
Empowering Consumer Decision: Decoding Incentive vs. Organic Reviews for Smarter Choices Through Advanced Textual Analysis

[Kate Kargozari](#)*, [Junhua Ding](#), [Haihua Chen](#)

Posted Date: 12 February 2024

doi: 10.20944/preprints202402.0687.v1

Keywords: incentive, organic, online reviews, review quality, sentiment analysis, semantic links, A/B testing, decision-making, recommendation



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Empowering Consumer Decision: Decoding Incentive vs. Organic Reviews for Smarter Choices Through Advanced Textual Analysis

Kate Kargozari *, Junhua Ding and Haihua Chen

Department of Information Science, University of North Texas 76203, Denton, Texas, United States; junhua.ding@unt.edu (J.D.); haihua.chen@unt.edu (H.C.)

* Correspondence: katekargozari@my.unt.edu

Abstract: In recent years, online review system attracts attention to assess seller-customer relations in the world of e-commerce. To address the quality concern of online review, especially incentivized ones, this study evaluates credibility and consistency based on reviews' volume, length, and content to distinguish the impact of incentives on customer review behavior, improving review quality, and purchase decision-making. Software product reviews collected from software review websites, including Capterra, Software Advice, and GetApp undergo Exploratory Data Analysis (EDA) to reveal critical features such as cost, support, usability, and product features. The indirect impact of companies' size, direct impact of users' experience, and different impacts of changing situations during years on the incentive reviews' volume are major findings of sentiment analysis. A/B testing results show a minimal to no impact of such reviews on purchasing decisions, highlighting discrepancies in credibility and consistency in volume, length, and content. Employing advanced techniques like Sentence-BERT (SBERT) and TF-IDF, the study explores semantic differences in reviews to improve recommendation systems for a more customized shopping experience. This approach seeks to establish a framework that discerns between review types and their effect on customer behavior. The findings contribute to developing more sophisticated and consistent e-commerce solutions, emphasizing the importance of authentic and reliable online reviews in influencing consumer choices.

Keywords: incentive; organic; online reviews; review quality; sentiment analysis; semantic links; A/B testing; decision-making; recommendation

1. Introduction

Recently social media revolutionized e-commerce by transforming how to assess seller-customer relationships. Among the critical factors in online purchasing decisions, including electronic word-of-mouth (eWOM), the price, and the website/business's reputation, online reviews, a form of eWOM, are specially crucial for shoppers decision-making [1]. Reviews categorize based on criteria. One categorization refers to ranking products or services by customer review sentiment including positive, negative, and neutral. Another categorization is based on reviews' writing criteria such as experience and monetary rewards denotes reviews as organic (no incentive or non-incentivized), incentive, or fake. Despite organic/non-incentivized reviews [2–4] that are based on real experiences and free from external motivation or incentives, some individuals may be tempted with rewards to write either incentivized reviews reflecting their actual product purchases experience [2,3,5–9] or fake reviews [7,10,11] lacking experiential foundation.

Delivering accurate information through the online review system is vital for informed purchases and reducing bias in existing seller-provided descriptions [5]. Online review systems face the major challenge of obtaining truthful and high-quality responses from agents [12]. Factors such as social presence can mediate the relationship between online review language style and consumers' purchase intention [13]. Although businesses can save money while receiving organic reviews [1], many

customers ignore posting the reviews. A direct relationship between the number of reviews and sale [6] encourages sellers to offer monetary rewards for honest reviews to boost both review count and product rating [14,15] while reducing bias. Incentivized reviews affect customer satisfaction [2]. Incentives can impact consumers' expressions and increase positive emotions in reviews [6], which influence purchase intention, trust and satisfaction [1].

However, the sensitivity of offering incentives may have a positive or negative effect [3] with possible results of negative reviews being seen as more credible reviews [16]. High volume of online reviews for a product can cause confusion, misinformation, and misleading [7] purchase decision, which harming trust and truthfulness of the reviews. Sellers' guidance for high-quality reviews has gaps. Not all positive/negative reviews are accurate, and customer satisfaction does not always align with review sentiment. Thus, customer behavior [17–19] toward posting product purchase reviews influences review quality. Improving review quality [3] enhances trust, assisting purchase decision-making and facilitating valuable contributions to the new review process.

Review quality, assessed from the standpoint of star rating, aids in product comparison, while the evaluation of individual products is concurrently influenced by the textual components of the reviews [20].

This study aims to assess the impact of incentives on customers' posting review behavior and review quality by examining the difference between incentive and organic reviews. Furthermore, the study will utilize both existing evidence on review quality along with information gained from this research to propose novel approaches for enhancing review quality reflecting on the credibility [21–24] and consistency [25–31] of the reviews while considering the impact of customers' purchase review behavior. Credible online reviews positively impact the hedonic brand image [22]. Despite the alarming message by negative reviews, they are not inherently more credible than positive reviews [24]. Source credibility moderates the relationship between review comprehensiveness and review usefulness [21,23]. Consistency can impact credibility of the review as consistent review can be either high or low quality review [29]. Consistency in content negatively impacts informational influence [25] and review helpfulness [26,27], while positively affects online reviews credibility [30]. Depends on the study, review consistency may positively impact review usefulness [28] and brands attitudes [31]. In this study, we focus on assessing online reviews credibility and consistency based on their volume, length, and content. To achieve this goal, we design two research questions:

1. What are the significant differences between incentive and organic reviews?
2. How do the incentive and organic reviews impact on customers' behavior on posting purchase's review, and as a result, on purchase's review quality, with impact on purchase decision-making?

We performed Exploratory Data Analysis (EDA) on various data set features pertaining to the "incentivized" status. Additionally, we conducted EDA analysis on sentiment analysis of review text to distinguish between incentivized and organic reviews. Moreover, we propose a comprehensive analysis using advanced techniques like Sentence-BERT (SBERT) and Term Frequency-Inverse Document Frequency (TF-IDF). These methods are chosen for their proficiency in capturing the semantic differences and frequency-based importance of terms within the reviews. Furthermore, We applied A/B testing on review rating scores, and as the ultimate goal, examined the impact of incentives on customers' purchase decision-making. We hypothesize that a deeper understanding of these reviews can notably boost the effectiveness of recommendation systems, leading to a more adapted and customer-centric shopping experience. This research intends to establish a methodological framework that not only distinguishes between different types of reviews but also assesses their relative influence on customer behavior, a vital step towards developing more reliable e-commerce solutions.

To the best of our knowledge, previous studies have not discovered the effect of company size and years of user experience as contributing factors. Moreover, our analysis of software reviews yields valuable insights for enhancing purchase reviews and more specific software reviews that generates more targeted guidelines to enhance overall review quality.

The article's structure is as follows: Section II presents related work, follows by Methodology in Section III. We present our results in Section IV, and further discussion in Section V. The article concludes in Section VI.

2. Related Work

Previous investigations have shown that trust in online reviews is equivalent to trust in friends' recommendations [32]. Therefore, this study reviews the variance between organic and incentivized reviews, their impact on customer behavior, review quality, and customer decision-making.

2.1. Incentive vs. Organic

Online reviews impact purchase intention; therefore, the relationship between online review stimulus and purchase intention response is important to explore [1]. In the short term, reviewers' contribution and readability levels rise; however, over time, review quality improves, leading to stabilization of their numerical rating behaviors [8]. Study demonstrates that incentives can enhance review quality [9]. Furthermore, in accordance with the social exchange theory (SET), incentives may motivate social behavior by considering the satisfaction of individual needs, such as encouraging customers to write online reviews [9]. Disclosing incentives maintains trust, reduces bias, boosts helpfulness and increases sales [10]. The disclosure of intrinsic communication motives for writing product reviews is more authentic and less betraying [33]. However, the impact of disclosing statements on product quality judgment depends on either customers' disclosure is integral or incidental [34]. For companies, incentives increase attracting customers' attention [3] and products' rate, reduces products' return, and contributes to companies' success [9].

2.2. Incentive and Purchase Decision-Making

Purchase decision-making considers complex situation using utility-driven systems to ease purchase decision-making by providing more details [35]. Incentivized reviews boost the effectiveness of efficient review signals to new customers [36]. In addition, incentives make users more active [37] and increase review writers by combining with social norm [4], which makes review writing more enjoyable [6], and increase review numbers [9]. Accordingly, incentives increase the volume and length of online reviews [4,5]; and consequently increase the volume of provided information to new customers for better purchase decisions [4]. Moreover, according to loss aversion theory, review valence is more influential than review usefulness in the decision-making process [38]; therefore, incentive reviews impact purchase decision-making as they increase valence [9] by increasing emotional words in customers' WOMs [5,6]. Disclosing incentives is crucial to prevent accuracy decrease and new consumer decision misguidance [10].

On the other hand, existing studies have investigated the importance of avoiding incentives. Offering and accepting incentives can decrease trust as follows market norms, not social norms that is a sign of human behavioral issues, raises moral concerns, increases review fraud that undermines review credibility, and establishes the interest between businesses and review posters [11]. In addition, incentives increase biased positive reviews [4]. Despite differing views [4,5], incentives may reduce user effort to write lengthy informative reviews [39]. Moreover, customers who are uncomfortable receiving incentives for their opinion may deliver negative reviews [4], which are valuable [40].

2.3. Approaches of Identifying and Analyzing Incentive Reviews

Multiple studies explore the effect of monetary incentives on online reviews' quality and value using unique approaches.

Incentive reviews were identified using data mining techniques considering the overall rating, helpfulness rate, review length, and other factors. VADER algorithm was used to improve model by incorporating review sentiment scores [5]. A difference-in-differences analysis reevaluated reviewers' behavior [8]. Counterfactual thinking approach investigated incentives' effect on online review

publication likelihood and valence in two experimental studies using a scenario approach, replicating and extending results by considering customers' satisfaction levels [9]. Machine-learning-based and dictionary-based approaches assessed the impact of sending efficient signals by review [36]. To analyze incentives' effect on product attention, data manipulation and statistical tests such as p-values and t-tests are used [3]. The stimulus-organism-response (S-O-R) framework developed, and a component-based structural equation modeling method (Smart PLS) used to assess online reviews and purchase intention relationships [1]. Extensive quantitative methods alongside the mixed-method experimental studies were used to evaluate review valence's impact on decision-making [38]. Regression models commonly used to assess reviews' impact on decision-making [35], that often incorporating methods like sentiment analysis [10]. A multi-methodological research design of two randomized experiments was utilized to test incentives' impact on reviews volume and length, and the potential bias in purchase decision-making [4].

3. Materials and Methods

We introduce approaches and techniques for data collection and analysis in this section.

3.1. Data Collection

Data was collected from software review websites, including Capterra¹, Software Advice², and GetApp³, that contain user-revealed experiences. We gathered information from review sections, including "Personal Information", "Itemized Scores", "Review time & source", and "Review text", Figure 1.

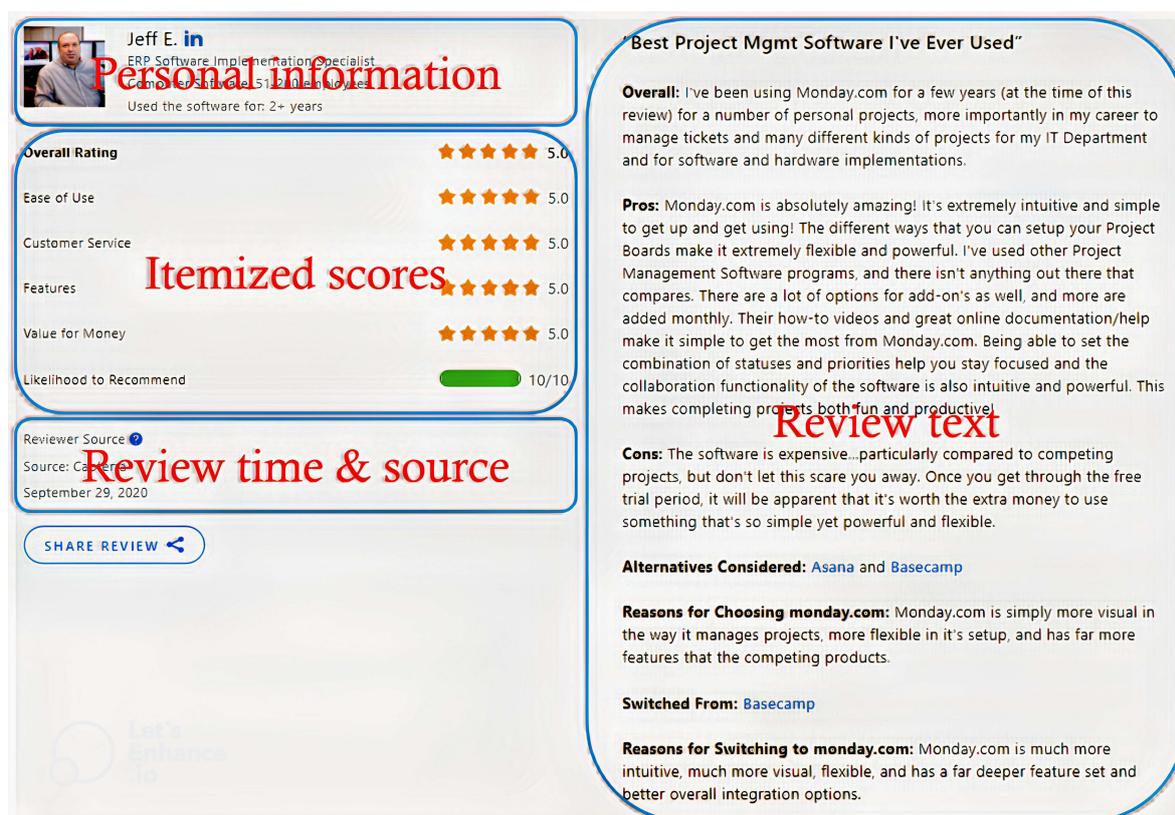


Figure 1. A glance of the data from CACOO reviews 2022 .

¹ <https://www.capterra.com/project-management-software>

² <https://www.softwareadvice.com/project-management>

³ <https://www.getapp.com/customer-management-software/crm>

“Personal information” may include the customer’s name, the name abbreviation or nickname, software use duration, and more. The main focus of this study centers around “Itemized scores”, “Review time & source”, and “Review text”. “Itemized scores”, which include Overall rating, ease of Use, features, value for money, and likelihood to recommend. “Review time & source” may include date and source; and the “Review text” mostly focuses on the review description, pros and cons. We scraped 1189 software product reviews from review websites using Python code, selenium, and BeautifulSoup. The collected review information includes title, description, pros and cons, ratings, and review details such as name, date, company, and prior product used. Overall 62,423 non-repetitive reviews were gathered and stored in a CSV file containing 43 attributes.

3.2. Data Pre-Processing

To pre-processed data for further analysis, we removed the “None” values from “incentivized” feature, leaving 49,998 instances in the dataset. We kept null values in other attributes to retain critical information. To binarize the “incentivized” feature, we categorized “NominalGift”, “VendorReferredIncentivized”, “NoIncentive”, “NonNominalGift”, and “VendorReferred”, into two groups. Respectively, the first two were classified as “Incentive” and the last three as “NoIncentive”. The new labels are stored in the column called “Incentivized”.

Following data pre-processing applied for sentiment analysis. Expanding contractions were used to replace the short versions of the words with their complete forms to ensure that each word is treated as separate tokens that can further be analyzed individually. Non-alphabetic or non-numeric characters, such as punctuation marks, were removed in the other pre-processing step. Lemmatization used to increase sentiment accuracy while decreasing text dimension by reducing words to their base form, known as dictionary form. Tokenization is used to break the text into words for accurate sentiment prediction. Stop words such as “a” and “an” removed to reduce computational resources, lower text dimensionality, and improve sentiment analysis accuracy.

The data pre-processing followed by EDA analysis, sentiment analysis, and A/B testing. Beside “incentivized” feature, this study considers other attributes such as “overAllRating”, “value_for_money”, “ease_of_use”, “features”, “customer_support”, “likelihood_to_recommend”, “year”, “company_size”, “time_used”, “preprocessed_pros”, “ReviewDescription_Sentiment”, “source”, “pros_Sentiment”, “preprocessed_cons”, “preprocessed_ReviewDescription”, “cons_Sentiment” and “Incentivized”.

3.3. Data Analysis

3.3.1. EDA Analysis

We conducted EDA analysis to extract information, understand dataset characteristics, and identify variables’ relationships. We leveraged the insights obtained from initial EDA analysis as guide for application of more advanced text analysis techniques.

3.3.2. Sentiment Analysis

Sentiment analysis was used to extract people’s opinions [41] and compare emotional tones of incentive and organic reviews.

We used the HuggingFaceTransformers⁴ for sentiment analysis. The model limitation of 200-char prevented us from combining all review texts (review description, pros, and cons). Therefore, to determine overall sentiment, We individually analyzed the sentiment of review texts and stored the results in the dataset as “ReviewDescription_Sentiment”, “pros_Sentiment”, and “cons_Sentiment”.

⁴ <https://github.com/huggingface/transformers>

As reviews rating scores reflect customers' satisfaction, Spearman's correlation coefficient is used to measure the correlation between incentive and organic review rating scores based on sentiment, considering the review description, "incentivized", and sentiment status. To ensure accuracy, we analyzed sample of 4,000 reviews per review category and determined the 95% confidence interval (CI) using the z-test.

3.3.3. Semantic Links

Semantic links were utilized as a method to uncover the conceptual connections between incentivized and organic reviews, thereby clarifying the relationships between these two categories of reviews was performed.

To extract deeper insights from the review text, we employed the TF-IDF (Term Frequency-Inverse Document Frequency) technique. This statistical measure assesses the significance of a word or phrase in a document relative to a corpus of documents. It identifies the most relevant words or phrases in the document by comparing their frequencies in that document against their frequencies across the entire corpus.

We randomly selected the 15000 reviews from each "Incentive" and "NoIncentive" category. The feature extraction process was applied to the "preprocessed_CombinedString" to generate trigrams for each review. These trigrams were found to be more meaningful than bigram. This feature analysis determined the frequency of each trigram's occurrence in each set of reviews. The TF-IDF scores (frequencies) exposed information of how important a trigram phrase is to a document in a corpus. To identify the top trigrams prevalent in both review categories, we compared and combined the frequency outcomes of trigrams between two categories and sorted them from the highest overall score to lowest. We used the results of feature extraction to calculate the cosine similarity with range from -1 (diametrically opposed vectors, dissimilar) to 1 (identical vectors), with 0 (no similar vectors) in between for further analyzing these frequencies. In this approach we also calculated the t-statistic (to determine whether the incentive or organic reviews are different) and the p-value (to determine whether the observed differences or similarities in this case are due a chance or based on statistically significant).

Furthermore, to discover the semantic links between incentivized and organic reviews, we implemented the SBERT (Sentence-BERT) model [42]⁵ an advanced Natural Language Processing (NLP) techniques. This is the modification of BERT (Bidirectional Encoder Representations from Transformers) with capability of mapping parts of the text to a 768-dimensional vector space. Despite BERT, which focuses on word-level embeddings, SBERT as modified version of BERT architecture generates semantically meaningful sentence-level embeddings. These embeddings allow for a more effective comparison of texts using cosine similarity to enhance our understanding of the semantic relationships between different types of reviews.

To deploy the model, data were segregated into two categories based on "Incentive" and "NoIncentive" status. The "sentence-transformer" package was imported followed by initiation of the "SentenceTransformer" class. This process automatically downloaded the "bert-base-nli-mean-tokens" model when its name is specified in the "Sentence-Transformer". Being a pre-trained model, it excels in capturing semantic meaning of sentences, which makes the model highly efficient for tasks involving semantic search. The reviews within each category were encoded into embeddings. Further, the script calculated the average embeddings for each category by summing all embeddings and dividing by the number of reviews, which provided a single vector, which represented each review's average semantic content. To assess and quantify the degree of similarity between two vectors, cosine similarity was calculated.

⁵ <https://huggingface.co/sentence-transformers/bert-base-nli-mean-tokens>

3.3.4. A/B Testing

A/B testing, a popular controlled experiment, known as split testing was conducted considering two alternatives, "Incentive" as (A) and "NoIncentive" as (B). Customer reviews were analyzed using "Incentive" and "NoIncentive" values to test the null hypothesis for significant difference between the two groups. The mean difference was measured between control and experimental groups using 10,000 repetitions. We ran six A/B tests comparing incentive and organic reviews across six rating attributes including "overAllRating", "value_for_money", "ease_of_use", "features", "customer_support", and "likelihood_to_recommend".

3.3.5. Recommendation

Recognizing more influence of "NoIncentive" reviews on customer decisions, as evidence by A/B testing results, our analysis of these reviews was geared towards developing recommendations tailored to meet customer needs more effectively. To achieve this goal, we utilized TF-IDF and Sentence-BERT techniques, enabling users to input their preferences as queries. These queries were matched with the "NoIncentive" reviews and their corresponding listing IDs, leading to the identification of the top 5 reviews most similar to each query. This approach aids customer decision-making by filtering the options.

To ensure the data is ready for analysis, we reevaluated the text preprocessing functions. Accurate labeling of ground truth data was essential for this task. As our dataset was sufficiently large and well-labeled, we randomly split it into two parts: 60% as the "main_data" and 40% as the "ground_truth_data". Subsequently, We extracted all "NoIncentive" reviews from both main and ground truth datasets.

To implement TF-IDF for content-base recommendation, after transforming the preprocessed main data text into numerical vectors using TF-IDF vectorization considering trigrams, we normalized vectors. For this purpose, we conduct Euclidean norm (L2 norm) to be sure that vector lengths do not influence the behavior of the model for calculation of similarity. We applied the same processes to ground truth data. We used "process_query_and_find_matches" function to use the same TF-IDF vectorizer for processing a user query and transforming that, and so computing the cosine similarities between the query vector and the TF-IDF vectors considering listing IDs.

Expanding on the previously mentioned capabilities of SBERT, this model is notably effective for semantic search. For this purpose the "sentence-transformers" library was installed and necessary packages were imported. Followed by splitting data and preprocessing mentioned previously. These steps succeeded by application of Sentence-BERT model ("bert-base-nli-mean-tokens") and generation of embeddings for reviews in both main and ground truth datasets that capture the semantic essence of the reviews. The given query was preprocessed and embedded using same Sentence-BERT model. The cosine similarity between query embedding and review embeddings was calculated to find top 5 similar reviews.

To evaluate the performance of both models, we calculated the metrics including precision, recall, F1 score, accuracy, match ratio, and Mean Reciprocal Rank (MRR). Execution of the models provided the list of top 5 most relevant review texts to a given query based on cosine similarity by matching with corresponding Listing IDs that each represent the specific product. In addition, along with these results the evaluation metrics' values were provided. Below are the definition of each metric value considered in these models.

$$Precision = \frac{\text{True Positive}}{\text{True Positives} + \text{False Positives}}$$

- Script-wise Precision: number of the top similar reviews identified by the model that are actually relevant

$$Recall = \frac{\text{True Positive}}{\text{True Positives} + \text{False Negatives}}$$

- Script-wise Recall: number of the relevant similar reviews were actually identified by the model

$$F1Score = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

- Script-wise F1 Score: balance between the precision and recall of the model.

$$Accuracy = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total number of Cases}}$$

- Script-wise Accuracy: number of reviews, relevant and irrelevant that correctly identified

$$MatchRatio = \frac{\text{Number of Matching Top Reviews}}{\text{Total Number of Top Reviews}}$$

- Script-wise Match Ratio: number of top reviews identified by the model that were actually present in the ground truth data (to evaluate the relevance of the model's predictions)

$$\text{Mean Reciprocal Rank (MRR)} = \frac{1}{\text{Number of Queries}} \sum \frac{1}{\text{Rank of First Relevant Answer}}$$

- Script-wise MRR: calculated the average reciprocal ranks of results for a set of queries (evaluating the performance of a ranking-based system where the order of the results is important)

4. Results and Analysis

4.1. EAD Analysis Results

After removing null values from “incentivized” feature, the EDA analysis revealed that among 49,998 remaining reviews, there are 44,255 Capterra, 3,485 Software Advice, and 2,258 GetApp reviews. The reviews were categorized into five groups including 29,466 “NominalGift”, 3,272 “VendorReferredIncentivized”, 16812 “NoIncentive”, 90 “NonNominalGift”, and 358 “VendorReferred”. The first two groups with 32738 reviews labeled “Incentive”, and the last three with 17,260 reviews labeled “NoIncentive”.

Listing IDs are the representation of different products. Three categories split 297 listing IDs to 253 listing IDs that include both “Incentive” and “NoIncentive” reviews, 31 listing IDs include just “Incentive” reviews, and 33 listing IDs include just “NoIncentive” reviews, Figure 2.

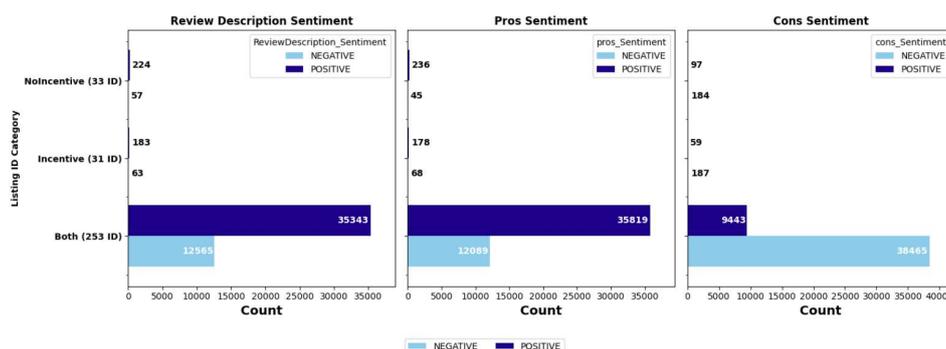


Figure 2. Number of reviews in each Listing ID Category by Sentiment Status

4.2. Sentiment Analysis Results

Assessing the “incentivized” status of rating scores revealed more incentivized than organic reviews for scores 2 and higher. Higher volume of zero scores for incentive reviews led to decreasing product recommendation based on cost and customer support, Figure 3.

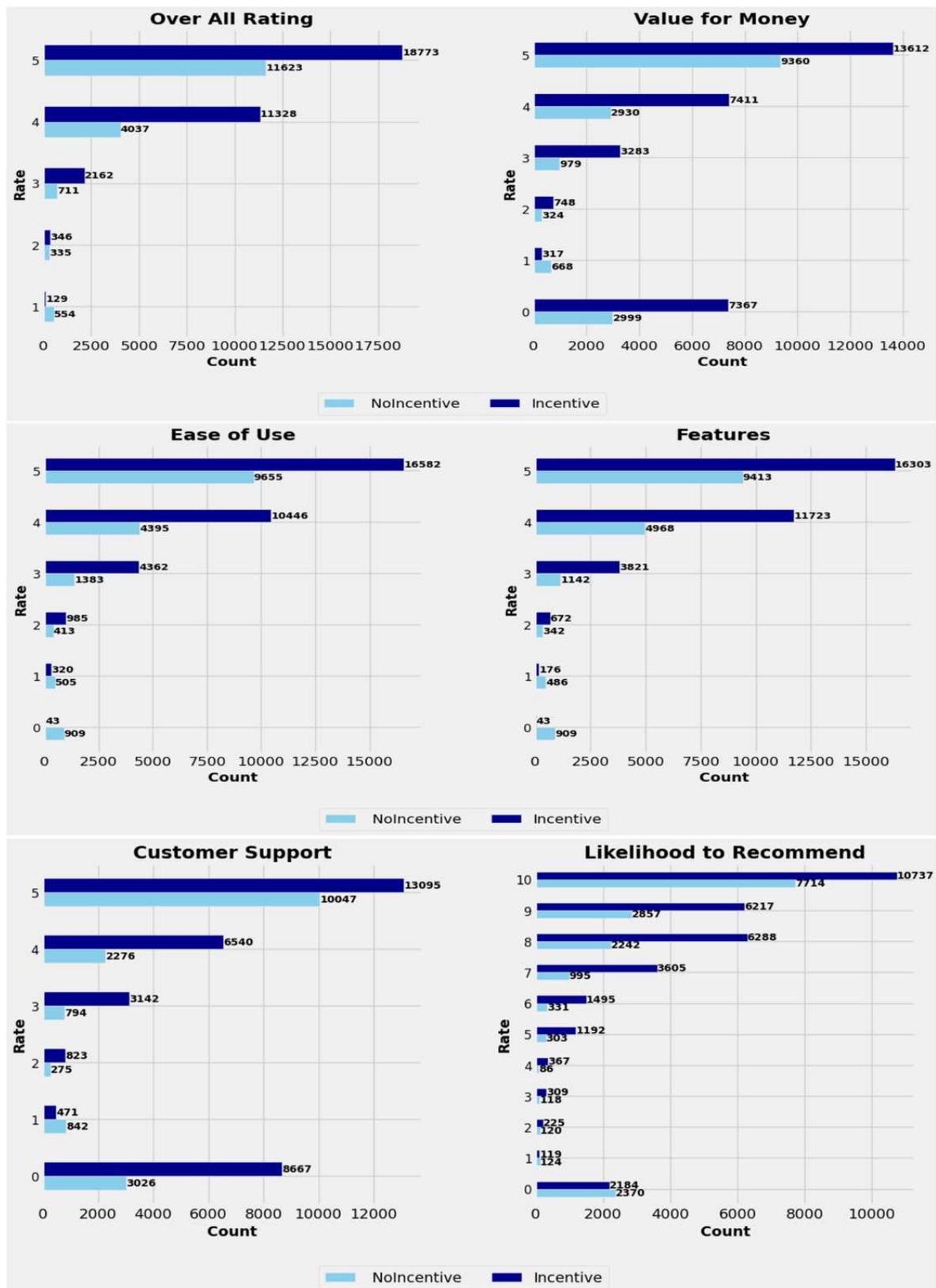


Figure 3. Number of reviews for rating scores based on incentivized status.

Figure 4 displays the tremendous escalation in software review volume, specific positive incentive reviews in 2018 and a sharp decline in 2020, revealing that changing circumstances have a greater impact on incentive than organic reviews. Rate increase may be due to increased social media

usage, greater rewards for reviews, and genuine feedback posting. Declining review volume may result from COVID-19, preventing incentivized reviews, and decreasing customer trust caused by growing awareness.

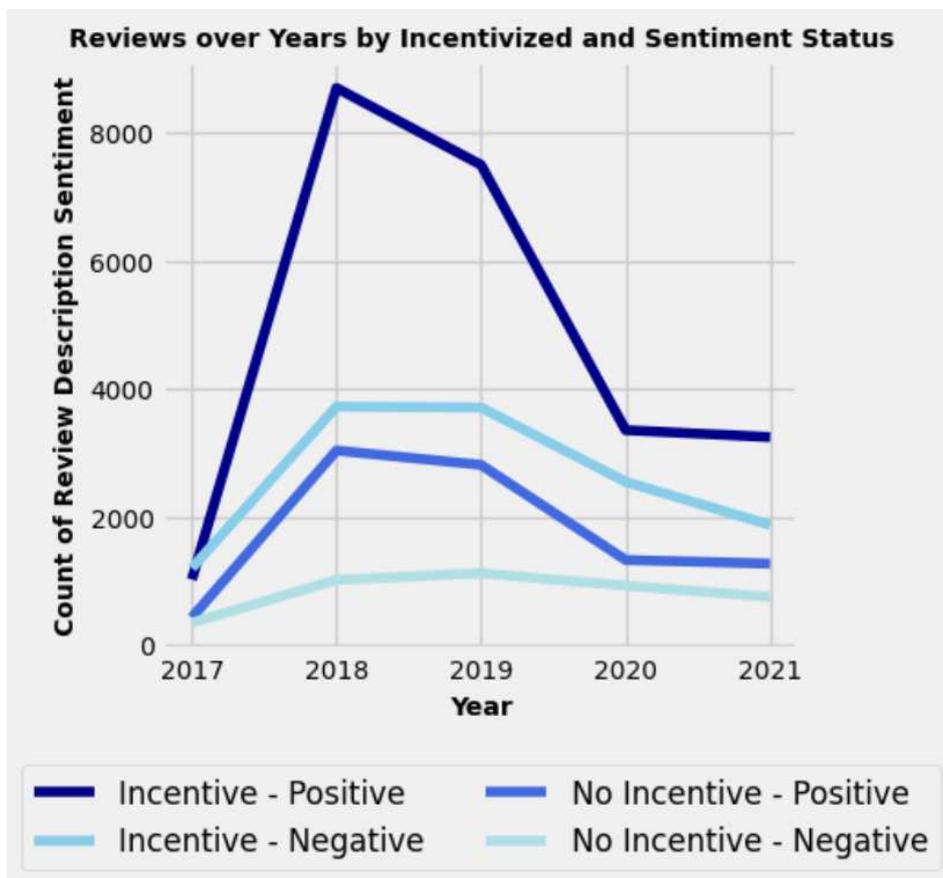


Figure 4. Reviews over years by incentivized and sentiment status.

Experienced users, more than two years of experience, using products tend to post positive and fewer negative incentive reviews, likely due to product familiarity and preference for benefits. Customers who use free trial post fewer reviews due to a lack of experience and confidence, they still contribute to post more incentive than organic reviews, Figure 5.

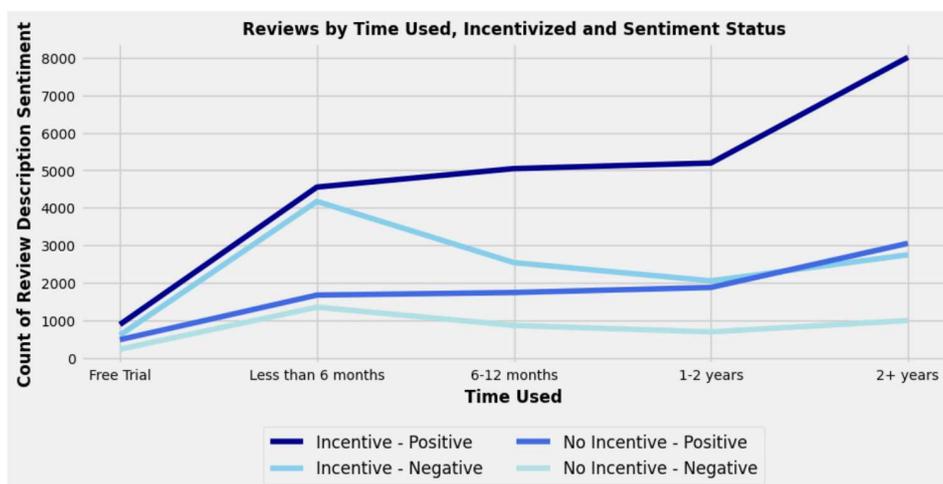


Figure 5. Distribution of incentivized reviews by time used and sentiment.

Our study highlights that small companies with 11-50 employees, specific positive incentive reviews, have more than 7000 reviews. However, companies with 5,001-10,000 employees have less than 510 reviews. The significant gap between the number of incentive and organic reviews for smaller companies compare to larger ones is due to easier establishment and a higher likelihood of posting reviews, Figure 6.

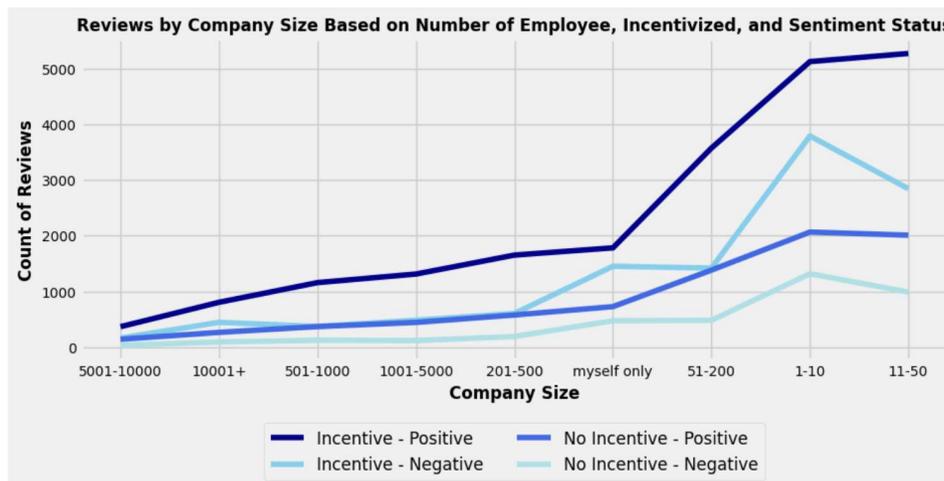


Figure 6. Distribution of incentivized reviews by company size and sentiment.

Analyzing review text sentiment considering “incentivized” status showed a higher incidence of positive sentiment in reviews’ descriptions and pros, and negative sentiment in cons. Regardless of the review sentiment, volume of incentivized reviews is larger than organic reviews, Figure 7.

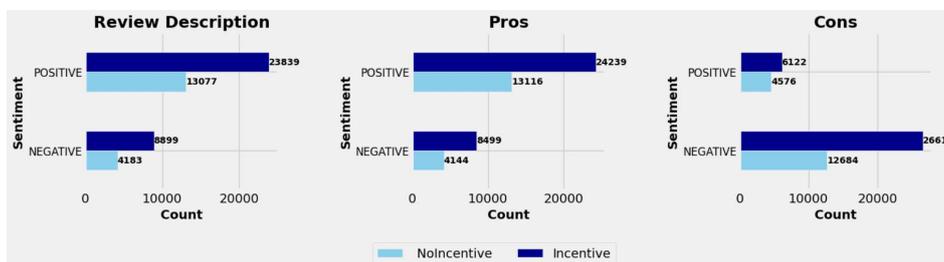


Figure 7. Number of review description, pros, and cons based on incentivized and sentiment status.

The word cloud is used to extract the top 20 words from each review text focusing on “incentivized”. Considering sentiment status, the words such as “great” and “good” were frequently used in positive incentive and organic reviews for different review text, in addition to negative incentive and organic reviews for review description and pros. However, the top 20 words for negative incentive and organic do not include any negative words as expected. This could be due to removing negative words such as “not” as stop words, possibly causing the omission of negative phrases like “not good”. Our results support prior research as incentives increase positive review length. Simultaneously, the volume of top 20 words is larger for positive incentives than organic and smaller for negative incentives than organic. Figure 8 represents these results for review description.

Furthermore, measuring the average length of incentive and organic review descriptions based on the number of characters and their sentiment status reveals longer negative organic reviews, 153.91, than negative incentive reviews, 125.17. However, positive incentive reviews are longer, 104.13, than positive organic reviews, 96.45. These results are consistent with previous studies, [4,5]. Overall these results for both negative incentive and organic reviews are higher than positive.

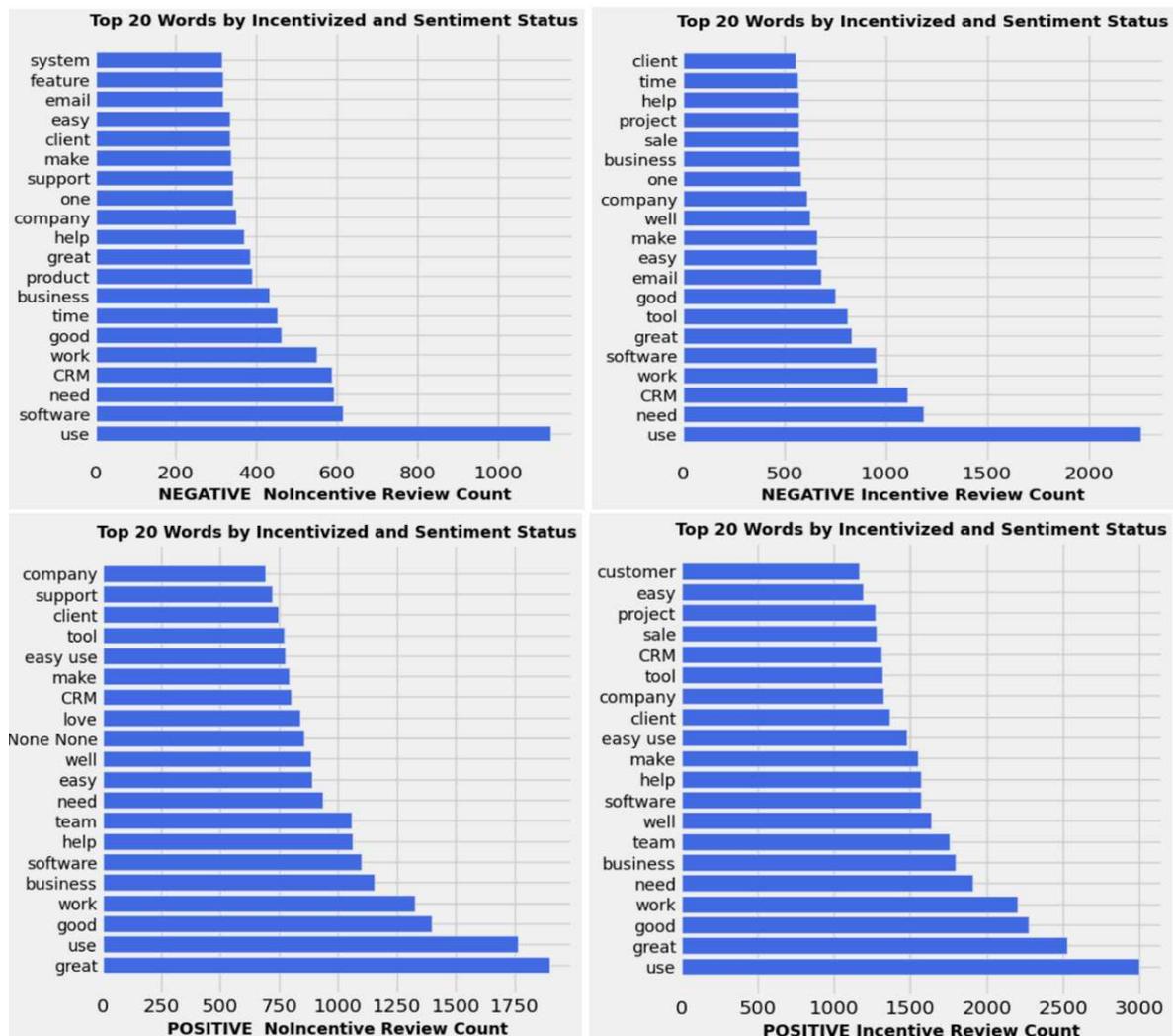


Figure 8. Review description top 20 words based on incentivized and sentiment status.

Our results of testing “Spearman’s rank correlation coefficient” on review rating scores, considering 95% confidence interval, indicate stronger correlation among organic reviews, and more specifically, negative organic reviews, Figure 9.

The highest correlation of 0.80 between “likelihood_to_recommend” and “overAllRating” appears to be influenced by high correlations between “overAllRating” with both “features” as 0.78 and “ease_of_use” as 0.76, in addition to the high correlation between “likelihood_to_recommend” with both “features” as 0.73 and “ease_of_use” as 0.72. Similar pattern with weaker positive correlations is seen in negative incentive reviews. Moreover, correlation between “features” and “ease_of_use”, considering various statuses, supports the need for user-friendly software that provides easier access to features. In addition, the significant correlation between “value_for_money” and “customer_support” for negative reviews denotes weaker correlation for negative incentive, 0.60, compare to negative organic reviews, 0.66. All z-test results show 95% confidence in significant correlations between review ratings due to p-values of zero.

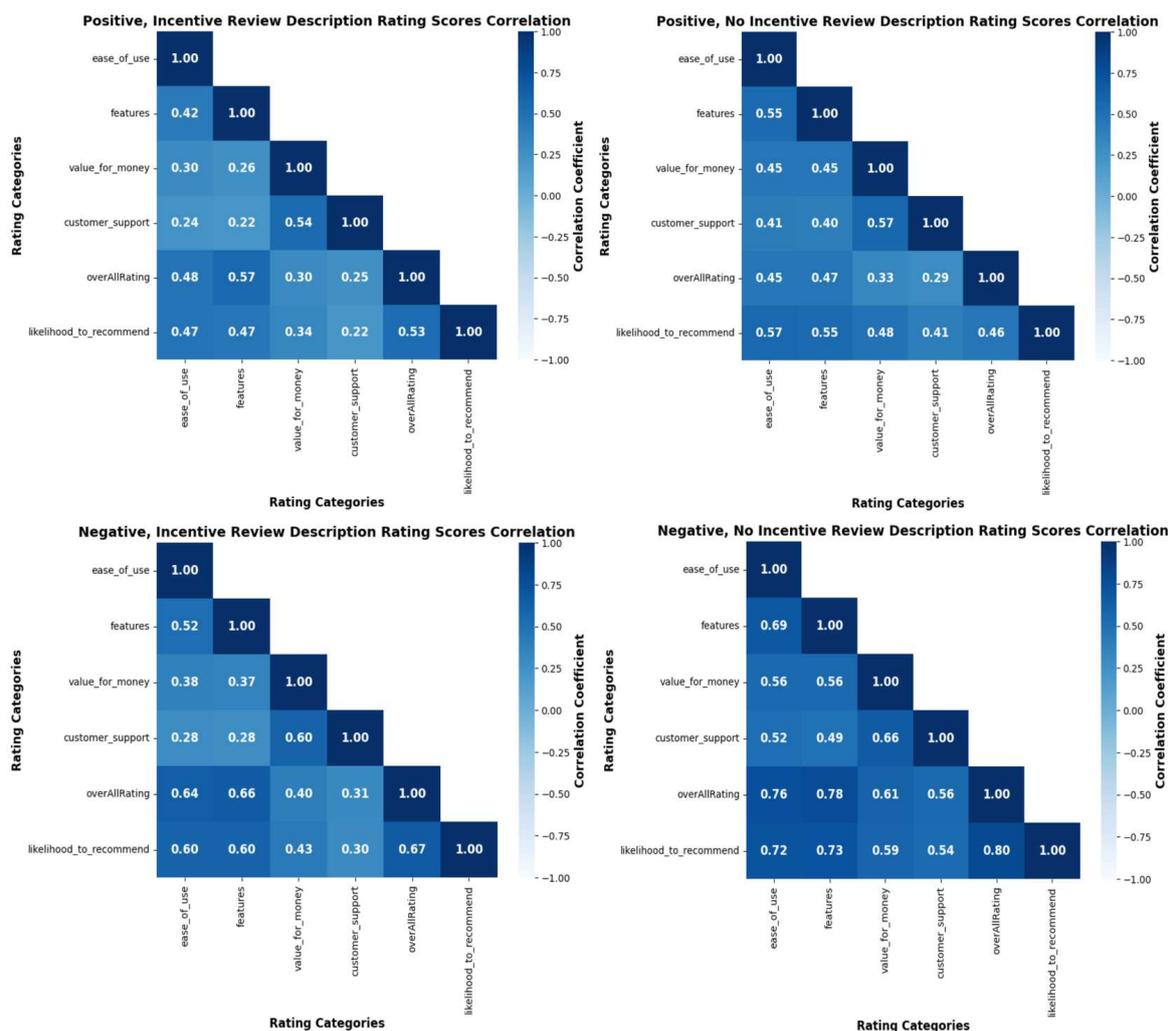


Figure 9. Correlation among review rating scores by incentivized and review description sentiment status.

4.3. Semantic Links Results

Semantic links compared incentive and organic reviews' contents using TF-IDF and SBERT methods.

4.3.1. Semantic Links Results Using TF-IDF

The two categories of reviews under examination highlight distinct differences, as evidence by the analysis of extracted trigrams and their respective frequencies through the application of TF-IDF technique. The "NoIncentive" reviews contain unique phrases such as "sensitive content hidden" and "everything one place". In contrast, the "Incentive" reviews are characterized by more frequent phrases like "project management tool" and "steep learning curve" more predominantly. Nonetheless, certain phrases like "great customer service" and "software easy use" are prominently featured across both categories with different frequencies. While there are numerous overlaps in the trigrams found in both categories, variations in the order and prevalence of these trigrams indicate the distinct priorities and areas of focus. For instance, organic ("NoIncentive" reviews tend to emphasize aspects like "software easy use" or similar phrases, whereas incentivized reviews are predominantly focus on elements such as "project management tool", Figure 10.

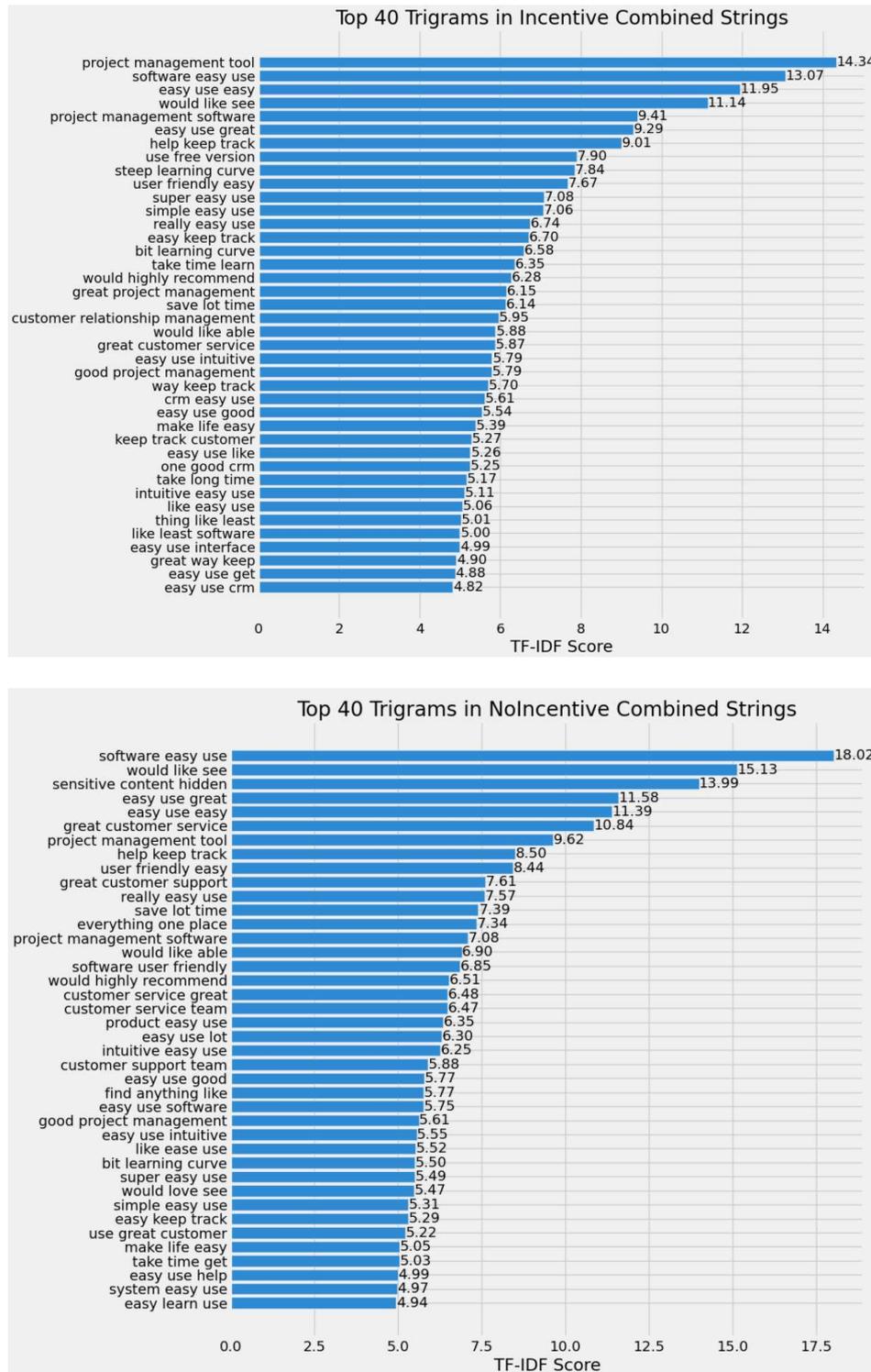


Figure 10. Top 40 Trigrams and Frequencies in Incentive and NoIncentive Reviews Texts.

The cosine similarity of 0.675 between incentive and organic reviews indicates moderate to high level of similarity between two review categories based on their trigram representations. This implies that concept of incentive and organic reviews are almost similar in terms of the most significant trigrams in each group. Therefore, it seems that offering the incentive for writing reviews, does not drastically change the language used in reviews. Although there are some sort of uniqueness in each set, the overlapping contents are significant.

We also found the t-test of -0.867, which negative points to the lower average TF-IDF score for incentive than organic reviews; however, due to the relatively small magnitude of the t-statistic, there is no large difference in means. In addition, the p-value of 0.389 is higher than the significance level of 0.01 or 0.05, which indicates that the observe difference in the average TF-IDF scores of these two review categories is not statistically significant. Therefore, the incentive and organic reviews do not have the significant differences based on their contents. This means any difference between “Incentive” and “NoIncentive” reviews could be due to the chance without the systematic root.

4.3.2. Semantic Links Results Using Sentence-BERT

We used the “SentenceTransformer” model to capture contextual meaning of the reviews using embeddings. The average embeddings for both review categories, that represented the mean vector of the embeddings and summarized each set of the reviews into single vectors, aid to measure the cosine similarity of 0.999. The result highlights that both types of reviews discuss similar topics or themes, which suggest that incentives may not have impact on the content of reviews. This reveals the two groups are almost identical in terms of topic, information and sentiments, and incentives do not alter the reviewers’ language or points of their focus.

4.4. A/B Testing Results

A/B test compared incentive and organic reviews for different rating scores, Table 1.

Table 1. The Statistical Values and Results of A/B Testing.

Attribute	Incentivized	Total	Sum Total Rating	Mean Value	Std	Std Error	Observed Difference	Empirical P
overAllRating	NoIncentive	17260	77620	4.497	0.913	0.007	-0.0227	0.0014
	Incentive	32738	146484	4.474	0.702	0.004		
value_for_money	NoIncentive	17260	62773	3.637	1.916	0.015	-0.2963	0.0000
	Incentive	32738	109366	3.341	1.965	0.011		
ease_of_use	NoIncentive	17260	71335	4.133	1.350	0.010	0.1455	1.0000
	Incentive	32738	140070	4.279	0.890	0.005		
features	NoIncentive	17260	71533	4.144	1.329	0.010	0.1744	1.0000
	Incentive	32738	141390	4.319	0.815	0.005		
customer_support	NoIncentive	17260	63113	3.657	1.954	0.015	-0.5050	0.0000
	Incentive	32738	103178	3.152	2.060	0.011		
likelihood_to_recommned	NoIncentive	17260	132317	7.666	3.431	0.026	0.1766	1.0000
	Incentive	32738	256756	7.843	2.685	0.015		

Organic review have a higher standard deviation (std) error for all rating scores despite higher total rating scores for incentive than organic reviews. For incentive reviews, lower std of “overAllRating” implies the consistency among ratings, and lower std error proposes a more precise estimate of the true mean. On the other hand, Organic reviews outperform significantly based on a p-value of 0.0014, with an overall impact on customer decision-making, indicated by the observed value of -0.0227. Cost-related ratings for organic reviews have a higher mean rating, and lower std that represents less variation in the rating. A P-value of 0.0000 indicates a significant difference between the two groups, and regarding the cost-related rating of the software reviews, incentives may not affect customers’ decisions. In terms of ease of use, incentive reviews are rated higher. However, there is no statistically significant difference between incentive and organic reviews due to a p-value of 1.0000. Therefore, the observed value may not reflect true values, indicating insignificant impact of incentives on customer decision-making. For software review features, incentive reviews have a higher mean and consistent rating. Although the observed value of 0.1744 points to the difference between the two groups, the p-value of 1.0000 indicates no statistically significant impact of incentive reviews on customer decision-making. As discussed, customer support is essential for any product and its rating score shows a negative observed difference, meaning organic reviews have a higher average rating than

expected. The significant difference between incentive and organic is because of the negative impact of incentive reviews on customer decision-making.

The results of the customer's willingness to recommend products indicate more variability in organic review ratings. However, regarding willingness to recommend, the p-value of 1.0000 reveals no significant difference between groups, so it may not impact decision-making.

4.4.1. Recommendation Results

To aid in customer decision-making, we followed the outcomes of our A/B testing and chose to concentrate on "NoIncentive" reviews, which have a greater impact on customer choices.

The queries used to evaluate the models are detailed below in Table 2. We selected 6 queries for model testing: the first three are unique, user-generated preferences, while the later three are derived from existing organic reviews. These include complete review, a partial of the previous complete review, and a variation where all the key terms in the review are substituted with their synonyms.

Comparing the results of top 5 listing IDs with corresponding similarity scores for each query, Table 3, reveals Sentence-BERT outperform the TF-IDF significantly for all top five selections in all the queries, except for the first listing ID of query 4 and 5 (Q 4 and Q 5), which are the seen data. This emphasizes the power of TF-IDF in identifying keywords and the limitation of this model in capturing the meaning and semantic of the data. The lower similarity scores for TF-IDF for majority of listing IDs in the queries, indicate lower degree of similarity between the query and most of the retrieved listings. In addition, significant numbers of zero similarity scores among TF-IDF results may be have several reasons. Lack of similarity between query and ground truth data, and mismatching the query's key-terms (such as bigram in Q 3) with model's configuration (such as trigram) are two major reasons. Sentence-BERT that is context-aware model performs well in seizing similarities based on the similarity score results. However, when the text has complex or very simple content like query 1 (Q 1) or query 3 (Q 3) it shows weaker performance respectively.

To compare the performance of the TF-IDF and the Sentence-BERT models, besides looking at their specific outcomes such as listing IDs and similarity scores, evaluation metrics could be considered, Table 4.

The first three queries demonstrate perfect precision (1.000), indicating that 100% of the recommended listing IDs are relevant. Moreover, the optimal MRR of 1.000 for both models suggests that the first relevant results is the top-listed ID. The SBERT model's higher similarity scores, coupled with its notable MRR, underscore its effectiveness in identifying the listing IDs closely related to the query. The lower precision (0.800) for seen queries for both models denotes 80% of recommended listing IDs are relevant. Despite the impressive accuracy of both models (0.994 for TF-IDF and 0.995 for SBERT), these values are not informative enough if there is a high chance of encountering irrelevant data. The models' low recall scores point to their limited ability to find all relevant listing IDs. However, the SBERT's slightly better recall values reveals a superior SBERT's performance in finding relevant listing IDs. The combination of high precision and low recall denotes the potential data imbalance in our dataset, which is reflected in low F1-scores, meaning that while most of the matching items the model identified are correct, the model struggles with recognizing a significant number of relevant items. This imbalance could arise from certain products having more reviews than others, skewing the model's learning toward listing IDs with more reviews. Moreover, the imbalanced distribution of sentiment (positive, negative or neutral) through dataset can impact the performance of the model while the point is focusing on the sentiment. Furthermore, the presence of detailed or complex reviews among more generic ones might impact the model's accuracy in matching reviews to queries. Perfect match ratio (1.000) for unseen queries (Q1-Q3) and good match ratio (0.800) for seen queries (Q4-Q6) reveal a similar level of accuracy in matching the top results with the ground truth data for both sets of the queries.

Table 2. Query Used for Recommendation.

Query	Abbreviation	Nature of Query	Query Text
Query 1	Q 1	Complex Customer Preferences	For my work I need the software to facilitate my work and gives me the will to recommend that to others as I am frustrated with other software I have used. I need the software to work well, no matter if it is complex or not as I like challenges, with good CRM, and good customer support, has enough features and I can work with that by my phone. The price is not that important.
Query 2	Q 2	Moderate Customer Preferences	I need the product with good features, that has low price, I can learn how to work with that fast and easily
Query 3	Q 3	Simple Customer Preference	I need Good CRM
Query 4	Q 4	One NoIncentive Review	Surprised Franklin Covey would even advertise think program would good could get work customer support beyond horrible there no pro point possibly layout great but would not know since can not get workI tired sync w ical with no success when you call support you route voice mailit take least hour someone call you back in sale hour later not in my office in front computer etc work out issue
Query 5	Q 5	Part of NoIncentive Review	Would not know since can not get workI tired sync w ical with no success when you call support you route voice mailit take least hour someone call you back in sale hour later not in my office in front computer etc work out issue
Query 6	Q 6	Synonyms Replacement in Review	Astonished would even publicize think program would decent could get work customer provision yonder awful there no pro opinion perhaps design countless but would not know since can not get workI exhausted synchronize w l with no achievement when you call support you way voice mailit take smallest hour someone call you back in transaction hour later not in my office in forward-facing computer etc. work out problem

Table 3. Similarity Scores of Top 5 Recommended Listing IDs.

Query	Model	Listing ID1	Similarity Score 1	Listing ID2	Similarity Score 2	Listing ID3	Similarity Score 3	Listing ID4	Similarity Score 4	Listing ID5	Similarity Score 5
Query 1	TF-IDF	113213	0.044	109395	0.027	9448	0.016	91202	0.015	10283	0.012
	Sentence-BERT	102517	0.863	91179	0.862	10317	0.850	2348	0.850	90844	0.850
Query 2	TF-IDF	90941	0.043	9908	0.008	106331	0.004	9920	0.000	104247	0.000
	Sentence-BERT	90844	0.892	106331	0.889	91734	0.886	10317	0.880	9908	0.880
Query 3	TF-IDF	9920	0.000	104247	0.000	100342	0.000	106331	0.000	102533	0.000
	Sentence-BERT	2035403	0.695	20406	0.694	10317	0.691	9929	0.690	20468	0.686
Query 4	TF-IDF	91817	1.000	90602	0.004	9908	0.002	106331	0.001	91734	0.001
	Sentence-BERT	91817	1.000	113901	0.919	91203	0.914	104287	0.912	113901	0.911
Query 5	TF-IDF	91817	0.747	9920	0.000	104247	0.000	100342	0.000	106331	0.000
	Sentence-BERT	91817	0.932	91203	0.916	113901	0.905	2348	0.905	109561	0.903
Query 6	TF-IDF	91817	0.476	9920	0.000	104247	0.000	100342	0.000	106331	0.000
	Sentence-BERT	91817	0.965	2348	0.919	90602	0.918	91203	0.917	91179	0.913

Table 4. Evaluation Metrics.

Query	Model	Precision	Recall	F1-Score	Accuracy	Match Ratio	Mean Reciprocal Rank
Query 1	TF-IDF	1.000	0.016	0.032	0.994	1.000	1.000
	Sentence-BERT	1.000	0.020	0.039	0.995	1.000	1.000
Query 2	TF-IDF	1.000	0.016	0.032	0.994	1.000	1.000
	Sentence-BERT	1.000	0.020	0.039	0.995	1.000	1.000
Query 3	TF-IDF	1.000	0.016	0.032	0.994	1.000	1.000
	Sentence-BERT	1.000	0.020	0.039	0.995	1.000	1.000
Query 4	TF-IDF	0.800	0.013	0.026	0.994	0.800	0.500
	Sentence-BERT	0.800	0.016	0.031	0.995	0.800	1.000
Query 5	TF-IDF	0.800	0.013	0.026	0.994	0.800	0.500
	Sentence-BERT	0.800	0.016	0.031	0.995	0.800	1.000
Query 6	TF-IDF	0.800	0.013	0.026	0.994	0.800	0.500
	Sentence-BERT	0.800	0.016	0.031	0.995	0.800	1.000

5. Discussion

In this section, we compare incentive and organic reviews, addressing the first research question, followed by answering the second research question by discussing the impact of incentives on customers' behavior toward posting reviews, review quality, and purchase decision-making, and finally we discuss which one of incentive or organic reviews can better assist customer decision.

5.1. Incentive vs. Organic

The analysis results prove that incentive reviews have more positive descriptions and pros, more negative cons, higher ratings, and minority with lower scores. Despite organic reviews, the volume of incentive reviews has changed dramatically over the years, revealing the dependency on various factors, including environmental situations (e.g., pandemics and economic problems). The incentive volume can grow by growing social platforms, improving customer experience, and expanding smaller companies. In addition, the overall results of semantic links between incentive and organic reviews imply even if two review groups are not identical but they share significant amount of common languages, which means incentives do not entirely alter the focus or sentiment of the reviews. Therefore, customers may treat incentive and organic reviews as having similar content characteristics. Referring to these findings, companies may shift their encouragement plan from offering incentives to focus on improved advertising, information sharing, consumer awareness, and distrust of review authenticity. Based on the result of A/B testing, incentive reviews have the higher sum of the "total rating" and lower std error for all rating scores. Overall rating is more consistent for incentive reviews.

5.2. Incentive Review and Customer Behavior

To answer “How do incentives impact customer behavior on posting purchase’s review?”, we rely on our findings from the first research question.

The analysis proves that incentives boost reviews, as the volume of incentive reviews is almost double to compare with organic reviews. Reviewers tend to rate the reviews positively, despite providing negative feedback. Higher sum of rating scores for incentive reviews compare to organic reviews indicate customers are more likely rewarded for posting positive reviews.

The dramatic alteration in the distribution of incentives over the years proves rewards as review posters’ motivation. Over time, factors like commerce, economy, social networks, environmental issues, and technology can reduce, restrict, or eliminate incentives from the business platform, causing users to post fewer reviews. Despite massive changes in volume over the years, incentive reviews consistently outnumber organic reviews indicating the impact of incentives on customers’ review behavior toward posting reviews. However, the study outcomes indicate that incentive may change the direction of customer thinking slightly, but does not alter that completely. Small businesses may incentivize individuals to write incentive or fake reviews to compete in the business world and increase profits. The better product understanding enhances incentive reviews quality and quantity. Furthermore, customer support quality impacts product cost satisfaction for many customers.

However, some users are discouraged from writing incentive reviews when they become aware of the potential for biased or suspicious content. On the other hand, recognizing the clear direction of the incentivized reviews toward businesses brands and profits, can guide new customers towards what might be more authentic review.

5.3. Incentive Review and Review Quality

Based on the analysis, higher volume of incentive reviews displays lower credibility and higher bias as they may contain non-experience-based information aimed at boosting review quantity and rating. A greater volume of incentive reviews for smaller companies may indicate bias and fake reviews, reducing credibility and consistency of the review quality. Although reviews from experienced users are more credible, those who receive incentives for their reviews tend to be less consistent in their rating compared to organic reviews, due to a significant increase in positive incentive and decrease in negative incentive reviews. In terms of cost and customer support, the significant number of zero rating scores for incentivized than organic reviews proves that incentives do not always increase positivity. Incentive and organic reviews show similar zero-rate volumes for recommendation likelihood, indicating greater consistency in organic reviews. Higher negative-to-positive cons ratio than positive-to-negative pros ratio, even for incentive reviews, suggests customer sensitivity to writing negative feedback, increasing the credibility of negative reviews. Negative software review ratings correlate more strongly than positive ratings, which may support that negative reviews are often more credible [16]. Based on the evidence, incentive reviews show inconsistent volume.

Furthermore, high volume of top 20 words in positive incentive reviews suggests a possible bias and reduced credibility compare to organic reviews. Additionally, the higher occurrence of phrases favoring companies in incentivized reviews, as oppose to higher occurrence of customer-favoring phrases in organic reviews, underscore the greater credibility of organic reviews. Offering rewards for incentive reviews reduces diversity and increases consistency in review content, despite organic reviews. Highlighting the priority of specific contents in the incentive reviews such as “project management tool” may indicate to the target promotion, which can impact the credibility of the reviews.

In comparison, the higher volume of these words in negative organic reviews indicates more detailed reviews, increasing credibility of organic reviews. Longer negative organic review descriptions that provide more information and detail reveals higher credibility and lower bias, outweighing incentive reviews.

To this point, our discussion on review credibility and consistency mainly focuses on reviews volume and length. Higher sum of rating scores for incentive reviews, considering different rating scores statistical values, may indicate motivated posting reviews for rewards, raising credibility concern.

5.4. Incentive Review and Purchase Decision-Making

Higher incentive review volume points to focusing more on the overall rating and quantity over review content.

Therefore, lack of accurate and comprehensive view of the products/services, cause less consistency in incentive reviews, resulting in lower credibility and uninformed purchase decisions. Willingness to post positive incentivized reviews, based on the higher sum of all rating scores, may indicate excessive positivity, impacting customers' purchase decision.

Regarding to the content of the reviews, Incentive review put more weight on project management tool, which can be sign of the strengthen brand recognition and it is mainly in favor of the businesses. Despite organic review that raise the bar for customer friendly contents such as "customer support" and "ease of use". Frequent repetition of content that is more favorable to businesses in incentive reviews can diminish the credibility, authenticity, and trustworthiness of the reviews for the new customers.

Referring to observed differences and p-values from the A/B testing results, incentive reviews have less impact on customer purchase decision based on their overall rating, may not impact customer purchase decision regarding software cost and software features, and have no significant impact on decision-making based on easiness of use. In addition, incentive reviews negatively affect customer decision-making, as shown by the customer support score evaluation.

5.5. Recommendation and Purchase Decision-Making

Based on the semantic links analysis results, which showed only a slight content difference between incentivized and organic reviews, we employed A/B testing for further analysis. This revealed that organic reviews have greater influence on customer decision-making. Utilizing SBERT and TF-IDF techniques, by acknowledging the greater effectiveness of organic reviews in shaping customer preferences, we developed a recommendation system. The system focused on organic reviews to designed the streamline for customer choices by presenting the top 5 most relevant products based on customer preferences as a query entry.

We initially used "seen queries" to evaluate the model for several key reasons: to verify its performance and functional integrity, ensure consistent responses in similar scenarios, and identify any potential failures in performance. While this approach with seen data proved beneficial for initial performance verification, we recognized the importance of extending our validation to unseen data. This crucial step was taken to moderate the risk of high performance metrics due to overfitting and to more accurately assess the model's ability to generalize. Further, our methodology involved utilizing review content that was detailed but not overly complex, targeting to optimize the performance of the recommendation model. This approach ensured that the model captured the essential distinctions in the reviews without being impacted by unnecessary complexity. Despite the challenges posed by an imbalanced dataset typical for online reviews, the SBERT model's high performance demonstrated the potential for achieving reliable and accurate outcomes with a judicious choice of the model.

5.6. Implications of the Study

Despite existing studies, our approach distinguishes our work by evaluating incentive reviews quality differently. We used EDA, sentiment analysis, semantic links and A/B testing to compare incentive and organic reviews quality and determine the impact of incentives on customers' behavior for posting reviews. Although incentive software reviews outnumber organic reviews by almost two-fold, this could be changed in either increasing or decreasing direction due to factors such as time, business platform and size, and user awareness and experience. Factors such as cost, software

features, ease of use, and customer support impact software product ratings of either incentive or organic reviews. This is due to the high correlation between these features or with software's overall rating and recommendation. Furthermore, our A/B testing shows that high volume and rating of incentive reviews may not significantly affect customer purchase decisions.

By representing the distinct impact of organic reviews on customer decision-making, and the capability of advanced models like SBERT and TF-IDF to distinguish and control this impact despite dataset imbalances, our research offers a pathway to more customer-centric and reliable recommendation algorithms. For experts in the field of e-commerce, integrating these perceptions can lead to the creation of systems that not only resonate more with consumer preferences but also surrogate trust and engagement through increased authenticity. Additionally, the methodological approach outlined, particularly the balance between seen and unseen data, sets a instance for future research and development in data-driven decision support systems. This study contributes to the broader discourse on how AI and machine learning can be intentionally used to enhance the user experience in digital marketplaces.

6. Conclusion and Future Work

Studies may have opposing results for the same situation due to differences in population under study, research methods, and approaches used. For instance, Woolley & Sharif (2021) [6] highlights more enjoyable writing reviews with incentives, while Garnefeld et al. (2020) [9] emphasize incentives role in increasing the review rate. However, Burtch et al. (2018) [4] remark the delivering negative review by incentivized customers. While our findings support some of the existing research, the emergence of these opposing findings underscores the need for further investigation across diverse populations of various sizes and cultures. It is imperative to examine different methods and approaches to reach broadly applicable outcomes to improve the quality of online review systems and recommendation system. Potential collaborations among companies can achieve such outcomes.

6.1. Strengths and Limitations

This study has broad applicability across various research domains and is not restricted to the fast-growing world of reviews. Increasing business growth and product diversity intensifies producer competition for sales to companies, which in turn sell products to consumers. This highlight the necessity of accessing high-quality reviews. This highlights the need to access high-quality reviews. Our study has the potential to enhance the performance of product review systems, boosting customer satisfaction and efficiency by saving time and money.

Moreover, our unique contribution to the review quality study that achieves by focusing on software review quality, specifically incentive reviews, distinguishes our work from existing research in the field.

While this work has several strengths, it also exhibits some weaknesses. Current methods cannot accurately determine review sentiment due to the subjective nature of reviews, which involves human emotion and expressions. A large dataset prevents using human power to annotate reviews sentiment. Even human annotation does not ensure result accuracy due to potential human error. On the other hand, due to inability to recognize emotions and expressions, automated annotation falls short in achieving higher accuracy.

Additionally, the model's constraint limited our sentiment analysis to analyzing only up to 200 characters. Therefore, our model may have missed important aspects of some reviews as their length exceeded this limitation.

6.2. Future Work

We analyzed purchase review differences and assessed review credibility and consistency to evaluate the influence of incentive reviews on customer decision-making. We assessed review quality before studying its influence on purchase decisions. We discussed research problems and our findings

through EDA on sentiment analysis and A/B testing, and as final approach to aid customers in decision-making, we designed the recommendation algorithm. However, for future work, we plan to survey software reviews to gather and analyze information from new users, considering the subjectivity of online review and purchase decisions. Additionally, We plan to explore review quality dimensions that affect purchasing decisions, specifically objectivity [43,44], depth [45,46], authenticity [47,48] and ultimately, helpfulness [49,50], taking into consideration the incentivized status of the reviews. This will allow us to compare users' perspectives on incentive reviews' quality and their impact on purchase decisions. Moreover, future work will focus on refining the recommendation system by application of advanced NLP techniques that can bring sentiment information into account and could provide deeper insight into customer reviews and their impact on e-commerce landscape.

Author Contributions: Conceptualization, K.K. and J.D.; methodology, K.K., J.D. and H.C.; software, K.K.; validation, K.K. and J.D.; formal analysis, K.K.; investigation, K.K.; resources, K.K. and H.C.; data curation, K.K.; writing—original draft preparation, K.K. and J.D.; writing—review and editing, K.K., J.D. and H.C.; visualization, K.K.; supervision, K.K. and J.D.; project administration, K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Acknowledgments: The author would like to thank Bhanu Prasad Gollapudi for his contribution to data collection and preparation.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhu, L.; Li, H.; Wang, F.; He, W.; Tian, Z. How online reviews affect purchase intention: a new model based on the stimulus-organism-response (S-O-R) framework. *Aslib J. Inf. Manag.* **2020**, *72*, 463–488.
2. Petrescu, M.; O'Leary, K.; Goldring, D.; Ben Mrad, S. Incentivized reviews: Promising the moon for a few stars. *J. Retail. Consum. Serv.* **2018**, *41*, 288–295.
3. Ai, J.; Gursoy, D.; Liu, Y.; Lv, X. Effects of offering incentives for reviews on trust: Role of review quality and incentive source. *Int. J. Hosp. Manag.* **2022**, *100*, 103101.
4. Burch, Gordon and Hong, Yili and Bapna, Ravi and Griskevicius, Vladas. Stimulating Online Reviews by Combining Financial Incentives and Social Norms. *Manag. Sci.* **2018**, *64*, 2065–2082.
5. Costa, A.; Guerreiro, J.; Moro, S.; Henriques, R. Unfolding the characteristics of incentivized online reviews. *J. Retail. Consum. Serv.* **2019**, *47*, 272–281.
6. Woolley, K.; Sharif, M. Incentives Increase Relative Positivity of Review Content and Enjoyment of Review Writing. *J. Mark. Res.* **2021**, *58*, 539–558.
7. Imtiaz, MN.; Ahmed, MT.; Paul, A. Incentivized Comment Detection with Sentiment Analysis on Online Hotel Reviews. *Authorea* 2020. Available online: <https://doi.org/10.22541/au.159559938.84764895> (accessed on 14 January 2024).
8. Zhang, M.; Wei, X.; Zeng, D. A matter of reevaluation: Incentivizing users to contribute reviews in online platforms. *Decis. Support Syst.* **2020**, *128*, 113158.
9. Garnefeld, I.; Helm, S.; Grötschel, AK. May we buy your love? psychological effects of incentives on writing likelihood and valence of online product reviews. *Electron. Mark.* **2020**, *30*, 805–820.
10. Cui, G.; Chung, Y.; Peng, L.; Zheng, W. The importance of being earnest: Mandatory vs. voluntary disclosure of incentives for online product reviews. *J. Bus. Res.* **2022**, *141*, 633–645.
11. Luca, M.; Zervas, G. Fake it till you make it: Reputation, competition, and yelp review fraud. *Manag. Sci.* **2016**, *62*, 3412–3427.
12. Kamble, V.; Shah, N.; Marn, D.; Parekh, A.; Ramchandran, K. The Square-Root Agreement Rule for Incentivizing Objective Feedback in Online Platforms. *Manag. Sci.* **2019**, *69*, 377–403.
13. Liu, Z.; Lei, S.H.; Guo, Y.L.; Zhou, Z.A. The interaction effect of online review language style and product type on consumers' purchase intentions. *Palgrave Commun.* **2020**, *6*, 1–8.

14. Le, L.T.; Ly, P.T.M.; Nguyen, N.T.; Tran, L.T.T. Online reviews as a pacifying decision-making assistant. *J. Retail. Consum. Serv.* **2022**, *64*, 102805.
15. Zhang, H.; Yang, A.; Peng, A.; Pieptea, L.F.; Yang, J.; Ding, J. A Quantitative Study of Software Reviews Using Content Analysis Methods. *IEEE Access.* **2022**, *10*, 124663–124672.
16. Kusumasondjaja, S.; Shanka, T.; Marchegiani, C. Credibility of online reviews and initial trust: The roles of reviewer's identity and review valence. *J. Vacat. Mark.* **2012**, *18*, 185–195.
17. Jamshidi, S.; Rejaie, R.; Li, J. Characterizing the dynamics and evolution of incentivized online reviews on Amazon. *Soc. Netw. Anal. Min.* **2019**, *9*, 1–15.
18. Gneezy, U.; Meier, S.; Rey-Biel, P. When and why incentives (don't) work to modify behavior. *J. Bus. Res.* **2011**, *25*, 191–210.
19. Chen, T.; Samaranayake, P.; Cen, X.; Qi, M.; Lan, Y.C. The importance of being earnest: Mandatory vs. voluntary disclosure of incentives for online product reviews. *Front. physiol.* **2022**, *13*, 2723.
20. ANoh, Y.G.; Jeon, J.; Hong, J.H. Understanding of Customer Decision-Making Behaviors Depending on Online Reviews. *Appl. Sci.* **2023**, *13*, 3949.
21. Aghakhani, N.; Oh, O.; Gregg, D.; Jain, H. How Review Quality and Source Credibility Interacts to Affect Review Usefulness: An Expansion of the Elaboration Likelihood Model. *Inf. Syst. Front.* **2022**, *25*, 1513–1531.
22. Chakraborty, U.; Bhat, S. Credibility of online reviews and its impact on brand image. *Manag. Res. Rev.* **2018**, *41*, 148–164.
23. Filieri, R.; Hofacker, C.F.; Alguezaui, S. What makes information in online consumer reviews diagnostic over time? The role of review relevancy, factuality, currency, source credibility and ranking score. *Comput. Hum. Behav.* **2018**, *80*, 122–131.
24. Mackiewicz, J.; Yeats, D.; Thornton, T. The Impact of Review Environment on Review Credibility. *J. Bus. Res.* **2016**, *59*, 71–88.
25. Hung, S.W.; Chang, C.W.; Chen, S.Y. Beyond a bunch of reviews: The quality and quantity of electronic word-of-mouth. *Inf. Manag.* **2023**, *60*, 103777.
26. Aghakhani, N.; Oh, O.; Gregg, D. Beyond the Review Sentiment: The Effect of Review Accuracy and Review Consistency on Review Usefulness. In Proceedings of the International Conference on Information Systems (ICIS), Seoul, Korea, 10–13 December 2017.
27. Xie, K.L.; Chen, C.; Wu, S. Online Consumer Review Factors Affecting Offline Hotel Popularity: Evidence from Tripadvisor. *J. Travel Tour. Mark.* **2016**, *33*, 211–223.
28. Aghakhani, N.; Oh, O.; Gregg, D.G.; Karimi, J. Online Review Consistency Matters: An Elaboration Likelihood Model Perspective. *Inf. Syst. Front.* **2021**, *23*, 1287–1301.
29. Zhao, K.; Stylianou, A.C.; Zheng, Y. Sources and impacts of social influence from online anonymous user reviews. *Inf. Manag.* **2018**, *55*, 16–30.
30. Tran, V.D.; Nguyen, M.D.; Luong, L.A. The effects of online credible review on brand trust dimensions and willingness to buy: Evidence from Vietnam consumers. *Cogent Bus. Manag.* **2022**, *9*, 2038840.
31. Wu, H.H.; Tippomut, P.; Chung, H.F.; Chu, W.K. The mechanism of positive emotions linking consumer review consistency to brand attitudes: A moderated mediation analysis. *Asia Pacific J. Mark. Logist.* **2020**, *32*, 575–588.
32. Bigne, E.; Chatzipanagiotou, K.; Ruiz, C. Pictorial content, sequence of conflicting online reviews and consumer decision-making: The stimulus-organism-response model revisited. *J. Bus. Res.* **2020**, *115*, 403–416.
33. Gerrath, M.H.; Usrey, B. The importance of being earnest: Mandatory vs. voluntary disclosure of incentives for online product reviews. *Int. J. Res. Mark.* **2021**, *38*, 531–548.
34. Du Plessis, C.; Stephen, A.T.; Bart, Y.; Goncalves, D. When in Doubt, Elaborate? How Elaboration on Uncertainty Influences the Persuasiveness of Consumer-Generated Product Reviews When Reviewers Are Incentivized. *SSRN Electron. J.* **2016**, *59*, Available online: <http://dx.doi.org/10.2139/ssrn.2821641>.
35. Yin, H.; Zheng, S.; Yeoh, W.; Ren, J. How online review richness impacts sales: An attribute substitution perspective. *J. Assoc. Inf. Sci. Technol.* **2021**, *72*, 901–917.
36. Siering, M.; Muntermann, J.; Rajagopalan, B. Explaining and predicting online review helpfulness: The role of content and reviewer-related signals. *Decis. Support Syst.* **2018**, *108*, 1–12.
37. Jamshidi, S.; Rejaie, R.; Li, J. Trojan horses in amazon's castle: Understanding the incentivized online reviews. In Proceedings of the 10th IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2018), Barcelona, Spain, 28–31 August 2018; pp. 335–342.

38. Jia, Y.; Liu, I.L. Do consumers always follow “useful” reviews? The interaction effect of review valence and review usefulness on consumers’ purchase decisions. *J. Assoc. Inf. Sci. Technol.* **2018**, *69*, 1304–1317.
39. Tang, M.; Xu, Z.; Qin, Y.; Su, C.; Zhu, Y.; Tao, F.; Ding, J. A Quantitative Study of Impact of Incentive to Quality of Software Reviews. In Proceedings of the 9th International Conference on Dependable Systems and Their Applications (DSA 2022), Wulumuqi, China, 4-5 August 2022; IEEE; pp. 54–63.
40. Li, X.; Wu, C.; Mai, F. The effect of online reviews on product sales: A joint sentiment-topic analysis. *Inf. Manag.* **2019**, *56*, 172–184.
41. Basiri, M.E.; Ghasem-Aghaee, N.; Naghsh-Nilchi, A.R. Exploiting reviewers’ comment histories for sentiment analysis. *J. Inf. Sci.* **2014**, *40*, 313–328.
42. Reimers, N.; Gurevych, I. Sentence-bert: Sentence embeddings using siamese bert-networks. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2019), Hong Kong, China, 3-7 November 2019.
43. Hair, M.; Ozcan, T. How reviewers’ use of profanity affects perceived usefulness of online reviews. *Mark. Lett.* **2018**, *29*, 151–163.
44. Luo, C.; Luo, X.R.; Xu, Y.; Warkentin, M.; Sia, C.L. Examining the moderating role of sense of membership in online review evaluations. *Inf. Manag.* **2015**, *52*, 305–316.
45. Bi, S.; Liu, Z.; Usman, K. The influence of online information on investing decisions of reward-based crowdfunding. *J. Bus. Res.* **2017**, *71*, 10–18.
46. Janze, C.; Siering, M. ‘Status Effect’ in User-Generated Content: Evidence from Online Service Reviews. In Proceedings of the 2015 International Conference on Information Systems: Exploring the Information Frontier (ICIS 2015), Fort Worth, Texas, USA, 13-16 December 2015; pp. 1–15.
47. Chatterjee, S.; Chaudhuri, R.; Kumar, A.; Wang, C.L.; Gupta, S. Impacts of consumer cognitive process to ascertain online fake review: A cognitive dissonance theory approach. *J. Bus. Res.* **2023**, *154*, 113370.
48. Campagna, C.L.; Donthu, N.; Yoo, B. Brand authenticity: literature review, comprehensive definition, and an amalgamated scale. *J. Mark. Theory Pract.* **2023**, *31*, 129–145.
49. Xu, C.; Zheng, X.; Yang, F. Examining the effects of negative emotions on review helpfulness: The moderating role of product price. *Comput. Hum. Behav.* **2023**, *139*, 107501.
50. Luo, L.; Liu, J.; Shen, H.; Lai, Y. Vote or not? How language mimicry affect peer recognition in an online social Q&A community. *Neurocomputing* **2023**, *530*, 139–149.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.