Amazon Fine Food Reviews Analysis

Team 9

Arabela Carceles, Diana Michelle Prieto Sanchez, Juan David Zuluaga, Lucas Sotkovszki Rodríguez and Raquel Silva

Data Set: Network, S. (2017). Amazon Fine Food Reviews. Kaggle.com. https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews

Executive Summary

- While overall sentiment is overwhelmingly positive (87.7% positive reviews), our analysis revealed specific pain points within highly-rated products, indicating unmet high expectations that require proactive attention.

- A small segment of highly active users consistently writes positive reviews, signifying strong brand loyalty and potential for a reward strategy to encourage more users to write more reviews.

- The current dataset over-represents positive experiences. Capturing more balanced feedback (negative/neutral) is crucial for comprehensive product development and risk mitigation. There is an opportunity for improvement in getting more people to write reviews, especially negative and neutral ones

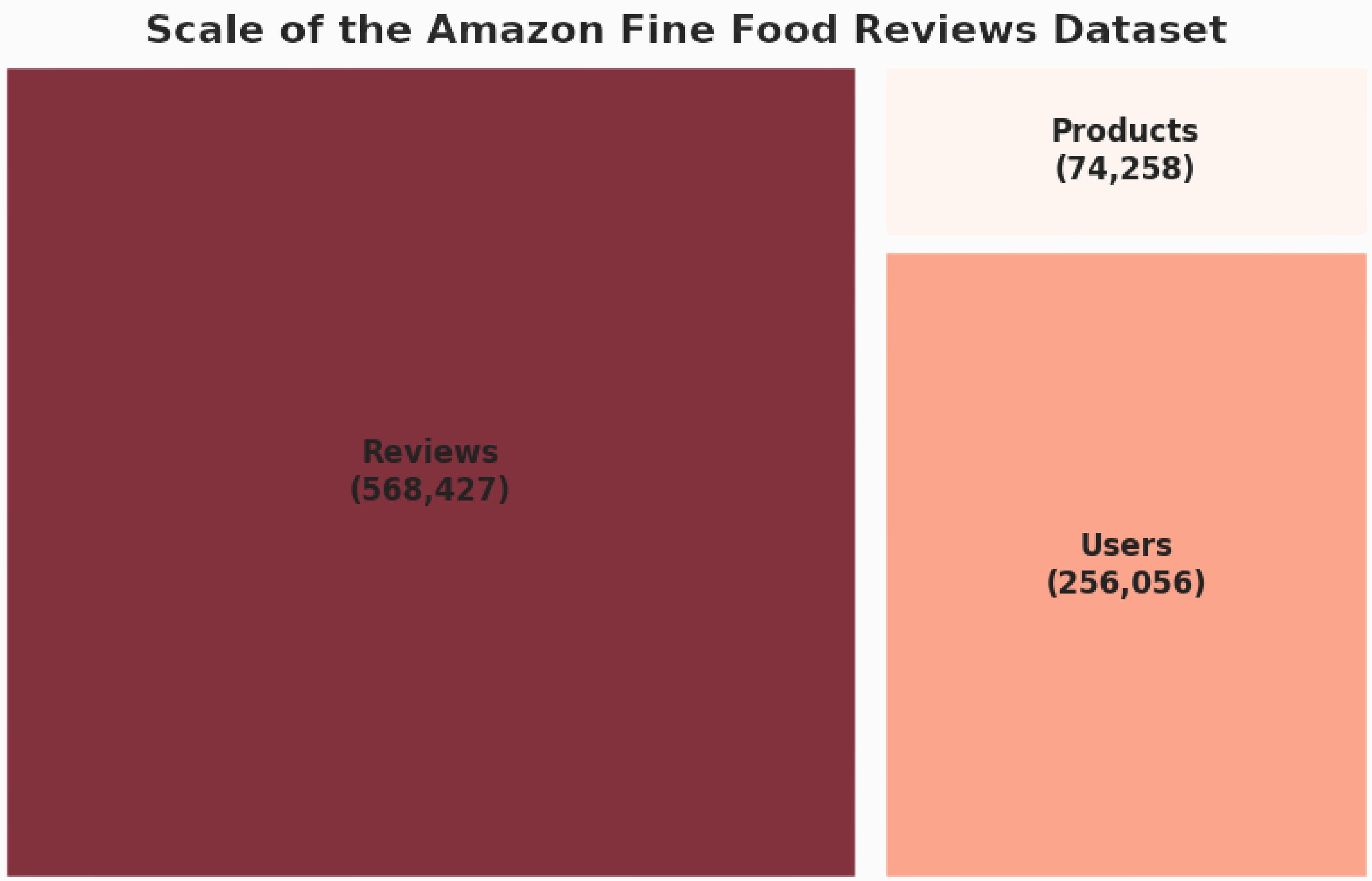
- Actionable Next Steps include implementing targeted feedback mechanisms, monitoring flagship products closely, and leveraging highly engaged users to drive continuous product and service enhancement.

Project Overview & Dataset Fundamentals

The main goal of this project is to analyze Amazon Fine Food Reviews to uncover customer sentiment, identify key topics, and provide actionable insights for Amazon and businesses in the fine food industry.

- Dataset: Amazon Fine Food Reviews
  - Source: Kaggle
  - Dates: October 1999 to October 2012
  - Volume: Over 568,000 reviews from 256,000 unique users on 74,000 products

- Key Columns: Score (1-5 rating), Summary (brief review), Text (full review), HelpfulnessNumerator/Denominator, Time (Unix timestamp)



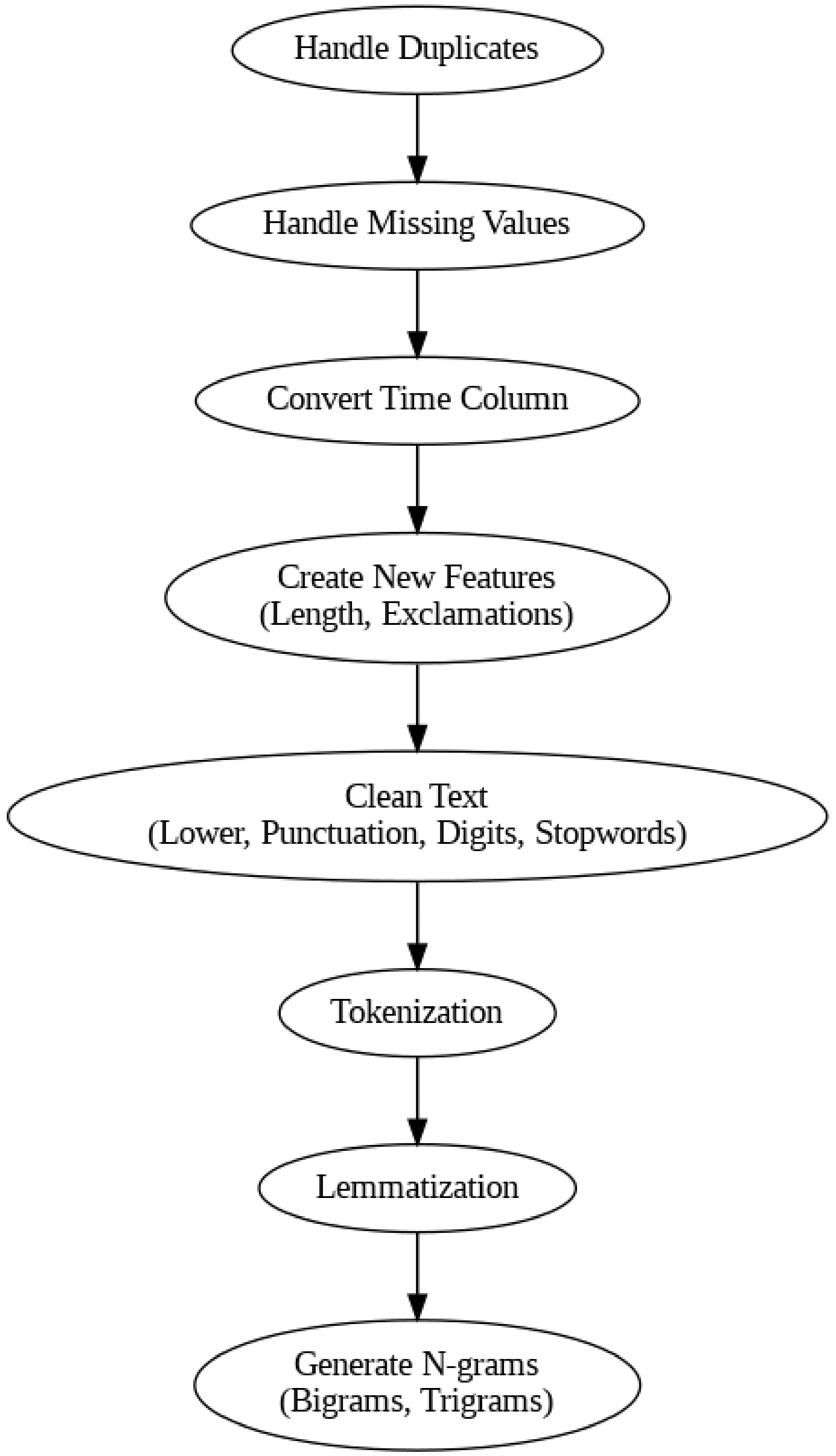
Data Import and Cleansing

Key Steps:

- Duplicate Elimination: Verified that no exact duplicate reviews were present.
- Missing Value Handling: Removed rows with missing 'Summary' to maintain data integrity for text analysis.
- Timestamp Conversion: Converted Unix timestamps to a readable datetime format for time-series analysis.
- Feature Engineering:
  - review_length: Word count for deeper text insights.
  - exclamations: Count of exclamation points to gauge emotional intensity
  - helpfulness_ratio: Calculated perceived usefulness of reviews.

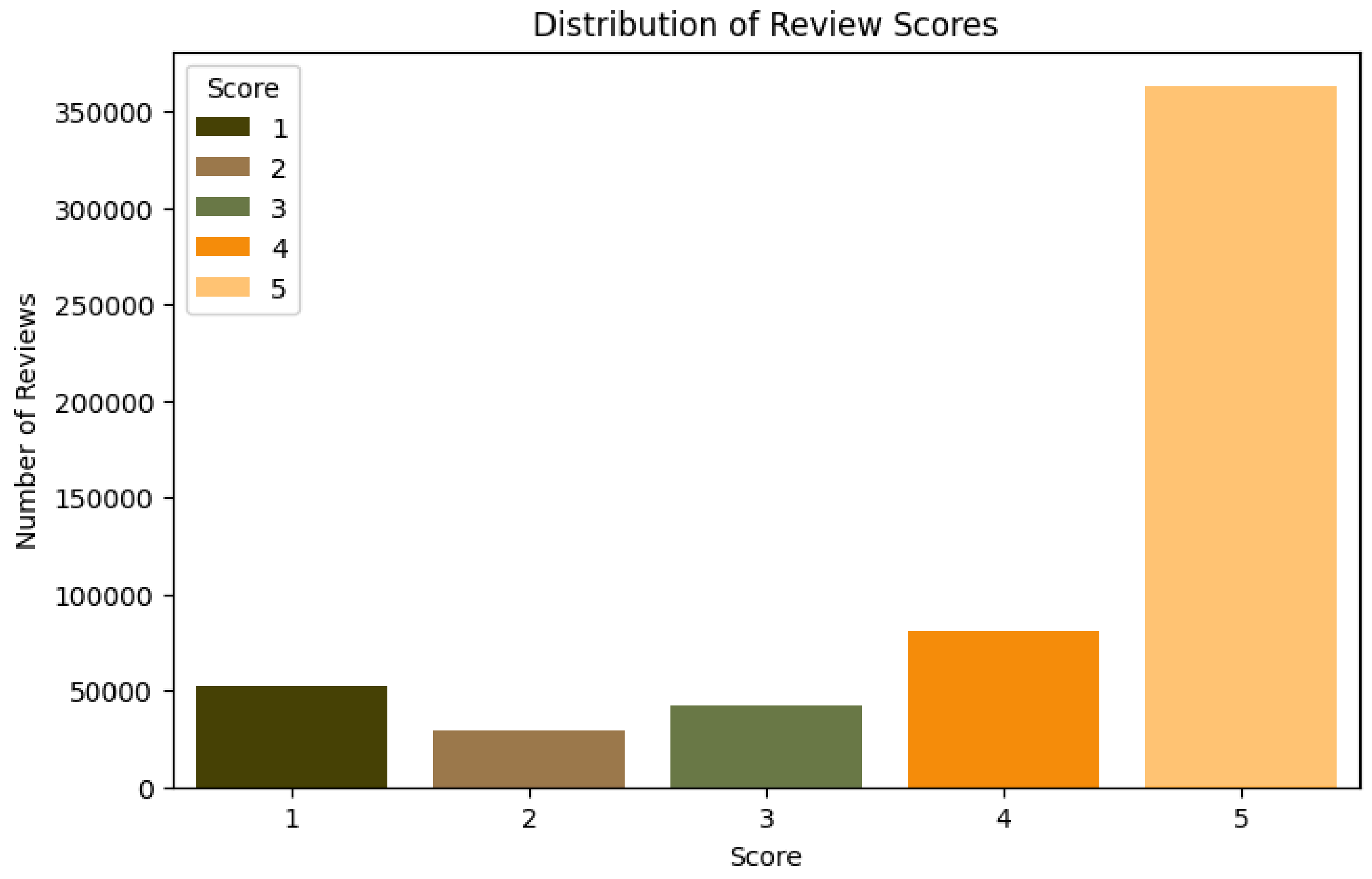
Text Preprocessing for NLP:

- Cleaning Function: Lowercasing, removal of non-alphanumeric characters, stopwords, and punctuation.
- Lemmatization: Reducing words to their base form (e.g., "running" to "run").
- N-gram Generation: Created Bigrams (2-word phrases) and Trigrams (3-word phrases) for capturing context (e.g., "great_coffee," "dog_love").

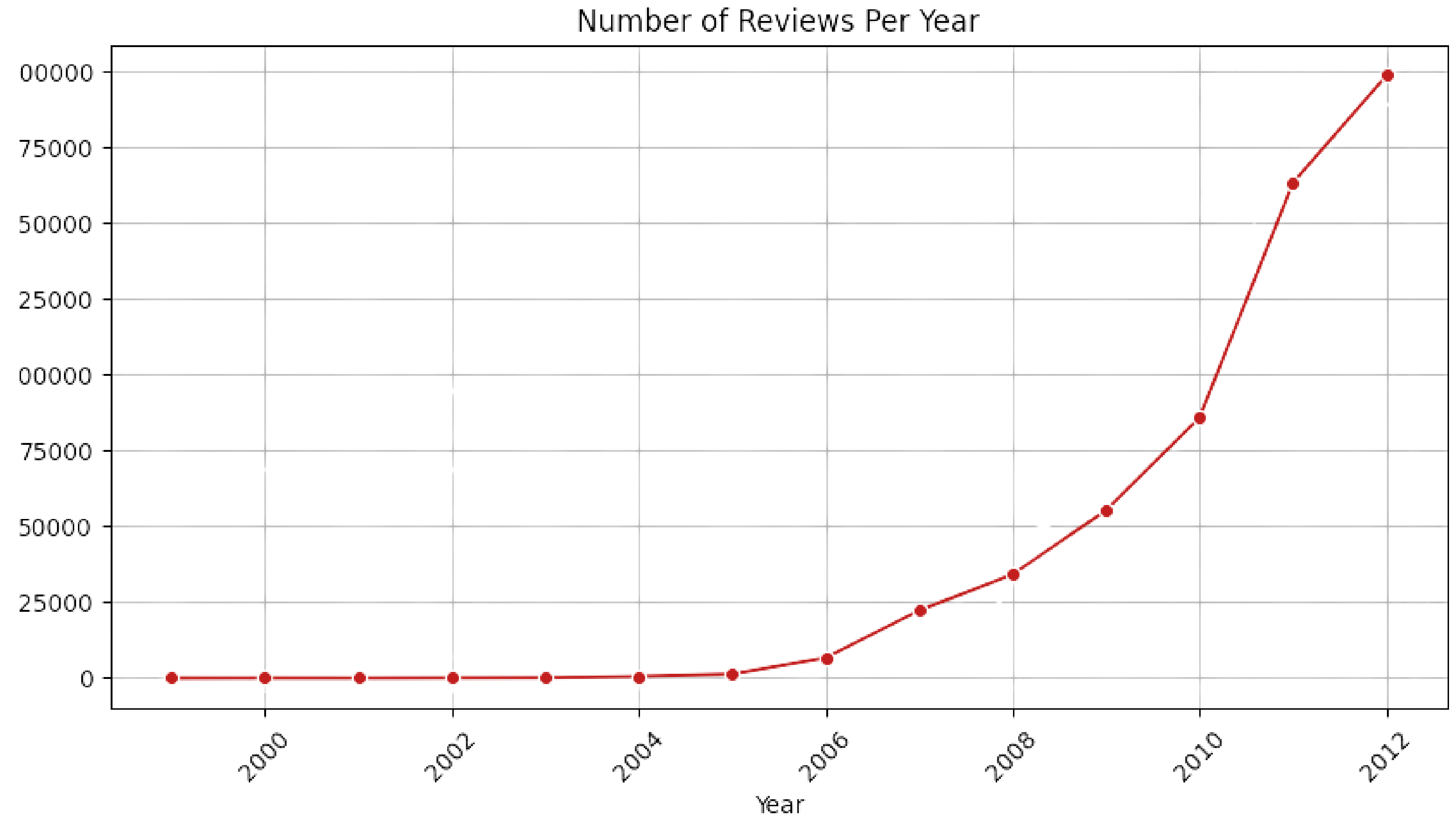


Descriptive Analysis

1 The majority of reviews are positive (4-5), with a few negative reviews (1-2)



2 An exponential increase in reviews from 2006 to 2012, aligning with Amazon's growth and potentially the launch of Amazon Fresh (2007), highlights increasing customer engagement over time.



3 Most users post only one review, but a small group of highly active users contribute significantly, potentially skewing overall sentiment.

Top Users by Engagement	Count
A3OXHLG6DIBRW8	448
A1YUL9PCJR3JTY	421
AY12DBB0U420B	389

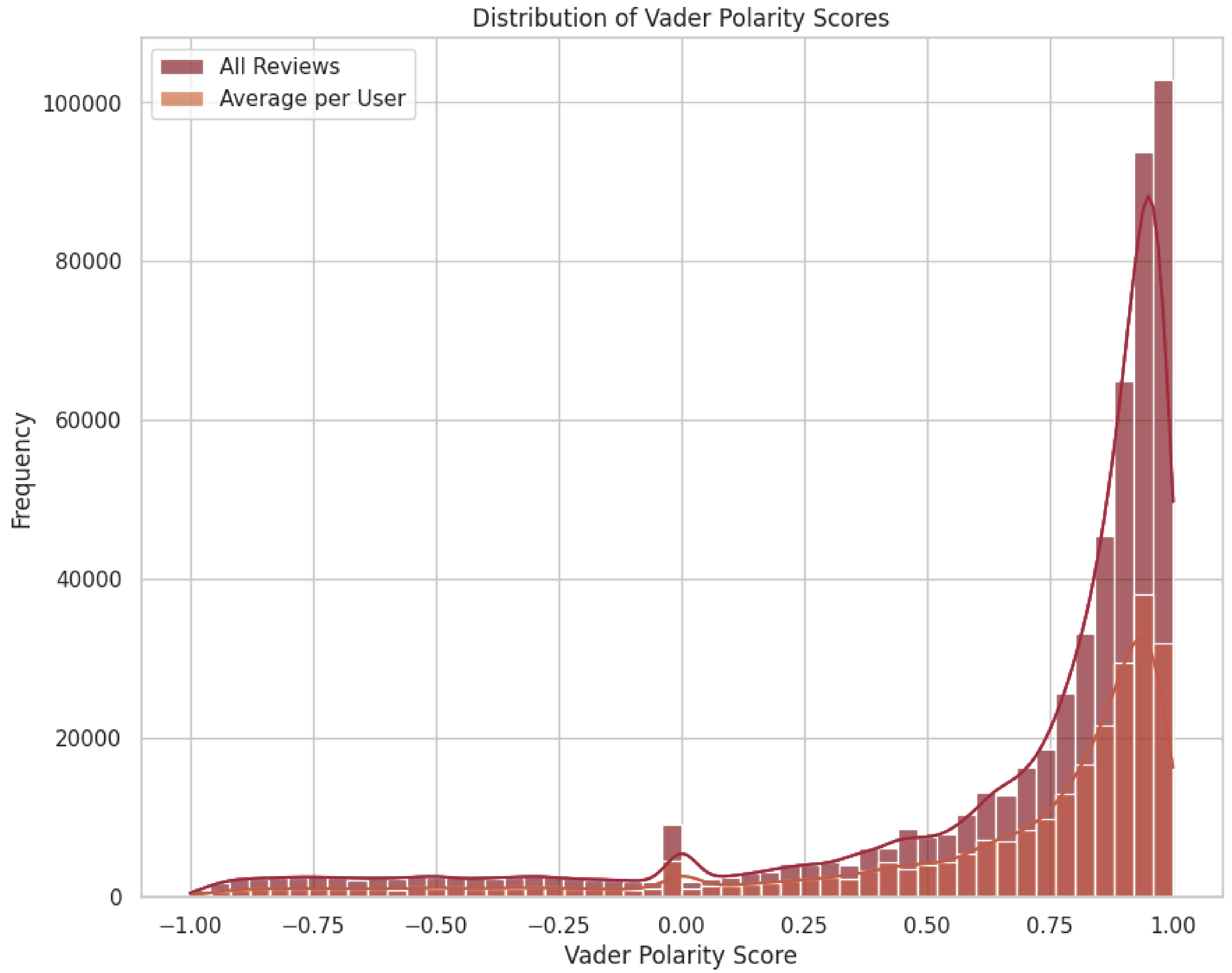
4 Calculating it from our helpfulness_ratio, we found ou that users find highest and lowest-rated comments most helpful when reviewing experiences.

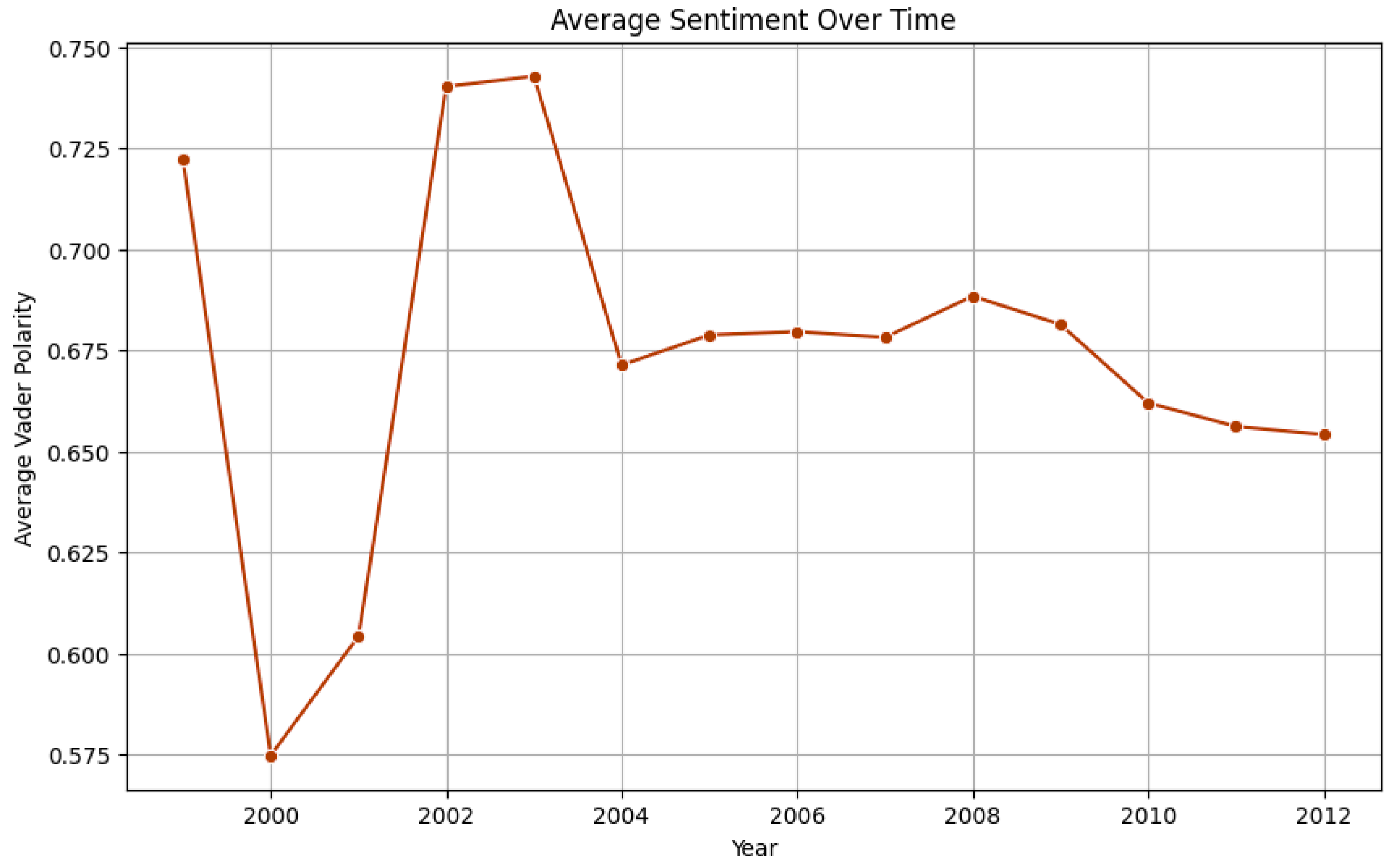
Sentiment Analysis

Utilized VADER for its ability to understand sentiment in social media text.

Overall Sentiment Distribution shows consistent with review scores, **87.7% of reviews express positive sentiment**, with only 9.3% negative and 2.9% neutral.

The distribution of average **VADER polarity per user** mirrors the **overall trend**, with **most users** exhibiting **positive sentiment**. However, analysis per user also reveals more neutral results, suggesting individual user patterns can be more nuanced.





Average sentiment shows a slight decrease in positivity over time, suggesting that initial "hype" for products may normalize. This could also relate to the internet's development, increase in number of products available and changing user review habits.

Topic Modeling: N-Grams

Leveraged NLTK for text preprocessing and wordcloud to visualize frequent words and phrases (n-grams) in review summaries. Stop word removal in this analysis might omit critical negative words (like "not"), so it must be interpret in conjunction with sentiment analysis.

Word Cloud

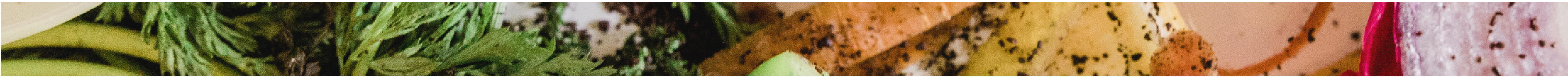
Primarily highlight positive sentiment (love, good, delicious) and product types (coffee, tea, dog).

Bigram Word Cloud

Reveal product-specific sentiment (great taste, gluten-free, dog food) indicating specific product attributes that are appreciated.

Trigram Cloud

Provide more context, sometimes highlighting negative sentiment when present, like don't waste money, despite the overall positive like love love love, or best coffee ever.



Topic Modeling: LDA

Topic 1
Good Products
("product," "great," "tea," "excellent")

Topic 2
Everyday Snacks & Pet Food
("delicious," "dog," "snack," "food")

Topic 3
Indulgences
("great," "yummy," "chocolate," "hot")

Topic 4
Pricing
("good," "great," "price," "stuff," "bad")

Topic 5
Best Products
("best," "love," "like," "buy")

Top Mod & Sentiment

Across all topics, sentiment remains largely consistent; however, Topic 5, focusing on "Best Products," shows a slightly elevated negative sentiment (10.26%), indicating that highly-rated products can evoke stronger negative reactions when customer expectations are not met, which aligns with the need to monitor such products; conversely, Topic 4, concerning "Pricing," exhibits slightly more neutrality (3.04%), suggesting that discussions around pricing tend to be less emotionally charged.

Strategic Implications for Amazon and Your Business

Cultivating Engaged Advocates through Rewards

The highly positive, active users are invaluable brand champions. Recognizing their consistent engagement can amplify their advocacy and provide deeper insights into your most loyal customer base.

Call to action

Implement structured loyalty programs or tiered reward systems that directly incentivize consistent, high-quality review contributions. This could include exclusive early product access, special discounts, or "brand ambassador" recognition, fostering deeper engagement and a continuous stream of valuable feedback from the most dedicated customers.

Product Strategy & Quality Assurance

The "best" and most loved products (Topic 5) unexpectedly draw proportionally more negative sentiment. This indicates specific unmet expectations or quality issues that could harm brand trust.

Call to action

Conduct a deep-dive analysis into negative reviews of top-selling products to pinpoint exact pain points like taste consistency, packaging, and delivery issues. Collaborate with R&D and QA teams for targeted product improvements and proactive customer service outreach.

Diversifying Feedback through Insistent Prompts

The current dataset over-represents positive experiences, leading to missed opportunities for critical feedback on product improvements. A more balanced view is crucial for product development.

Call to action

Integrate automated, simple, and slightly more insistent review prompts directly into the post-delivery or post-consumption user journey (like Uber's model). Specifically, design these prompts to encourage immediate, balanced feedback, paying particular attention to capturing neutral and negative experiences for a more comprehensive understanding of product performance.

Customer Insights for Product Development

Topic modeling reveals common product attributes and desires.

Call to action

Leverage these insights to inform new product development, tailor marketing to emphasize valued attributes like healthy properties, and identify emerging market trends for product descriptions. Furthermore, by identifying specific customer complaints, such as price concerns, targeted actions like sending personalized coupons can be implemented.

Conclusion and Next Steps

Our analysis provides a robust understanding of customer sentiment, key discussion topics, and behavioral patterns within the fine food review landscape.

Key Takeaway: Proactive engagement with all forms of customer feedback – especially critical insights from top-rated products and one-time purchasers – is essential for maintaining brand trust and driving sustainable growth in the competitive fine food market.

Future Opportunities for Your Company:

- Link review sentiment and volume to actual sales data to quantify the business impact of online feedback.
- Explore building models to predict product success or failure based on early review patterns.
- Utilize more sophisticated NLP techniques for aspect-based sentiment analysis to pinpoint sentiment about specific product features, like talking positively about the taste of the product but negatively about the packaging. Especially what could be improved according to customers.