Deciphering Satisfaction: Analyzing the Spectrum of Customer Feedback at Pizza Hut, Sri Lanka*

Chenyiteng Han

April 1, 2024

This paper employs statistical models to analyze customer review data from a Pizza Hut outlet in Sri Lanka, specifically examining how star ratings correlate with the presence and length of textual feedback. Our analysis reveals that customers providing lower star ratings are more likely to leave detailed text reviews. These findings suggest that customer reviews, particularly those with negative feedback, are a rich but underutilized resource for businesses aiming to enhance customer satisfaction and service quality. The paper thus sheds light on consumer behavior patterns, underscoring the importance of nuanced feedback analysis in the continual improvement of customer experience.

Table of contents

---

*Code and data from this analysis are available at: https://github.com/Hhnxxxxxx/Pizza-Hut-Reviews.git.

1 Introduction

This paper delves into the empirical world of customer feedback, with a sharp focus on discerning patterns within star ratings and text reviews from a prominent Pizza Hut branch in Sri Lanka. Amidst the deluge of customer ratings and reviews that businesses receive daily, we aimed to understand the implicit messages conveyed by customers through their chosen modes of feedback. Specifically, we investigated two core relationships: one, the probability of customers leaving text reviews alongside their star ratings, and two, the extent of detail in those reviews as gauged by their length.

The estimand of our study is the effect of customer star ratings on the probability of leaving textual feedback and the length of such feedback at a Pizza Hut outlet in Sri Lanka. Through the application of a logistic regression model and a multilevel negative binomial model, we analyzed the tendencies of customers to leave text reviews and their verbosity across different levels of satisfaction. Section 3 Models delves into the technical aspects of these models and their underpinnings. The results, as elaborated in Section 4 Results, pointed to a clear inverse relationship between the star ratings and both the likelihood and length of text reviews. Customers giving lower star ratings were significantly more inclined to leave text reviews, and their reviews tended to be longer. This finding is pivotal as it emphasizes the tendency of less satisfied customers to actively voice their feedback, potentially providing businesses with critical information for improvements.

The implications of this study are significant for the retail and service industries as they emphasize the need for a nuanced approach to customer feedback analysis. The remainder of this paper, detailed in Section 2 Data through Section 5 Discussion, explores the data's raw and processed forms, the analytical methods applied, the insights gleaned from the results, and reflects upon the limitations and future directions of this research, including the potential integration of natural language processing for deeper qualitative analysis. By understanding these findings within the broader context of customer behavior, businesses can better tailor their services to meet customer needs and expectations, ensuring a more satisfactory customer experience overall.

2 Data

The foundation of our analysis on Pizza Hut customer feedback in Sri Lanka was a dataset obtained from Kaggle, accessed through the Kaggle API (Kaggle (2022)) which allows for automated and direct data retrieval. In our R (R Core Team (2022)) environment, the dataverse package (Leeper (2022)) was instrumental in organizing and managing the dataset. Data preparation, including cleaning and transformation, was handled adeptly with the tidyverse collection of packages (Wickham and al. (2019)), particularly utilizing dplyr (Wickham and al. (2022)) for its powerful data manipulation capabilities.

Visualization and interpretation of the data were greatly enhanced by ggplot2 (Wickham (2022)) for creating informative graphics, while RColorBrewer (Neuwirth (2022)) provided aesthetic color schemes to enrich these visual representations. The knitr package (Xie (2014)) facilitated the seamless integration of R code within our dynamic report, allowing for an interactive presentation of results.

For the advanced statistical modeling, we employed rstanarm (Team (2022)), which provided a suite of tools for Bayesian inference in generalized linear models. This was complemented by tidybayes (Kay (2022b)), which enabled a tidy workflow when working with Bayesian models. The arrow package (Foundation (2022)) was pivotal in efficient data reading and writing, ensuring quick access to our dataset.

The modelsummary package (Arel-Bundock (2022)) allowed us to create detailed and comprehensive statistical tables, effectively summarizing the results of our complex models. To visualize distributions and express uncertainty in our data, ggdist (Kay (2022a)) proved invaluable. For handling the outputs from mixed models, broom.mixed (Bolker et al. (2022)) was utilized, allowing for the tidying and interpretation of mixed effects models. Collectively, the synergy of these R packages provided a robust framework for our research, ensuring a high standard of accuracy and reproducibility.

2.1 Raw Data

This study utilizes the "Pizza Hut Reviews - Insights from One of Sri Lanka's Pioneer Branches" dataset, extracted from Kaggle. The dataset is a treasure trove of customer sentiments, encompassing 4000 entries of evaluative feedback for the Pizza Hut branch located at Union Place, Colombo, which is noted for being among the first to establish Pizza Hut's presence in Sri Lanka.

As depicted in Table 1, the dataset provides a snapshot into the customers' appraisals and narratives. Each record consists of three columns that collectively offer a comprehensive picture of consumer opinion.

- title: This column consistently identifies the venue as "Pizza Hut - Union Place," situating the reviews within the specific context of this branch's performance.

- stars: Reflecting the customer's satisfaction, this field quantifies their experience on a 1 to 5 scale, with 1 indicating dissatisfaction and 5 signifying an excellent service encounter.

- text: Perhaps the most telling of the columns, this contains the actual text of the customer's review. It ranges from concise accolades to detailed narratives, providing qualitative depth beyond the numerical rating.

While alternative datasets from different branches or regions could have been considered, the choice to focus on this specific dataset was intentional. The Union Place branch's historic significance within the Sri Lankan context presents a unique opportunity to understand customer feedback dynamics where brand establishment is deep-rooted.

In presenting the raw data in Table 1, readers are invited to witness the unfiltered voice of the customer, which will later be dissected and analyzed to glean actionable insights. This approach ensures a thorough examination of the data points, setting the stage for a nuanced analysis that respects the complexity and authenticity of customer feedback.

Table 1: Sample of the Raw Dataset

Title	Stars	Text
Pizza Hut - Union Place	5	
Pizza Hut - Union Place	5	
Pizza Hut - Union Place	4	

Table 1: Sample of the Raw Dataset

Title	Stars	Text
Pizza Hut - Union Place	5	
Pizza Hut - Union Place	4	
Pizza Hut - Union Place	5	
Pizza Hut - Union Place	5	Good place to visit to have some food
Pizza Hut - Union Place	5	Had a personal pan chicken tandoori, it was quite on point just like the pizza hut one looks forward to. The staff was courteous and had parking space too.
Pizza Hut - Union Place	5	
Pizza Hut - Union Place	4	

2.2 Cleaned Data

Following the data curation process, we have derived a cleaner and more analytically suitable dataset titled "clean_pizza_hut_reviews," specifically sculpted to facilitate quantitative analysis. The original verbose feedback has been distilled into quantitative variables that retain the essence of customer ratings while allowing for numerical manipulation.

As demonstrated in Table 2, the refined dataset retains the granularity of ratings through the 'stars' variable, indicative of customer satisfaction levels, while introducing two new variables: 'text_indicator' and 'words.'

- review_id: Assigned as a unique identifier to each review, this variable serves a structural purpose, facilitating reference and management within the dataset rather than bearing any intrinsic meaning.

- stars: Preserved from the raw data, this metric remains unchanged, symbolizing the customer's rating, ranging from 1 for dissatisfaction to 5 for exemplary service.

- text_indicator: A binary variable, it denotes the presence (1) or absence (0) of textual commentary accompanying the star rating. This distinction is pivotal for subsequent analyses that differentiate between ratings supplemented with text and those that are not.

- words: This numeric variable quantifies the length of the text accompanying the review. It operates in tandem with the 'text_indicator'—if a review is textually barren ('text_indicator' is 0), the 'words' count will also be 0. Conversely, when a review includes text, the 'words' variable captures the word count, adding a layer of depth to the quantitative assessment.

This cleaned version facilitates a more straightforward approach to understanding the patterns and tendencies in customer feedback without the convolution of qualitative data interpretation.

Table 2: Sample of the Cleaned Dataset

Review ID	Stars	Text Indicator	Words
1	5	0	0
2	5	0	0
3	4	0	0
4	5	0	0
5	4	0	0
6	5	0	0
7	5	1	8
8	5	1	29
9	5	0	0
10	4	0	0

2.3 Data Summary

Two figures have been constructed to offer visual summaries of the cleaned dataset, elucidating patterns in customer reviews at a glance.

2.3.1 Frequency of Text Reviews Across Star Categories

The histogram in Figure 1 categorizes reviews based on the star ratings they correspond to and whether the reviews include text. Reviews with text are indicated in blue, and those without in red. A conspicuous majority of reviews across all star ratings do not include textual feedback, as the prevalence of red bars suggests. This is particularly pronounced in the 5-star category, which also features the highest number of reviews, indicating the store's generally favorable reception.

Notably, as the star rating decreases, the proportion of text reviews (blue bars) increases relative to those without text. This trend culminates in the 1-star category, where the number of text reviews surpasses that of non-text reviews, suggesting that customers who are dissatisfied are more likely to provide written explanations of their experiences. This observation will be quantitatively examined through logistic regression modeling to confirm the hypothesized relationship between the star rating and the propensity to leave textual feedback.
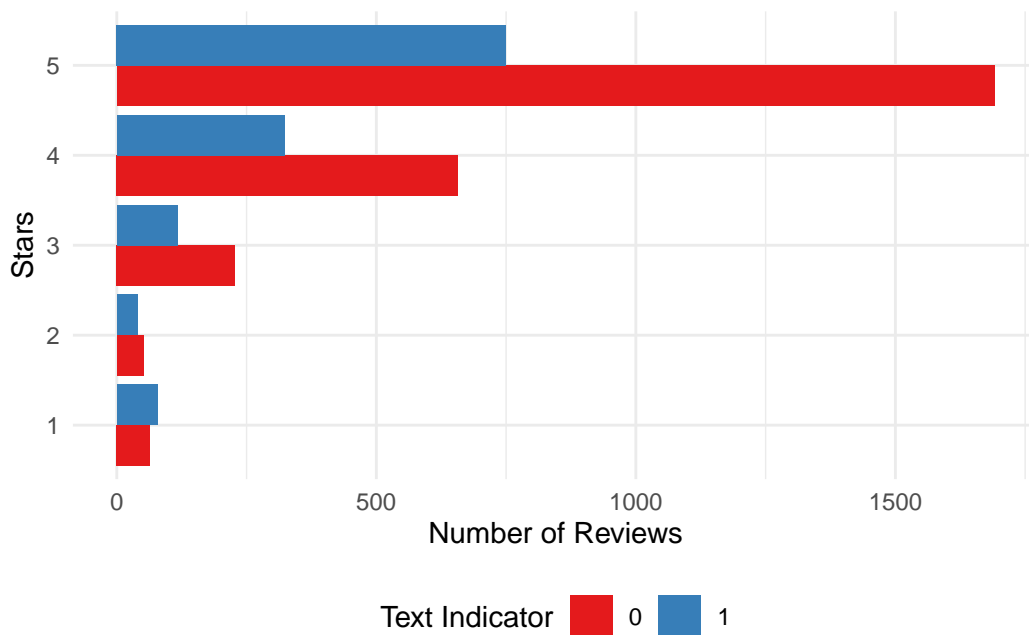

Figure 1: Different Frequency of Text Reviews Across Different Star Categories

2.3.2 Overview of Text Length in Reviews by Star Rating

In Figure 2, we present a box plot that examines the length of text in reviews, filtered by star rating, for those containing written commentary (text indicator=1). The logarithmic scale is employed on the y-axis to manage outliers and maintain readability. This transformation allows us to observe the central tendency and dispersion of text lengths without distortion from extreme values.

The plot suggests a tendency for reviews with higher star ratings to have shorter text lengths, although the trend does not exhibit a stark contrast across the star categories. The presence of outliers is apparent, particularly in the lower star categories. To further investigate and model the relationship between the star rating and the length of text, a multilevel regression analysis will be employed, taking into account the variability within and between the different levels of star ratings.
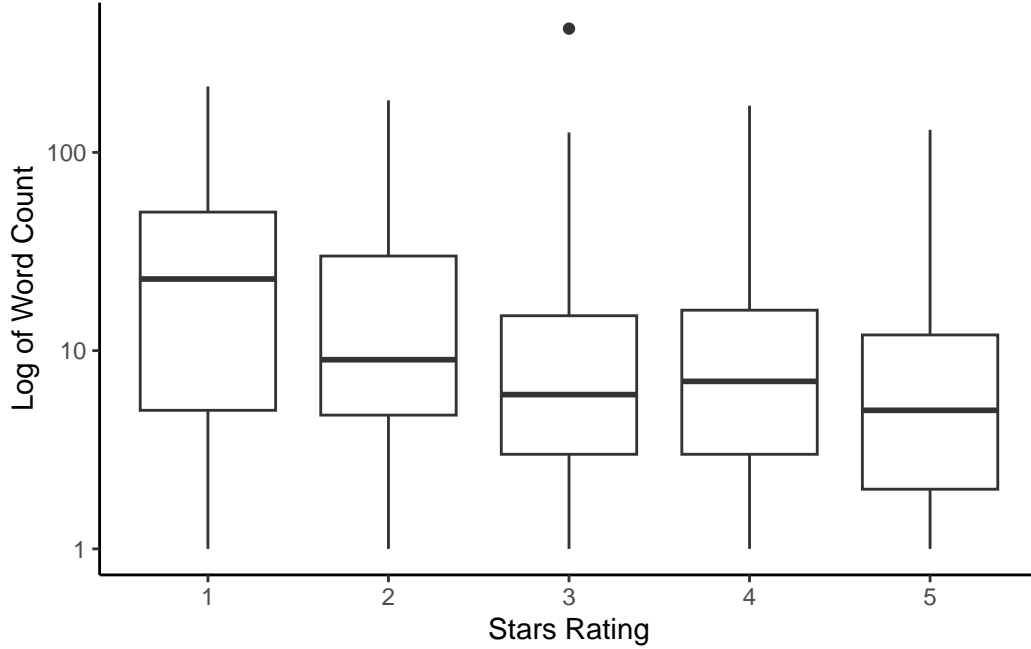
Figure 2: Different Frequency of Text Reviews Across Different Star Categories

The synthesis of these visuals into our analysis provides preliminary insights into the nature of the reviews, laying the groundwork for more nuanced statistical modeling. They serve to highlight the initial observations that customers tend to write longer texts when they are less satisfied and tend to leave reviews without text when more satisfied, with the possible implications of these trends to be explored in subsequent modeling efforts.

2.4 Measurement

The translation of the real-world phenomenon of customer experience at a Pizza Hut outlet into an entry in our dataset involves several structured steps:

- Experience Recording: The initial interaction is the customer's visit to the Pizza Hut branch, where their experience is influenced by various factors such as food quality, service, ambiance, and overall satisfaction.

- Feedback Mechanism: Post-visit, customers are invited to rate their experience on a scale of 1 to 5 stars, with the option to provide additional textual feedback. This mechanism serves as a direct channel for customers to express their satisfaction or dissatisfaction.

- Data Entry: The feedback received—comprising of star ratings and written reviews—is systematically entered into the raw dataset. Ratings are recorded numerically, while

textual comments are transcribed, capturing the richness of customer sentiment.

- Quality Assurance: To ensure the integrity of the data, the entries undergo a cleaning process. This step addresses issues such as missing values, duplicates, or irrelevant entries, thereby standardizing the dataset for consistent analysis.

- Variable Construction: From the cleaned data, new variables are constructed. The 'text_indicator' variable is derived to indicate the presence (1) or absence (0) of textual feedback, and the 'words' variable measures the length of the written comment if present.

- Dataset Compilation: These steps culminate in the compilation of the cleaned dataset. Each entry now systematically represents the quantified aspect of a customer's experience through the 'stars' rating, the 'text_indicator', and the 'words' count.

- Analytical Readiness: The dataset, now in a clean and structured form, is poised for various statistical analyses. These analyses will explore the relationships within the data, such as the correlation between star ratings and the propensity to leave textual feedback or the influence of customer sentiment on the length of reviews.

The path from a customer's dining experience to a data point within this dataset exemplifies a carefully curated process that maintains the essence of their feedback while transforming it into a format conducive to rigorous data analysis.

3 Models

Model Diagnostics are included in Appendix A.

3.1 Logistic Regression Model for Text Indicator and Star Ratings

Our analytical endeavor aims to decipher the underlying patterns in customer feedback for a Pizza Hut outlet, specifically examining the propensity of customers to leave textual comments alongside their star ratings. The central thesis of our model is to understand the relationship between the quantitative measure of customer satisfaction (the star ratings) and the qualitative aspect (textual feedback).

The binary nature of our outcome of interest, `text_indicator`, which signifies the presence or absence of text in a review, necessitates a modeling approach that can handle dichotomous data effectively. Logistic regression is particularly suited for this task as it predicts the probability of a binary response based on one or more predictor variables. This model provides the probability that a review contains textual feedback for each level of star rating, which is crucial for our investigation into the nuances of customer reviews.

Mathematically, the logistic regression model can be expressed in the form of an equation for the log-odds of the probability P of a review containing text ($Y = 1$):

$$\log \left(\frac{P(Y = 1)}{1 - P(Y = 1)} \right) = \beta_0 + \beta_1 \cdot X_1 + ... + \beta_n \cdot X_n \qquad (1)$$

In equation (1):

- Y is the binary response variable (text_indicator), with 1 indicating the presence of text in a review, and 0 indicating its absence.

- β_0 is the intercept term of the model.

- $\beta_1, ..., \beta_n$ are the coefficients for the predictor variable, representing the impact of each star rating on the log-odds of text_indicator being 1.

- $X_1, ..., X_n$ are the encoded categorical variables derived from the stars predictor, with each level (star rating) having its associated coefficient in the model.

Through this logistic regression framework, we aim to quantify the extent to which the star rating influences the probability of textual feedback being provided. This quantification will enable us to discern whether customers who are more satisfied (higher stars) or less satisfied (lower stars) are more inclined to provide detailed textual feedback, thus offering valuable insights into consumer behavior and satisfaction.

By integrating the logistic regression model into our analysis, we expect to unveil patterns that could inform strategies to enhance customer engagement and satisfaction, potentially guiding operational improvements and customer service protocols.

3.2 The Multilevel Model for Text Length and Star Ratings

Our objective extends to examining the detail provided in customer reviews, specifically the length of text in reviews that include textual feedback. To achieve this, we built a multilevel model that allows us to assess how the star ratings, representing different levels of customer satisfaction, relate to the verbosity of the reviews.

We opted for a multilevel model approach due to the inherent group structure within our data: reviews can be naturally divided into groups based on the star rating. This model type accommodates the non-independence of observations within the same group (star rating category) and enables us to capture both the fixed effects of the star ratings and the random effects that account for the variability within each star rating level.

Additionally, the length of the review text, counted as the number of words, typically exhibits overdispersion, where the variance exceeds the mean. The negative binomial distribution is chosen over the simpler Poisson distribution to model such count data with overdispersion. The negative binomial model is particularly suited for this kind of data as it introduces an extra parameter to account for the overdispersion.

The multilevel negative binomial model is formulated as follows, focusing on the logarithm of the expected count of words in the reviews:

$$\log(\text{E[words|stars]}) = \beta_0 + u_{0j} + \beta_1 \cdot X_1 + ... + \beta_n \cdot X_n \tag{2}$$

In equation (2):

- E[words|stars] is the expected count of words in the text reviews given the star rating.

- β_0 is the overall intercept across all groups.

- u_{0j} represents the random effects capturing the variations in word count within each level of star ratings.

- $\beta_1...\beta_n$ are the fixed-effect coefficients for the predictor variable, which in this case, is the star rating treated as a categorical factor.

- $X_1...X_n$ are the encoded categorical variables derived from the stars predictor.

By focusing on the subset of data where text indicator is 1, we concentrate on those reviews that actually contain text. This filtering ensures that the model is trained on the most informative data for our research question—understanding the depth of feedback in relation to customer satisfaction levels.

Through this multilevel negative binomial model, we seek to elucidate the patterns that may exist between the satisfaction indicated by the star rating and the length of textual feedback provided. The analysis will explore whether customers who report different levels of satisfaction tend to leave more or less detailed reviews, as reflected in the word count.

By adopting this modeling approach, we aim to extract substantive insights from the text reviews, which are particularly valuable to the store for understanding customer experiences in depth and leveraging those insights for service improvement.

4 Results

4.1 The Relationship Between Text reviews and Star Ratings

The results of our logistic regression model are shown in Table 3 and Figure 3, which together help explain the relationship between the star ratings given by customers and their likelihood of leaving a text review The relationship between.

Table 3 shows the estimated coefficients of the logistic regression model. The intercept as a baseline comparison shows a positive coefficient, although this coefficient is not statistically significant given the accompanying standard errors. The coefficients for stars=2 through

stars=5 are all negative, with factors stars=3, factor stars=4, and factor stars=5 showing more significant coefficients (indicated by smaller standard errors).

The negative coefficient on star rating indicates that customers who give higher star ratings (2 to 5) are less likely to leave text reviews compared to the baseline (implied to be 1 star because not shown). The size of these coefficients increases with star rating, indicating that the likelihood of leaving a text review decreases with increasing satisfaction (star rating).

Figure 3 shows the 90% confidence intervals for the model coefficients. These intervals provide a range of values within which we can be 90% confident about the true value of the coefficient. If the interval of the coefficient does not cross the zero line, it indicates a significant effect at the 10% significance level.

The confidence intervals for stars=3, stars=4, and stars=5 are well below zero, which reinforces the conclusion in Figure 3 that as the star rating increases, the likelihood of a text review is lower. The intervals for stars=2 overlap to zero, indicating less certainty in the effect of a 2-star rating compared to the baseline.

Taken together, these results support the assertion that customers who provide lower star ratings are more likely to leave text reviews, while customers who provide higher star ratings are more likely to leave reviews without text. This pattern can be observed by increasingly negative coefficients for higher stars and their corresponding confidence intervals.

The R-squared values in Table 3 indicate that the proportion of variance in the response variable explained by the model is very low, suggesting that factors other than star ratings may influence the decision to leave a text review.

In summary, logistic regression analysis provides evidence to support the notion that customer satisfaction, as measured by star ratings, is inversely related to the probability of leaving text feedback. This insight is valuable for understanding customer engagement and can inform strategies to encourage more comprehensive feedback from highly satisfied customers.

4.2 The Different Length of Text Reviews Between Different Stars

The multilevel negative binomial model's findings are encapsulated in Table 4 and Figure 4. These visualizations allow us to interpret the relationship between the star ratings and the length of the text reviews provided by customers.

Table 4 contrasts the coefficients from the negative binomial model with those from the multilevel negative binomial model. Notably, all coefficients for stars=2 through stars=5 are negative, indicating that as star ratings increase, the expected count of words in reviews tends to decrease. The coefficients are statistically significant, as suggested by the small standard errors, particularly for ratings of 3 stars and above in both models. This aligns with the hypothesis that lower star ratings are associated with longer, presumably more detailed, text reviews.

Table 3: Whether a Customer is Likely to Give a Text Review Based on the Star Rating he Gave

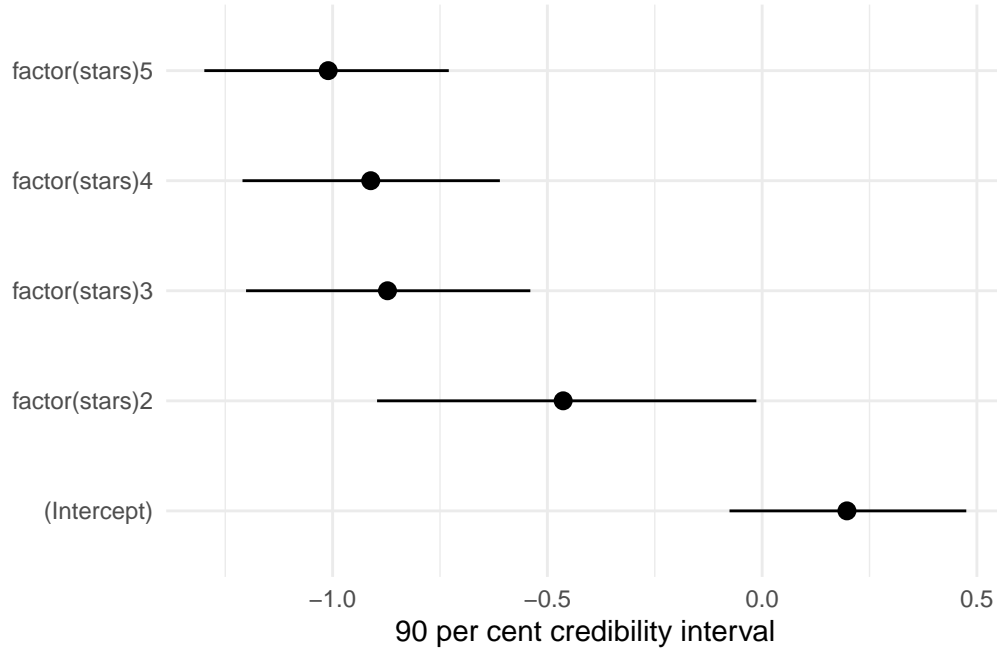|	Give a Text Review
(Intercept)	0.197
	(0.166)
factor(stars)2	−0.463
	(0.271)
factor(stars)3	−0.872
	(0.204)
factor(stars)4	−0.911
	(0.180)
factor(stars)5	−1.010
	(0.170)
Num.Obs.	4000
R2	0.011
Log.Lik.	−2508.704
ELPD	−2513.8
ELPD s.e.	22.2
LOOIC	5027.5
LOOIC s.e.	44.5
WAIC	5027.5
RMSE	0.47

Figure 3: Credible Intervals for Predictors of Giving a Text Review

Figure 4 offers a visual representation of the distribution of the length of text reviews across different star ratings. The plots suggest a clear trend: the group with a 1-star rating shows a distribution that extends towards higher word counts, indicative of longer reviews. As the star ratings increase, the distributions shift leftward, denoting shorter reviews. This pattern is consistent across all star levels and corroborates the conclusion drawn from the coefficients in Table 4.

The combined evidence from the coefficients and the distribution plots substantiates the conclusion that customers leaving lower star ratings tend to provide more extended textual feedback in their reviews, while those with higher star ratings are inclined to leave briefer comments. This trend may reflect a tendency for less satisfied customers to elaborate on the reasons for their dissatisfaction.

The consistency between the coefficients' significance and the visual trends in review lengths across star ratings enhances the credibility of this conclusion. The models collectively indicate a meaningful and inverse relationship between satisfaction levels, as expressed by star ratings, and the extent of customer engagement in providing textual feedback.

Table 4: Whether a Customer is Likely to Give a Text Review Based on the Star Rating He
Gave

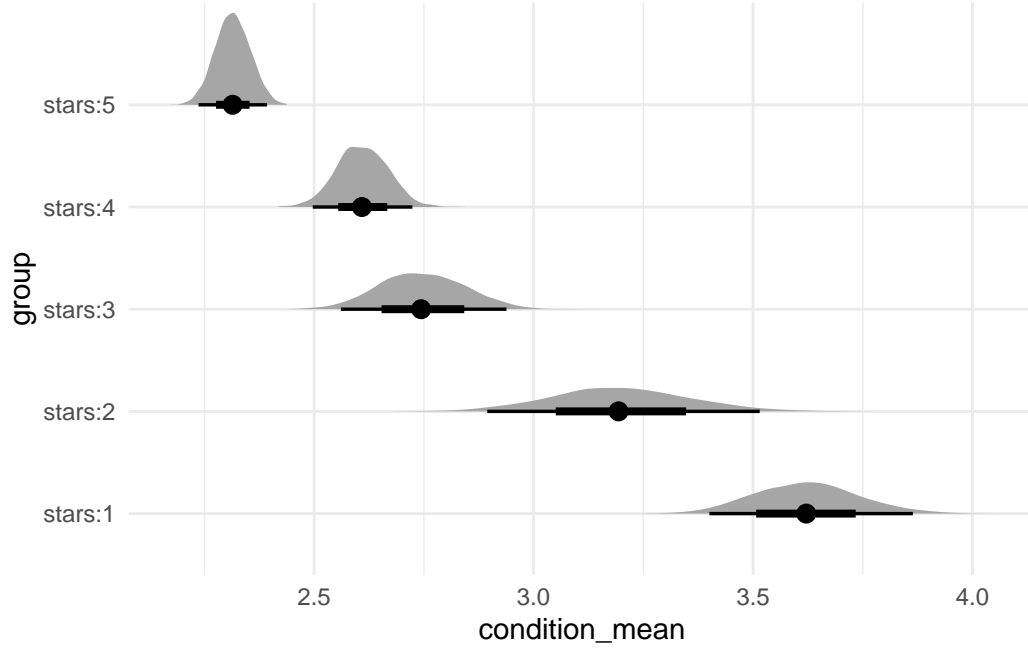|	Neg binomial	Multilevel Neg Binomial
(Intercept)	3.647	2.854
	(0.116)	(0.260)
factor(stars)2	−0.413	
	(0.205)	
factor(stars)3	−0.909	
	(0.161)	
factor(stars)4	−1.042	
	(0.132)	
factor(stars)5	−1.335	
	(0.125)	
Num.Obs.	1308	1308
Log.Lik.	−4669.934	−4669.796
ELPD	−4680.4	−4679.8
ELPD s.e.	55.6	55.4
LOOIC	9360.8	9359.6
LOOIC s.e.	111.3	110.7
WAIC	9360.7	9360.4
RMSE	22.73	22.73

Figure 4: Examining the Distribution of Length of Text for Different Stars

5 Discussion

5.1 Overview of Works

In this article, we conduct a comprehensive analysis of customer feedback at Pizza Hut stores in Sri Lanka, focusing on the relationship between star ratings and customers' propensity to leave text reviews and the length of these reviews. Using Kaggle's dataset, we carefully cleaned and prepared the data to build two different statistical models. The first model is a logistic regression designed to evaluate the likelihood of a customer leaving text feedback along with the star rating they provide. The second model is a multilevel negative binomial model used to understand the relationship between star ratings and text length in reviews containing written feedback. By analyzing the data through these models, we found that customers who gave lower star ratings were more likely to give reviews with text, and the texts they gave were longer and more detailed than those of customers who gave high star ratings. From this, we seek to uncover the nuances of customer satisfaction and engagement with the aim of providing actionable insights that enhance customer experience and business practices.

5.2 Insights into Societal Behaviors

The findings in this paper shed light on societal behaviors within the customer service domain, revealing a clear trend in consumer feedback patterns. It shows that customers who experience dissatisfaction are more inclined to provide detailed textual feedback, potentially as a means to express and substantiate their discontent. Conversely, customers who are highly satisfied tend to express their approval through ratings alone, rather than elaborating in text. This behavior highlights a societal tendency where individuals are more motivated to articulate negative experiences as a call for acknowledgement and change, whereas positive experiences are often succinctly acknowledged. This insight underscores the importance for businesses to closely examine negative feedback, as it may contain critical information for service improvement and customer satisfaction.

This study also teaches us about the quieter side of customer satisfaction. While unhappy customers often take the time to detail their grievances, happy customers may simply leave a high rating without further comment. This tendency means businesses might not be hearing from many satisfied customers, potentially giving a skewed picture of overall service satisfaction. It's a reminder that not all feedback is vocal, and positive experiences can go underreported in customer reviews.

5.3 Reflecting on the Limitations

One of the limitations in our approach is the assumption that the length of the text reviews is directly proportional to the level of detail and richness of the content. However, a longer review does not always equate to more substantive feedback. Some lengthy reviews may be verbose without offering concrete insights, while shorter reviews could be concise yet rich in valuable information. This paper's methodology does not account for the semantic depth or the specific qualitative insights that might be present in the text, which requires a more nuanced analysis, potentially through natural language processing techniques.

Additionally, our analysis is constrained to a single Pizza Hut branch, which limits the generalizability of the findings across different locations, cultures, or even other service industries. Customer behavior can be highly context-specific, and what holds for one branch in Sri Lanka may not be true elsewhere. Also, the scope of the data does not capture the non-linear complexities of customer experiences—factors such as customer expectations, individual differences, and specific circumstances surrounding each visit are beyond the reach of this study. These elements represent areas for further research to build on the preliminary findings presented here.

5.4 Future Directions and Unexplored Avenues

The path forward from this study points to the integration of natural language processing (NLP) to unravel the qualitative depth of customer reviews. Future research could deploy sentiment analysis to gauge the emotional tone and extract specific themes from textual feedback, providing a more granular understanding of customer satisfaction. Expanding the dataset to encompass multiple locations and a broader range of service industries would also offer a more comprehensive view of customer feedback behaviors across different contexts. Additionally, investigating the impact of individual customer characteristics, such as loyalty and prior experiences, could yield insights into the nuanced drivers of review behaviors. Pursuing these avenues will enhance our understanding of customer feedback mechanisms and inform the development of more targeted strategies for improving customer experience and engagement.

Appendix

A Model Diagnostics

A.1 Diagnostics for the Logistic Regression Model

Figure 5a is a trace plot for the logistic regression model discussed in Section 3.1. It illustrates the sampled values across multiple iterations of the Markov Chain Monte Carlo (MCMC) process for each model parameter. The plot shows that the chains for each parameter are well-mixed and stable, without noticeable drifts or divergences, suggesting good convergence. This indicates that the sampling process is reliable and the estimates for the model parameters are robust.

Figure 5b is an Rhat plot for the logistic regression model in Section 3.1. It displays the Rhat statistic for each chain, which is a measure of convergence. An Rhat value of 1 indicates perfect convergence, while values greater than 1 suggest that more sampling might be necessary. The plot shows that all Rhat values are close to 1, none exceeding the threshold of 1.1, which suggests that convergence has been achieved, and the chains have sampled from the posterior distribution effectively.
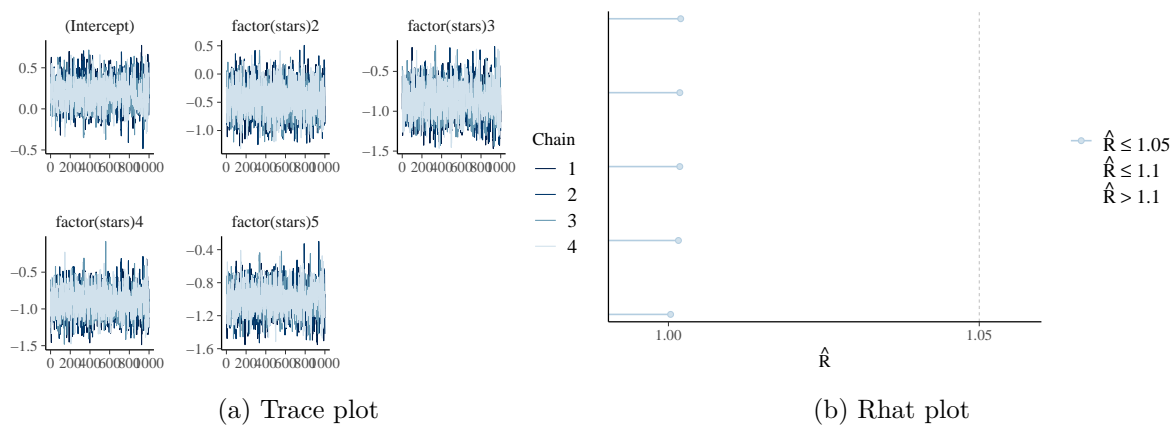

(a) Trace plot   (b) Rhat plot

Figure 5: Checking the convergence of the MCMC algorithm

A.2 Diagnostics for the Multilevel Model

Figure 6a is a trace plot for the multilevel model referenced in Section 3.2. This graph shows the behavior of the sampled values across iterations for the model's parameters. We observe that while most parameters have traces that suggest convergence, with dense and intermingled lines, there are a few, notably the 'reciprocal_dispersion' parameter, that exhibit less stable behavior, with wider fluctuations. This instability suggests potential issues with convergence

19

for some parameters, indicating that the model may require further tuning or additional iterations to ensure reliable parameter estimates.

Figure 6b is an Rhat plot for the multilevel model in Section 3.2. It displays Rhat values for the model's parameters, with values ideally close to 1.00 indicating satisfactory convergence. However, the plot indicates that one of the parameter's Rhat values exceeds the acceptable threshold, represented by a value greater than 1.10. This is a clear indication that the model has not converged well for this particular parameter, signaling the need for additional diagnostic checks or extended sampling to achieve convergence.
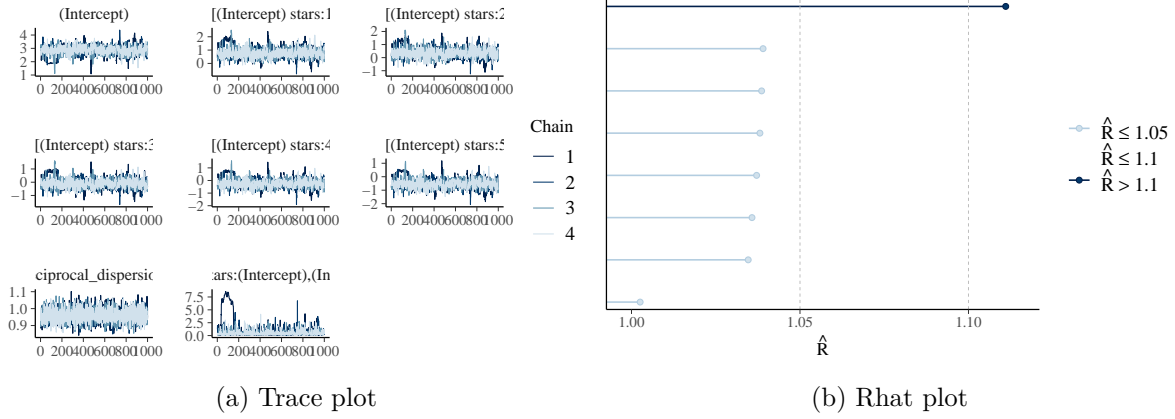

(a) Trace plot   (b) Rhat plot

Figure 6: Checking the convergence of the MCMC algorithm

References

Arel-Bundock, Vincent. 2022. *Modelsummary: Summary Tables and Plots for Statistical Models and Data.* https://CRAN.R-project.org/package=modelsummary.

Bolker, Ben et al. 2022. *Broom.mixed: Tidying Methods for Mixed Models.* https://CRAN.R-project.org/package=broom.mixed.

Foundation, Apache Software. 2022. *Arrow: Integration to 'Apache Arrow'.* https://CRAN.R-project.org/package=arrow.

Kaggle. 2022. "Kaggle API." https://www.kaggle.com/docs/api.

Kay, Matthew. 2022a. *Ggdist: Visualizations of Distributions and Uncertainty.* https://mjskay.github.io/ggdist/.

―――. 2022b. *Tidybayes: Tidy Data and Geoms for Bayesian Models.* https://mjskay.github.io/tidybayes/.

Leeper, Thomas. 2022. *Dataverse: Client for Dataverse 4 Repositories.* https://CRAN.R-project.org/package=dataverse.

Neuwirth, Erich. 2022. *RColorBrewer: ColorBrewer Palettes.* https://CRAN.R-project.org/package=RColorBrewer.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Team, Stan Development. 2022. *Rstanarm: Bayesian Applied Regression Modeling via Stan.* https://mc-stan.org/rstanarm.

Wickham, Hadley. 2022. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.* https://CRAN.R-project.org/package=ggplot2.

Wickham, Hadley, and et al. 2019. *Tidyverse: Easily Install and Load the Tidyverse.* https://CRAN.R-project.org/package=tidyverse.

———. 2022. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2014. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://yihui.org/knitr/.