Romeo vs. Juliet: Investigating Gender-Based Linguistic Variations in Online Reviews
A Sentiment Analysis Approach

Maria Martin Barbera, Giami Huynh, Hannah Hommel
Amsterdam University College
Text Mining
Jelke Bloem
May 31st, 2024

Introduction

In the digital age, online reviews play a crucial role in shaping consumer decisions, especially for products as ubiquitous as smartphones. Understanding the nuances behind these reviews can offer valuable insights for manufacturers and marketers alike. Given the widespread reliance on smartphones, it is beneficial to understand how different demographics, particularly men and women, express their opinions and differences. Our project utilizes sentiment analysis to explore and compare how men and women review the iPhone XR. Sentiment analysis is the process of using natural language processing and machine learning to identify and quantify the emotional tone expressed in written text, gauging public opinion. By analyzing patterns in language, sentiment analysis can reveal whether a piece of text is positive, negative, or neutral, and highlight the key emotions conveyed.

A study conducted by Ravula et al. (2023) found that "women (vs. men) are more authentic, less analytical, more positive-affective, and less negative-affective" (p.1). Yu (2009) further supported these findings, noting that women are more influenced by the quality of information provided hence giving more structural reviews. Communication styles often vary by gender, with women typically engaging in more supportive, participatory, and personal communication, while men often communicate in a more direct, assertive, and competitive manner (Merchant, 2012). These differences can influence how each gender expresses opinions and emotions in their reviews. Research from Theses & Cinardo (2011) reported that feminine communication often emphasizes equality, support, and responsiveness; conversely, masculine communication aims to establish higher status among peers through directness and lack of emotional response (p.6).

Based on findings from the studies, by analyzing differences and similarities in the language patterns of men and women, we hope to provide insights that can help improve product design, marketing strategies, and customer satisfaction. The aim is to identify noticeable differences in the way males and females review mobile phones and to assess the potential reasons for any gender biases in review content and helpfulness ratings. By understanding gender-specific feedback, organizations can develop more tailored and effective approaches to address consumer needs and enhance the overall user experience.

Methodology

Pre-Processing

In this project, we used a dataset of iPhone XR reviews that includes ratings and reviews from Indian customers regarding their experience and opinions that have been translated into English (*data.world*). Mobile phone review is chosen because smartphones are essential for men and women, ensuring a balanced number of reviews from men and women. The dataset was stored in a JSON file so to access it we used the JSON library. There are several attributes in the dataset including profile name, helper count, review rating, review text, and review title. To get the gender of the reviewer, we applied filters to get their first name isolated since the corpora only include first names. Initially, many people were referred to as "Amazon Customer" or "Anonymous", hence had to be filtered first to have identifiable profile names. The next step was to remove contractions (e.g.: Dr.) to classify their gender, however, others such as Mr. and Ms. were used to classify gender. The data continues to require further cleaning as there were some profile names containing symbols (ex. @), capitalization inconsistencies, and punctuation which was handled by using multiple regular expressions (via the re library).

After filtering contractions, symbols, and others, the number of names was reduced to 4,368 from 5,010. The subsequent step is identifying the users' gender by utilizing several corpora. However, the issue was that we had to consider the existence of unisex names, specifically Indian unisex names. In this case, we were able to find a list of 35 unisex names to filter out (Sant, 2023), leading to 26 individuals being eliminated from the dataset. The following action was to find male and female Indian names corpora where the gender would be saved in a new column in the dataset. The largest corpus for each gender that we were able to find contained around 14,000 names (Milos Bejda 2015a & Milos Bejda 2015a). Then we applied a smaller list of female and male names (Sant, 2023). After applying these corpora we were able to gender around 1,700 individuals which were saved in a text file with csv to use for the sentimental analysis.

Sentimental analysis

The code performs sentiment analysis on review texts, classifying them into three categories: positive, negative, and neutral, and storing this classification in a new column labelled sentiment_label. This classification is achieved using the TextBlob library to calculate the sentiment polarity of each review text and the NLTK library to remove common stopwords. The data is grouped by gender and sentiment label to calculate the counts of each sentiment category for each gender and determine the average sentiment scores by gender, along with the mean and mode ratings. A for-loop identifies the number of unique words and repeated words, which is then used to calculate the repetition ratio.

Additionally, the code identifies and counts common positive and negative sentiment words by gender. These common words are visualized using bar graphs that report the top 20 most common words in the reviews, along with their frequency. This is done by creating two separate functions to identify common positive words and negative words. These functions iterate over the review data, filtering out common stopwords and specified exclude words. For each filtered word, the sentiment polarity is calculated using the TextBlob library. If a word's sentiment polarity is positive or negative, it is counted or subtracted and added to a counter specific to that gender. This analysis is executed separately for male and female reviews, and the top 20 most common negative words for each gender are extracted for potential visualization.

The analysis concludes by calculating the total helpful votes by gender and sentiment label and evaluates the accuracy of sentiment classification against manually categorized data with a limited sample of 100 reviews.

Results & Discussion

In our project, we aim to explore potential differences in how men and women review mobile phones. To achieve this, we performed over several tests such as sentiment analysis, sentiment score calculation, word frequency analysis, review length analysis, etc. These examinations help us understand the emotional and evaluative language employed by men and women, providing a nuanced view of consumer sentiment. By focusing on the most frequently used positive and negative words, we sought to uncover any distinct patterns or trends in the language used by each gender. This comparison allows us to determine whether there are notable distinctions in how men and women express satisfaction or dissatisfaction with their mobile phones. In the sections that follow, we present the expected and actual findings of our investigation. We start by listing the top 20 positive words for both men and women, followed by the top 20 negative words. Subsequently, we provide a comparative analysis to highlight significant similarities and differences in the sentiment expressed by each gender. Our results lay the groundwork for a detailed discussion on the implications of these findings, offering insights into consumer behaviour and potential marketing strategies.

Expected Results

Our hypotheses were grounded in existing research on gender differences in communication styles. We anticipated that female reviews would be more lengthy and detailed compared to male reviews. Prior studies have shown that women tend to use more words and provide more elaborate descriptions in their writing, reflecting a preference for contextual and comprehensive communication. We expected this tendency to manifest in longer, more detailed mobile phone reviews from women. Additionally, we hypothesized that female reviews would be less harsh in their comments. Research indicates that women are more likely to soften their criticisms and use polite language, aiming to maintain positive social interactions even when providing negative feedback (Mulac & Lundell, 1994). Therefore, we expected women to employ a more measured and less critical tone in their reviews. In contrast, we expected male reviews to be more straightforward, simple, and repetitive. Men's communication style often focuses on direct evaluations and factual information, with less emphasis on elaboration and nuance (Colley & Todd, 2002). We anticipated that men would write more concise reviews, repeating key points for emphasis, and using a more direct and unembellished language.

Moreover, we predicted that women would use more emotionally expressive language in their reviews. Women tend to share their feelings more openly and use a higher frequency of adjectives and adverbs that convey positive or negative emotions (Newman et al., 2008). Consequently, we expected female reviews to feature more emotionally charged words, reflecting a greater emotional engagement with the product. Conversely, we anticipated that men might focus more on technical aspects and performance-related terms, reflecting a practical and utilitarian approach to reviewing mobile phones. This aligns with research suggesting that men often prioritize functionality and performance in their evaluations.

Results

An interesting finding from our analysis showed that reviews from men appear to be more helpful in comparison to reviews from women in regards to smartphones such as the iPhone XR. In Table 1.1, male reviews have been found to be upvoted for helpfulness sixteen times more than female reviews. When looking further into the language used by both genders to understand the gap in finding reviews from one gender more 'useful' than the other, the results revealed something contrary to our expectations: the language used by men and women in their mobile phone reviews is strikingly similar This may have been due to the lack of identifiable females in the reviews being 40%. We anticipated more distinct differences in the length and emotional expressiveness of the reviews, as well as variations in the harshness of criticisms. However, the findings show significant overlap in both positive and negative words, suggesting that our hypothesis was incorrect.

Furthermore, contrary to our expectations, male and female reviews are shown to be commonly the same length, and that there are some longer reviews from men (Figures 3 & 4). However, male reviews appear to range in length that can reach nearly 160 words, whereas female reviews show more consistency within a certain text length of less than 100 words (with some exceptions). Men may be more prone than women to lose their minds over trivial matters, which could account for the higher upvotes for reviews that are helpful because the males' comments tend to be more insightful and lengthy with their reviews.

The positive words used by both men and women show significant overlap, indicating common aspects that both genders appreciate in mobile phones (Figure 1). Words like awesome, best, great, excellent, and others appear in the top 20 for both genders. This indicates a shared positive sentiment towards the overall quality, performance, and value of mobile phones among men and women. However, subtle differences might reflect gender-specific preferences or expressions such as men tend to use words like much, little, and thanks, which might indicate a focus on aspects like value for money (much, little) and appreciation or gratitude (thanks). On the other hand, women use words like super, new, and heavy. The inclusion of “super” and “new” might suggest an emphasis on novelty and exceeding expectations, while “heavy” could reflect a positive connotation of durability or build quality, depending on the context of the reviews.

Negative words, on the other hand, used by men and women in their mobile phone reviews also show considerable overlap, indicating common issues and frustrations with mobile phones (Figure 2). This suggests that both sexes have similar criticisms regarding certain aspects of mobile phones. However, there are some differences in the negative words used by each gender which might reflect distinct areas of concern or expression. Our results show that men tend to use words like long, expensive, pathetic, fake, etc., which may indicate specific frustrations with the longevity, cost, quality, and usability of mobile phones. These terms indicate that men might be more focused on practical and technical shortcomings. Women, on the other hand, include words like poor, least, center, disappointing, badly, and confused. The use of “poor” and “disappointing” indicates a general dissatisfaction with overall performance or quality, while “least” and “center” may reflect concerns about specific features or design aspects. The word “confused” suggests issues with usability or user interface that might lead to a frustrating user experience.

Additionally, mean ratings, mode ratings, average sentiment scores, and word usage statistics provide further insights into the differences between genders in their mobile phone reviews. In Table 2, men tend to give higher mean ratings (M = 4.51, SD = N/A) compared to women (M = 4.27, SD = N/A). Additionally, average sentiment scores are slightly higher for men (M = 0.46, SD = N/A) than for women (M = 0.41, SD = N/A), indicating a slightly more positive sentiment in men's reviews. However, women use a greater variety of unique words in their reviews (F = 1366, M = 1800), with a higher repetition ratio for men (F = 2.71, M = 3.43) suggesting more repeated words in men's reviews (F = 3697, M = 6172) (Table 3). These statistical differences complement the qualitative analysis, providing a comprehensive understanding of gender differences in mobile phone reviews.

Conclusion

Our analysis of the iPhone XR reviews using sentiment analysis has revealed interesting insights into the similarities and differences in how men and women express their opinions about these products. Contrary to our initial hypotheses, the language used by both genders shows significant overlap in both positive and negative sentiments. This suggests a common appreciation and shared frustrations regarding the quality, performance, and value of mobile phones.

Positive reviews from both men and women indicate a high value placed on overall quality and performance, using similar language to express their satisfaction. Subtle deviations exist, such as men emphasizing aspects like value and appreciation, while women focus more on novelty and build quality. For instance, men might use words that reflect practical benefits, whereas women highlight the enjoyment of new features and sturdiness based on top common review words. While negative reviews also show considerable overlap, men often discuss technical issues and cost concerns, while women express general dissatisfaction and usability problems. This reflects slight variations in focus, with men potentially more attuned to specific technical shortcomings and pricing, while women might be more concerned with the overall user experience. Quantitative metrics reveal some distinctions: male reviewers tend to use a broader vocabulary and have a higher repetition ratio, suggesting a more concise but varied use of language. Female reviewers, on the other hand, provide more detailed and less repetitive feedback. These findings highlight that while men and women share many common perspectives on mobile phones, there are nuanced differences in their reviews that reflect distinct priorities and expressions.

Contrary to our initial hypothesis and existing literature, our analysis revealed no substantial differences in how males and females review mobile phones. This finding challenges the belief that gender significantly influences review content and sentiment, suggesting that the specific product category of mobile phones might contribute to this unexpected result. It is possible that mobile phones, being a ubiquitous and essential device, elicit similar responses across genders, minimizing noticeable differences in review patterns. For future research, we suggest broadening the scope to include various product categories, as different types of products might reveal more pronounced gender-based differences in online reviews. Additionally, incorporating a larger and more diverse dataset could provide more comprehensive insights such as age, cultural background, and technology proficiency might also help uncover more nuanced patterns in online reviews. This multi-faceted approach could enhance our understanding of consumer behaviour and inform more effective marketing strategies.

Index

	Negative	Neutral	Positive	Total
Female	69	117	507	693
Male	68	162	774	1004
Total	138	279	1281	1697

Table 1: Summary of sentiment categorized by gender

	Negative	Neutral	Positive	Total
Female	82	27	335	444
Male	1763	5124	406	7293

Table 1.1: "Helpful votes" by gender sentiment

	Male	Female
Mean Ratings	4.514940	4.271284
Mode Ratings	5	5
Average Sentiment Scores	0.410025	0.456076

Table 2: Mode and Mean ratings, and average sentiment scores

	Male	Female
Total unique words	1800	1366
Total repeated words	6172	3697
Repetition ratio	3.43	2.71

Table 3: Summary of unique and repeated words, and repetition ratio

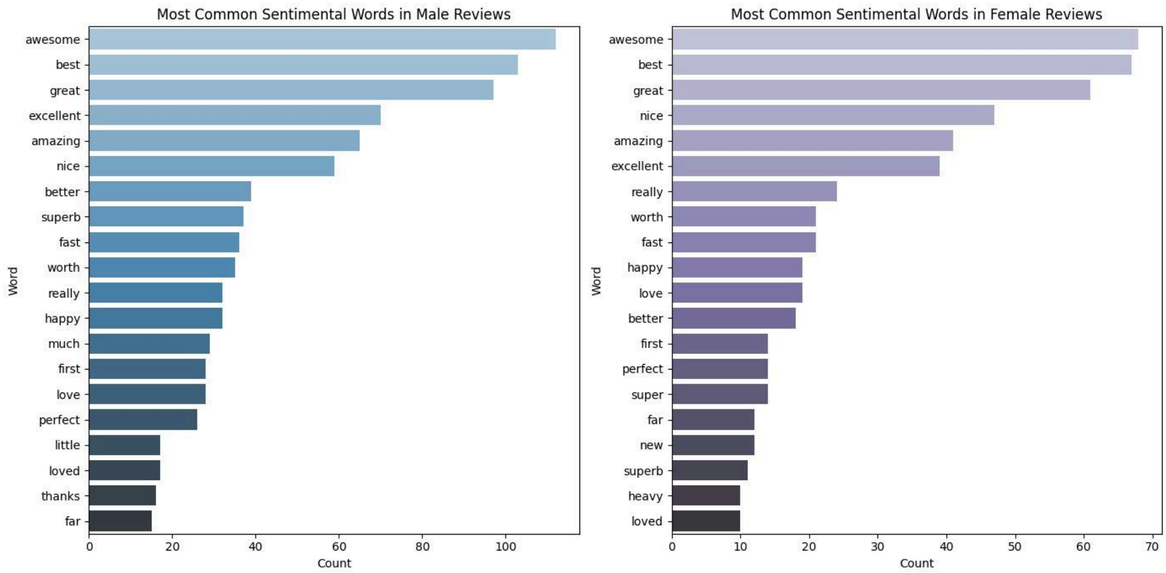


Figure 1: Most common sentimental words in male and female review

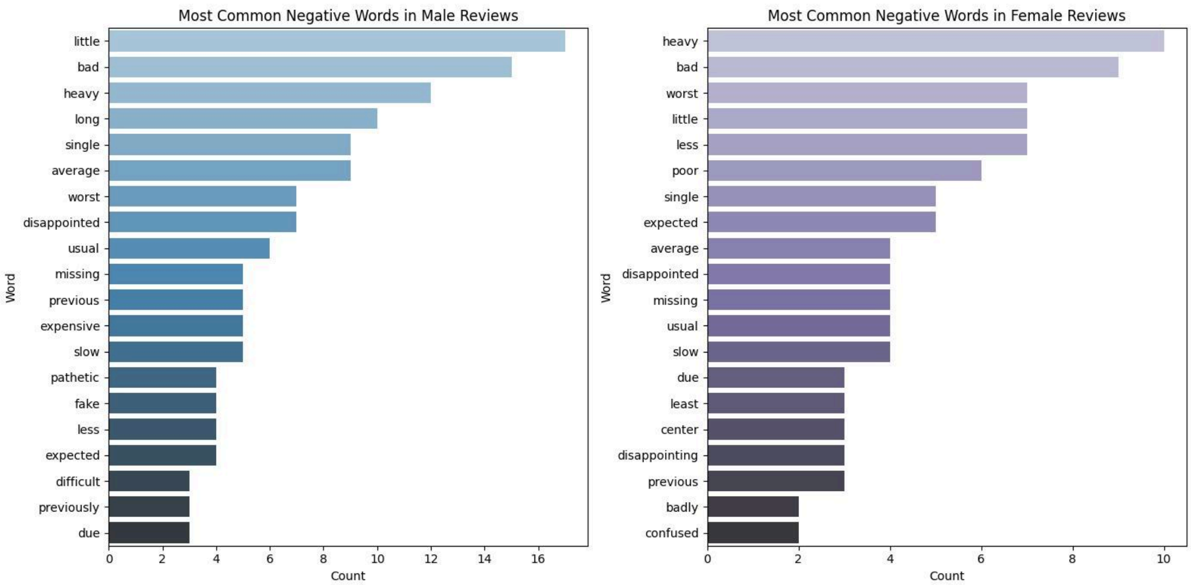
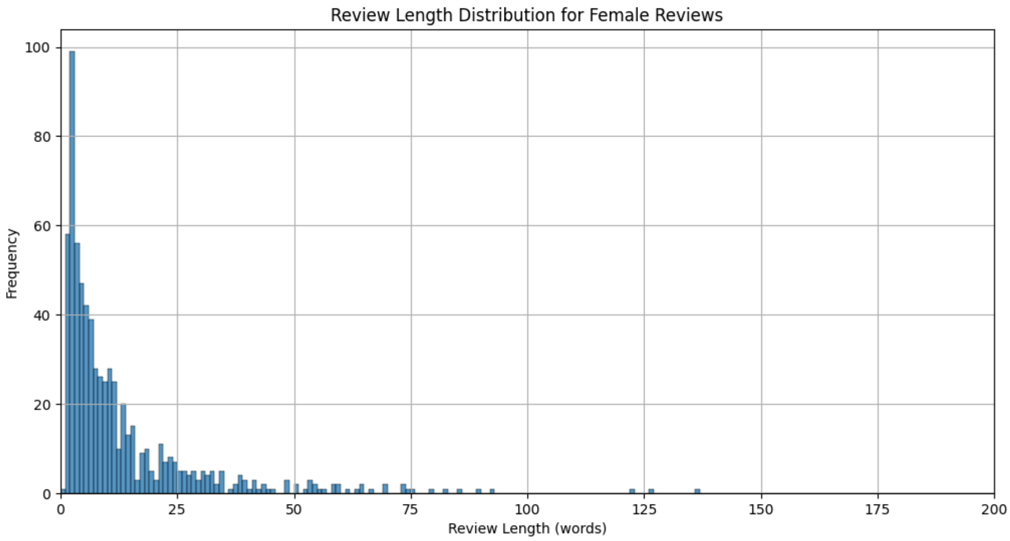
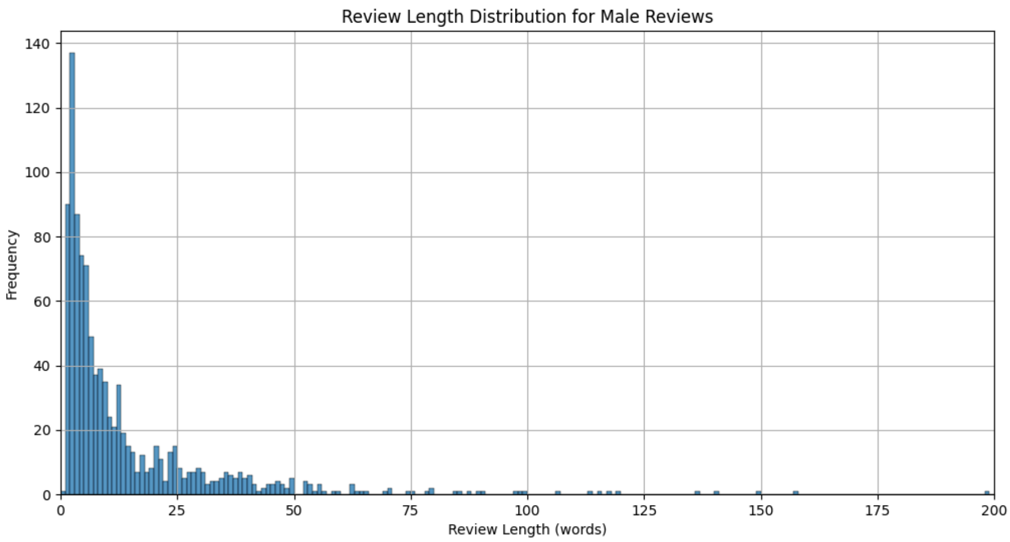


Figure 2: Most common negative words in male and female reviews



<u>Figure 3:</u> Review length distribution for female reviews



<u>Figure 4:</u> Review length distribution for male reviews

References

Bejda, M. (2015a). *Dataset of ~14,000 Indian female names for NLP training and analysis. The names have been retrieved from public records. (name,gender,race)*. Gist. https://gist.github.com/mbejda/9b93c7545c9dd93060bd

Bejda, M. (2015b). *Dataset of ~14,000 Indian male names for NLP training and analysis. The names have been retrieved from public records. (name,gender,race)*. Gist. https://gist.github.com/mbejda/7f86ca901fe41bc14a63

Colley, A., & Todd, Z. (2002). Gender-Linked Differences in the Style and Content of E-Mails to Friends. *Journal of Language and Social Psychology*, *21*(4), 380–392. https://doi.org/10.1177/026192702237955

data.world. (n.d.). Data.world. Retrieved June 1, 2024, from https://data.world/opensnippets/iphone-11-reviews/workspace/file?filename=apple_iphone_11_reviews.json

Merchant, K. (2012). *How Men and Women Differ: Gender Differences in Communication Styles, Influence Tactics, and Leadership Styles*. https://scholarship.claremont.edu/cgi/viewcontent.cgi?article=1521&context=cmc_theses

Mulac, A., & Lundell, T. L. (1994). Effects of gender-linked language differences in adults' written discourse: Multivariate tests of language effects. *Language & Communication*, *14*(3), 299–309. https://doi.org/10.1016/0271-5309(94)90007-8

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender Differences in Language Use: An Analysis of 14,000 Text Samples. *Discourse Processes*, *45*(3), 211–236. https://doi.org/10.1080/01638530802073712

Ravula, P., Bhatnagar, A., & Gauri, D. K. (2023). Role of gender in the creation and persuasiveness of online reviews. *Journal of Business Research*, *154*, 113386. https://doi.org/10.1016/j.jbusres.2022.113386

Sant, V. (2023, June 2). *vsant/indian-name-classifier*. GitHub. https://github.com/vsant/indian-name-classifier

Theses, H., & Cinardo, J. (2011). *CCU Digital Commons Male and Female Differences in Communicating Conflict*. https://digitalcommons.coastal.edu/cgi/viewcontent.cgi?article=1087&context=honors-theses#:~:text=According%20to%20Wood%20and%20Holmstrom

Yu, T. (2009). *Do Male and Female Favor Online Product Review the Same Do Male and*

Female Favor Online Product Review the Same.
https://core.ac.uk/download/pdf/301390761.pdf