Journal of
Information
Systems
Education

Volume 37
Issue 2
Spring 2026

Teaching Tip

Generative AI in Excel: A Hands-on Module for Text Mining With Microsoft Copilot

Shwadhin Sharma

Recommended Citation: Sharma, S. (2026). Teaching Tip: Generative AI in Excel: A Hands-on Module for Text Mining With Microsoft Copilot. *Journal of Information Systems Education*, 37(2), 218-242. https://doi.org/10.62273/KRAV4122

Article Link: https://jise.org/Volume37/n2/JISE2026v37n2pp218-242.html

Received:	September 29, 2025
First Decision:	November 7, 2025
Accepted:	December 12, 2025
Published:	June 15, 2026

Find archived papers, submission instructions, terms of use, and much more at the JISE website:
https://jise.org

ISSN: 2574-3872 (Online) 1055-3096 (Print)

Journal of Information Systems Education, 37(2), 218-242, Spring 2026
https://doi.org/10.62273/KRAV4122

Teaching Tip
Generative AI in Excel: A Hands-on Module for Text Mining With Microsoft Copilot

Shwadhin Sharma
College of Business
California State University, Monterey Bay
Seaside, CA 93955, USA
ssharma@csumb.edu

ABSTRACT

This teaching tip presents a step-by-step module that enables information systems students to perform text mining directly in Microsoft Excel using Microsoft Copilot. The approach lowers technical barriers by allowing natural language prompts to generate structured outputs, while maintaining a focus on interpretation and managerial communication. The module scaffolds eight techniques, progressing from word frequency analysis and sentiment classification to visualization, topic modeling, anomaly detection, correlation analysis, semantic similarity, and simple predictive modeling. Activities use a classroom-sized subset of the *Amazon Customer Reviews Dataset*, providing authentic text with rich metadata. Ready-to-use prompts, screenshots, and instructor notes highlight common pitfalls and demonstrate how prompt refinement improves results. The module offers an accessible pathway for integrating AI-enabled text analytics into IS courses, fostering both technical competence and critical evaluation of AI outputs.

Keywords: Microsoft Copilot, Text mining, Generative AI in education, Information systems pedagogy, Predictive modeling

1. INTRODUCTION

The ability to analyze unstructured text has become a central competency in information systems (IS). Organizations across sectors collect massive volumes of text data, including customer reviews, social media posts, call center transcripts, and survey responses. This information holds strategic value for understanding customer sentiment, detecting emerging issues, and guiding decision-making (Chowdhury, 2024; Miner et al., 2012). As a result, text mining and natural language processing (NLP) are widely recognized as essential skills for graduates entering data-intensive roles (Sharda et al., 2020).

Despite this importance, integrating text mining effectively into IS curricula remains challenging. Traditional approaches often rely on programming environments such as Python or R, which require technical prerequisites that many business-oriented students lack (Leidig & Salmela, 2021). At the same time, employers increasingly expect graduates to have foundational exposure to analytics involving unstructured data, emphasizing the ability to link technical results to managerial insights (Brynjolfsson et al., 2023). This underscores the need for pedagogical approaches that lower barriers to entry while exposing students to advanced analytics concepts.

Generative AI offers new opportunities to meet this need. Microsoft Copilot, integrated directly into Microsoft Excel, enables students to apply NLP and machine learning techniques through natural language prompts (Microsoft, 2024). These capabilities reflect a broader trend toward embedding AI directly within productivity software, allowing business users to perform advanced analytics without coding (Jorzik et al., 2024). Copilot can modify workbooks, add columns, and create visualizations, such as charts and dashboards, without requiring coding. By allowing students to perform tasks such as sentiment classification, summarization, and similarity search within a familiar environment, Copilot serves as both a facilitator and a scaffold for experiential learning (Kolb, 1984; Zhang et al., 2024).

For instructional purposes, the module utilizes the publicly available *Amazon Customer Reviews Dataset* (McAuley et al., 2015), which contains millions of product reviews across categories. Embedding AI-enabled text mining into Excel aligns with research emphasizing the importance of applied and experiential learning in IS education (Panda & Padhy, 2025). It supports motivation and self-efficacy by reducing technical barriers (Abid & Baxter, 2024).

This teaching tip presents a step-by-step module introducing eight text mining techniques in Microsoft Excel using Microsoft Copilot. These techniques are organized into three levels of difficulty: (1) basic tasks such as word frequency analysis, sentiment classification, and visualization and dashboards; (2) intermediate tasks, including topic modeling and predictive modeling; and (3) advanced tasks such as anomaly detection, correlation analysis, and semantic similarity. Together, they guide students through the full text mining pipeline, from pre-processing and exploration to predictive and diagnostic modeling.

The contribution of this teaching tip is twofold. First, it demonstrates a practical pathway for introducing advanced text mining concepts without programming prerequisites by leveraging a widely used IS tool. Second, it encourages critical reflection on AI-generated outputs, preparing students to apply these tools effectively while maintaining managerial judgment and ethical awareness. Compared with programming-based environments such as Python or R, the Copilot-in-Excel approach emphasizes accessibility and conceptual understanding rather than code syntax or library configuration. While traditional environments provide greater parameter control and reproducibility, Copilot allows students, especially those without programming experience, to perform comparable analyses through natural-language prompting, making an advanced text mining approachable within business-oriented courses.

2. PEDAGOGICAL FOUNDATIONS AND BACKGROUND

The design of this teaching tip rests on three interconnected pedagogical foundations: experiential learning, constructivist and cooperative learning, and the need for curricular alignment with industry expectations. Together, these principles guide the progression from basic to advanced text mining techniques and frame the learning objectives that anchor the module.

2.1 Experiential Learning Foundations

Experiential learning theory views learning as a cyclical process in which knowledge is constructed through experience, reflection, conceptualization, and experimentation (Kolb, 1984). For analytics education, this model is particularly relevant because abstract concepts, such as topic modeling or semantic similarity, are difficult to grasp without practical application. The Copilot-in-Excel module follows this cycle. Students begin by loading the dataset and applying prompts for tasks such as sentiment classification or frequency analysis. They then reflect on inconsistencies or errors in outputs, abstract lessons from these experiences, and apply more advanced techniques such as topic modeling and semantic similarity. In doing so, they see how AI-generated analytics can support managerial decision-making through iterative experimentation.

2.2 Constructivist and Cooperative Learning

Constructivist pedagogy emphasizes that learners build meaning through interaction with content, peers, and instructors (Piaget, 1970; Vygotsky, 1978). While this module is assessed individually, in-class discussion allows students to share interpretations, critique Copilot's reasoning, and recognize that AI outputs require human oversight (Mienye & Stewart, 2025; Love et al., 2025). The module's staged difficulty also reflects Vygotsky's (1978) zone of proximal development. Students begin with familiar tasks, such as frequency counts and visualizations, progress to more abstract tasks, like topic clustering, and conclude with complex, AI-driven analyses. This sequence challenges students appropriately and promotes confidence and deeper understanding.

2.3 Curricular Alignment and Industry Relevance

Excel has long served as a foundational tool for teaching analytics and data-driven decision making, with established instructional resources demonstrating its effectiveness in modeling, visualization, and introductory data mining (Powell & Baker, 2014; Willis, 2016). Because Excel is already embedded in many IS curricula, it provides a natural platform for introducing more advanced analytical thinking without requiring students to learn a new programming environment. This continuity is especially important in programs where students may take Information Systems as a concentration and may not have the technical background for advanced programming languages.

Accrediting bodies such as ABET and AACSB emphasize applied, project-based learning and expect graduates to demonstrate both technical and communication skills (ABET, 2022; AACSB, 2020). The IS 2020 curriculum similarly highlights data competencies, experiential learning, and the ability to translate analytical results into managerial insight (Leidig & Salmela, 2021). Extending Excel with AI-enabled features such as Copilot aligns with these expectations by enabling instructors to introduce text mining, machine learning concepts, and data storytelling within a tool that students already know.

At the same time, industry adoption of generative AI is accelerating. AI-enabled tools such as Microsoft Copilot, Power BI Copilot, and Azure OpenAI are increasingly used for text summarization, customer feedback analysis, sentiment monitoring, and automated reporting. Organizations in retail, customer service, and e-commerce routinely apply generative AI to classify reviews, extract themes from support tickets, and generate managerial summaries. Microsoft has documented enterprise use cases where Copilot assists with call-center transcript analysis, social media monitoring, and automated dashboard creation in Excel and Power BI. Similar applications are found across sectors, including telecommunications, consumer electronics, and financial services, where AI-assisted text analytics accelerates insight generation for business analysts without programming expertise.

As AI capabilities reshape data analysis workflows, employers are increasingly valuing graduates who can collaborate with generative AI tools to effectively interpret and communicate insights (Brynjolfsson et al., 2023). By aligning the module with these emerging industry practices, the teaching approach ensures that students gain experience with tools and workflows that increasingly mirror real-world analytics environments. This directly supports these goals by blending technical execution with managerial framing. Students learn to critique results, assess risks, and summarize outputs for decision-making audiences, reinforcing the dual emphasis on analytical skill and professional judgment.

3. MODULE PREPARATION AND REQUIREMENTS

3.1 Prerequisites and Pedagogical Requirements

The successful implementation of the Copilot-in-Excel text mining module requires that students enter the activity with certain baseline skills and adequate course preparation. Faculty should ensure that students have completed prior coursework or training that provides the following:

- *Excel proficiency*. Students should already be comfortable with Excel basics, including entering formulas, creating pivot tables, and generating simple charts.
- *Foundational statistics*. A working knowledge of descriptive statistics (mean, variance, correlation) and exposure to regression concepts are recommended.
- *Understanding data quality*. Students should be introduced to the importance of cleansing and shaping data before analysis. This includes recognizing null values, filtering invalid entries, and appreciating how incomplete or inconsistent data can bias results.
- *Capacity for critical reflection*. Students should be prepared to treat AI-generated results as starting points rather than definitive answers. Faculty are encouraged to create space for discussion after

each activity, allowing students to critique errors, inconsistencies, and the risks associated with over-reliance on automated analytics.

- *Flexibility for extensions*. Depending on the level of the course, faculty may extend the module beyond the basics. Introductory courses may conclude with sentiment classification and visualization, while advanced courses may incorporate anomaly detection, predictive modeling with additional features, or comparisons to outputs from Python or R.
- *Familiarity with authentic datasets*. While the provided Amazon reviews subset is a manageable entry point, students should be aware that text mining can also be applied to other domains. Faculty may substitute institutional survey data, customer service records, or social media comments, provided data privacy and ethics are addressed.

3.2 Copilot Requirements

Microsoft Copilot in Excel is an AI assistant that accepts natural language prompts and generates structured outputs such as new columns, tables, or charts. In this module, students use it for sentiment labeling, clustering, frequency extraction, similarity analysis, and predictive modeling. Copilot operates entirely within Excel's environment and does not access external NLP libraries such as NLTK or spaCy. Its analytical outputs should be viewed as AI-assisted transformations rather than configurable machine learning pipelines, as parameter tuning and reproducibility are limited. To successfully complete the tutorial activities, students require access to Microsoft 365 Copilot in Excel, available through both the desktop (Windows or macOS) and cloud-based Excel for the Web environments. A stable internet connection and an active institutional or individual Microsoft 365 subscription with Copilot enabled are necessary. Copilot features are not available in non-subscription or offline Excel versions.

3.3 Dataset Preparation

For this module, we utilized the publicly available *Amazon Customer Reviews Dataset* (McAuley et al., 2015), hosted by UCSD (University of California, San Diego) and documented at https://nijianmo.github.io/amazon/index.html. The complete dataset is one of the largest open collections of consumer product reviews, comprising approximately 100 GB of files and containing 233.1 million reviews across 29 product categories. Each review record contains structured metadata, including review ID, product ID, reviewer ID, star rating (1-5), review text, helpfulness votes, and timestamp, along with category-specific product information such as product title and category path. This richness makes the dataset suitable for both supervised and unsupervised text mining tasks.

To ensure feasibility in a classroom setting, we extracted a random data sample from four categories of datasets: *Electronics, Furniture, Grocery,* and *Gift Cards*. The teaching dataset contains a total of 40,232 rows**.** These categories were selected because they represent a broad range of product types, generating diverse vocabulary and sentiment expressions that are especially useful for demonstrating text mining techniques. The sample size strikes a balance between authenticity and manageability: large enough to reveal meaningful frequency patterns, co-occurrences, and topic clusters, yet small enough to remain responsive in Excel with Copilot during in-class exercises.

For instructional use, the subset was pre-processed into a comma-separated values (CSV) file containing 15 columns. For instructors who would like to adopt the exact subset used in this teaching tip, the prepared dataset can be accessed directly at https://drive.google.com/file/d/19pil13yNPeI8syXBo19-pou7OzOxvwnV/view?pli=1. For instructors who prefer to work with the complete dataset or explore other categories, the raw files are available at the GitHub/UCSD repository cited above. These files are provided in compressed **.tsv.gz** format and can be opened in Excel following the steps below:

1. Unzip the files from the downloaded folder using 7-Zip or any other decompression tool.
2. Open Excel**,** choose the *File tab,* then *Open*, and select the unzipped .tsv file.
3. When prompted, set the delimiter to "Tab."
4. Excel will display the dataset as a table. Choose the *File tab,* and then *Save As* and export the file as CSV (Comma-delimited, *.csv*).

5. If desired, merge multiple category files to create a larger dataset that better fits your instructional objectives.

4. DETAILED TEACHING APPROACH

The detailed teaching approach unfolds in a sequence of activities that gradually increase in complexity. Instructors are encouraged to guide students not only through the mechanics of generating outputs but also through reflection on why certain results appear, how to refine prompts when needed, and what business insights can be derived from the analysis.

4.1 Word Frequency Analysis

Students begin the module by loading the dataset into Excel and confirming that the customer reviews are stored in the *review_body* column. Figure 1 shows what the initiation of the data in Copilot looks like.

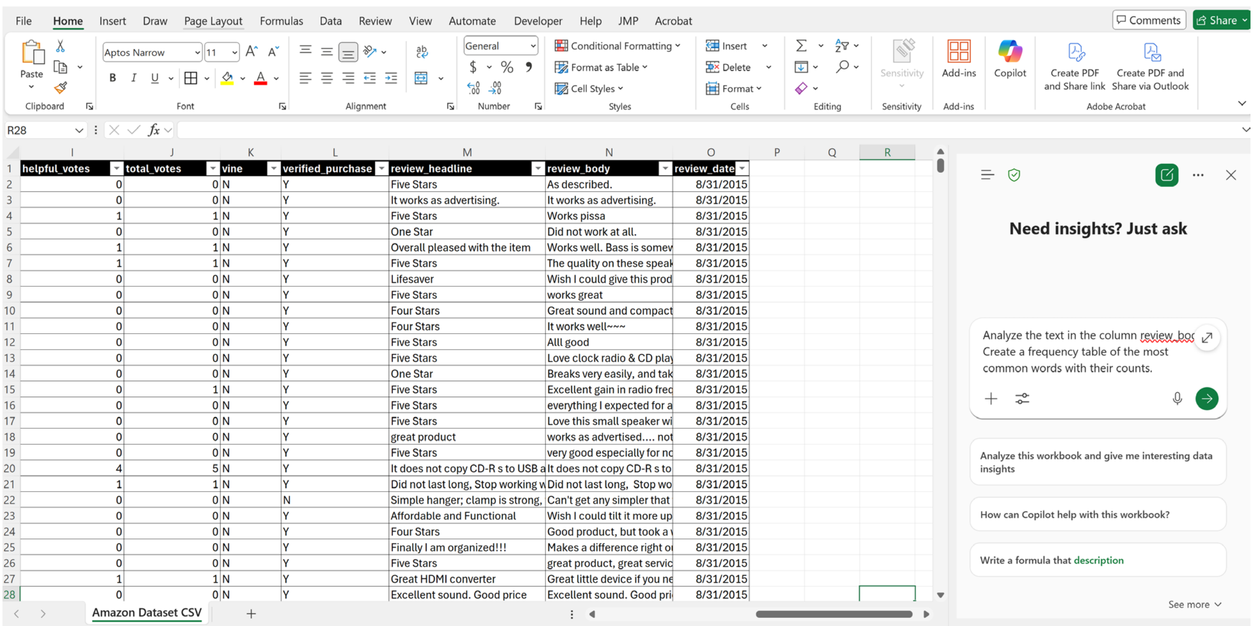

Figure 1. Initiation of the Data in Copilot Using Prompts

They then apply a basic Copilot prompt as follows: "*Analyze the text in the column review_body. Create a frequency table of the most common words with their counts.*"

As shown in Appendix A - Table A-1, the initial output is dominated by common words such as *it, to, for, of, with, and, very*. Although these words occur frequently, they carry little analytical value. Students typically see filler words overwhelming the results. The instructor uses this as a teaching point to explain that these are referred to as *stopwords* in text mining. Recognizing the limitations of a naïve approach helps students understand that text pre-processing is a necessary step before meaningful analysis can occur.

Some students may attempt to improve the prompt by asking Copilot to exclude stopwords using external NLP libraries. For example, they might type: "*Analyze the review_body column and return a frequency table of the most common words, excluding all standard English stopwords using NLTK or spaCy. That means you are excluding common stopwords such as the, is, and, was, my, this, that and generic adjectives like good, great, nice, bad, awesome, excellent.*"

When they do so, Copilot often responds with an error message: "*NLTK stopwords couldn't be downloaded due to a connection error. spaCy's English language model (en_core_web_sm) is not available in the current environment.*"

This is used as a teaching moment. The instructor explains that Copilot, like other generative AI tools, sometimes attempts to use external resources and libraries. When those are not available, the system defaults to simpler methods. This demonstrates that AI assistants are not self-contained black boxes but are constrained by their environment. To achieve better results, the students are asked to refine their instruction with a more specific prompt: "*Analyze the review_body column and return a frequency table of the top-most common words. Focus on nouns or product-related terms strictly. Read the whole table before you give the output.*"

With this refinement, Copilot produces a more useful list of terms, including *speaker, headphones, TV, cable, battery, charger,* and *price*. These results, illustrated in Appendix A - Table A-2, highlight product attributes and experiences rather than filler words. The transition from meaningless to meaningful terms reinforces the importance of pre-processing and careful instruction in text mining.

Students then extend the analysis by asking Copilot to "*break down these frequencies by product category.*" For example, audio-related terms such as *speaker* and *headphones* appear almost exclusively in the Electronics category, while *Gift Cards* surface fewer product-specific mentions. Appendix A - Table A-3 presents an example of this category-level comparison.

As a possible future extension, students who wish to explore further may filter by star_rating to generate separate frequency tables for high- and low-rated reviews. This extension highlights how negative reviews (1-2 stars) often contain terms such as *return* or *poor quality*, while positive reviews (4-5 stars) often contain terms such as *love* or *fast delivery*.

4.2 Sentiment Classification

After exploring word frequencies, students continue their analysis by applying sentiment classification to the same set of customer reviews. They begin by instructing Copilot with the following prompt: "*Classify each review in the column review_body as Positive, Neutral, or Negative. Insert the results in a new column.*"

Copilot generates a new column that labels each review with one of the three sentiment categories. Appendix A – Table A-4 shows an approximate output that may show up with the prompt above. At first glance, the results appear reasonable, as many five-star reviews are assigned a Positive Rating and one-star reviews are marked as Negative. However, a closer inspection reveals that the classifications are not always accurate. Some neutral reviews, such as "*Item arrived on time,*" are marked as Positive, while other vague statements like "*It works*" may be misclassified depending on the context. Reviews with mixed sentiment, such as "*The sound is good, but the battery does not last,*" are particularly difficult for Copilot to classify correctly because they contain both praise and criticism. These inconsistencies are highlighted in the discussion, and the instructor uses them to remind students that sentiment analysis is often approximate and that AI outputs must always be critically evaluated.

To improve the quality of results, the instructor directs students to issue a more explicit prompt: "*Analyze the text in review_body. Classify each review as Positive, Neutral, or Negative based strictly on sentiment toward the product. Do not classify reviews as Positive if they contain negative experiences. If a review contains both praise and criticism, classify it as Neutral. Insert the result in a new column called Sentiment_Refined.*"

With this refinement, Copilot produces a more balanced classification, as can be seen in Appendix A – Table A-5. Reviews that are clearly favorable (e.g., "*Love the sound quality and design*") are labeled Positive, those that express dissatisfaction (e.g., "*Battery stopped charging after a week*") are labeled Negative, and those with mixed tone are assigned to Neutral. Students then compare these refined classifications against the existing star_rating column using the following prompt: "*Compare the Sentiment_Refined column with star_rating. Generate a summary table showing the number of Positive, Neutral, and Negative classifications for each star_rating (1 through 5).*"

Copilot produces a crosstab table that allows students to see how closely textual sentiment aligns with structured star ratings (see Appendix A – Table A-6). Typically, four- and five-star reviews cluster under "Positive," one- and two-star reviews cluster under "Negative," and three-star reviews scatter between "Neutral" and "Positive."

The instructor asks students to interpret this comparison. They note that while sentiment classifications generally track star ratings, discrepancies exist because text comments capture nuance that star ratings cannot. For example, some three-star reviews may be classified as Positive if the text emphasizes features the customer liked. In contrast, certain four-star reviews may fall into Neutral if the customer expressed reservations. These patterns spark a discussion of why sentiment analysis can be valuable as a complement to numerical ratings, providing managers with deeper insights into customer experiences.

4.3 Visualization and Dashboards

After completing sentiment classification, students continue by transforming those categorical outputs into visuals suitable for managerial communication. They begin with a direct instruction to Copilot that focuses attention on overall sentiment proportions. The student typed: "*Using the Sentiment column, create a pie chart showing the proportion of Positive, Neutral, and Negative reviews.*" Copilot renders a chart that displays the distribution across the three classes. The instructor briefly discusses whether a pie chart is the most effective representation for this purpose and encourages students to consider alternatives as part of their data storytelling (see Appendix B – Figure B-1).

To investigate whether attitudes shift across time, students then ask Copilot to show temporal dynamics with a minimal prompt: "*Visualize sentiment trends over time.*" Copilot typically responds with a line or clustered column chart that aggregates Positive, Neutral, and Negative counts (or percentages) by review date. Students interpret whether periods of heightened negative sentiment coincide with known events, such as product launches or shipping constraints, and are asked to articulate what follow-up questions a manager might ask upon observing those patterns (see Appendix B – Figure B-2).

The class next explores whether verification status is associated with different sentiment profiles. Students instruct Copilot to produce side-by-side visuals with explicit formatting guidance using the exact command from the draft: "*Compare verified vs. non-verified reviews. Give me two charts. Have these charts side by side. Add data labels/values.*"

Copilot places two visuals adjacent to one another so students can compare distributions at a glance. The instructor invites students to note differences in Positive/Negative proportions and to consider plausible explanations (e.g., verified purchases may skew toward more authentic experiences), while also highlighting the need for caution when inferring causality from descriptive displays (see Appendix B – Figure B-3).

Because many stakeholders also want to understand how ratings vary by category and verification status, students extend the view with a categorical summary by issuing: "*Give me a chart that reviews the ratings by product category based on verified vs. unverified purchases.*"

Copilot produces a comparative view that cross-references star_rating with product_category while splitting by verification status. Here, the instructor emphasizes how visualizing structured fields (ratings, categories, verification) alongside text-derived outputs (sentiment) enables a richer, more managerial narrative about customer experience (see Appendix B – Figure B-4).

Students now "*Visualize the breakdown of word frequency by category in a chart. Add data labels/values for the frequency bar*" using Copilot. The output could look something as shown in Appendix B – Figure B-5.

The class will now build an integrated dashboard of all the charts generated by the prompts discussed above: "Create a dashboard combining all the charts created above. Arrange them with clear titles and a professional layout." The output could look something as shown in Appendix B – Figure B-6.

As an optional exercise, students can filter by product category (Electronics, Furniture, Grocery, Gift Cards) and create category-specific dashboards. For example, they can filter rows where product_category = "Electronics," re-run the sentiment pie chart and frequency bar chart, and place them on a dashboard titled "Electronics Reviews Only." This demonstrates how dashboards can reveal differences across product types.

4.4 Simple Predictive Modeling

The final activity in the basic sequence introduces students to the concept of supervised machine learning by demonstrating how text-derived features can be utilized to predict structured business outcomes. Here, the outcome of interest is the review star rating, and the predictive feature is the sentiment classification generated in the previous step. The goal is for students to recognize the link between qualitative text and quantitative ratings while also confronting the limitations of simple models.

Students begin by identifying the key variables. The feature is the *Sentiment* column (Positive, Neutral, Negative), while the target is the *star_rating* column (ranging from 1 to 5). The instructor emphasizes that students should expect some alignment immediately. Positive reviews will often correspond to higher ratings, while Negative reviews should align with lower ones.

To confirm this baseline relationship, students first prompt Copilot: "*Compare Sentiment with star_rating. Create a summary table showing the number of reviews and the average star_rating for each sentiment category.*"

Copilot produces a summary table similar to the one shown in Appendix A – Table A-7. This result confirms a clear correlation: Positive sentiment is strongly associated with higher star ratings, while Negative sentiment aligns with lower ones.

After that, students subsequently conduct predictive modeling exercises. They provide Copilot with the instruction: "*Using column named 'Sentiment' as the predictor and 'star_rating' column as the target, build a simple predictive model. Insert predicted star ratings in a new column called Predicted_Rating. Report the overall accuracy of the model as a percentage.*"

Copilot responds by inserting a *Predicted_Rating* column into the spreadsheet and reporting an accuracy score. In one run, the model achieved an accuracy of 83.9% (see Appendix A – Table A-8). Students can download and review the updated file, which includes predicted star ratings in a new column.

The instructor then guides students to probe further by requesting diagnostic information. They type: "*Show the confusion matrix for the model.*"

Copilot generates a table where rows represent the actual star ratings, columns represent predicted ratings, and each cell contains the number of reviews that fall into that actual-predicted pair (see Appendix B – Figure B-7). The instructor explains how to read the confusion matrix and discusses what it reveals about the model's strengths and weaknesses.

At this stage, the class reflects on limitations. The instructor asks: "Why might a review classified as Positive still have only three stars?" Students recognize that mixed or qualified reviews (e.g., "Good quality, but overpriced") and vague comments (e.g., "It's fine") are often misaligned with a simple Positive/Negative categorization. The activity reinforces that predictive modeling of sentiment is powerful but also reductive.

For those seeking an optional extension, students may enrich the model by incorporating additional features. They are asked to type: "*Build a predictive model using both Sentiment and the top 10 frequent words as predictors to estimate star_rating. Report whether this improves accuracy.*"

Copilot responds by retraining the model with multiple predictors, introducing students to the idea of multi-feature models while still keeping the activity manageable within Excel. This extension offers a preview of more advanced modeling concepts, eliminating the need for programming or external tools.

4.5 Topic Modeling (Theme Clustering)

Students now progress from examining individual words and phrases to identifying broader themes across the dataset. The goal is to help them see how thousands of reviews can be clustered into a handful of meaningful categories that summarize customer concerns. Topic modeling is introduced as a method to reveal patterns that are not immediately apparent from individual terms, with the instructor noting that more advanced techniques, such as Latent Dirichlet Allocation (LDA), are commonly employed in research for this purpose.

Students begin by ensuring the dataset is open in Excel and that the free-text reviews are located in the *review_body* column. They then issue an initial clustering instruction: "*Cluster the reviews in the* *column review_body into 4-6 major themes or topics. Assign each review a topic label in a new column called Topic.*"

Copilot responds by creating new columns with labels such as *"Value & Quality," "Ease of Use & Convenience,"* and *"Brand Loyalty."* The output could look something like what is shown in Appendix A – Table A-9.

Some topic assignments were conceptually coherent, whereas others were less consistent across reviews. For instance, a review about easy setup was assigned to *Product Satisfaction & Performance*. The instructor explains that topic modeling is inherently probabilistic, meaning reviews may address multiple themes, but Copilot simplifies them into a single label. This opens a discussion about the trade-offs inherent in clustering (see Appendix A – Table A-10).

To better understand the relative size of each cluster, students next ask Copilot: "*Summarize the number of reviews in each Topic cluster. Create a bar chart showing the size of each cluster.*"

Copilot produces a summary table and a corresponding bar chart. An example output may look like the following (see Appendix A – Table A-11). Also, the bar chart showing the distribution across these clusters is shown in Appendix B – Figure B-8. This result indicates that many reviews remain unclassified; however, the identified clusters still capture a significant portion of the dataset.

The instructor then encourages students to interpret the findings. They are asked: *Which themes dominate the reviews? What do these themes suggest about what customers care most about?* A discussion follows about the limitations of assigning reviews to a single theme when they may address multiple issues; for example, a comment like *"pricey but fast shipping"* could logically belong in both *the Price/Value and Delivery Experience* categories. The teaching point emphasizes that clustering is a way of reducing complexity: instead of reading 40,232 reviews individually, managers can focus on a handful of major themes that summarize customer concerns.

4.6 Predictive Modeling and Anomaly Detection

The next stage of the module introduces students to predictive modeling and anomaly detection as methods for uncovering patterns that go beyond descriptive analysis. These techniques enable students to estimate outcomes, such as helpfulness or ratings, and to flag suspicious cases where structured and unstructured signals do not align. The aim is to help students appreciate the predictive power of review features while also recognizing the limitations of simple models.

4.6.1 Anomaly Detection. Students first explore anomaly detection by identifying reviews in which the star rating and textual sentiment appear to conflict. They prompt Copilot with: "*Identify suspicious reviews (e.g., high rating but negative text).*"

Copilot highlights 1,139 reviews with this mismatch, flagging them as potentially anomalous. Students then summarize common issues by requesting: "Summarize common issues in suspicious reviews." The most frequently mentioned keywords in these reviews include: broken, returned, waste, disappointed, cheap, problem, poor, bad, stopped, and junk. These terms suggest that product failures and poor quality are frequent in anomalous reviews despite high star ratings. The instructor uses this output to discuss possible explanations, such as rating errors, sarcasm, or incentivized reviews.

4.6.2 Predictive Modeling. Building on this, students apply predictive modeling to estimate outcomes. They are guided to use features such as review length, sentiment polarity, and product category to predict helpfulness votes. Copilot is instructed: "*Using features like review length, sentiment polarity, and product category, build a model to predict helpful_votes.*"

Copilot responds with a Random Forest Regressor, reporting a mean squared error (MSE) of 30.48 and an R^2 score of -0.49. The instructor emphasizes that the negative R^2 indicates poor predictive performance and explains potential reasons, such as high variance in helpful votes and the absence of features like reviewer reputation or review time.

Students then turn to predicting star ratings as an outcome variable. They first built a regression model with the same features, obtaining an MSE of 1.17 and an R^2 score of 0.24. The model shows moderate predictive power. Sentiment and review length do help estimate ratings, but there is room for improvement.

Copilot then switches to a classification approach (instead of regression) at the students' request: "*Predict star ratings as discrete categories (1 to 5 stars).*"

The model achieves 67% accuracy, with high precision and recall for 5-star reviews, but weaker performance for 1-4-star reviews due to class imbalance (see Appendix A – Table A-12). The instructor asks students why the model struggles with minority classes and introduces balancing techniques as possible solutions for future exploration. Features such as review length, sentiment polarity, verified purchase, and product category do help, but more nuanced features (e.g., review text embeddings, reviewer history) could further improve performance.

Next, students explore simpler regression relationships. To examine how review length influences voting behavior, they provide the instruction: "*Explore how review length influences total_votes using regression.*"

Copilot applies an ordinary least squares (OLS) model with total_votes as the dependent variable and review_length as the predictor. The model produces an R^2 of 0.063, indicating that about 6.3% of the variation in total votes is explained by review length. The coefficient of 0.0048 suggests that longer reviews attract slightly more votes, although other factors drive most of the variation. The relationship is statistically significant (p-value < 0.001). A scatter plot with a regression line is generated to illustrate this modest but statistically significant trend (see Appendix B – Figure B-9).

As shown above, there is a positive and statistically significant relationship between review length and the total number of votes. However, the effect size is small, and most of the variation in votes is due to other factors not captured by review length alone.

Finally, students examine whether sentiment polarity predicts helpful votes. They type: "*Analyze if more positive or negative reviews get more helpful votes using regression.*"

Copilot builds a model with *helpful_votes* as the dependent variable and *sentiment_score* as the predictor. The model achieves an R^2 of 0.004, with a negative coefficient (-1.0718). Only 0.4% of the variation in helpful votes is explained by the sentiment score. Additionally, this suggests that negative reviews are slightly more likely to be marked as helpful, although the effect size is small. The relationship is statistically significant, though the effect size is small (p-value < 0.001). The instructor uses this finding to stimulate discussion about why readers may perceive critical reviews as more trustworthy or informative (see Appendix B – Figure B-10).

Based on Appendix B – Figure B-10, negative reviews are slightly more likely to be marked as helpful. This could reflect a tendency for users to find critical feedback more informative or trustworthy. However, the low R^2 value suggests that many other factors influence helpful votes (e.g., review length, product type, verified purchase status).

Through this sequence of activities, students are introduced to the predictive and diagnostic power of machine learning models applied to customer review data. They also learn that predictive performance varies by outcome and feature selection, and that limitations such as class imbalance, variance, and omitted variables must always be acknowledged. Anomaly detection highlights the importance of validating model outputs against real-world behavior. At the same time, predictive modeling demonstrates how text-derived features can be operationalized to forecast outcomes that matter to managers.

4.7 Descriptive Relationship and Consistency Analysis

In addition to predictive tasks, students also examine descriptive techniques that uncover relationships among review features and test logical consistency between different text fields. These methods emphasize exploration and validation rather than prediction, reinforcing the importance of diagnosing how variables interact and whether review content is coherent.

The activity begins with correlation analysis. Students are asked to explore relationships between quantitative and categorical features such as star ratings, helpful votes, verification status, and review length. Using Copilot, they provide the following prompt: "*Explore relationships between star_rating, helpful_votes, verified_purchase, and review length. Create a correlation heatmap.*"

Copilot generates a heatmap along with a correlation matrix (see Appendix B – Figure B-11). The correlation between star rating and helpful votes was negative (r = –0.07), indicating that lower-rated items tended to receive slightly more helpful votes. The correlation between star rating and verified purchase was weakly positive (r = 0.06), and the correlation between star rating and review length was slightly negative (r = –0.12), indicating that longer reviews were often associated with lower ratings. Students also noted that helpful votes and review length were positively correlated (r = 0.25), while verified purchases and review length were negatively correlated (r = –0.22). The instructor guided a discussion about how even small correlations can reveal behavioral tendencies, while cautioning students not to overinterpret weak relationships.

The next exercise focuses on text entailment, where students check whether review headlines logically follow from the review body. They issue the instruction: "*Check if the review headline logically follows from the review body. Report a similarity score between 0 and 1, and flag potential mismatches.*"

Copilot responds by creating a *HeadlineBodySimilarity* column with similarity scores (which is a score from 0 to 1) and a *Potential_Mismatch* column flagged as True when similarity is below 0.2 (see Appendix A – Table A-13).

Students observe that many mismatches occur when short or vague headlines, such as "Great!," are paired with more nuanced review bodies, or when headlines emphasize one feature while the body text highlights another. The instructor explains that entailment is a method for validating whether structured or short text fields accurately reflect the substance of the longer review, which is crucial when headlines are used in recommendation systems or summaries.

The instructor then invites students to reflect on the managerial implications of these analyses. Correlation analysis reveals patterns such as longer reviews attracting more helpful votes, while entailment analysis highlights inconsistencies that may reduce trust in review platforms. Together, these descriptive methods emphasize that analytics is not always about prediction; sometimes, the value lies in identifying subtle relationships and ensuring the coherence of data inputs.

5. EVALUATION AND EVIDENCE OF EFFECTIVENESS

To assess the effectiveness of the Copilot-in-Excel text mining module, we collected end-of-semester survey responses and student reflections from three sections across one course (n = 63). The module was implemented in an upper-level Information Systems and Analytics course offered in a hybrid format at a mid-sized public university. The course focuses on the end-to-end data pipeline, including extract–transform–load (ETL) processes, exploratory data analysis, sentiment analysis, and hypothesis testing. Students were asked to rate each of the techniques introduced in the module in terms of ease of use, contribution to learning, and overall value for understanding text mining in Excel.

Table 1 summarizes mean responses. All items were measured on a 1–5 Likert-type scale, where 1 represents Strongly Disagree and 5 represents Strongly Agree. M represents the sample mean, and SD represents the standard deviation. Results indicate that students rated the module highly across all techniques, with sentiment classification and visualization receiving the highest marks for ease of use. At the same time, semantic similarity and topic modeling were viewed as the most valuable for learning. Predictive modeling and anomaly detection were noted as more challenging, but students still recognized their value for linking unstructured data to business insights.

Students' qualitative feedback reinforced these findings. One student commented, *"I didn't expect Excel to handle tasks like topic modeling or semantic similarity. Copilot made it approachable."* Another noted, *"Sometimes the labels weren't perfect, but that actually helped me think more critically about how AI works."* Several students remarked that Copilot was sometimes inconsistent with prompts but mostly easy to use. In contrast, others emphasized that the skills they practiced would be valuable to highlight on their résumés. A number of students also observed that visualization and dashboards helped them "*see the bigger picture.*" At the same time, predictive modeling gave them "*a taste of what machine learning looks* *like in practice.*" Collectively, these reflections suggest that the module not only improved technical competence but also encouraged students to critically assess AI outputs and recognize the professional value of these techniques.

Technique	Ease of Use (M, SD)	Learning Contribution (M, SD)	Overall Value (M, SD)
Word frequency	4.40 (0.55)	4.20 (0.61)	4.30 (0.58)
Sentiment classification	4.45 (0.52)	4.35 (0.60)	4.42 (0.57)
Visualization/dashboards	4.50 (0.49)	4.28 (0.63)	4.38 (0.55)
Predictive modeling	3.90 (0.70)	4.05 (0.66)	4.00 (0.62)
Topic modeling	4.05 (0.65)	4.40 (0.59)	4.32 (0.61)
Anomaly detection	3.85 (0.72)	4.10 (0.64)	4.05 (0.67)
Correlation analysis	4.20 (0.61)	4.18 (0.62)	4.25 (0.60)
Semantic similarity	3.95 (0.68)	4.50 (0.55)	4.40 (0.58)

Table 1. Student Evaluations of Techniques in the Copilot-in-Excel Module (n = 63)

Overall, the evaluation demonstrates that the Copilot-in-Excel module provided both accessibility and depth. Simpler tasks, such as word frequency and sentiment analysis, built confidence. Advanced tasks, such as semantic similarity and topic modeling, offered students exposure to techniques they had not encountered previously, underscoring the pedagogical value of combining breadth with critical reflection.

6. DISCUSSION, LIMITATIONS, AND CONCLUSION

The Copilot-in-Excel text mining module provided students with a structured and accessible way to engage with advanced analytics concepts, including sentiment classification, visualization, predictive modeling, topic modeling, anomaly detection, correlation analysis, and semantic similarity. In contrast to conventional text-mining instruction using Python or R, the Copilot-in-Excel module trades some analytical precision for ease of adoption and classroom feasibility. This comparison helps instructors recognize that the module's primary value lies in lowering technical barriers while fostering critical reflection on AI-generated results rather than in replicating full-scale programming workflows. Evaluation results demonstrated that students rated the techniques highly in terms of ease of use, learning contribution, and overall value, with particularly strong gains in advanced methods that had previously been inaccessible to them. Qualitative feedback confirmed that students found the Copilot approachable and relevant to their professional development, even as they recognized its occasional inconsistencies and limitations.

In addition to supporting conceptual understanding, the module also provides opportunities for students to practice evaluating and validating AI-generated analytics. Because Copilot occasionally produces outputs that require human interpretation, instructors can encourage students to perform quick verification checks as part of their workflow. These include comparing sentiment labels with star ratings, reading a small sample of reviews within each topic cluster to assess coherence, and reviewing a few anomaly cases to determine whether they reflect meaningful inconsistencies. Framing these steps as part of the analytical process reinforces the broader learning objective of developing students' critical thinking and judgment when working with generative AI tools.

For instructors, the module demonstrates how emerging AI tools can be integrated into traditional platforms, such as Excel, to bridge the gap between technical depth and classroom feasibility. By scaffolding activities from simple (word frequency, sentiment) to complex (topic modeling, semantic similarity), the module enabled students to build confidence while also being exposed to techniques that connect directly to business analytics practice. Deliverables, such as dashboards and reflections, ensured that students not only performed analyses but also interpreted their outputs critically and communicated results in a professional format. This Copilot-in-Excel text mining module is particularly well-suited for courses that emphasize introductory or intermediate data analytics, business intelligence, or applied information systems. Because the module does not require students to use a programming language, it provides an accessible entry point for courses focused on data preparation, exploratory analysis, text mining fundamentals, or decision-support applications. This makes the approach especially valuable in business, management, and information systems curricula where Excel remains a primary analytical tool.

At the same time, several limitations should be noted. Copilot's responses were occasionally inconsistent, especially when handling prompts that required stopword removal or nuanced classification. Because Copilot's large language model evolves over time, outputs may also vary slightly across users, sessions, or software versions, introducing challenges for reproducibility. Instructors may need to demonstrate refinements and verification strategies, such as manually cross-checking AI outputs, to ensure reliable results. The dataset used in this module was drawn from a sample of Amazon reviews, and while authentic, it does not represent the full diversity of text data students might encounter in other business domains. In addition, the evaluation was conducted across three sections within one institutional context, which may limit the generalizability of findings. In future research, the sample size can be expanded by incorporating additional course sections and multiple semesters to improve the generalizability of the findings. The module also does not provide a direct comparison between Copilot and programming-based tools such as Python or R, which future studies could explore to quantify relative pedagogical and analytical benefits. Finally, access to Copilot in Excel requires institutional licensing, which may not be uniformly available across universities, and the extent of industry adoption for text mining applications remains an emerging area of evidence.

Despite these challenges, the module offers an effective model for integrating generative AI into IS education. Future extensions could include incorporating additional features into predictive modeling, exploring other forms of anomaly detection, or expanding the module to compare Copilot results with outputs from Python or R-based text mining. Overall, this teaching tip demonstrates that AI-driven tools can democratize access to advanced analytics, enabling students to learn not only how to perform techniques but also how to question, evaluate, and apply them in managerial contexts.

7. REFERENCES

AACSB. (2020). *2020 Guiding Principles and Standards for Business Accreditation*. https://www.aacsb.edu/educators/accreditation/business-accreditation/aacsb-business-accreditation-standards

ABET. (2022). *Criteria for Accrediting Computing Programs*. https://www.abet.org/wp-content/uploads/2023/05/C001_CAC-Criteria_2023-2024.pdf

Abid, A., & Baxter, S. L. (2024). Breaking Barriers in Behavioral Change: The Potential of Artificial Intelligence-Driven Motivational Interviewing. *Journal of Glaucoma*, 33(7), 473-477. https://doi.org/10.1097/IJG.0000000000002382

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at Work. *National Bureau of Economic Research*. https://doi.org/10.3386/w31161

Chowdhury, R. H. (2024). Sentiment Analysis and Social Media Analytics in Brand Management: Techniques, Trends, and Implications. *World Journal of Advanced Research and Reviews*, 23(2), 287-296. https://doi.org/10.30574/wjarr.2024.23.2.2369

Kolb, D. A. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Prentice Hall.

Leidig, P. M., & Salmela, H. (2021). *A competency model for undergraduate programs in information systems*. Association for Computing Machinery. https://doi.org/10.1145/3460863

Love, A. S., Niu, C., & Labay-Marquez, J. (2025). Artificial Intelligence in Public Health Education: Navigating Ethical Challenges and Empowering the Next Generation of Professionals. *Health Promotion Practice*. https://doi.org/10.1177/15248399251320989

Jorzik, P., Klein, S. P., Kanbach, D. K., & Kraus, S. (2024). AI-Driven Business Model Innovation: A Systematic Review and Research Agenda. *Journal of Business Research*, 182, 114764. https://doi.org/10.1016/j.jbusres.2024.114764

McAuley, J. J., Pandey, R., & Leskovec, J. (2015). Inferring Networks of Substitutable and Complementary Products. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). https://doi.org/10.1145/2783258.2783381

Microsoft. (2024, November 19). *Copilot in Excel: Transforming Data Analysis*. Microsoft Community Hub. https://techcommunity.microsoft.com/blog/excelblog/copilot-in-excel-transforming-data-analysis/4303611

Mienye, I. D., & Swart, T. G. (2025). ChatGPT in Education: A Review of Ethical Challenges and Approaches to Enhancing Transparency and Privacy. *Procedia Computer Science*, 254, 181-190. https://doi.org/10.1016/j.procs.2025.02.077

Miner, G., Elder, J., Fast, A., Hill, T., & Nisbet, R. (2012). *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications.* Academic Press.

Panda, S. P., & Padhy, A. (2025). *Business Intelligence With Power BI and Tableau: Cloud-Based Data Warehousing, Predictive Analytics, and Artificial Intelligence-Driven Decision Support.* Deep Science Publishing. https://doi.org/10.70593/978-93-7185-811-3

Piaget, J. (1970). *The Science of Education and the Psychology of the Child* (D. Coltman, Trans.). Orion Press.

Powell, S. G., & Baker, K. R. (2014). *Management Science: The Art of Modeling With Spreadsheets* (4th ed.). Wiley.

Sharda, R., Delen, D., & Turban, E. (2020). *Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support* (11th ed.). Pearson.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes.* Harvard University Press.

Willis, V. F. (2016). A Model for Teaching Technology: Using Excel in an Accounting Information Systems Course. *Journal of Accounting Education*, 36, 87-99. https://doi.org/10.1016/j.jaccedu.2016.05.002

Zhang, X., Zhang, P., Shen, Y., Liu, M., Wang, Q., Gašević, D., & Fan, Y. (2024). A Systematic Literature Review of Empirical Research on Applying Generative Artificial Intelligence in Education*.* *Frontiers of Digital Education*, 1(3), 223-245. https://doi.org/10.1007/s44366-024-0028-5

AUTHOR BIOGRAPHY

Shwadhin Sharma is an Associate Professor in the College of Business at California State University, Monterey Bay. He received his Ph.D. in Management and Information Systems from Mississippi State University. His research interests include big data analytics, privacy and security, technology empowerment, and dispositional factors in IT. His work has been published in the *Journal of Computer Information Systems*, *Government Information Quarterly*, *Information Technology & People*, *Behaviour & Information Technology*, *Electronic Commerce Research and Applications*, and *Computers & Security*. He also serves on the editorial boards of several journals and conferences**.**

APPENDICES

Appendix A. Sample Output Tables

Word	Count
It	34,384
To	31,675
For	17,675
Of	15,802
Br (an HTML tag as in)	9,878

Table A-1. Word Frequency of the Sample Data (Shorter Screenshot)

Word	Frequency
Product	3755
Price	3155
Quality	3038
Sound	2341
Speaker	1100
Headphones	1077
Tv	1021

Word	Frequency
cable	792
battery	598
earbuds	266
mount	223
charger	208
adapter	156
mic	85

Table A-2. Refined Word Frequency of the Sample Data

Category	Electronics	Furniture	Gift Card	Grocery
Product	1,303	1,049	131	1,272
Price	1,125	1,369	20	641
Quality	1,347	1,283	35	373
Sound	2,309	26	4	2
Speaker	1,091	9	0	0
Headphones	1,076	1	0	0
Tv	791	219	3	8
Cable	763	29	0	0
Battery	589	7	1	1
Earbuds	266	0	0	0
Mount	189	34	0	0
Charger	206	2	0	0
Adapter	154	0	0	2
Mic	85	0	0	0

Table A-3. Word Frequency by Product Category

review_body	review_date	sentiment
As described.	8/31/2015 0:00	Positive
It works as advertising.	8/31/2015 0:00	Positive
Works pissa	8/31/2015 0:00	Positive
Did not work at all.	8/31/2015 0:00	Negative
Works well. Bass is somewhat lacking but is present. Overall pleased with the item.	8/31/2015 0:00	Positive
The quality on these speakers is insanely good and doesn’t sound muddy when adjusting bass. Very happy with these.	8/31/2015 0:00	Positive
Wish I could give this product more than five stars. Lifesaver.	8/31/2015 0:00	Positive
works great	8/31/2015 0:00	Positive
Great sound and compact. Battery life seems good. Happy with this product.	8/31/2015 0:00	Positive

Table A-4. Sentiment Classification (Shorter Screenshot)

review_body	review_date	Sentiment_Refined
As described.	8/31/2015 0:00	Neutral
It works as advertising.	8/31/2015 0:00	Neutral
Works pissa	8/31/2015 0:00	Neutral
Did not work at all.	8/31/2015 0:00	Negative
Works well. Bass is somewhat lacking but is present. Overall pleased with the item.	8/31/2015 0:00	Neutral
The quality on these speakers is insanely good and doesn't sound muddy when adjusting bass. Very happy with these.	8/31/2015 0:00	Positive
Wish I could give this product more than five stars. Lifesaver.	8/31/2015 0:00	Neutral
works great	8/31/2015 0:00	Positive
Great sound and compact. Battery life seems good. Happy with this product.	8/31/2015 0:00	Positive

Table A-5. Refined Sentiment Classification (Shorter Screenshot)

Star Rating	Negative	Neutral	Positive
1	3,133	0	0
2	1,560	0	0
3	0	2,382	0
4	0	0	4,934
5	0	0	28,223

Table A-6. Sentiment Classification Against Star Rating

Sentiment	Number of Reviews	Average Star Rating
Negative	4,693	1.33
Neutral	2,382	3.00
Positive	33,157	4.85

Table A-7. Average Star Rating by Sentiment Classification

sentiment	Sentiment_encoded	Predicted_Rating
Positive	2	5
Positive	2	5
Positive	2	5
Negative	0	1
Positive	2	5
Positive	2	5
Positive	2	5

Table A-8. Sentiment Classification (Shorter Screenshot)

review_body	Topic	Topic_Name
As described.	4	Ease of Use & Convenience
It works as advertising.	0	Product Satisfaction & Performance
Works pissa	0	Product Satisfaction & Performance
Did not work at all.	4	Ease of Use & Convenience
Works well. Bass is somewhat lacking but is present. Overall pleased with the item.	4	Ease of Use & Convenience
The quality on these speakers is insanely good and doesn't sound muddy when adjusting bass. Very happy with these.	1	Value & Quality
Wish I could give this product more than five stars. Lifesaver.	4	Ease of Use & Convenience
works great	0	Product Satisfaction & Performance
Great sound and compact. Battery life seems good. Happy with this product.	1	Value & Quality
It works well~~~	0	Product Satisfaction & Performance
All good	1	Value & Quality
Love clock radio & CD player. Easy to operate.	3	Brand Loyalty & Preferences
Breaks very easily, and takes a while to load music.	4	Ease of Use & Convenience

Table A-9. Creation of Themes Based on Reviews (Shorter Screenshot)

review_body	review_theme
Breaks very easily, and takes a while to load music	product features
Excellent gain in radio frequency reception over the stock antenna	product features
everything I expected for a great price	price/value
Love this small speaker with loud volume, great for the beach	product features
works as advertised.... nothing else to really say about it	product features
very good especially for notebooks with micro hdmi and your monitor	product features
It does not copy CD-R's to USB as advertised. My CD's were unreadable	delivery experience
Did not last long, Stop working within a year.	product features
Can't get any simpler that this. And it works just about anywhere	product features
Wish I could tilt it more up & down, but it works fine especially for car use	product features
Good product, but took a week to receive.	delivery experience
Makes a difference right out of the box. I have 4 mini usb's.	Product features
great product, great service.	customer service

Table A-10. Creation of Refined Themes Based on Reviews (Shorter Screenshot)

Review Theme	Number of Reviews
Product Features	5709
Delivery Experience	4762
Price/Value	3150
Customer Service	939
Packaging	623
Other (Unclassified)	25049

Table A-11. Summary Table of Review Themes and Review Count (Shorter Screenshot)

Star Rating	Precision	Recall	F1-Score	Support
1	0.40	0.38	0.39	632
2	0.14	0.08	0.10	321
3	0.14	0.07	0.10	503
4	0.17	0.11	0.13	957
5	0.78	0.89	0.83	5634

Table A-12. Predictive Model Performance Accuracy by Rating

review_headline	review_body	review_date	headline_sentiment	body_sentiment	Sentiment_Mismatch
Cute headphones	Great quality for the price. I believe th	8/31/2015 0:00	0.5	0.266666667	FALSE
Great Value and good to look	Good looking over the ear headphones t	8/31/2015 0:00	0.8375	0.290909091	FALSE
Easy installation and setup -	Appears to work just great. Used it to co	8/31/2015 0:00	0.433333333	0.046875	FALSE
Five Stars	Works great in Charlotte, NC. Received a	8/31/2015 0:00	0	0.3245	FALSE
Headphones Case is great	This case is perfect for my Symphonized	8/31/2015 0:00	0.8	0.583333333	FALSE
Not happy at all	We bought this for my daughter for her h	8/31/2015 0:00	-0.4	0.080539773	TRUE
fast delivery	works as expected; well made cable.	8/31/2015 0:00	0.2	-0.1	TRUE
Best price and fast shipping	Best price and fast shipping. Great speak	8/31/2015 0:00	0.6	0.6275	FALSE
Great noise cancelling headp	Great pair of headphones for the price. I	8/31/2015 0:00	1	0.252678571	FALSE

Table A-13. Match Between Headline and Review Body

Appendix B. Sample Output Figures

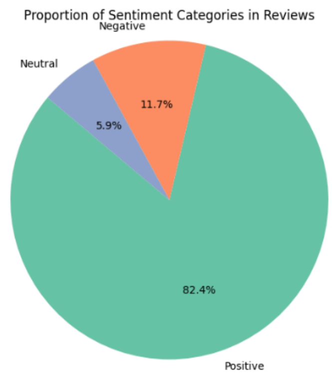

Figure B-1. Sentiment Categories in Pie Chart

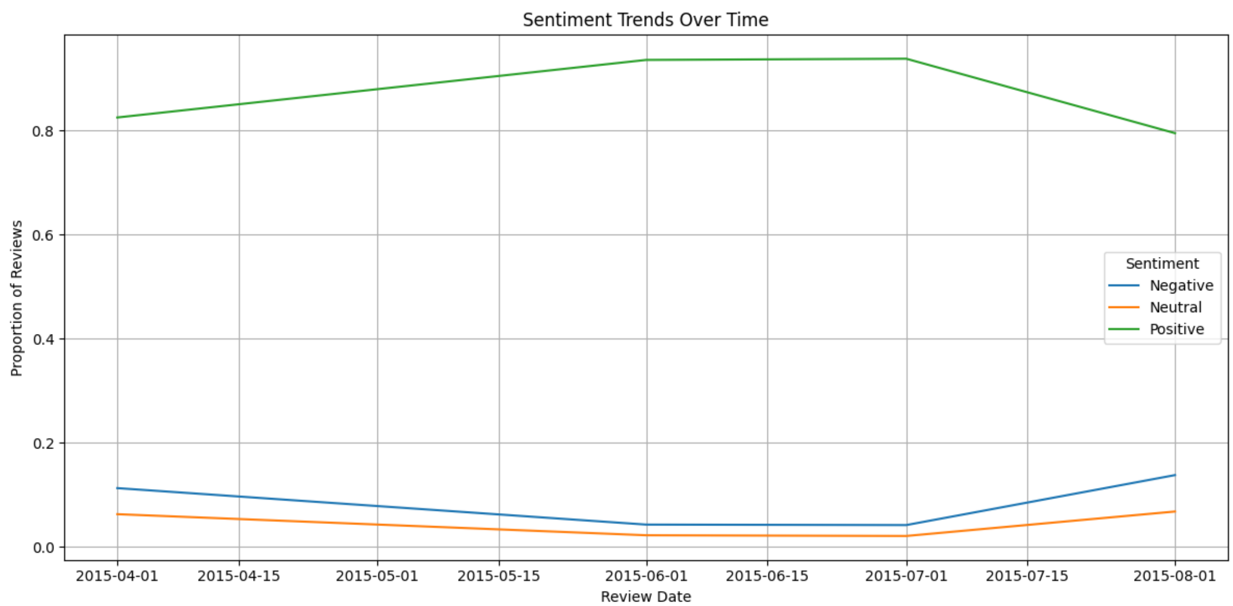

Figure B-2. Sentiment Trends Over Time

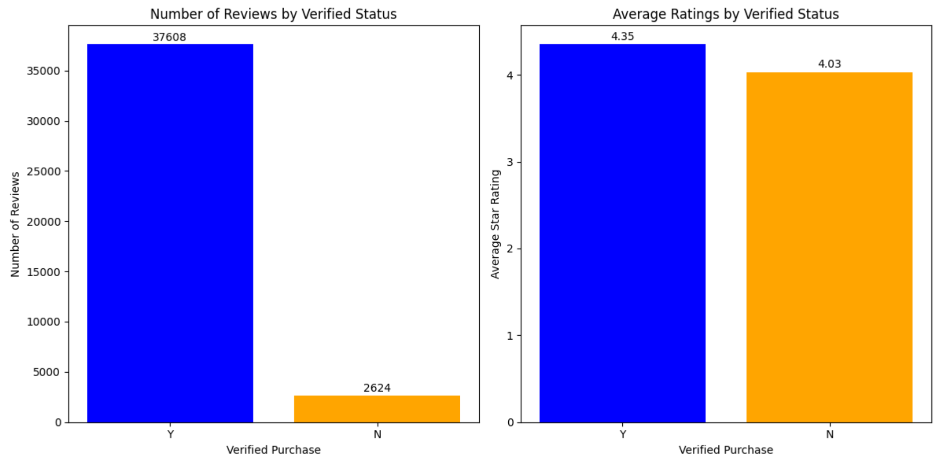

Figure B-3. Number of Reviews and Avg. Ratings by Verified Purchase

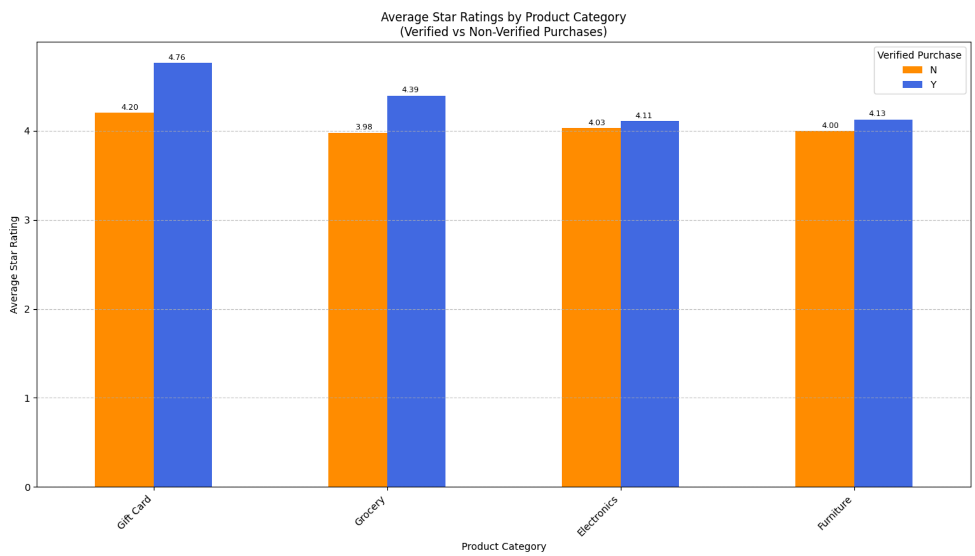

Figure B-4. Average Star Ratings by Product Category and Verified Purchase

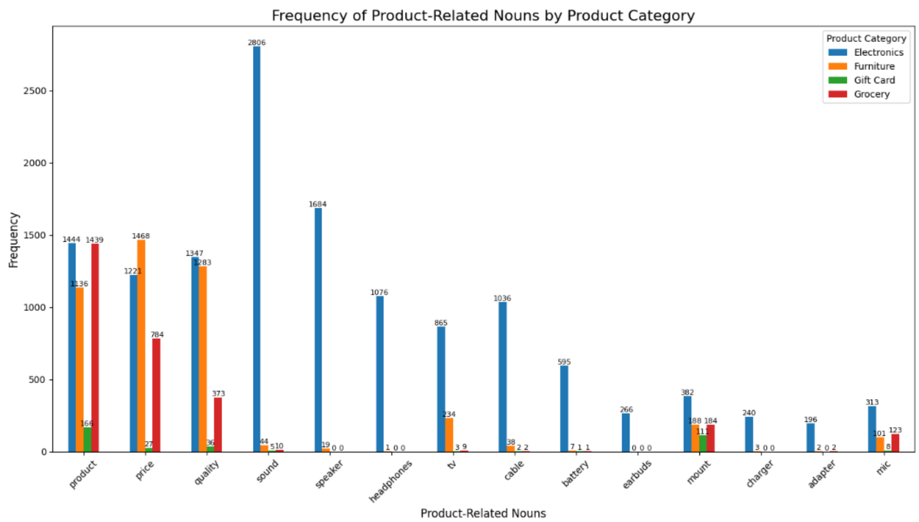

Figure B-5. Frequency by Product Category

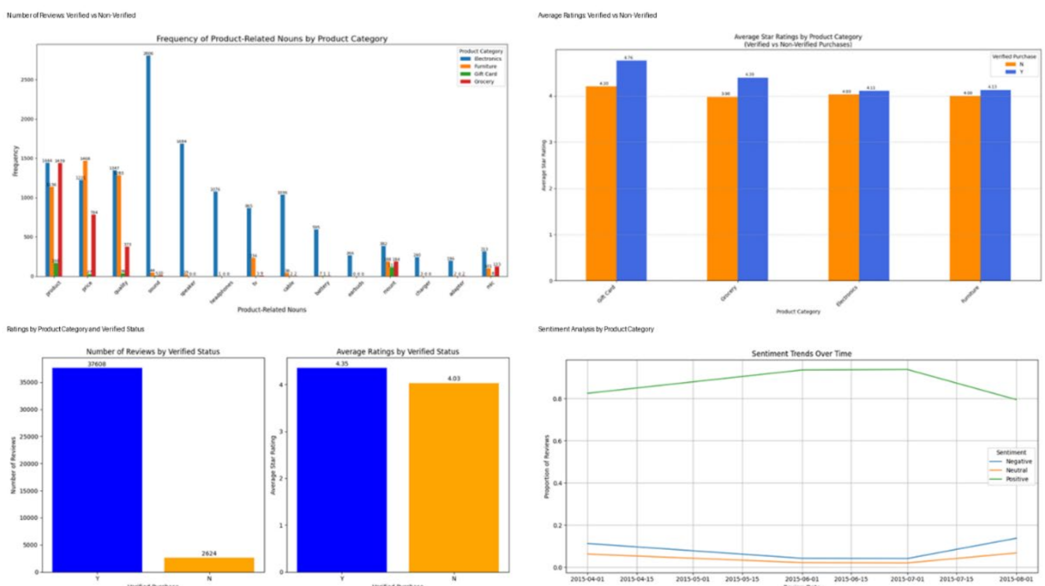

Figure B-6. Dashboard for Visualization

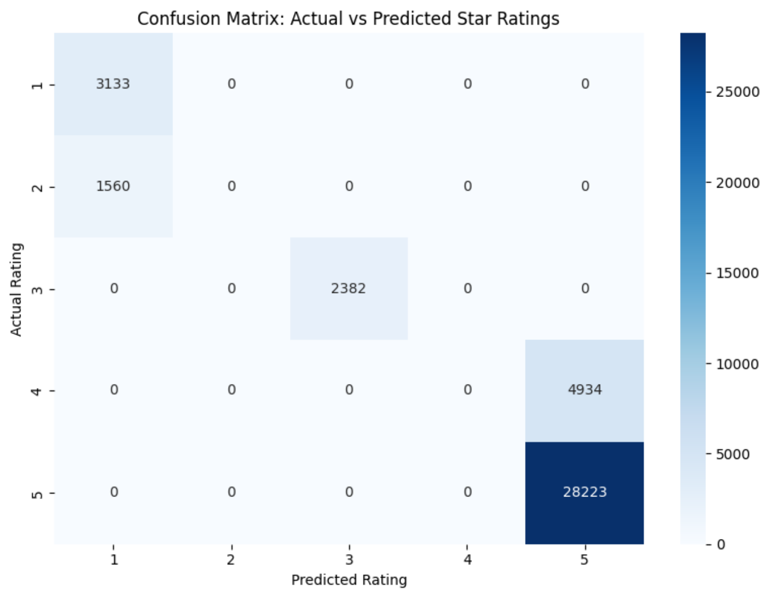

Figure B-7. Confusion Matrix of Actual vs. Predicted Star Ratings

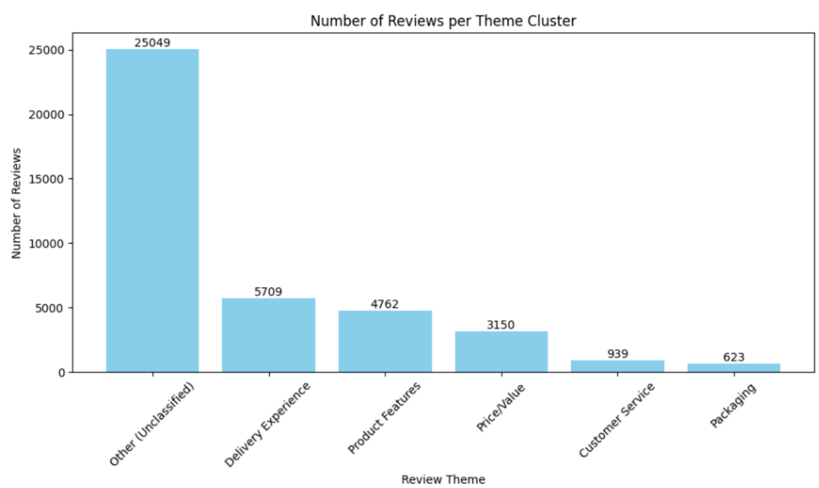

Figure B-8. Bar Chart of Review Themes and Reviews Count (Shorter Screenshot)

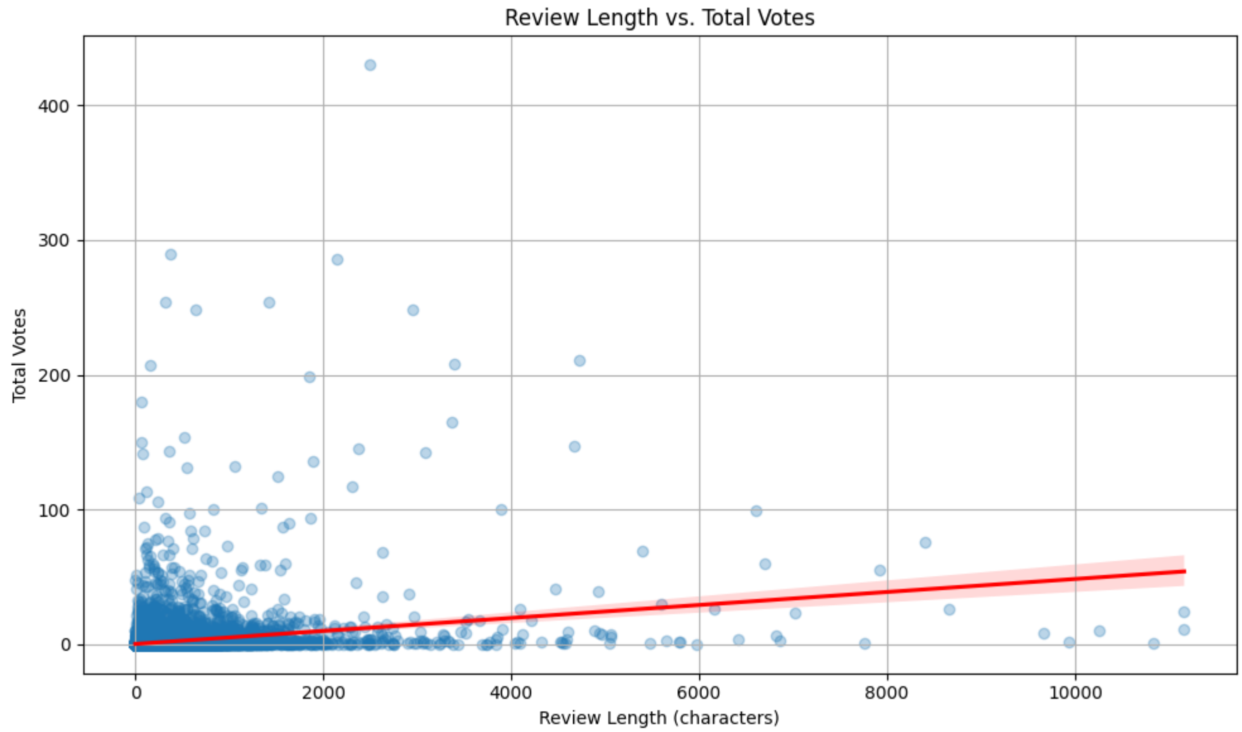

Figure B-9. Impact of Review Length on Voting Behavior

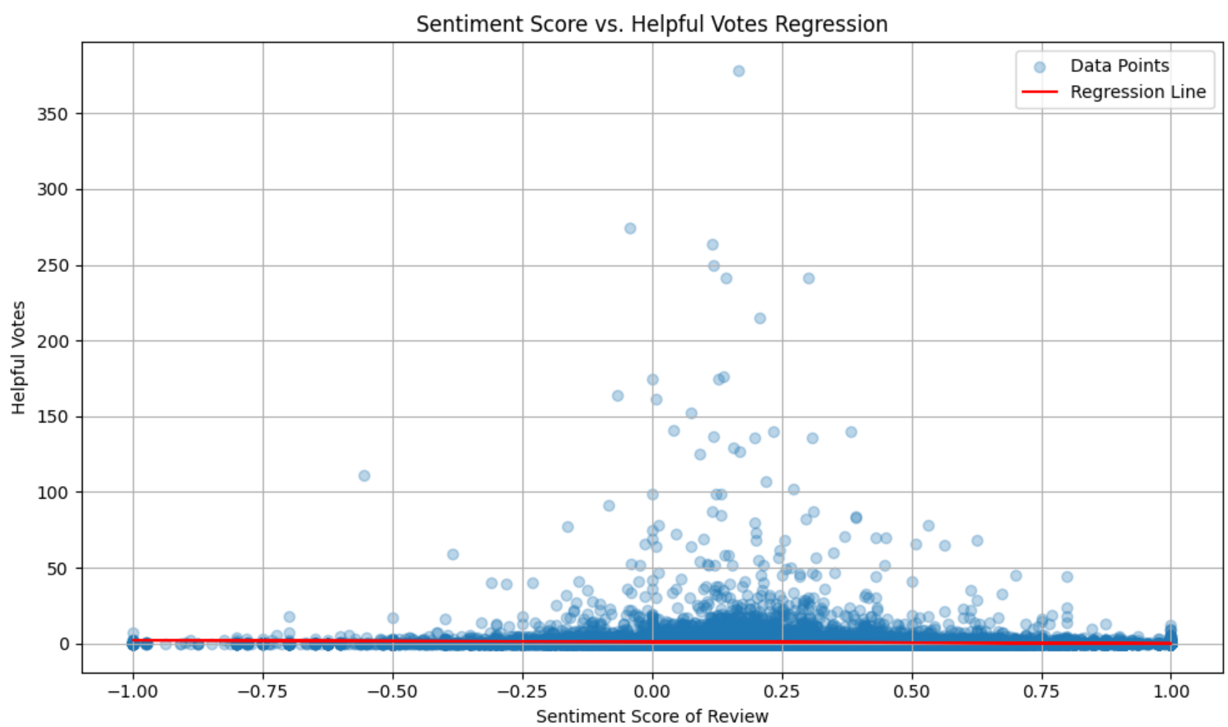

Figure B-10. Impact of Sentiment Polarity on Helpful Votes

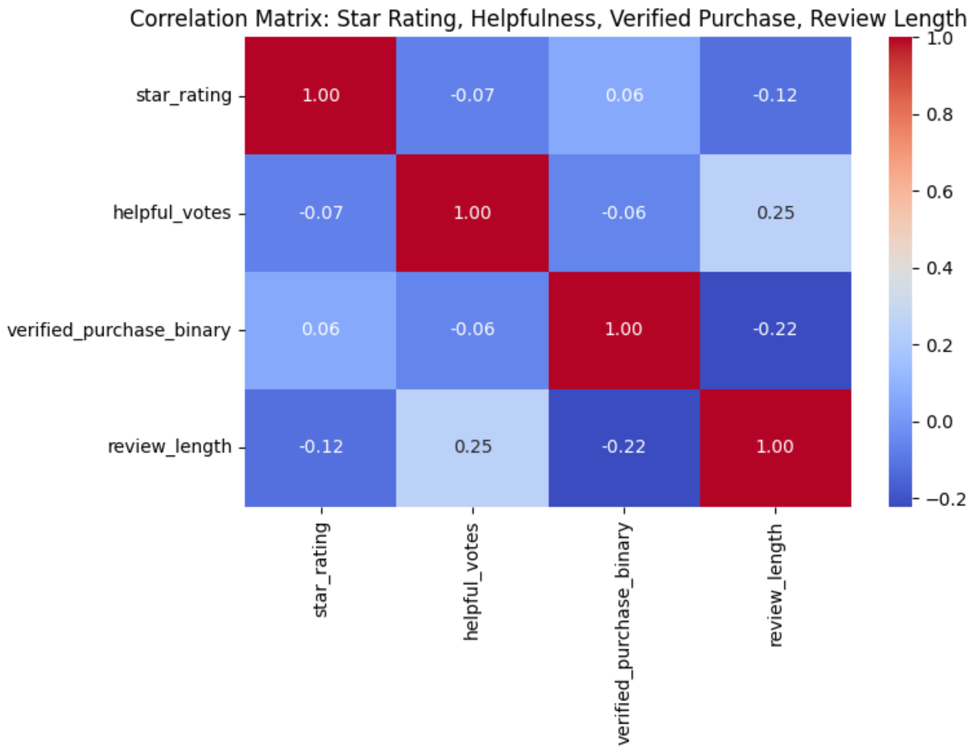

Figure B-11. Correlation Matrix

INFORMATION SYSTEMS & COMPUTING ACADEMIC PROFESSIONALS

STATEMENT OF PEER REVIEW INTEGRITY

All papers published in the ***Journal of Information Systems Education*** have undergone rigorous peer review. This includes an initial editor screening and double-blind refereeing by three or more expert referees.

ISSN: 2574-3872 (Online) 1055-3096 (Print)