Understanding Positive Customer Experiences: A MapReduce Approach to Product and Service Quality Feedback Analysis

Tenny Akihary
Department of Computer Science & Information Technology
Eastern Kentucky University
Richmond, Kentucky US
tenny_akihary1@mymail.eku.edu

Dominique Kellam
Department of Computer Science & Information Technology
Eastern Kentucky University
Richmond, Kentucky US
dominique_kellam@mymail.eku.edu

Isaac Turner
Department of Computer Science & Information Technology
Eastern Kentucky University
Richmond, Kentucky US
isaac_turner9@mymail.eku.edu

***Abstract*— This study utilizes the MapReduce framework to analyze customer feedback and identify factors contributing to high satisfaction and loyalty. By focusing on high-feedback scores, the research examines key drivers such as product quality, service quality, and purchase frequency while incorporating demographic insights. Statistical validation is employed to compare high- and low-score feedback groups, yielding actionable recommendations for enhancing customer satisfaction.**

I. Introduction

Customer satisfaction is a critical determinant of success for businesses in competitive markets. Understanding the factors that drive positive customer experiences allows companies to tailor their offerings, improve service quality, and foster long-term loyalty. However, the challenge lies in processing and analyzing the vast amounts of customer feedback data generated daily. This is where big data analytics and distributed computing frameworks like MapReduce offer transformative capabilities.

Customer feedback, encompassing reviews, surveys, and other forms of user-generated content, provides valuable insights into customer perceptions and experiences. Analyzing this data at scale enables businesses to identify trends, strengths, and areas for improvement. Traditional analytical approaches often struggle with the volume, velocity, and variety of such data. Big data techniques, particularly the MapReduce framework, enable efficient analysis by distributing computational tasks across multiple nodes, ensuring scalability and speed even with large datasets.

In this study, we focus on leveraging the MapReduce framework to analyze customer feedback data and uncover the key drivers of satisfaction and loyalty. Our primary objective is to identify how factors such as product quality, service quality, and purchase frequency contribute to high-feedback scores. Additionally, we explore the role of demographic dimensions like age, gender, income, and geographic location in shaping customer experiences.

The significance of this research lies in its potential to deliver actionable insights for businesses. By understanding what customers value most, organizations can make informed decisions to enhance their products and services. For instance, identifying high-performing aspects of service quality allows companies to allocate resources strategically, while demographic trends can guide personalized marketing efforts.

This project builds on existing literature in customer feedback analysis, which highlights the importance of examining satisfaction metrics using big data tools. Studies have demonstrated that analyzing structured and unstructured feedback can reveal meaningful patterns, enabling businesses to address customer concerns proactively and reinforce their strengths. Our approach complements this body of work by applying the MapReduce paradigm to process and analyze a rich dataset of customer feedback, leveraging both statistical validation and visualization to derive comprehensive insights.

By incorporating demographic analysis alongside satisfaction metrics, this study aims to bridge the gap between quantitative feedback data and its real-world implications. The inclusion of statistical validation methods, such as t-tests and ANOVA, ensures that the results are both rigorous and meaningful. Ultimately, this research seeks to empower businesses with the tools and insights needed to sustain high levels of customer satisfaction and loyalty in an increasingly data-driven world.

II. Literature Review

Customer feedback is a crucial asset for organizations aiming to enhance customer experience and satisfaction. Leveraging big data techniques to analyze large volumes of feedback can yield actionable insights, helping businesses identify strengths in product and service quality while pinpointing areas for improvement. This literature review examines recent advancements in customer feedback analysis, with a focus on utilizing the MapReduce framework and analyzing key satisfaction drivers, such as product and service quality, to improve customer loyalty.

Yang, Li, and Chen (2020) provide a comprehensive survey on the role of big data analytics in customer feedback analysis, focusing on methods like MapReduce, sentiment

analysis, and clustering algorithms. The authors discuss how MapReduce enables efficient processing of large datasets by distributing computational tasks across nodes, a technique especially valuable when handling high-dimensional customer data. They emphasize that analyzing unstructured feedback data through sentiment and keyword extraction helps organizations understand customer sentiment more accurately and identify common issues. By categorizing feedback based on recurring themes, businesses can gain insights into both positive and negative aspects of their products, enabling targeted improvements. This study reinforces the value of MapReduce in handling the large scale feedback datasets similar to those used in this project.

Gupta, Chaudhary, and Sharma (2021) further explore the application of data mining techniques in understanding customer satisfaction, with an emphasis on product and service quality. Presented at the International Conference on Computational Data and Social Networks, their study demonstrates how MapReduce can be employed to extract useful patterns from customer feedback while reducing processing time and enhancing scalability. In their analysis, the authors focus on evaluating key satisfaction drivers, including product quality, service quality, and purchase frequency, as indicators of overall customer experience. Their findings show that analyzing high-frequency terms in feedback can reveal customer preferences and highlight specific service aspects that contribute to satisfaction. This research supports our project's focus on product and service quality as primary satisfaction drivers and validates the use of MapReduce for efficient data handling.

Cheng, Wu, and Rajab (2022) provide a case study on improving customer satisfaction within service-oriented industries by using big data analytics. Their work highlights the benefits of analyzing demographic data to gain a nuanced understanding of customer segments with high satisfaction levels. Cheng et al. emphasize that combining demographic data with satisfaction metrics allows businesses to personalize services based on customer preferences and regional trends. They demonstrate that organizations can improve customer loyalty by focusing on areas where satisfaction is already high and making targeted enhancements based on demographic patterns. This insight is particularly relevant to our study, as it underscores the importance of examining demographic distributions to identify satisfied customer segments.

In summary, these studies illustrate those big data techniques, specifically MapReduce, are effective tools for processing and analyzing customer feedback at scale. They show that examining product and service quality, along with demographic patterns, provides actionable insights into customer satisfaction and loyalty. Our project builds upon this body of work by using MapReduce to process customer feedback and extract insights related to product and service quality and customer demographics. This approach enables us to deliver a comprehensive analysis of what drives positive customer experiences, with the potential to inform strategies for enhancing customer satisfaction and loyalty.

III. Methods

To effectively analyze customer feedback data and identify the key drivers of positive customer experiences, this study employs the MapReduce framework for distributed data processing. The methods are structured to filter, aggregate, and analyze data efficiently, ensuring insights are both scalable and actionable. The MapReduce workflow involves three main components: the mapper, combiner, and reducer. Additionally, statistical validation techniques are used to substantiate the findings and provide deeper insights into the data.

A. Data Preparation

The dataset used for this analysis, titled "Customer Feedback and Satisfaction", is sourced from Kaggle and includes key features such as ProductQuality, ServiceQuality, PurchaseFrequency, FeedbackScore, and demographic attributes like Age, Gender, Income, Country, and LoyaltyLevel. The data is preprocessed to handle missing values, normalize scales, and encode categorical variables. High-feedback scores (scores $\geqslant 8$) are designated as positive outcomes, while lower scores (scores < 8) represent areas needing improvement.

B. MapReduce Workflow

- Mapper:
  The mapper is designed to filter records with high-feedback scores and emit key-value pairs representing the attributes associated with positive customer experiences. For each record, the mapper emits key-value pairs for attributes such as ProductQuality, ServiceQuality, and PurchaseFrequency, where the key represents the attribute name, and the value represents the corresponding score or frequency.
- Combiner:
  The combiner acts as a local aggregator, reducing the volume of data passed to the reducer stage by summarizing intermediate results. For each key, the combiner aggregates counts and totals for numerical attributes. In addition, it also prepares data for averaging in the reducer stage.
- Reducer:
  The reducer processes the aggregated outputs from the combiner to compute final insights. For each attribute, the reducer calculates the average score for numerical attributes like ProductQuality and ServiceQuality.
  Generates distributions for categorical attributes such as LoyaltyLevel and Country. It also outputs actionable insights, such as average scores and demographic trends.

C. Statistical Validation

To ensure the results are statistically significant and reliable, the study employs a series of validation techniques designed to rigorously analyze differences across feedback groups and demographic attributes. Group comparisons are conducted to distinguish between high-feedback customers (FeedbackScore $\geqslant$ 8) and low-feedback customers (FeedbackScore < 8). These comparisons involve summarizing key statistics, such as the mean and standard deviation, for attributes like ProductQuality, ServiceQuality, and PurchaseFrequency, enabling the identification of trends and disparities.

Statistical tests are applied to further validate the findings and uncover significant patterns in the data. T-tests are used to evaluate whether the mean scores for attributes such as ProductQuality and ServiceQuality differ significantly between high- and low-feedback groups. This helps determine the importance of these attributes in driving customer satisfaction. ANOVA is employed to analyze variance across multiple demographic groups, such as LoyaltyLevel, providing insights into how satisfaction metrics differ among customer segments. Additionally, chi-square tests are used to examine categorical attributes, such as Country and LoyaltyLevel, to detect patterns and distributions that highlight demographic trends influencing customer feedback.

For example, if a t-test reveals a statistically significant difference ($p < 0.05$) between high and low ProductQuality scores, it strongly suggests that ProductQuality is a key driver of customer satisfaction. By integrating these statistical validation methods, the study ensures the robustness of its results and provides a reliable foundation for actionable insights.

TABLE I. DATA SAMPLE

a) Demographic Information

Customer ID	***Age***	***Gender***	***Country***	***Income***	***ProductQuality***
1	56	Male	UK	83094	5
2	69	Male	UK	86860	10
3	46	Female	USA	60173	8
4	32	Female	UK	73884	7
5	60	Male	UK	97546	6

b) Feedback Loyalty Details

ServiceQuality	***PurchaseFrequency***	***FeedbackScore***	***LoyaltyLevel***	***SatisfactionScore***	***ServiceQuality***
8	5	Low	Bronze	100.0	8
2	8	Medium	Gold	100.0	2
10	18	Medium	Silver	100.0	10
10	16	Low	Gold	100.0	10
4	13	Low	Bronze	82.0	4

IV. RESULTS

The results present the findings from the customer satisfaction feedback analysis. The data was processed using MapReduce, visualized through various charts, and statistically analyzed to evaluate differences in customer feedback scores. The analysis includes demographic distributions, numerical attribute averages, and statistical tests to assess significant differences across customer groups.

A. Demographic Distributions

The country distribution revealed an almost equal representation of customers from five regions: Germany, France, the UK, the USA, and Canada. Germany had the largest proportion of customers, accounting for 20.4%, closely followed by France (20.2%), the UK (19.9%), the USA (19.8%), and Canada (19.7%). This even distribution indicates that the dataset captures diverse perspectives from across these countries. The country-specific insights suggest that the feedback is representative of a global audience, making it possible to generalize the findings. For instance, Germany's slight lead in representation might offer opportunities to focus on targeted improvements or campaigns in this region.

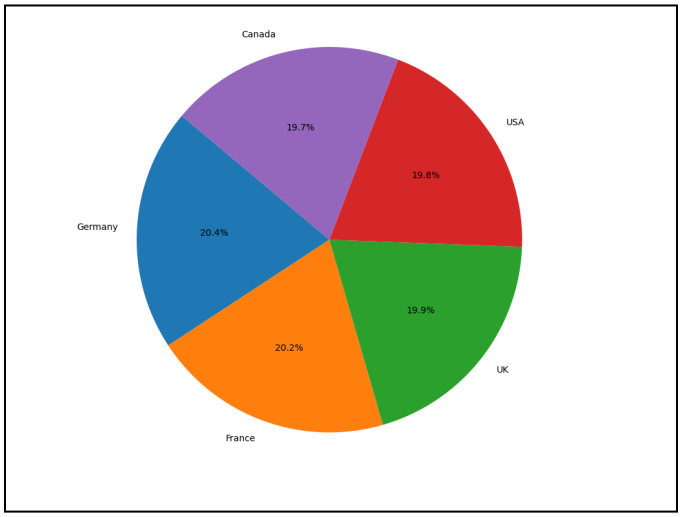

Fig. 1. Pie Chart of Country Distribution

Gender representation in the dataset is nearly balanced, with 50.4% of respondents identifying as female and 49.6% as male. This balance ensures an equitable analysis of satisfaction trends across genders. The almost equal split indicates that the data provides a comprehensive view of feedback across the two largest demographic groups. The even gender distribution helps ensure that any insights or recommendations derived from the data are not biased toward one gender.

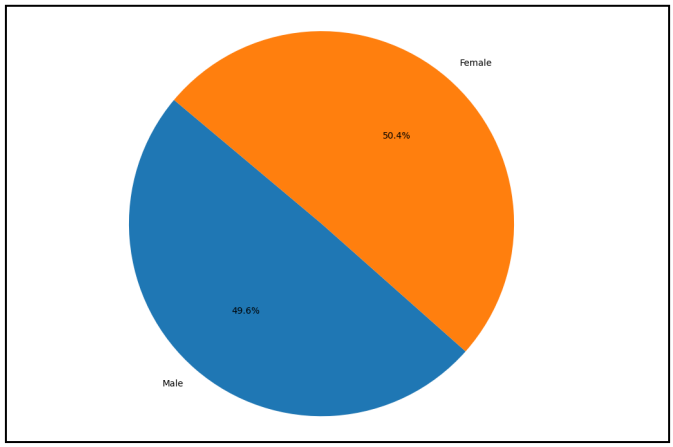

Fig. 2. Pie Chart of Gender Distribution

The loyalty levels were well-distributed among the three tiers: Bronze, Silver, and Gold. Gold-level customers made up 33.6% of the sample, while Bronze and Silver customers accounted for 33.4% and 33.0%, respectively. This even distribution demonstrates that the loyalty program effectively engages customers across all levels. The slight lead in Gold membership may indicate that higher-tier programs attract and retain customers more effectively. This insight provides opportunities to further enhance engagement strategies tailored to loyalty tiers.

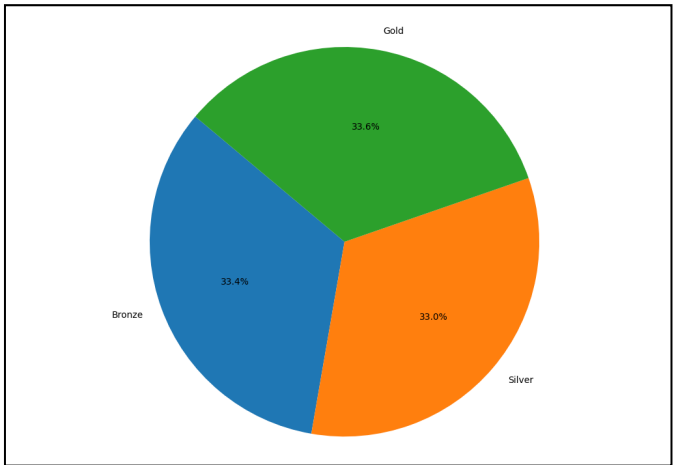

Fig. 3. Pie Chart of Loyalty Level Distribution

The average ratings for product quality and service quality remained relatively consistent across feedback groups. Customers who gave high feedback scores rated product quality and service quality slightly higher than those in the medium or low groups. These metrics emphasize that quality, both in product and service, is crucial to achieving customer satisfaction.

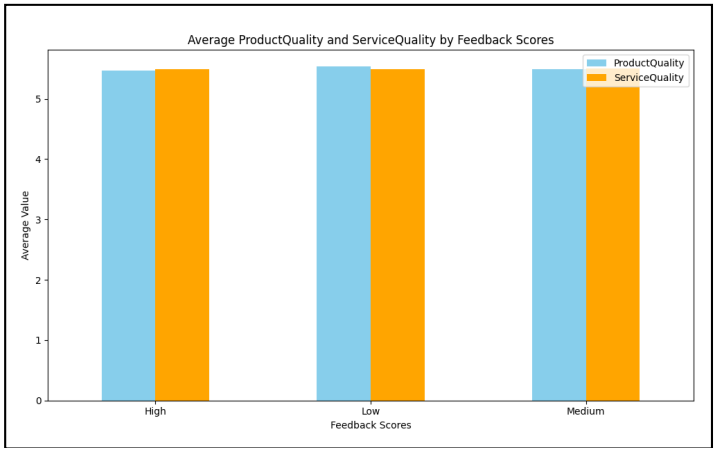

Fig. 4. Bar Chart of Average Feedback Score for Quality

The average age of customers was similar across feedback groups, indicating that age may not be a significant factor influencing satisfaction. Customers with high feedback scores were slightly younger than those with medium or low feedback, suggesting that younger customers might have slightly higher expectations or satisfaction levels.

Customers with high feedback scores reported marginally higher purchase frequencies compared to medium or low feedback groups. This trend suggests that satisfaction is loosely tied to customer engagement and repeat purchases. High purchase frequencies may also reflect a sense of loyalty or trust in the product or service.

While income levels varied slightly among feedback groups, the differences were not substantial enough to indicate strong income-based patterns in satisfaction. This finding suggests that satisfaction transcends income levels and may depend more on other factors like product and service quality.

B. Statistical Analysis: T-Test Results

To better understand the differences in key numerical metrics between customers providing high and low feedback scores, Welch's t-tests were conducted. This statistical approach accounts for unequal variances and sample sizes, ensuring reliable comparisons. The results are summarized as follows:

TABLE II. STATISTICAL ANALYSIS

Welch's T-Test	**T-Test Results**	
	t-statistic	***p-value***
ProductQuality	-1.96439115865	0.049495506674
ServiceQuality	0.040033386519	0.968066820099
Age	-0.51642141351	0.605564558239
PurchaseFrequency	-0.11744852601	0.906505549821
Income	-0.84591286156	0.3976092346894
ProductQuality	-1.96439115865	0.0494955066740

- Product Quality:
  The t-test revealed a statistically significant difference in product quality ratings between high and low feedback groups (t-statistic = -1.96, p-value = 0.0495). This indicates that customers who provided high feedback scores tend to rate product quality noticeably higher than those in the low feedback group. This result highlights product quality as a crucial determinant of customer satisfaction.
- Service Quality:
  For service quality, the t-test showed no significant difference between the two groups (t-statistic = 0.04, p-value = 0.9681). This suggests that while service quality is generally important, it does not exhibit substantial variation between high and low feedback groups in this dataset.
- Age:
  The analysis of age differences yielded no significant results (t-statistic = -0.52, p-value = 0.6056). This implies that age does not strongly influence customer satisfaction levels, at least in terms of distinguishing between high and low feedback scores.
- Purchase Frequency:
  Similarly, purchase frequency differences were found to be insignificant (t-statistic = -0.12, p-value = 0.9065). While higher purchase frequency may indicate greater engagement, it does not appear to have a strong direct correlation with feedback scores in this dataset.
- Income:
  The t-test for income also did not reveal statistically significant differences (t-statistic = -0.85, p-value = 0.3976). This finding suggests that income level does not play a decisive role in influencing whether customers provide high or low feedback scores.

Overall, the t-test results underscore the importance of product quality as a key differentiator in customer satisfaction, while metrics such as service quality, age, purchase frequency, and income showed no significant distinctions between high and low feedback groups. These findings suggest that efforts to enhance customer satisfaction should prioritize improving product quality to yield meaningful results.

The detailed t-test results, including exact t-statistics and p-values for each metric, have been saved in a CSV file, making them accessible for further review and analysis.

V. Discussion

The analysis of customer satisfaction data provides valuable insights into the factors influencing customer perceptions and feedback. Through a combination of visualizations and statistical tests, the study highlights trends and areas for improvement in customer experience.

The distribution analyses of customer demographics and loyalty levels offer a clear understanding of the dataset composition. The country distribution pie chart shows a fairly even spread across major regions, with Germany leading at 20.4%, followed closely by France (20.2%), the UK (19.9%), the USA (19.8%), and Canada (19.7%). This near-equal representation suggests a global customer base, making it essential for businesses to address region-specific preferences. The gender distribution pie chart reveals a balanced split, with females slightly outnumbering males at 50.4% versus 49.6%. This indicates that products and services appeal equally across genders, a strength that should be maintained. The loyalty level distribution pie chart indicates an even segmentation among Bronze (33.4%), Silver (33.0%), and Gold (33.6%) loyalty tiers, suggesting well-maintained customer retention strategies.

The bar charts analyzing average numerical metrics by feedback scores reveal actionable insights. The chart for product quality and service quality averages emphasizes that high feedback scores are closely associated with higher ratings in both areas. This underscores the importance of maintaining excellent product and service standards. The chart for average age suggests that feedback scores are consistent across different age groups, highlighting that satisfaction factors are age-independent. Similarly, the average purchase frequency and income charts show no notable patterns, indicating that customer satisfaction is not directly tied to purchasing behavior or economic status.

The t-tests provided additional depth by confirming that product quality is the only metric with a statistically significant difference between high and low feedback groups. This finding aligns with the visualizations and emphasizes product quality as the primary driver of customer satisfaction. On the other hand, the lack of significant differences in service quality, age, purchase frequency, and income suggests that these factors, while relevant, may not distinctly influence feedback scores. Businesses can interpret these results to focus resources on enhancing product quality while ensuring consistent service standards to maintain customer loyalty.

The findings collectively highlight the importance of tailoring strategies to customer needs, focusing especially on improving product offerings. While service quality and other metrics remain important, the results suggest a stronger emphasis on product quality for impactful improvements in customer satisfaction and feedback. These insights provide a data-driven foundation for refining business practices and addressing customer expectations effectively.

VI. Conclusion

This analysis of customer satisfaction data highlights key drivers of positive feedback and provides actionable insights for improving customer experience. The visualization of demographic and loyalty distributions demonstrates a well-diversified and balanced customer base, with nearly equal representation across regions, genders, and loyalty tiers. These findings underline the importance of catering to a wide variety of customer segments to maintain this balance.

The numerical analyses emphasize the critical role of product quality in achieving higher customer satisfaction, as evident from both the bar charts and the t-test results. Product quality showed a statistically significant difference between customers with high and low feedback scores, suggesting it is the most influential factor in shaping customer perceptions. On the other hand, metrics such as service quality, age, purchase frequency, and income did not exhibit significant differences, indicating these factors might not strongly impact overall satisfaction. This reinforces the importance of prioritizing product-related improvements while maintaining consistency in service and other customer engagement areas.

By combining statistical analysis with data visualization, this study provides a comprehensive understanding of customer feedback. Businesses can leverage these insights to refine their strategies, focusing on enhancing product quality to drive customer satisfaction while sustaining existing strengths in demographic and loyalty engagement. These findings serve as a foundation for targeted efforts to improve customer experience and foster long-term loyalty.

References

[1] J. Paliwal, "Customer Feedback and Satisfaction," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/jahnavipaliwal/customer-feedback-and-satisfaction. [Accessed: Nov. 9, 2024].

[2] T. M. Cheng, P. Wu, and M. S. Rajab, "Utilizing big data to improve customer satisfaction in service industries: A case study," *International Conference on Service-Oriented Computing (ICSOC)*, Springer, 2022.

[3] A. Gupta, S. Chaudhary, and R. Sharma, "Enhancing customer experience through data mining and analysis of customer feedback," *Proceedings of the International Conference on Computational Data and Social Networks (CSoNet)*, Springer, 2021.

[4] D. Yang, B. Li, and Y. Chen, "Big data analytics for customer feedback: A survey," *IEEE Access, IEEE BigData Conference*, 2020.