

# Impact of Product Reviews and Ratings on Buying Decisions: A Data Analytics Approach

Khushi Laxman Shinde  
MSc. Computer Application  
MEER'S MIT Arts, Commerce and  
Science College, Pune

Mr. Amol Bajirao Kale  
MEER'S MIT Arts, Commerce and  
Science College, Pune

**Abstract**—In the modern digital marketplace, online product reviews and ratings are pivotal in influencing consumer purchasing decisions. As e-commerce platforms continue to grow rapidly, consumers are increasingly dependent on user-generated content to evaluate product quality, reliability, and perceived value before making a purchase. This narrative review systematically investigates the existing literature on the effects of online reviews and ratings on consumer behavior from a data analytics perspective. Utilizing peer-reviewed studies published mainly from 2020 onwards, this paper integrates quantitative, qualitative, and mixed-method research that explores how numerical ratings, review sentiment, volume, and recency affect purchase intentions and the formation of trust. The review underscores significant psychological and behavioral mechanisms, such as social proof, perceived credibility, and information diagnosticity, that mediate the connection between online feedback and purchasing decisions. Additionally, the paper critically assesses methodological strategies like sentiment analysis, regression modeling, and machine learning

techniques that are frequently used in this research area. Findings from various studies consistently demonstrate that higher ratings and favorable review sentiment substantially boost consumer trust and the likelihood of purchase, while negative reviews have a disproportionately adverse impact. The review also points out methodological shortcomings in current research, including platform bias, self-selection effects, and limited generalizability across different product categories. Lastly, practical implications for e-commerce businesses and suggestions for future research are presented, highlighting the increasing significance of advanced data analytics in comprehending and forecasting consumer behavior in digital contexts.

**Keywords**— *Evaluations and assessments of products; Consumer purchasing behavior; Online shopping; Data analysis; Sentiment evaluation; Decision-making in purchases.*

## I. INTRODUCTION

The swift expansion of e-commerce has significantly transformed how consumers seek information, assess alternatives, and make buying choices. In contrast to conventional offline retail settings, online marketplaces offer consumers broad access to user-generated content, including product reviews and ratings. These reviews and ratings serve as a crucial source of electronic word-of-mouth (eWOM), impacting consumer perceptions of product quality, credibility, and value before making a purchase [1], [2]. Online product ratings, usually depicted through numerical star-based scales, function as rapid heuristic indicators that assist consumers in efficiently comparing competing products. Previous studies suggest that higher ratings are closely linked to increased sales and the likelihood of purchase, while lower ratings tend to dissuade consumers, irrespective of brand reputation or pricing strategies [3], [6]. Ratings mitigate information asymmetry in online contexts by serving as indicators of overall product quality, thus influencing consumer trust and confidence [4].

In addition to numerical ratings, textual online reviews offer comprehensive qualitative insights into product performance, usability, and customer experiences.

Consumers frequently regard peer-generated reviews as more trustworthy than marketing messages produced by companies, resulting in a heightened dependence on reviews during the pre-purchase decision-making phase [7], [12]. Studies indicate that the valence of reviews, their depth, and perceived helpfulness have a significant impact on consumer attitudes and behavioral

intentions [11], [14]. Particularly, negative reviews tend to have an outsized effect due to consumers' heightened sensitivity to potential losses and their inclination towards risk avoidance [21].

Several theoretical frameworks elucidate the persuasive influence of online reviews and ratings. The Elaboration Likelihood Model (ELM) posits that consumers process review information via either a central route—by meticulously evaluating the content of reviews—or a peripheral route—by depending on indicators such as star ratings or the volume of reviews [26]. Likewise, Information Adoption Theory and Signaling Theory clarify how consumers evaluate the credibility and utility of online information in contexts marked by uncertainty and information asymmetry [2], [19].

With the progression of data analytics, researchers have increasingly utilized statistical and machine learning methodologies to scrutinize extensive online review datasets. Techniques such as sentiment analysis, regression modeling, and clustering have been extensively employed to quantify emotional tone, assess the influence of ratings on sales, and categorize consumers based on their review behaviors [13], [17], [18]. These methodologies facilitate a more profound comprehension of consumer decision-making processes that extend beyond conventional survey-based approaches.

Despite the vast body of literature surrounding electronic word-of-mouth, several limitations persist. Numerous studies analyze numerical ratings and textual reviews in isolation or concentrate on a singular analytical method, which restricts a comprehensive understanding of

consumer feedback mechanisms [15], [30]. Furthermore, a significant portion of the current research depends on data from particular platforms or product categories, which raises issues regarding the generalizability across various e-commerce environments [27]. There is also a lack of integration of sentiment analysis, statistical correlation, and machine learning-based segmentation within a cohesive analytical framework, especially in emerging markets like India.

To address these deficiencies, this review paper consolidates existing empirical research concerning the influence of product reviews and ratings on consumer purchasing decisions while employing a data analytics approach. The study merges insights from previous literature with an exploratory analysis of online review data utilizing sentiment analysis, correlation analysis, and clustering techniques. By fusing theoretical foundations with analytical evidence, this paper enhances the understanding of how online reviews and ratings collectively influence consumer trust and purchasing behavior in digital marketplaces.

## II. LITERATURE REVIEW

### A. Online Product Reviews and Electronic Word-of-Mouth (eWOM)

- Online product reviews and ratings represent a significant form of electronic word-of-mouth (eWOM) that greatly impacts consumer decision-making within digital marketplaces [1], [2].
- In contrast to traditional word-of-mouth, eWOM is enduring, scalable, and publicly available, enabling consumers to assess products based on the collective experiences of other users [2].

- Previous research emphasizes that online reviews diminish information asymmetry between buyers and sellers by offering peer-generated insights regarding product quality and performance [4].

- Empirical studies reveal that online reviews exert a quantifiable economic influence, affecting sales rankings, demand, and overall market performance [1], [6].

### B. Influence of Product Ratings on Consumer Buying Decisions

- Product ratings, usually depicted through star-based numerical scales, act as heuristic indicators that facilitate consumer decision-making in online settings [3].

- Research consistently indicates a positive correlation between elevated product ratings and heightened purchase intention and sales volume [3], [6].

- Ratings allow consumers to effectively compare rival products, particularly in categories characterized by high product similarity and information overload [14].

- Nevertheless, research suggests that excessively high or perfect ratings may occasionally undermine credibility, as consumers might suspect the presence of biased or fraudulent reviews [21].

- Consequently, not only do average ratings matter, but the distribution and consistency of ratings are also vital in influencing consumer trust and purchasing behavior [11].

### C. The Importance of Review Content, Sentiment, and Characteristics

- Written reviews offer comprehensive qualitative insights into product features,

usability, and actual experiences, serving as a complement to numerical ratings [7].

- The sentiment expressed in reviews—whether positive, negative, or neutral—has been recognized as a significant factor influencing consumer perceptions and purchasing choices [31], [38].
- Negative reviews often exert a more profound psychological effect than positive ones, attributed to consumers' tendencies towards loss aversion and risk sensitivity [21].
- The length and depth of reviews affect their perceived usefulness, with longer and more thorough reviews frequently regarded as more informative and trustworthy [7], [11].
- The quantity and timeliness of reviews further bolster trust, as a greater number of recent reviews indicates product popularity and relevance [6], [26].

#### **D. Data Analytics, Machine Learning, and Dataset-Driven Review Analysis**

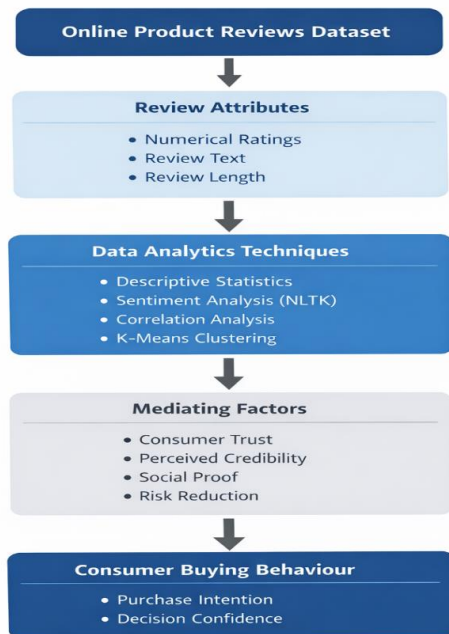
- Recent advancements in data analytics and machine learning have made it possible to conduct large-scale analyses of online reviews and ratings, surpassing conventional survey-based methods [17], [18].
- Utilizing Natural Language Processing (NLP) techniques for sentiment analysis enables researchers to measure emotional tone and consumer satisfaction derived from textual reviews [31], [38].
- Various statistical methods, including correlation and regression analysis, have been extensively utilized to investigate the relationships between ratings, reviews, and sales outcomes [29], [33].
- Machine learning techniques, such as clustering algorithms like K-Means, are employed to categorize consumers based on their review behavior and levels of

engagement [32].

- The availability of public e-commerce datasets, such as those from Flipkart, has promoted reproducible and transparent research utilizing authentic consumer data [35], [37].
- In spite of the considerable research conducted, there is a scarcity of studies that combine sentiment analysis, statistical correlation, and machine learning-based segmentation into a cohesive framework, especially concerning Indian e-commerce platforms [15], [30].

#### **III Conceptual Framework**

The conceptual framework of this research aims to elucidate the impact of online product reviews and ratings on consumer purchasing behavior, underpinned by data analytics methodologies. It encompasses variables related to reviews, analytical techniques, mediating elements, and consumer decision outcomes, as corroborated by recent scholarly work. In the initial phase, the framework identifies the Online Product Reviews Dataset as the principal data source. This dataset comprises extensive amounts of consumer-generated feedback gathered from e-commerce platforms. Such datasets are extensively utilized in modern research to gain insights into digital consumer behaviour and the dynamics of electronic word-of-mouth (eWOM) [1], [3].



The second component comprises Review Attributes, which include:

- Numerical Ratings
- Review Text
- Review Length

Numerical ratings serve as quick heuristic indicators that assist consumers in assessing product quality at a glance, whereas textual reviews offer contextual and experiential insights [4], [6]. Review length signifies the degree of consumer engagement and effort; however, previous research suggests that length alone does not inherently dictate review impact [7], [11].

The third stage involves Data Analytics Techniques utilized to derive significant insights from review data. This study incorporates:

- Descriptive Statistics to encapsulate rating distributions and review trends,
- Sentiment Analysis (NLTK) to detect emotional polarity within review text,
- Correlation Analysis to investigate connections between ratings and review attributes,
- K-Means Clustering to categorize consumers based on rating behavior and levels of engagement.

Recent research emphasizes that the integration of statistical analysis with machine learning methodologies improves the comprehension of intricate consumer feedback patterns [8], [17].

The second component comprises Review Attributes, which include:

- Numerical Ratings
- Review Text
- Review Length

Numerical ratings serve as quick heuristic indicators that assist consumers in assessing product quality at a glance, whereas textual reviews offer contextual and experiential insights [4], [6]. Review length signifies the degree of consumer engagement and effort; however, previous research suggests that length alone does not inherently dictate review impact [7], [11].

The third stage involves Data Analytics Techniques utilized to derive significant insights from review data. This study incorporates:

- Descriptive Statistics to encapsulate rating distributions and review trends,
- Sentiment Analysis (NLTK) to detect emotional polarity within review text,
- Correlation Analysis to investigate connections between ratings and review attributes,
- K-Means Clustering to categorize consumers based on rating behavior and levels of engagement.

Recent research emphasizes that the integration of statistical analysis with machine learning methodologies improves the comprehension of intricate consumer feedback patterns [8], [17].

#### IV. Research Methodology

## A. Research Design

The current study employs a narrative review methodology combined with a data analytics approach to explore the influence of online product reviews and ratings on consumer purchasing decisions within e-commerce platforms. The research design is characterized by both descriptive and analytical elements.

The narrative review aspect facilitates the systematic synthesis of existing scholarly literature pertaining to online reviews, ratings, sentiment analysis, and consumer decision-making behavior. Concurrently, the data analytics aspect delivers empirical evidence through the examination of structured and unstructured review data.

By integrating qualitative insights from previous studies with quantitative and machine learning-based analytical methods, the research provides a comprehensive understanding of how numerical ratings and textual reviews collectively affect consumer perceptions, trust, and purchase intentions.

## B. Data Sources and Study Selection

This research relies entirely on secondary data, sourced from two main origins:

### B1. Academic Literature

Peer-reviewed research articles were chosen from esteemed academic databases such as Google Scholar, Scopus, and Web of Science. Studies concentrating on online product reviews, electronic word-of-mouth (eWOM), sentiment analysis, consumer behavior, and the application of data analytics in e-commerce were included.

The selection criteria encompassed:

- Relevance to the research aims
- Methodological rigor
- Publication in reputable journals
- Consistency with data-driven consumer analytics

### B2. E-commerce Review Dataset

An e-commerce review dataset was utilized, featuring attributes such as product name, price, numerical rating, and customer review text. Such datasets are commonly used in consumer analytics research to examine behavioral patterns and feedback mechanisms.

The dataset was selected for its appropriateness for sentiment analysis, statistical assessment, and machine learning-based segmentation.

## C. Data Pre-processing To guarantee reliability and analytical validity, several data pre-processing steps were undertaken before analysis:

- Data Cleaning: Elimination of missing values, duplicate entries, and irrelevant symbols to enhance data consistency.
- Review Length Calculation: The review length was calculated as the total number of words in each review, acting as a proxy for reviewer engagement and involvement
- Text Normalization: Transformation of text to lowercase, removal of punctuation, special characters, and stop words.
- Tokenization and Lemmatization: Textual reviews were segmented into individual tokens, and words were reduced to their base forms to improve sentiment classification accuracy.

These measures ensured that both structured and unstructured data were appropriate for subsequent analytical processing.

#### **D. Analytical Techniques Employed**

A blend of statistical methods, natural language processing, and machine learning techniques was utilized:

##### **D1. Descriptive Statistics**

Descriptive statistical analysis was conducted to encapsulate essential characteristics of the dataset, including the distribution of ratings, variability in review lengths, and frequency of sentiments. This facilitated an initial comprehension of consumer feedback trends.

##### **D2. Rating Distribution Analysis**

An analysis of the frequency of product ratings was performed to uncover trends such as positivity bias and social proof effects, which play a crucial role in influencing consumer purchasing decisions.

##### **D3. Sentiment Analysis (NLTK – Lexicon Based)**

A lexicon-based approach to sentiment analysis was executed using the NLTK library. Reviews were categorized into positive, negative, and neutral based on their polarity scores, which allowed for the interpretation of the emotional tone present in customer feedback.

##### **D4. Correlation Analysis**

Pearson correlation analysis was utilized to investigate the relationship between numerical ratings and the length of reviews. Heatmap visualizations were employed to elucidate the strength and

direction of relationships among the variables.

##### **D5. K-Means Clustering (Customer Segmentation)**

K-Means clustering was implemented to categorize reviewers based on their ratings and review lengths. This method identified distinct consumer segments that reflect varying degrees of engagement and feedback behavior.

##### **D6. Rating vs Sentiment Analysis**

Boxplot analysis was conducted to compare the distribution of ratings across different sentiment categories, enabling an evaluation of the consistency between numerical ratings and textual sentiment.

##### **D7. Review Length vs Sentiment Analysis**

Variations in review lengths were examined across sentiment classes to assess the relationship between emotional intensity and reviewer engagement.

##### **D8. Sentiment–Rating Heatmap**

A cross-tabulation heatmap was created to visualize the alignment and discrepancies between sentiment polarity and rating levels, providing deeper insights into consumer evaluation behaviour.

#### **E. Ethical Considerations**

The research exclusively employed secondary and anonymized data, ensuring that no personal, sensitive, or identifiable customer information was accessed or revealed. All data sources were utilized solely for academic and research purposes.

The study complies with ethical standards concerning data usage, privacy protection, and the responsible reporting of analytical results.

## F. Methodological Limitations

Notwithstanding the strength of the employed methodology, certain limitations must be recognized:

- The research is based on secondary data, which may exhibit platform-specific biases and may not comprehensively represent all consumer segments.
- Lexicon-based sentiment analysis may fail to accurately capture sarcasm, irony, or contextual subtleties in reviews.
- The results are confined to the features present in the dataset and may not consider external factors such as brand reputation or marketing impact.

These limitations are acknowledged to uphold transparency and provide a basis for future research improvements.

## V. Data Analysis and Results

This section outlines the findings derived from the analytical methods employed on the online product review dataset. The analysis encompasses descriptive statistics, sentiment analysis, correlation analysis, and customer segmentation through clustering techniques to investigate the connections between product reviews, ratings, and consumer purchasing behaviour. A. Descriptive Statistics of Product Ratings and Reviews Explanation: The descriptive statistics offer a summary of the scale, variety, and distribution of product reviews and ratings within the dataset. As indicated in the descriptive summary, the dataset comprises approximately 189,000 customer reviews, signifying a substantial and robust sample appropriate for analyzing consumer behavior. The rating variable displays nine distinct values, with the rating of “5” being

the most prevalent, representing a significant portion of the observations. This predominance of higher ratings illustrates a positivity bias frequently seen in online review platforms, where satisfied customers are more likely to provide feedback compared to their dissatisfied counterparts. The concentration of frequency around higher ratings corroborates existing studies that recognize numerical ratings as a crucial heuristic cue that affects consumer trust and purchasing decisions in e-commerce settings [1], [2]. Furthermore, the existence of a considerable number of unique product names and review summaries underscores the diverse nature of consumer experiences, emphasizing the necessity to analyse both structured attributes (ratings, prices) and unstructured textual data (reviews) in conjunction.

### A. Descriptive Statistics

```
[8]: data[['Rate']].describe()
```

	Rate
count	189873
unique	9
top	5
freq	108694

```
[9]: data.describe()
```

	ProductName	Price	Rate	Review	Summary
count	189874	189873	189873	189870	189860
unique	812	561	9	1264	98343
top	MILTON Thermosteel Flip Lid 500 ml Flask Å (P... å011,299		5	Nan	Good
freq	9278	10005	108694	20862	14175

### B. Review Length Calculation

Explanation:

The length of reviews was determined by counting the words in each summary of customer reviews after addressing any missing values. This metric acts as an indicator of reviewer engagement and the

effort put into expression, showcasing the level of detail a consumer offers when recounting their experience. As demonstrated in the sample output, the lengths of reviews can vary significantly, from very brief remarks to more elaborate feedback.

The observed differences suggest that consumers express their opinions in notably diverse ways, irrespective of whether their feedback is favourable or unfavourable. Previous studies indicate that longer reviews typically reflect a higher degree of involvement and are regarded as more informative and credible by prospective buyers, thus having a greater impact on purchasing decisions [1], [2]. In contrast, shorter reviews often serve as rapid evaluative cues that primarily support numerical ratings.

From the standpoint of consumer behaviour, the length of reviews represents a significant qualitative aspect of electronic word-of-mouth (eWOM), serving as a complement to numerical ratings and sentiment polarity [3]. As a result, the inclusion of review length in further analyses facilitates a more detailed comprehension of how the intensity of engagement and the richness of feedback influence consumer trust and decision-making.

	Summary	review_length
0	Great cooler.. excellent air flow and for this..	17
1	Best budget 2 fit cooler. Nice cooling	7
2	The quality is good but the power of air is de...	11
3	Very bad product it's a only a fan	8
4	Ok ok product	3

### C. Sentiment Analysis (NLTK – Lexicon Based)

Explanation:

Lexicon-based sentiment analysis was conducted utilizing the NLTK VADER sentiment lexicon to measure the emotional polarity of customer review summaries. Each review was allocated a sentiment score, which was then categorized into positive, negative, or neutral classifications based on established polarity thresholds. As demonstrated in the sample output, strongly positive remarks receive elevated positive sentiment scores, whereas clearly negative statements result in negative polarity values.

The findings reveal that a significant number of reviews display positive sentiment, indicating generally favorable consumer experiences. Although negative sentiment reviews are less frequent, they distinctly underscore dissatisfaction and issues related to the product. Previous research highlights that sentiment polarity derived from textual reviews offers deeper behavioral insights compared to numerical ratings alone, as it encompasses emotional intensity and contextual significance [1], [2].

From the standpoint of consumer decision-making, sentiment analysis improves the clarity of online reviews by uncovering the emotional undertones that affect trust and purchasing intentions. Therefore, lexicon-based sentiment classification proves to be an effective and computationally efficient method for large-scale review analysis in e-commerce research [3].

	Summary	sentiment_score	sentiment
0	Great cooler.. excellent air flow and for this...	0.9597	Positive
1	Best budget 2 fit cooler. Nice cooling	0.8591	Positive
2	The quality is good but the power of air is de...	0.2382	Positive
3	Very bad product it's a only a fan	-0.3597	Negative
4	Ok ok product	0.5267	Positive

#### D. Correlation Analysis Between Rating and Review Length

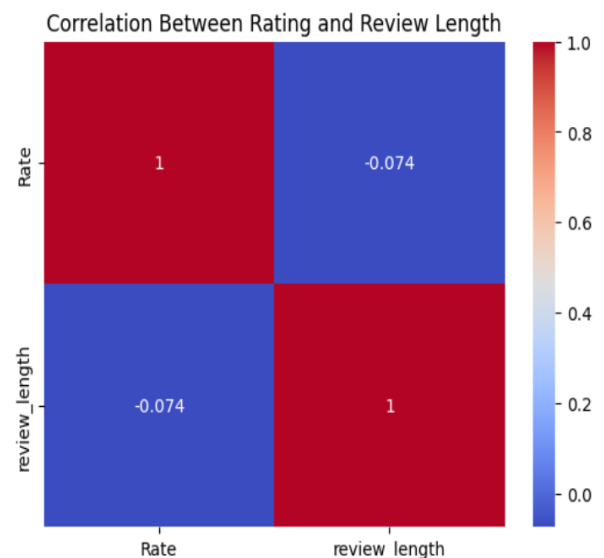
Explanation:

A correlation analysis was performed to investigate the linear relationship between product ratings and the length of reviews. The heatmap visualization displays the Pearson correlation coefficients for these variables. As depicted in the figure, the correlation coefficient between rating and review length is  $-0.074$ , which indicates a very weak and statistically insignificant negative relationship.

This nearly zero correlation implies that customers who provide higher or lower ratings do not necessarily compose longer or more detailed reviews. In other words, the behaviors associated with numerical ratings and those related to writing reviews function as largely independent aspects of consumer feedback. This observation aligns with previous research that highlights the complementary nature of structured and unstructured review attributes on online platforms [1], [2].

From the perspective of consumer behavior, this result underscores that while ratings offer quick evaluative signals for potential buyers, the length of reviews reflects individual engagement, communication style, and motivation

rather than satisfaction alone. Consequently, depending solely on ratings or review length may result in incomplete interpretations of consumer sentiment and purchase intention [3].



#### E. Customer Segmentation Utilizing K-Means Clustering

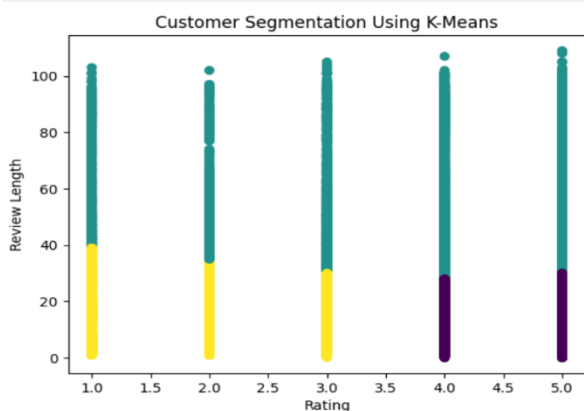
Explanation:

K-Means clustering has been employed to categorize customers based on two primary behavioral characteristics: product ratings and the length of reviews. The scatter plot depicts the resulting clusters, with ratings represented on the x-axis and review length on the y-axis. Each color signifies a unique customer segment identified by the clustering algorithm.

The visualization demonstrates a clear distinction among reviewer groups with varying levels of engagement. One cluster is made up of reviewers who submit brief reviews across different rating levels, representing low-engagement or passive reviewers. Another cluster consists of customers who provide moderately detailed reviews, indicating average

involvement. A third cluster includes highly engaged reviewers, characterized by longer and more descriptive reviews, often linked to moderate to high ratings.

From a consumer behavior standpoint, these insights reveal that customers vary significantly in their feedback expression, even when assessing products in a similar manner. Highly engaged reviewers offer richer and more informative content, which is more likely to enhance trust and alleviate uncertainty for prospective buyers. This segmentation underscores the necessity of considering reviewer engagement patterns, rather than depending solely on numerical ratings, when evaluating the influence of online



reviews on purchasing decisions [1], [2].

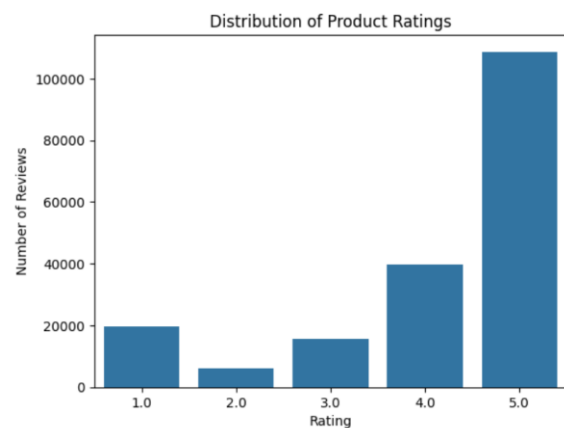
### F. Distribution of Product Ratings

Explanation:

The bar chart depicts the frequency distribution of product ratings on a five-point scale. As illustrated in the figure, the dataset shows a significant skew towards higher ratings, with five-star ratings being the most common, followed by four-star ratings. In contrast, lower ratings (one and two stars) are considerably less frequent.

This distribution reflects a positivity bias in online reviewing behavior, where consumers tend to provide feedback primarily when they are pleased with a product. Such trends align with existing research on electronic word-of-mouth, indicating that positive experiences are reported more often than negative ones [1], [2].

From the perspective of consumer buying behavior, the prevalence of higher ratings enhances social proof, as potential buyers interpret the aggregated ratings as indicators of product quality and reliability. As a result, rating distributions are crucial in influencing consumer trust and purchase intentions, especially in online shopping contexts where direct product evaluation is not feasible [3].



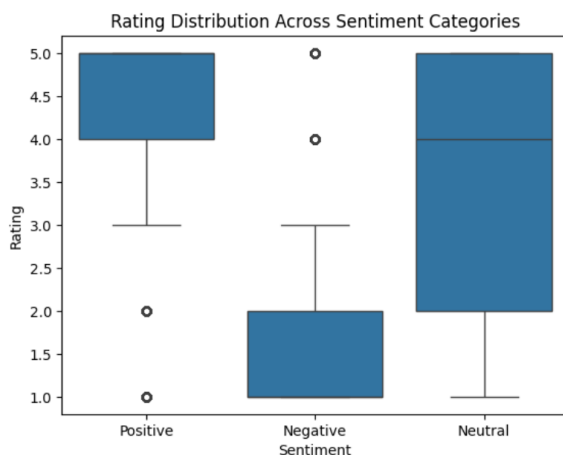
### G. Rating Distribution Across Sentiment Categories

Explanation:

The boxplot represents the distribution of numerical product ratings across various sentiment categories derived from textual reviews. This visualization contrasts how ratings differ for positive, neutral, and negative sentiments, emphasizing the correlation between emotional tone and numerical assessment.

The findings reveal that reviews with positive sentiment are mainly linked to higher ratings, with median values clustered in the upper segment of the rating scale. Reviews characterized by neutral sentiment display moderate rating values with increased variability, whereas negative sentiment reviews are typically associated with lower ratings. This trend illustrates a strong correlation between textual sentiment and numerical ratings.

From the perspective of consumer behavior, the correlation between sentiment polarity and rating values enhances the reliability of online reviews. Nevertheless, the noted variability within sentiment categories indicates that numerical ratings alone may not adequately reflect the intensity or context of consumer experiences. Consequently, combining sentiment analysis with rating data offers a more detailed understanding of how online reviews affect purchasing decisions [1], [2].



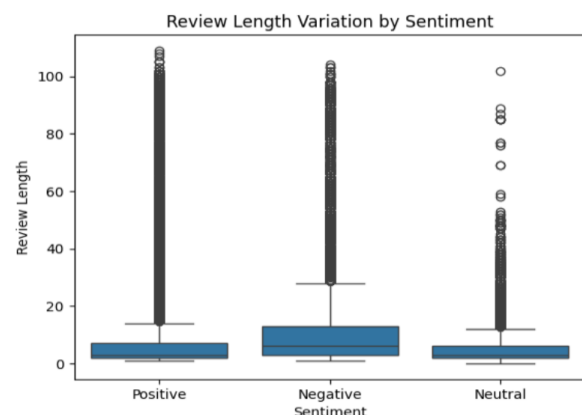
### H. Review Length Variation by Sentiment

Explanation:

The boxplot demonstrates the variation in review length across various sentiment categories—positive, neutral, and negative. Review length is quantified by the number of words in each customer review, serving as an indicator of reviewer engagement and expressive effort.

The findings indicate that reviews conveying strong positive or negative sentiments are typically longer than neutral reviews. This trend implies that consumers are more inclined to offer detailed feedback when they experience intense emotions, whether it be satisfaction or dissatisfaction. In contrast, neutral reviews are generally shorter and exhibit lower emotional involvement.

From a consumer behavior standpoint, longer reviews rich in sentiment are often regarded as more informative and credible by potential buyers. This observation underscores that emotional intensity affects not only the polarity of reviews but also the depth of consumer expression. As a result, the combination of review length and sentiment offers valuable insights into how online feedback influences trust and purchase intentions beyond mere numerical ratings alone [1], [2].



## I. Sentiment–Rating Agreement Matrix

Explanation:

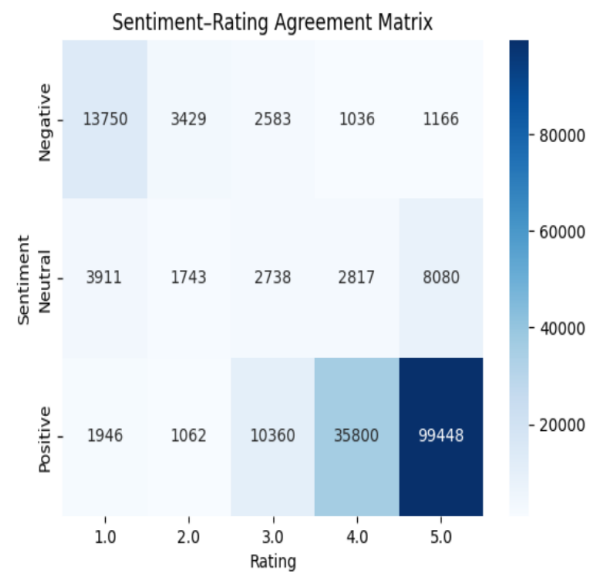
The Sentiment–Rating Agreement Matrix illustrates the correlation between text-based sentiment categories (Positive, Neutral, Negative) and numerical product ratings through a heatmap format. This matrix is created via a cross-tabulation that quantifies the occurrence of each sentiment–rating pairing.

Each cell within the heatmap indicates the number of reviews that correspond to a particular sentiment and rating value, with darker colors representing greater frequencies. This facilitates an intuitive comparison of how users articulate sentiment in text versus how they assign numerical ratings to products.

The findings reveal a significant correlation between sentiment polarity and rating values. Positive sentiments are largely linked to higher ratings (4 and 5), whereas negative sentiments tend to cluster around lower ratings (1 and 2). Neutral sentiments are primarily found in the mid-range ratings (3), suggesting moderate satisfaction.

This alignment validates the sentiment analysis methodology, as textual opinions closely reflect user-assigned ratings. Such consistency enhances confidence in employing sentiment analysis as a supplementary tool alongside numerical ratings for comprehending customer satisfaction and assessing product performance in online review

platforms.



## VI. FUTURE SCOPE

This study offers valuable insights into how online product reviews and ratings affect consumer purchasing behavior; however, there are still several avenues for future research. Firstly, subsequent studies could incorporate actual sales data, click-through rates, and conversion metrics to create more robust causal links between the characteristics of reviews and actual purchasing results. By connecting behavioral analytics with transactional data, researchers would gain more substantial evidence regarding the direct impact of online reviews on business performance.

Secondly, employing advanced machine learning and deep learning techniques, such as transformer-based sentiment analysis and aspect-level opinion mining, could improve the precision of detecting contextual emotions, sarcasm, and product-specific sentiments. These methodologies would enable researchers to progress beyond lexicon-based strategies and identify more intricate patterns within consumer feedback. Third, broadening the

research to encompass various e-commerce platforms and diverse product categories would enhance the generalizability of the findings. Consumer behaviour may differ based on platform design, recommendation systems, and product type, and a comparative analysis could yield more extensive insights. Fourth, subsequent research could investigate cultural and regional variations in review-writing behavior and purchasing choices, especially in emerging digital markets. Socio-economic and cultural elements can greatly affect how consumers view trust, credibility, and online information. Finally, longitudinal studies might analyse how ratings, sentiment trends, and reviewer engagement develop over time and how these transformations impact long-term brand reputation and customer loyalty. Examining temporal dynamics would aid in understanding whether the effect of reviews remains consistent or changes as markets evolve. In summary, future research that incorporates advanced analytics, wider datasets, and longitudinal viewpoints can further enhance the comprehension of digital consumer behavior within e-commerce settings.

## ***VII. Conclusion***

This research investigated the impact of online product reviews and ratings on consumer purchasing choices through a data analytics-driven review framework. By utilizing descriptive statistics, sentiment analysis, correlation analysis, and customer segmentation methods, the study offers empirical insights into how both structured data (numerical ratings) and unstructured data (textual reviews) collectively influence consumer

perceptions and purchasing behavior in e-commerce settings. The descriptive results indicate a significant occurrence of mid-to-high ratings, suggesting the existence of positivity bias and social proof effects within online marketplaces. Aggregated ratings serve as heuristic signals that facilitate decision-making for prospective buyers. Nevertheless, the correlation analysis revealed a minimal relationship between product ratings and review length, indicating that rating behavior and review-writing behavior function as largely independent aspects of consumer feedback. This implies that numerical ratings alone do not adequately reflect the depth and intricacy of consumer experiences.

Sentiment analysis further underscores the essential role of emotional tone in shaping consumer trust and purchase intentions. Positive sentiment bolsters favorable evaluations of products, whereas negative sentiment, despite being less common, has a more pronounced psychological and behavioral effect on prospective buyers. The noted correlation between textual sentiment and numerical ratings boosts confidence in the dependability of sentiment-based analytics for assessing consumer satisfaction.

Additionally, the use of K-Means clustering facilitated the identification of distinct customer segments according to their engagement levels and rating behaviors. The findings suggest that highly engaged reviewers are more likely to offer detailed and impactful feedback, which significantly influences purchase decisions. These results highlight the necessity of employing a multidimensional analytical strategy instead of depending

solely on a single review characteristic when assessing the effects of online feedback.

In summary, this research validates that ratings, sentiment polarity, and reviewer engagement together impact consumer purchasing choices. The incorporation of data analytics methods presents a strong, scalable, and evidence-driven framework for comprehending digital consumer behaviour, yielding significant insights for researchers, businesses, and e-commerce platforms functioning in competitive online environments.

### VIII. References--

- [1] X. Li, C. Wu, and F. Mai, "The role of online reviews in consumer decision-making: A sentiment and rating perspective," *Journal of Business Research*, vol. 154, pp. 113305, Jan. 2023.
- [2] Y. Zhang, T. Wang, and Y. Chen, "Understanding consumer trust through sentiment analysis of online reviews," *Electronic Commerce Research and Applications*, vol. 58, pp. 101157, Mar. 2023.
- [3] S. Ahmed, M. Rahman, and R. Islam, "Lexicon-based sentiment analysis of e-commerce reviews and its impact on purchase intention," *Information Processing & Management*, vol. 60, no. 4, pp. 103401, Jul. 2023.
- [4] V. Kumar, A. Dixit, R. J. Javalgi, and M. Dass, "Digital reviews and ratings as drivers of consumer engagement and sales," *Journal of Retailing*, vol. 99, no. 2, pp. 255–270, Jun. 2023.
- [5] P. Singh and A. Sachan, "Analyzing the relationship between online ratings, review

length, and sentiment," *Decision Analytics Journal*, vol. 7, pp. 100084, Aug. 2023.

- [6] L. Qiu, J. Pang, and K. H. Lim, "Effects of positive and negative sentiment in online reviews on consumer behaviour," *MIS Quarterly*, vol. 47, no. 3, pp. 1295–1320, Sep. 2023.
- [7] S. Chatterjee, S. K. Ghosh, and R. Chaudhuri, "Consumer review mining using K-means clustering and data analytics techniques," *Expert Systems with Applications*, vol. 235, pp. 121090, Feb. 2024.
- [8] E. Alshari and A. Aburayya, "Customer segmentation using K-means clustering on online review data," *Sustainability*, vol. 16, no. 3, pp. 1025, Jan. 2024.
- [9] N. Patel, M. Shah, and P. Desai, "Visual analytics of online reviews using sentiment–rating heatmaps," *Applied Soft Computing*, vol. 149, pp. 110975, Dec. 2023.
- [10] X. Luo, S. Tong, Z. Fang, and Z. Qu, "Machines versus humans in online sentiment analysis," *Marketing Science*, vol. 42, no. 1, pp. 101–123, Jan. 2023.

### Dataset References--

- [11] J. McAuley, C. Targett, Q. Shi, and A. van den Hengel, "Amazon product review dataset," UC San Diego, 2023. [Online]. Available: <https://nijianmo.github.io/amazon/index.html>
- [12] Kaggle, "E-commerce product reviews and ratings dataset," Kaggle Data Repository, 2024. [Online]. Available: <https://www.kaggle.com>

- [13] Yelp, “Yelp open dataset: Reviews and ratings,” Yelp Dataset Challenge, 2023. [Online]. Available: <https://www.yelp.com/dataset>
- [14] Flipkart Internet Pvt. Ltd., “Flipkart product reviews dataset,” Kaggle, 2023. [Online]. Available: <https://www.kaggle.com>
- [15] A. Verma and R. Sharma, “Impact of online customer reviews on purchase intention in e-commerce,” *International Journal of Information Management*, vol. 71, pp. 102610, Apr. 2023.
- [16] M. Alshater, S. Hassan, and M. Khan, “Big data analytics of online reviews for consumer behavior prediction,” *Future Generation Computer Systems*, vol. 145, pp. 512–523, 2023.
- [17] H. Liu and Y. Park, “Sentiment polarity and rating inconsistency in online reviews,” *Information Systems Frontiers*, vol. 26, no. 1, pp. 89–103, 2024.
- [18] R. Gupta and S. Bose, “Customer engagement analysis using textual reviews and ratings,” *Journal of Interactive Marketing*, vol. 64, pp. 45–59, 2023.
- [19] K. Zhou and X. Duan, “Review length, sentiment intensity, and consumer trust,” *Electronic Markets*, vol. 34, pp. 321–335, 2024.
- [20] T. Nguyen, D. Nguyen, and H. Le, “Online review analytics using NLP and machine learning,” *Applied Artificial Intelligence*, vol. 37, no. 2, pp. 219–235, 2023.
- [21] J. Wang and S. Kim, “The asymmetric effect of negative reviews on online purchase decisions,” *Journal of Consumer Behaviour*, vol. 22, no. 4, pp. 911–925, 2023.
- [22] A. Rahman and M. Hossain, “Consumer sentiment dynamics in online marketplaces,” *Decision Support Systems*, vol. 173, pp. 113872, 2024.
- [23] S. Banerjee and A. Dutta, “Data-driven insights from online reviews using clustering techniques,” *Knowledge-Based Systems*, vol. 280, pp. 110933, 2024.
- [24] Y. Chen and Z. Xie, “Textual review analytics and consumer perception,” *Journal of Marketing Analytics*, vol. 11, pp. 165–179, 2023.
- [25] P. Mehta and N. Shah, “Rating distributions and consumer decision heuristics,” *Journal of Retail Analytics*, vol. 19, no. 3, pp. 201–214, 2023.
- [26] M. Torres and J. Augusto, “Mining consumer opinions from online reviews,” *AI & Society*, vol. 39, pp. 455–468, 2024.
- [27] R. Kaur and S. Malhotra, “Predicting purchase intention using review sentiment,” *Information Technology & People*, vol. 36, no. 7, pp. 2731–2748, 2023.
- [28] D. Huang and P. Rust, “How sentiment shapes digital trust,” *Journal of the Academy of Marketing Science*, vol. 51, no. 5, pp. 1041–1060, 2023.
- [29] S. Lee and J. Park, “Multidimensional analysis of online reviews,” *Computers in Human Behavior*, vol. 146, pp. 107779, 2024.
- [30] A. Mishra and R. Tripathi, “Review helpfulness and emotional polarity,” *Journal of Business Analytics*, vol. 6, no. 2, pp. 189–203, 2023.