ISOM5160 Group Report

Dataset: amazon_food_reviews.csv

Group7

CAO, Xi	21271664	LI, Heyi	21265689
LIAO, Jingyu	21262106	LIN, Chuwei	21237955
YE, Chenwei	21199517	ZHANG, Ziyang	21266920

October 13, 2025

Background & Objective

Project background

- Amazon product reviews contain a large amount of redundant and noisy information.
- It's hard for merchant to get valuable insights and learn how to improve their product

Objective

- Organize Amazon reviews and extract key informations (include sentiment, timeline information and goods key features extract) so merchants know how to improve their products

Data Cleaning

Dirty data

- Missing values in key fields (e.g., Text, Score, ProductId)
- Duplicate reviews from same user /multiple Good reviews same product
- Noisy , inconsistent text ,stop words will redundant the later works

Steps

- Duplicate and blank value processing: Make product reviews unique and valid
- Unify text format and remove stop words: to facilitate subsequent sentiment analysis and keyword extraction

Sentiment Analysis

Why sentiment analysis?

- The sentiment of some reviews is different from the actual review scores
- Provides merchants with actionable insights for product improvement

Steps

- Model: Pre-trained BERT
- Download the pre-trained BERT model, download the data set for local training, and then score the original data for sentiment tendency

Sentiment Analyze Result

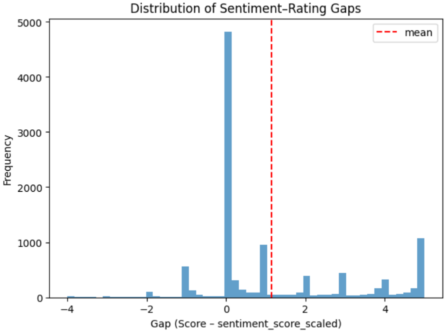
keywords	sentiment_score
['product', 'exactly', 'adv	0.00123052822891
['favorite', 'glutenfree', '	0.99981278181076
['usually', 'love', 'kettle'	0.00468845758587
['tried', 'many', 'varietie	0.99964845180511
['oki', 'read', 'reviews', '	0.98734623193740
['husband', 'love', 'chips',	0.99963760375976
['chips', 'good', 'bad', 'c	0.00788646098226
['chips', 'remind', 'long',	0.01704546995460
['took', 'get', 'used', 'ch	0.99965572357177
['kettle', 'chips', 'spicy',	0.99847334623336
['perhaps', 'worst', 'chips'	0.00042024036520
['totally', 'orgasmic', 'ch	0.93455797433853
['saltfree', 'product', 'pu	0.00043656674097
['someone', 'brought', 'par	0.99852883815765
['backyard', 'bbq', 'kettle'	0.52436339855194
['opening', 'numerous', 'ba	0.00021553144324
['getting', 'worried', 'rea	0.01317740604281
['kettle', 'brand', 'chips',	0.93911117315292
['chips', 'nasty', 'thought'	0.00136297626886

Added sentiment columns and keyword collections

Sentiment-Rating Correlation & Outlier Detection

1. Methodology

- Adjusted sentiment scores to match 5-point rating scale
- Computed Sentiment-Rating **Gap** = **Rating** - **Sentiment_Score_Scaled**
- Detected outliers with $|Z| > 2$: 483/10819 ($\approx$ 4.5%)
- **General Trend:** Positively related; ratings are **systematically higher** than sentiment scores.



Sentiment-Rating Correlation & Outlier Detection

2. Insights: Why Sentiment ≠ Rating

a. TextSentiment Model Bias

- Key words like "unfortunately", "so good", "used to be good" confuse sentiment detection.
- Context-dependent irony or comparison not captured by model.

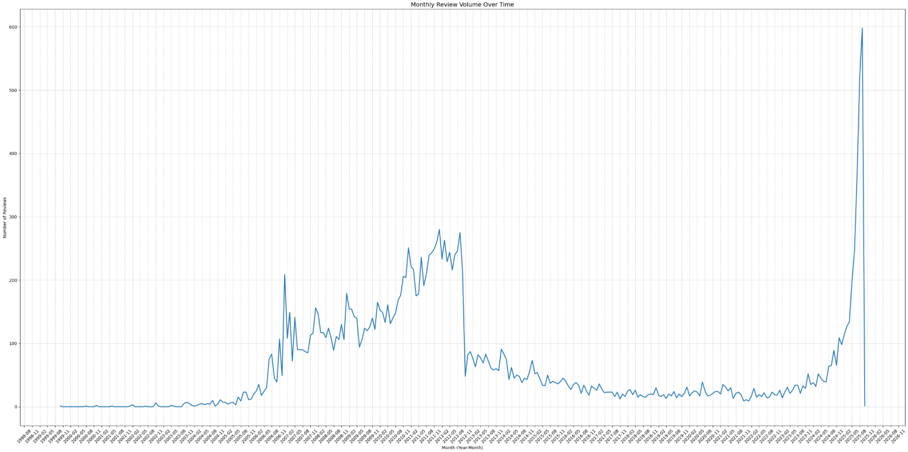
Example of the model bias:
"the product was exactly **as advertised and fresh. unfortunately** i keep them in a candy dish in the office and they are going fast. we need to reorder to keep up with demand"

b. Human Behavior Factors

- Users express disappointment **politely**, often masking negative emotions.
- Positive service experience (e.g. refund) leads to positive feedback despite low product rating.
- Social courtesy bias: buyers avoid leaving harsh ratings or comments.

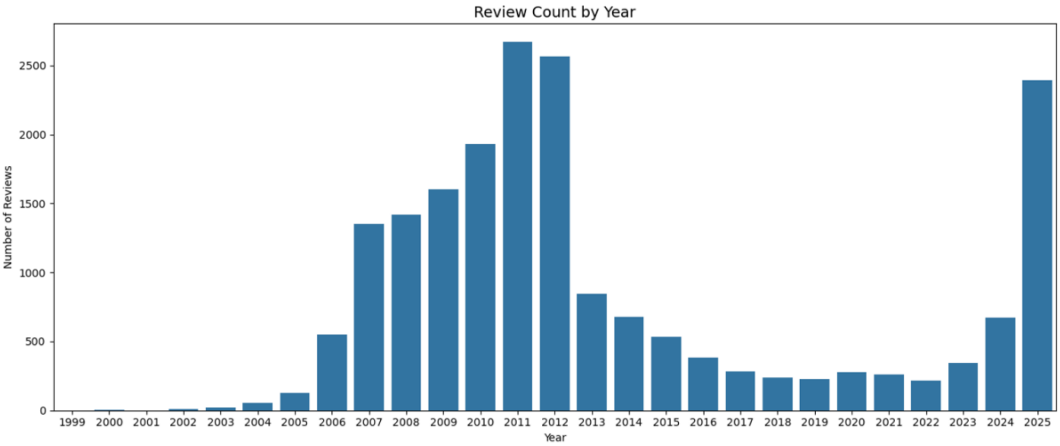
Example of human behavior bias:
"i was very angry about this but jr mushrooms has said they will **refund** me for the truffles and even let me keep them. so i **give jr credit for excellent responsiveness** and customer service, although i still feel they should not be labeled black winter truffles"

Comment Time Series Analysis



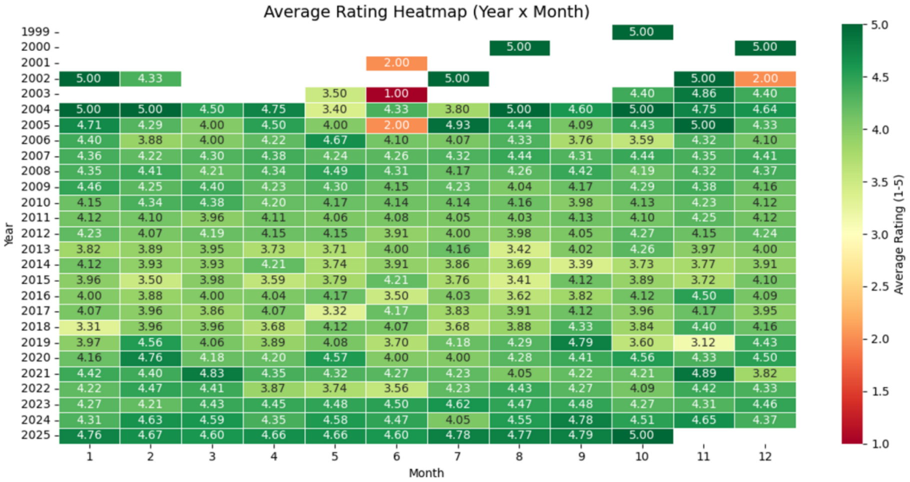
Review volume shows long-term growth, reflecting rising customer engagement.

Comment Time Series Analysis



Yearly trends can help evaluate product lifecycle and marketing effectiveness.

Comment Time Series Analysis



Ratings remain consistently positive, showing strong brand reputation.

Negative Review Analysis

Why need negative review analysis?

- Assist in finding overall improvement directions (through negative review analysis, it is possible to identify the parts of the entire system that need improvement, and determine which specific aspect of product quality, transportation time, and other factors caused the negative reviews, in order to optimize them)
- Assist in analyzing the reasons for the existence of specific products (by analyzing negative comments, feedback can be given to product suppliers for self correction)

Steps

- Read the comment data after data cleaning
- Classify comments through logical reasoning
- Count the frequency of negative reviews

Negative Review Analysis

Negative evaluation criteria

- Accurately locate real negative reviews through dual filtering criteria
- Objective rating: Score <= 2 (considered low on a 5-point scale)
- Subjective sentiment: Sentimentscore < 0.4 (sentiment tendency score generated by data_cleaning.py, below 0.4 is considered negative)

Keyword filtering

- Map high-frequency keywords to 8 common negative cause categories in the food industry, and define the classification rules through the category_mapping dictionary:
- Taste: such as taste, bitter, bland, etc
- Quality: such as poor, cheap, terrible, etc
- Expired (expired/spoiled): such as expired, rotten, mold, etc
- Other categories: packaging, price, delivery, quantity, smell
- Keywords that do not match the preset category are classified as' other ', ensuring that all keywords are categorized accordingly

Correlation Between Ratings and Product Descriptions - Overview

Steps

1. Scrape supplementary data from amazon.com
2. Analyse correlation for both visual & text information
3. **Correlation computation:** We use spearmanr to compute Correlation

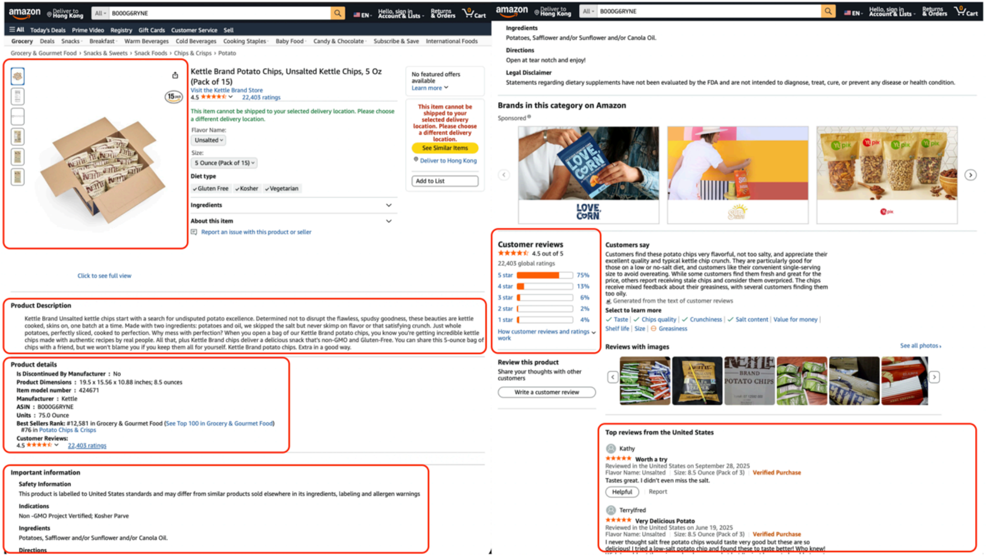
Part1. Visual Information Correlation

- Correlation with the number of sample images

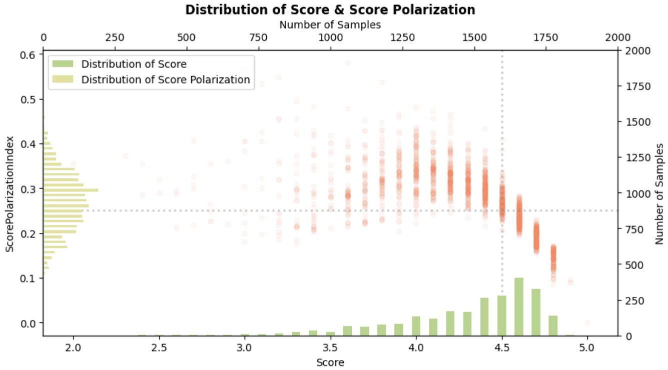
Part2. Text Information Correlation

- Correlation with description length
- Correlation with reading ease
- Correlation with the marketing tone of the description
- Correlation with product description items

Supplementary data (Scraped from Amazon's website)



Rating features

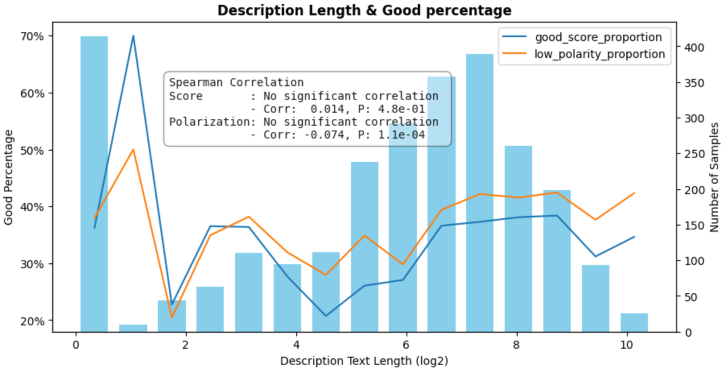


1. **Mean Score:** The score displayed on the product detail page.
2. **Score Polarization:** Indicates whether the ratings for this product are polarized.
3. **Standards of Good Score:** Score >= 4.5; Polarization <= 0.25

Correlation with the number of sample images

Conclusion: The is a **positive correlation** in product ratings and descriptions. The more images included in the description, the greater the likelihood of the product receiving positive reviews.

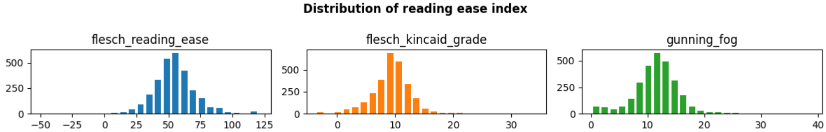
Correlation with description length



Conclusion: There is **no significant correlation** between product ratings and the length of product descriptions.

Correlation with reading ease

ease_index	compare_with	corr	p_value	conclusion
flesch_reading_ease	Score	0.0163	0.3863	No significant correlation
flesch_reading_ease	Polarization	-0.0187	0.3214	No significant correlation
flesch_kincaid_grade	Score	0.0123	0.5135	No significant correlation
flesch_kincaid_grade	Polarization	-0.0130	0.4900	No significant correlation
gunning_fog	Score	0.0066	0.7270	No significant correlation
gunning_fog	Polarization	-0.0076	0.6870	No significant correlation

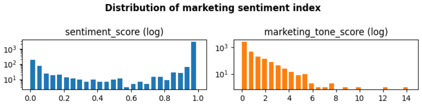


- **flesch reading ease:** The higher the score, the easier it is to read.
- **flesch kincaid grade:** The required grade level to read; The higher the number, the more difficult the reading level.
- **gunning fog:** Long word ratio; The higher the ratio, the harder it is to understand.

Conclusion: Product ratings show <u>**no significant correlation**</u> with reading difficulty.

Correlation with the marketing tone of the description

sentiment_type	compare_with	corr	p_value	conclusion
marketing_tone_score	Score	0.0588	0.0007	No significant correlation
marketing_tone_score	Polarization	-0.0469	0.0066	No significant correlation
sentiment_score	Score	0.0788	0.0000	No significant correlation
sentiment_score	Polarization	-0.0568	0.0010	No significant correlation



Types of marketing tone (Measuring Criteria):

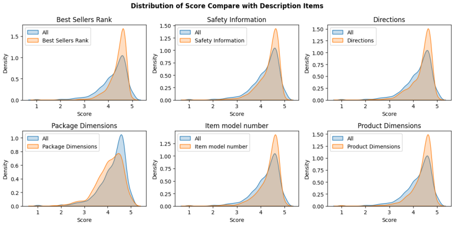
1. **marketing tone score:** The proportion of exaggerated marketing language in product descriptions. The higher the score, the more exaggerated the marketing claims become.
2. **sentiment score by language model**: model is distilbert-base-uncased-finetuned-sst-2-english, The higher the score, the more positive the emotion.

Conclusion:

- Product ratings **show no significant correlation** with marketing tone/sentiment.
- *However, since the metrics used to measure the emotional orientation of product descriptions do not follow a normal distribution, conclusions drawn from this basis may be unreliable.

Correlation with product description items

item	compare_with	corr	p_value	conclusion
Best Sellers Rank	Score	0.3809	2.9e-116	Highly correlated
Directions	Score	0.2211	2.12e-38	Correlated
Item model number	Score	0.2721	4.95e-58	Correlated
Package Dimensions	Score	-0.2473	6.66e-48	Correlated
Package Dimensions	Polarization	0.1717	1.32e-23	Slightly correlated
Product Dimensions	Score	0.2715	9.67e-58	Correlated
Safety Information	Score	0.1618	4.03e-21	Slightly correlated



Products with the following description items might have a better rating:

- **Best Sellers Rank:** Only high-quality goods carry this label.
- **Item model number & Product Dimensions:** Products with model numbers may be more formal and reliable.
- **Directions:** Products accompanied by directions may reduce negative reviews caused by users' inability to use or misuse the product.
- **Safety Information:** Products with safety information may reduce negative reviews caused by user allergies.

Products with the following description items might have a lower rating:

- **Package Dimensions:** May receive negative reviews due to logistics-related issues.

Review Weighted Analysis

Remove outliers where HelpfulnessNumerator > HelpfulnessDenominator

$$r = \frac{\text{HelpfulnessNumerator} + a}{\text{HelpfulnessDenominator} + a + b} \qquad t = 1 + \log(1 + \text{HelpfulnessDenominator})$$

Comment Weight:

$$\text{weighted\_mean} = \frac{\sum(\text{Score} \times \text{weight})}{\sum \text{weight}}$$

Example: If there are 10 five-star ratings with no votes and 1 one-star rating with 50 people finding it useful the weighted mean will lean closer to one star.

Review Weighted Analysis

Bayesian Mean:

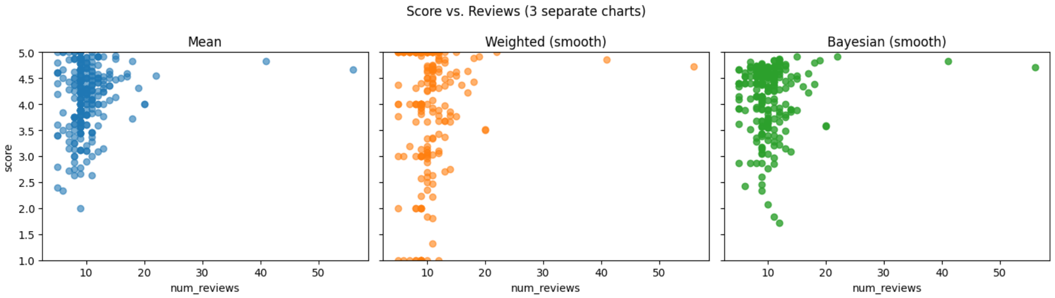
$$WeightedScore = \frac{v}{v+m} \cdot R + \frac{m}{v+m} \cdot C$$

Optimized:

$$\frac{m \times \text{global\_mean} + \text{weighted\_mean} \times \sum \text{weight}}{m + \sum \text{weight}}$$

Conclusion: Based on the weighted mean, it converges toward the global average to avoid small-sample extremes. The fewer the reviews, the closer the Bayesian mean approaches the global average; Products with many reviews the Bayesian mean approaches its own weighted mean.

Review Weighted Analysis



Conclusion: The raw mean is highly susceptible to bias when the sample size is small, easily skewed by a single extreme review. Weighted Smooth reduces the impact of noisy reviews. Bayesian Smooth applies confidence contraction to products with small sample sizesbringing ratings closer to the global mean to prevent misrepresentation.

Overall Takeaways

Overall Takeaways for Merchants

- **Listen Beyond the Stars** - Text sentiment often reveals issues hidden behind high scores.
- **Focus on Quality Signals** - Address taste, packaging, and logistics to curb negative reviews.
- **Professional Product Overview** - Include detailed, credible product information and visuals.
- **Leverage Weighted Analytics** - Use smoothed scores for fairer product comparisons.
- **Continuously Monitor Feedback Trends** - Use time-series tracking to align improvement with customer behavior.

Technical Contributions

- **End-to-end review analysis pipeline** combining data cleaning, sentiment modeling, and statistical validation.
- **Hybrid NLP-statistical approach** ensuring both semantic accuracy and analytical robustness.
- **Practical insight generation** helping sellers understand reviews and improve their products.

Q & A

Thank you for listening !

Group 7

CAO, Xi	21271664	LI, Heyi	21265689
LIAO, Jingyu	21262106	LIN, Chuwei	21237955
YE, Chenwei	21199517	ZHANG, Ziyang	21266920