Decision Making with Sentiment Analysis in R

Sanika Parakatil Joseph
Fresenius University of Applied Science

Data Analysis for Decision-Making (WS 2025/26)

Prof. Dr. Stephan Huber

2025-12-10

Author Note

Correspondence concerning this article should be addressed to Sanika Parakatil Joseph, Email: parakatil_joseph.sanika@stud.hs-fresenius.de

Abstract

Sentiment analysis transforms unstructured text into measurable indicators of positive or negative emotion, enabling organizations to understand how stakeholders feel about products, services, or events. Using R and the tidytext framework, sentiment scoring can be performed efficiently through tokenization, lexicon matching, and visualization. This handout demonstrates how sentiment analysis supports evidence-based decision-making by revealing patterns in customer opinions and identifying early warning signals in user feedback. A practical workflow and reproducible R code illustrate how emotional data can guide managerial actions and strategic choices. The included exercise reinforces the techniques discussed.

Decision Making with Sentiment Analysis in R

1 Introduction

Sentiment analysis is a text-mining method that identifies and quantifies the emotional tone expressed in written language, allowing unstructured text to be transformed into measurable indicators of positive or negative sentiment. This approach is valuable for decision-making because it provides managers with insight into customer perceptions, emerging issues, and reputation trends that traditional numerical data may not capture. Negative sentiment can signal dissatisfaction or operational risks, while positive sentiment may reveal strengths and opportunities for improvement or marketing.

By reducing uncertainty and offering real-time feedback, sentiment analysis supports evidence-based decisions across product development, customer service, and strategic planning. This handout introduces the fundamentals of sentiment analysis in R using the tidytext framework, outlines the workflow for processing and scoring text data, and presents an applied example with accompanying R code. A short exercise is included to reinforce the techniques and demonstrate how sentiment insights can inform managerial action.(Hu & Liu, 2004).

2 Methods in R

This section describes the procedure used to perform sentiment analysis in R using a small sample of Amazon Fine Food–style product reviews. The workflow follows a lexicon-based approach implemented with the tidytext framework and relies on the Bing sentiment lexicon to classify words as positive or negative. All steps are implemented in reproducible R code within the Quarto document.(Silge & Robinson, 2017).

2.1 Data Import and Preparation

The dataset `amazon_fine_food_review.csv` contains a subset of product reviews with the variables `Id` (review identifier), `ProductId`, `UserId`, `Score` (star rating from 1 to 5), `Time` (review date), `Summary`, and `Text` (full review). The file is stored in the same directory as the Quarto document and imported using readr. Basic preprocessing converts the date and calculates the length of each review.(McAuley & Leskovec, 2011).

```r
library(tidyverse)
library(tidytext)
library(lubridate)
```

```
library(ggplot2)



reviews <- read_csv("amazon_fine_food_review.csv")


reviews <- reviews %>%
  mutate(
    Time = as.Date(Time),
    review_length = str_count(Text, "\\S+")
  )


head(reviews)
```

```
# A tibble: 6 x 8
     Id ProductId  UserId Score Time       Summary          Text  review_length
  <dbl> <chr>      <chr>  <dbl> <date>     <chr>            <chr>         <int>
1     1 B001E4KFG0 U1         5 0001-01-20 Fantastic cookies Thes~           18
2     2 B001E4KFG0 U2         4 0003-01-20 Good but a bit s~ The ~           16
3     3 B001E4KFG0 U3         2 0005-01-20 Too dry          I fo~            12
4     4 B008J4RP1U U4         1 0007-01-20 Terrible quality The ~            11
5     5 B008J4RP1U U5         3 0010-01-20 Average snack    Noth~            13
6     6 B000HDKODC U6         5 0012-01-20 Excellent coffee Rich~            12
```

2.2 Tokenization and Lexicon Matching

To enable sentiment analysis, the review text is first tokenized into individual words. The `unnest_tokens()` function from tidytext transforms each review into multiple rows with one word per row. These tokens are then matched with the Bing sentiment lexicon, which labels words as positive or negative. This approach allows each word to carry a sentiment value, enabling the calculation of review-level sentiment scores.

```
bing_lexicon <- get_sentiments("bing")


tokens <- reviews %>%
```

```
  unnest_tokens(word, Text)

sentiment_words <- tokens %>%
  inner_join(bing_lexicon, by = "word")

head(sentiment_words)
```

```
# A tibble: 6 x 9
     Id ProductId  UserId Score Time       Summary review_length word  sentiment
  <dbl> <chr>      <chr>  <dbl> <date>     <chr>           <int> <chr> <chr>
1     1 B001E4KFG0 U1         5 0001-01-20 Fantas~            18 amaz~ positive
2     1 B001E4KFG0 U1         5 0001-01-20 Fantas~            18 fresh positive
3     1 B001E4KFG0 U1         5 0001-01-20 Fantas~            18 happy positive
4     2 B001E4KFG0 U2         4 0003-01-20 Good b~            16 good  positive
5     2 B001E4KFG0 U2         4 0003-01-20 Good b~            16 sweet positive
6     3 B001E4KFG0 U3         2 0005-01-20 Too dry            12 disa~ negative
```

2.3 Sentiment Scoring

Sentiment scores are computed at the level of entire reviews. Each positive word is assigned a value of +1 and each negative word a value of −1. These values are summed per review to obtain a single sentiment score. Reviews with missing scores (no matched sentiment words) are assigned a value of zero.

```
# Assign +1 to positive, -1 to negative

review_sentiment <- sentiment_words %>%
mutate(score = if_else(sentiment == "positive", 1, -1)) %>%
group_by(Id) %>%
summarise(sentiment_score = sum(score), .groups = "drop")

review_sentiment
```

```
# A tibble: 12 x 2
```

```
      Id sentiment_score
   <dbl>           <dbl>
 1     1               3
 2     2               2
 3     3              -1
 4     4              -3
 5     5              -1
 6     6               4
 7     7               1
 8     8               0
 9     9               2
10    10               0
11    11              -3
12    12               1
```

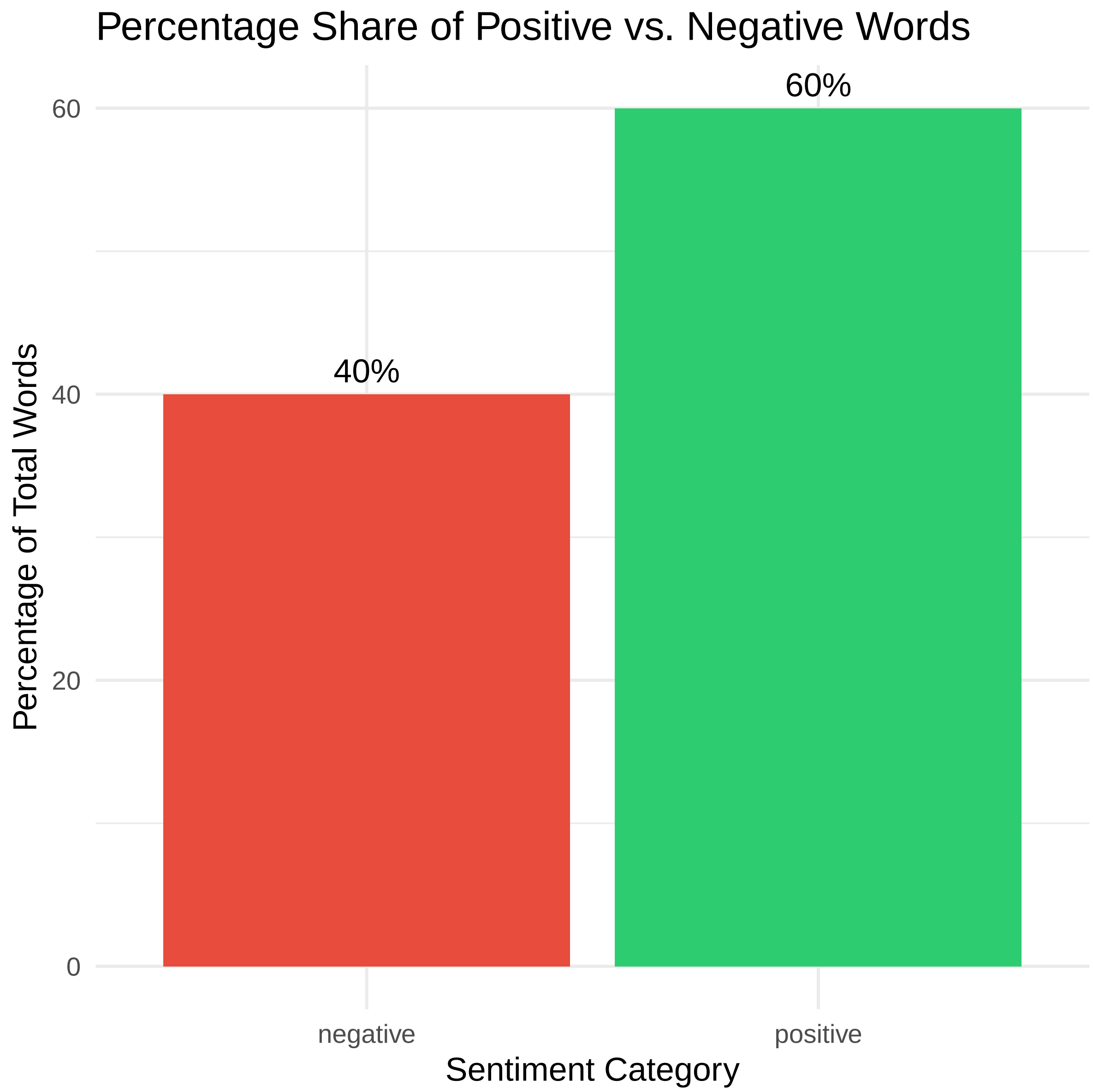
```r
# Merge sentiment scores back to the original reviews

reviews_scored <- reviews %>%
left_join(review_sentiment, by = "Id") %>%
mutate(sentiment_score = replace_na(sentiment_score, 0))

reviews_scored
```

```
# A tibble: 12 x 9
      Id ProductId  UserId Score Time       Summary          Text  review_length
   <dbl> <chr>      <chr>  <dbl> <date>     <chr>            <chr>         <int>
 1     1 B001E4KFG0 U1         5 0001-01-20 Fantastic cooki~ Thes~            18
 2     2 B001E4KFG0 U2         4 0003-01-20 Good but a bit ~ The ~            16
 3     3 B001E4KFG0 U3         2 0005-01-20 Too dry          I fo~            12
 4     4 B008J4RP1U U4         1 0007-01-20 Terrible quality The ~            11
 5     5 B008J4RP1U U5         3 0010-01-20 Average snack    Noth~            13
 6     6 B000HDK0DC U6         5 0012-01-20 Excellent coffee Rich~            12
 7     7 B000HDK0DC U7         4 0014-01-20 Good but expens~ The ~            14
```

```
 8      8 B000HDK0DC U8          2 0016-01-20 Bitter and burnt The ~            13
 9      9 B000SQNQIQ U9          5 0018-01-20 Perfect tea      This~            14
10     10 B000SQNQIQ U10         3 0019-01-20 Decent but weak  The ~            15
11     11 B000SQNQIQ U11         1 0021-01-20 Worst tea ever   This~            19
12     12 B002HQ0Z1U U12         4 0023-01-20 Tasty granola    The ~            15
# i 1 more variable: sentiment_score <dbl>
```

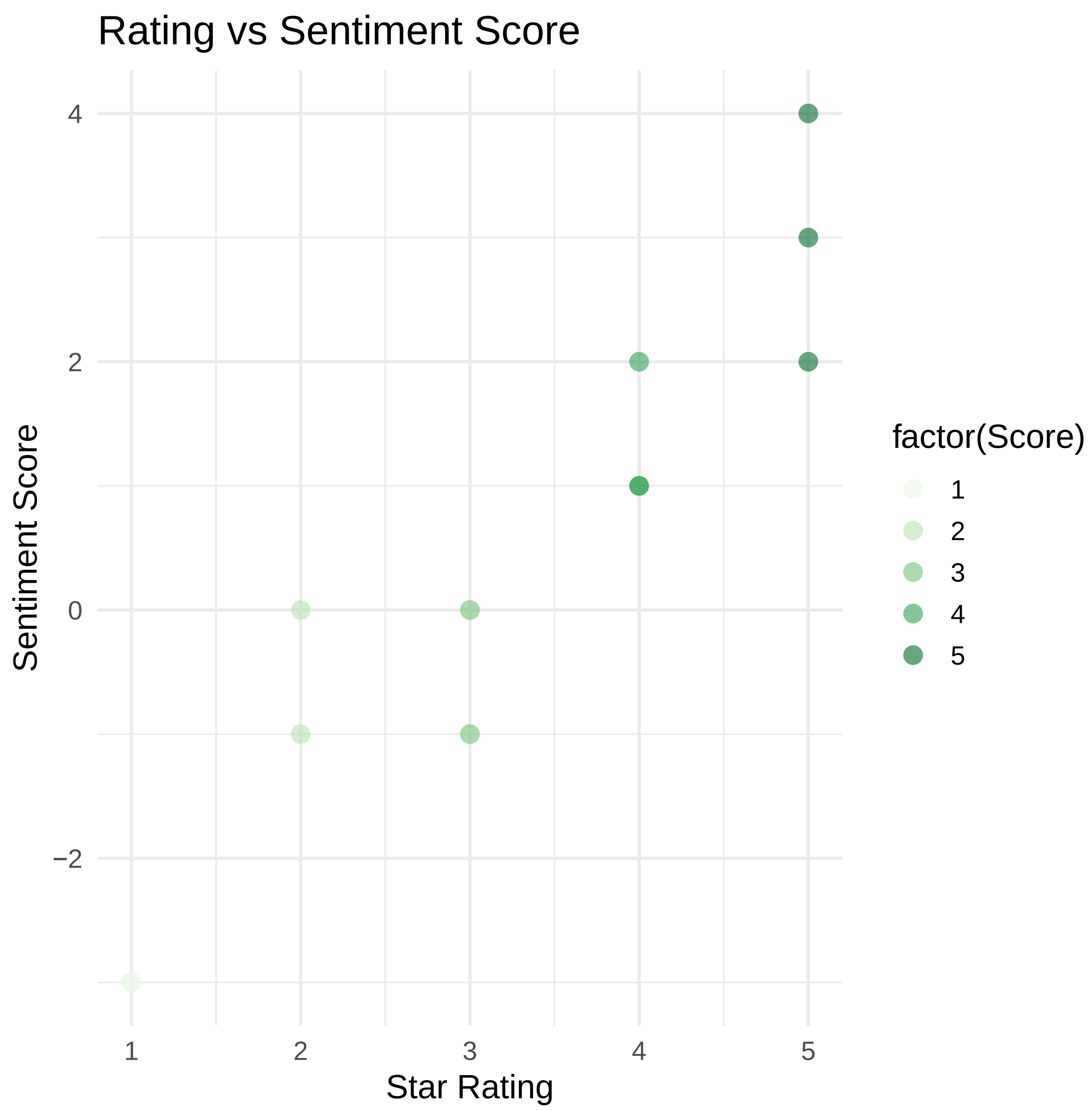
2.4 Visualisation

Several plots are generated to summarise the sentiment patterns and illustrate how they can support decision-making. First, a bar chart shows the number of positive and negative words across all reviews.

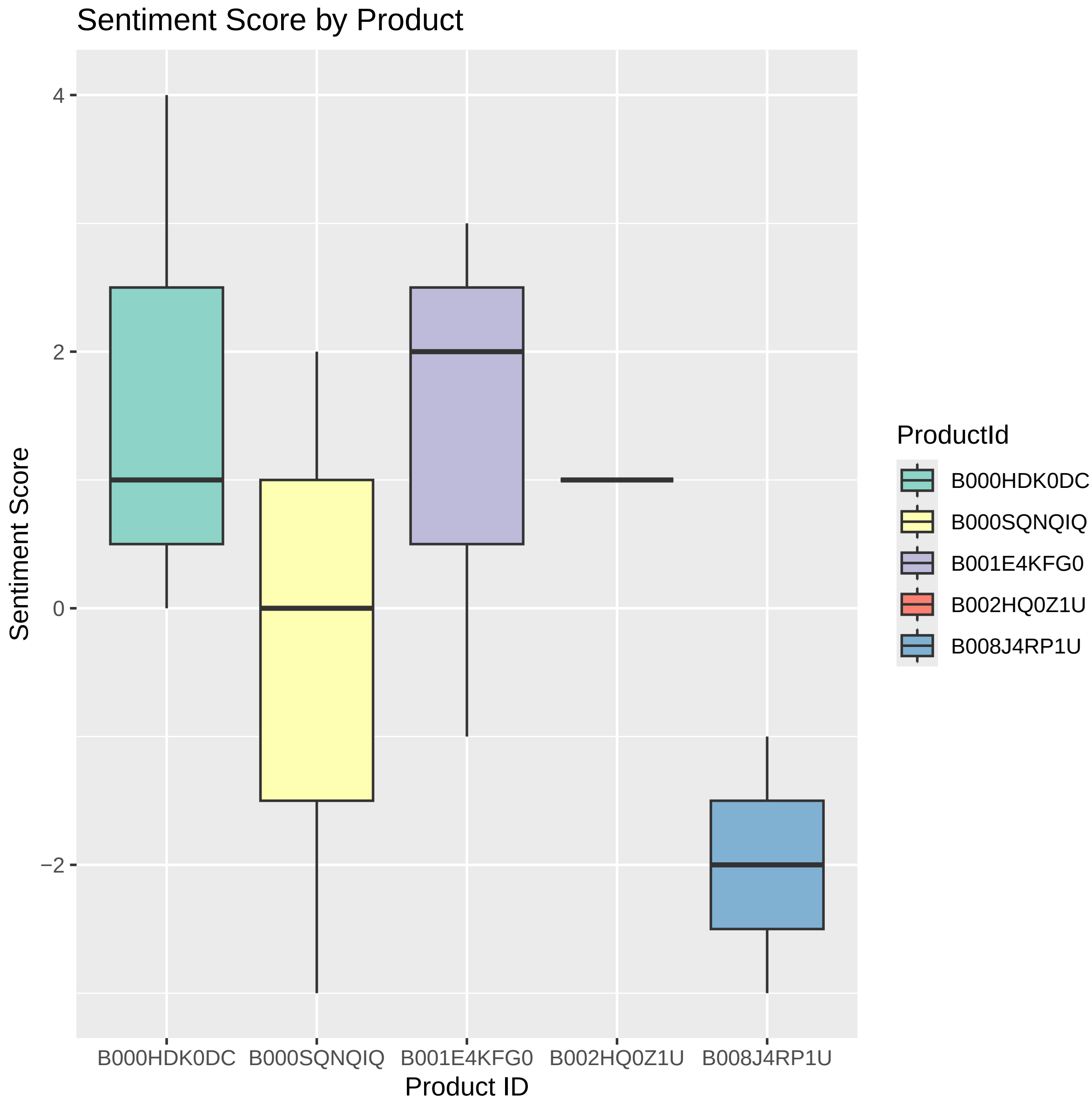


Second, a scatter plot examines the relationship between star ratings and sentiment scores.

```
Warning: Unknown palette: "blues"
```



Finally, a box plot compares sentiment scores across different products, which can help identify items associated with more negative feedback.



3 Results

The results of the sentiment analysis reveal several clear patterns in the Amazon Fine Food sample. The bar plot shows that positive words appear more frequently than negative ones, suggesting that most customers describe their experiences in favorable terms, even if occasional issues are mentioned. The scatter plot further supports this by showing that higher star ratings tend to align with higher sentiment scores, meaning that what people write generally matches the score they give. The box plot shows differences between products, with some generating consistently positive sentiment while others receive more mixed or negative reactions. Together, these results provide a balanced picture of how customers feel across different items.

4 Discussion

These findings highlight how sentiment analysis can deepen our understanding of customer feedback beyond simply looking at star ratings. The overall positive tone in the bar plot suggests strong product performance, but the presence of negative words reminds managers that even well-rated items can have recurring complaints worth addressing. The scatter plot's alignment between text sentiment and numerical ratings shows that written reviews offer reliable emotional cues. Meanwhile, the box plot helps identify which products are consistently appreciated and which may require closer attention due to variability in sentiment. This combination of insights helps organisations focus on specific issues rather than relying solely on averages or surface-level summaries.

5 Conclusion

Overall, the analysis demonstrates how sentiment analysis in R can turn written feedback into clear, useful insights for decision-making. By evaluating both the emotional tone of the text and its relationship to ratings, managers can better understand customer satisfaction, identify problem areas, and prioritise improvements. Even with a small dataset, visualising sentiment reveals patterns that would be easy to miss in raw reviews. As businesses increasingly depend on customer feedback to guide their decisions, sentiment analysis becomes an important and practical tool for interpreting emotions at scale.

6 Affidavit

I hereby affirm that this submitted paper was authored unaided and solely by me. Additionally, no other sources than those in the reference list were used. Parts of this paper, including tables and figures, that have been taken either verbatim or analogously from other works have in each case been properly cited with regard to their origin and authorship. This paper either in parts or in its entirety, be it in the same or similar form, has not been submitted to any other examination board and has not been published.

I acknowledge that the university may use plagiarism detection software to check my thesis. I agree to cooperate with any investigation of suspected plagiarism and to provide any additional information or evidence requested by the university.

6.1 Checklist

- ☒ The handout contains 11 pages of text.
- ☒ The submission contains the Quarto file of the handout.
- ☒ The submission contains the Quarto file of the presentation.
- ☒ The submission contains the HTML file of the handout.
- ☒ The submission contains the HTML file of the presentation.
- ☒ The submission contains the PDF file of the handout.
- ☒ The submission contains the PDF file of the presentation.
- ☒ The title page of the presentation and the handout contain personal details (name, email, matriculation number).
- ☒ The handout contains a bibliography, created using BibTeX with an APA citation style.
- ☒ Either the handout or the presentation contains R code that demonstrates coding expertise.
- ☒ The filled out Affidavit.
- ☒ The link to the presentation and the handout published on GitHub.

[Sanika Parakatil Joseph,] [10 Dec 2025,] [Cologne]

7 References

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–177.

McAuley, J., & Leskovec, J. (2011). *Amazon fine food reviews dataset*. https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews

Silge, J., & Robinson, D. (2017). *Tidytext: Text mining using dplyr, ggplot2, and other tidy tools*. https://cran.r-project.org/package=tidytext