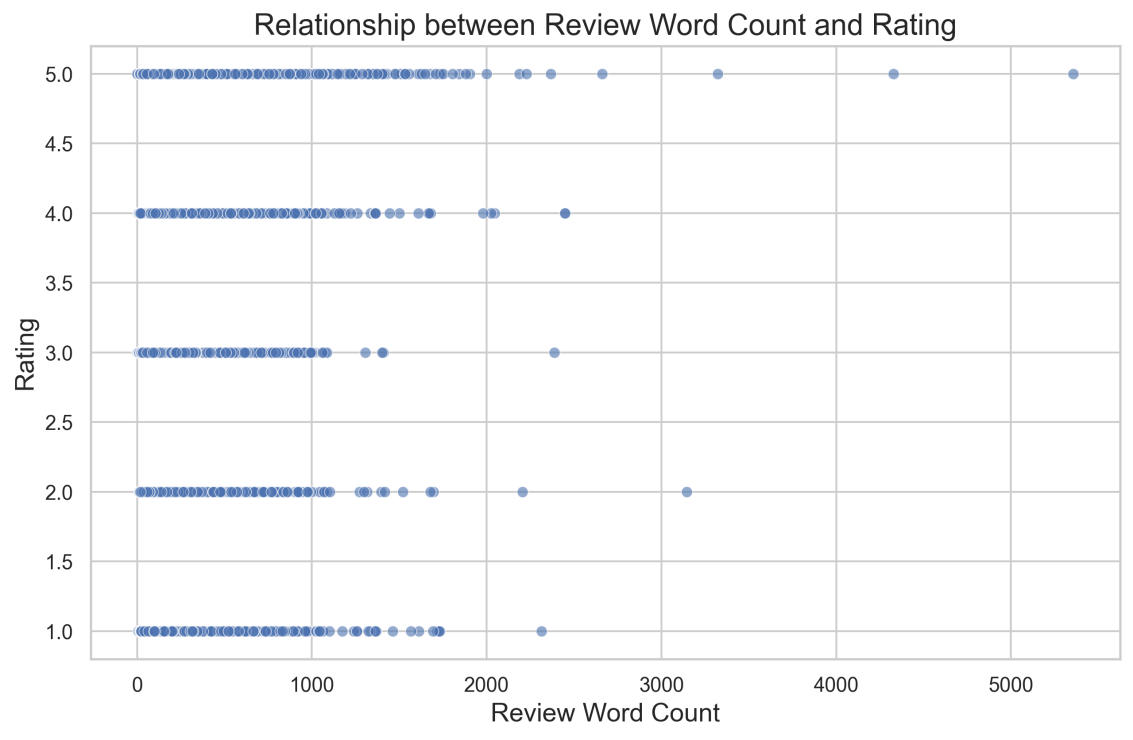
****Amazon Book Reviews****

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

**Our goal is to gather user review information data on a wide range of books from Amazon. We collect essential information, including book ID, book title, along with review details like reviewer name, rating, the full review text and so on. Once the data is collected, we clean it by removing unnecessary columns, handling missing values, and ensuring data types are consistent.**

****Data Analyze****

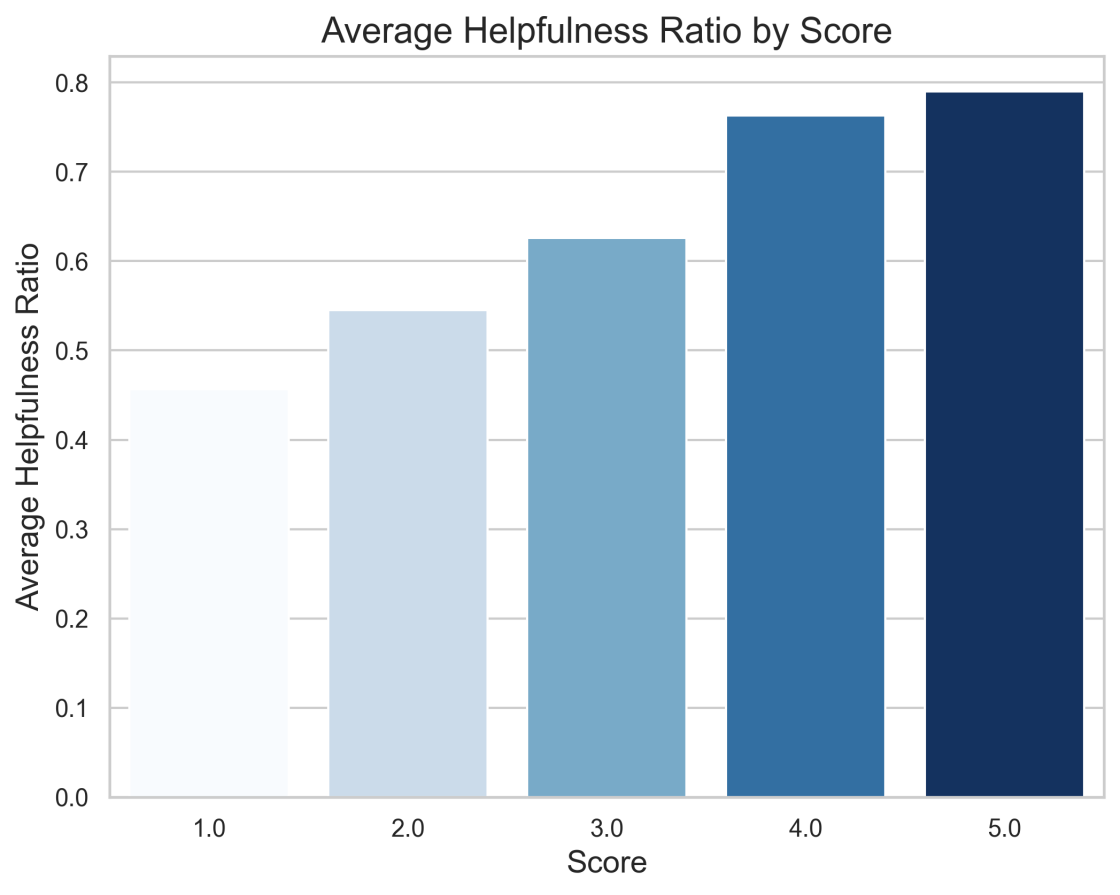
**In all, we collected about 1500 books and related 160000 reviews or so. Here are some examples of the analysis we made.**

****

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 generally have higher word counts. Only reviews with a rating of 5 exceed 4000 words, indicating that very detailed reviews are more common among highly satisfied users.

This pattern implies that longer reviews might reflect stronger satisfaction or dissatisfaction. Extremely detailed reviews are often associated with high satisfaction, as seen in the case of 5 star ratings. To understand more, analyzing the feelings or details in the reviews could show how review length matches the user's 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, such as 4 or 5 stars, consistently show a higher average helpfulness ratio. In contrast, lower-rated reviews, like those with 1 or 2 stars, tend to have lower helpfulness ratios.

This may suggest that users generally find positive reviews more helpful or reliable, while negative feedback may be seen as less constructive or more subjective. It highlights that the perceived value of a review is closely tied to its positivity and sentiment.