



Sentiment Analysis on Customer Reviews and Analysis of Factors Impacting Customer Satisfaction

By Samiksha Deb(B22CS029), Gunnu Lavanya(B22CS035), Vanshika Sarraf(B22CS037)

Supervised by Dr. Soumen Moulik, NIT Meghalaya

Introduction & Sentiment Analysis



Project Focus

Sentiment analysis on Amazon Fine Food Reviews & Topic Modelling.



Importance

90% consumers influenced by reviews



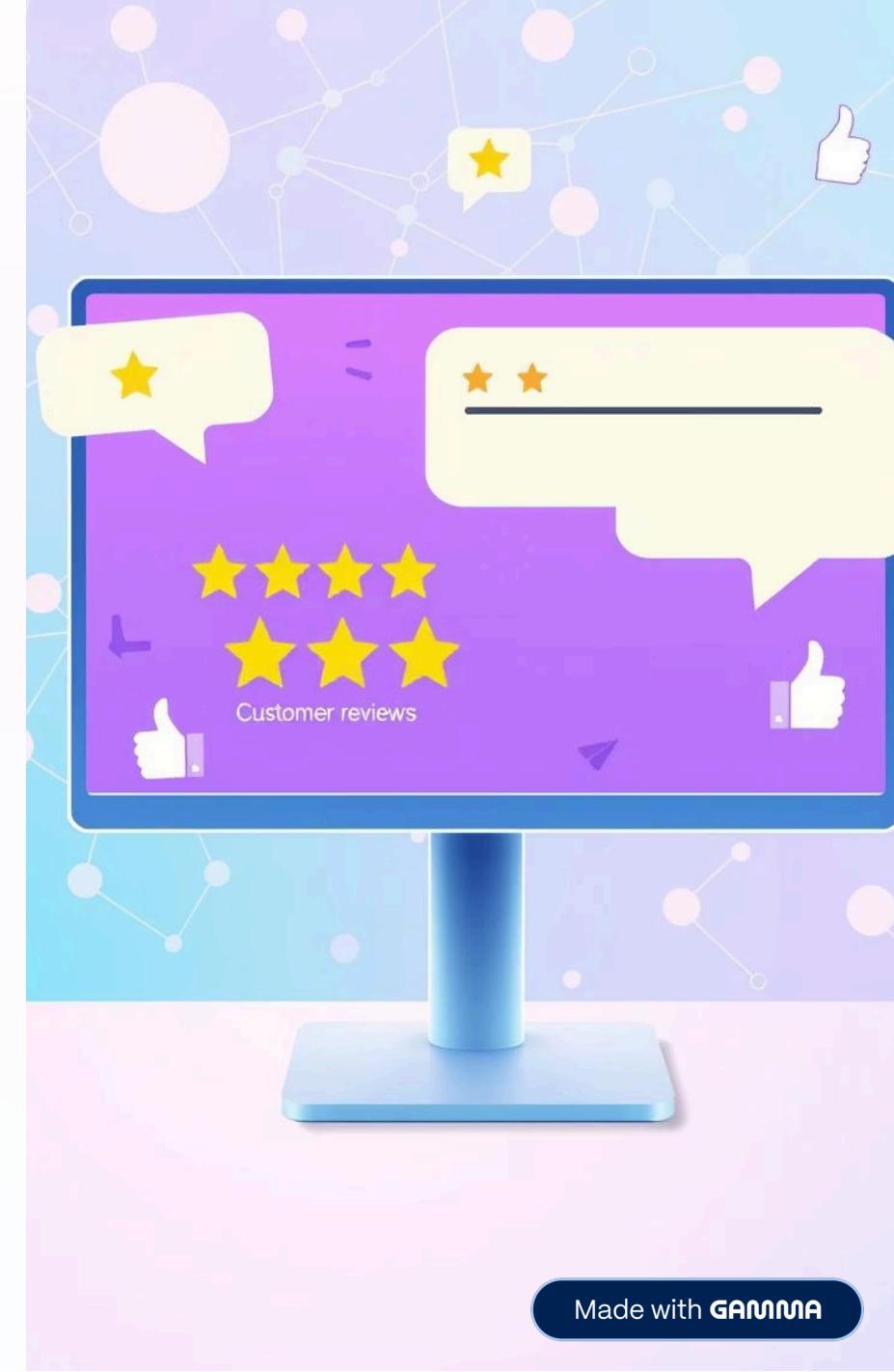
Sentiment Analysis

Classify reviews as positive, negative, or neutral



Method

Used VADER for fast, rule-based sentiment scoring



Project Scope & Approach

Dataset

First 100,000 Amazon Fine Food Reviews

Techniques

- Text preprocessing
- Topic modeling (LDA)
- Sentiment analysis (VADER)
- Coherence Analysis
- Probability Distribution

Outcomes

Sentiment pipeline & key satisfaction factors

Data Preparation & Preprocessing



Steps

HTML removal, punctuation cleanup, lowercasing



Tokenization

Split text into words



Stopword Removal

Filter common words



Lemmatization

Reduce words to base form



Feature Extraction

Bag of Words model for vectorization

Topic Modeling with LDA

What is Topic Modeling?

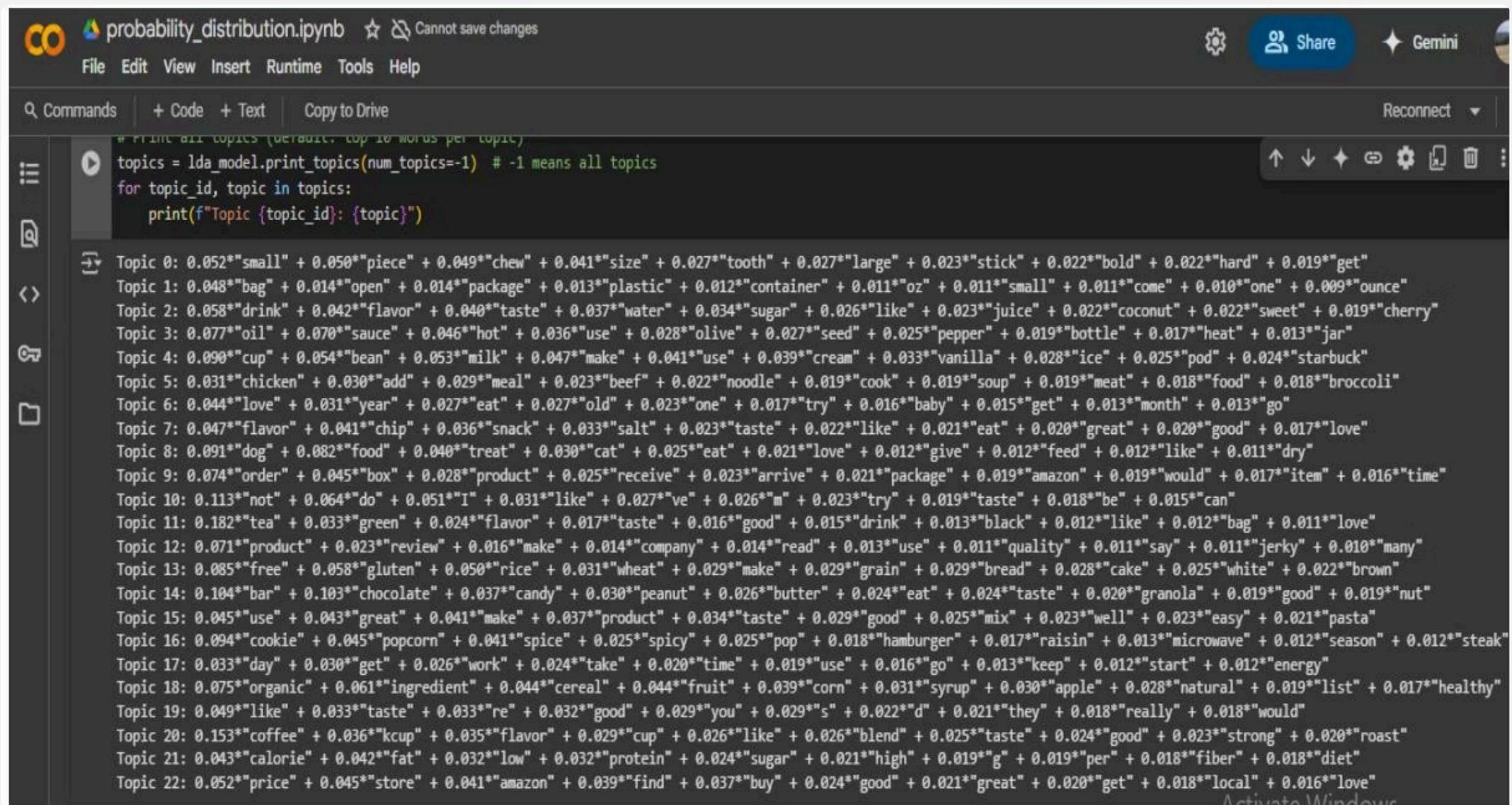
Unsupervised discovery of themes in documents

Why LDA?

Interpretable topics with word distributions

Usage

Assigned dominant topic per review



The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** probability_distribution.ipynb
- Toolbar:** File, Edit, View, Insert, Runtime, Tools, Help
- Left Sidebar:** Includes icons for file operations like New, Open, Save, and a search bar labeled "Commands".
- Code Cell:** Contains Python code to print LDA topics.

```
# PRINT TOPICS (DEFAULT: TOP 10 WORDS PER TOPIC)
topics = lda_model.print_topics(num_topics=-1) # -1 means all topics
for topic_id, topic in topics:
    print(f"Topic {topic_id}: {topic}")
```
- Output Cell:** Displays the generated topics as a list of words and their weights, such as "Topic 0: 0.052*\"small\" + 0.050*\"piece\" + 0.049*\"chew\" + 0.041*\"size\" + 0.027*\"tooth\" + 0.027*\"large\" + 0.023*\"stick\" + 0.022*\"bold\" + 0.022*\"hard\" + 0.019*\"get\"".
- Right Sidebar:** Includes "Share" and "Gemini" buttons.
- Bottom Status Bar:** Shows "Activate Windows".

Coherence Analysis

Purpose

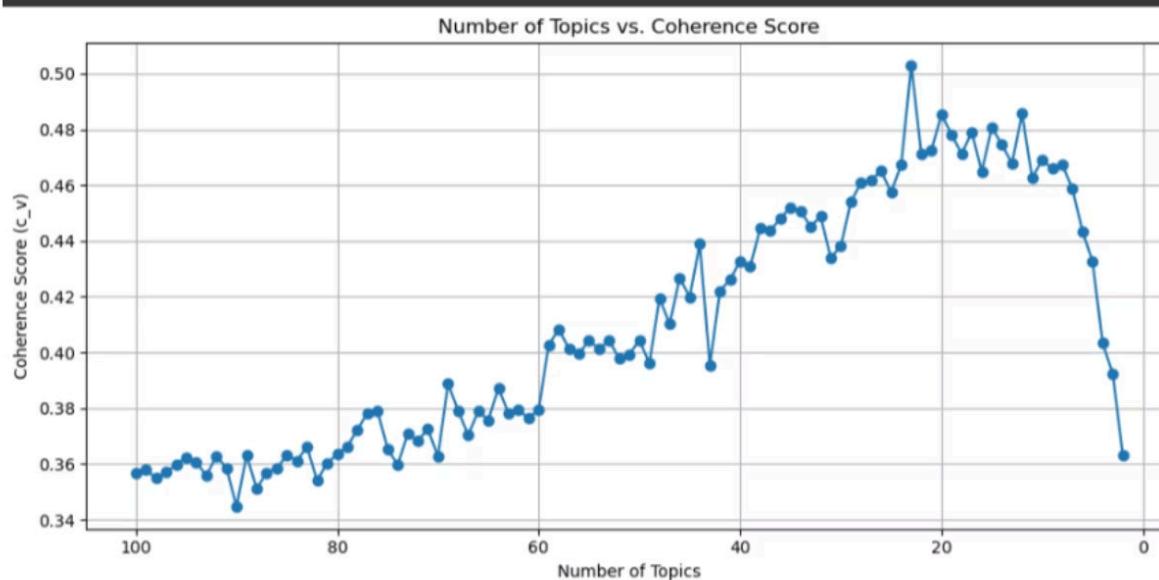
Evaluate topic quality and interpretability

Method

Tested multiple topic counts
(k=100 to 2)

Outcome

Selected model with highest coherence score



77%	76/99 [7:45:40<1:27:35, 228.48s/it]Topics: 25, Coherence Score: 0.4575
78%	77/99 [7:49:21<1:23:01, 226.44s/it]Topics: 24, Coherence Score: 0.4673
79%	78/99 [7:53:07<1:19:11, 226.27s/it]Topics: 23, Coherence Score: 0.5031
80%	79/99 [7:56:57<1:15:46, 227.33s/it]Topics: 22, Coherence Score: 0.4711
81%	80/99 [8:00:36<1:11:09, 224.72s/it]Topics: 21, Coherence Score: 0.4727
82%	81/99 [8:04:25<1:07:47, 225.96s/it]Topics: 20, Coherence Score: 0.4852
83%	82/99 [8:08:08<1:03:47, 225.16s/it]Topics: 19, Coherence Score: 0.4783
84%	83/99 [8:11:48<59:37, 223.56s/it]Topics: 18, Coherence Score: 0.4712
85%	84/99 [8:15:26<55:27, 221.85s/it]Topics: 17, Coherence Score: 0.4791
86%	85/99 [8:19:02<51:21, 220.14s/it]Topics: 16, Coherence Score: 0.4650
87%	86/99 [8:22:41<47:39, 219.96s/it]Topics: 15, Coherence Score: 0.4807
88%	87/99 [8:26:16<43:41, 218.43s/it]Topics: 14, Coherence Score: 0.4745
89%	88/99 [8:29:53<39:56, 217.84s/it]Topics: 13, Coherence Score: 0.4678
90%	89/99 [8:33:38<36:41, 220.18s/it]Topics: 12, Coherence Score: 0.4858
91%	90/99 [8:37:19<33:03, 220.38s/it]Topics: 11, Coherence Score: 0.4626
92%	91/99 [8:40:53<29:07, 218.47s/it]Topics: 10, Coherence Score: 0.4691
93%	92/99 [8:44:32<25:31, 218.73s/it]Topics: 9, Coherence Score: 0.4659
94%	93/99 [8:48:07<21:44, 217.44s/it]Topics: 8, Coherence Score: 0.4673
95%	94/99 [8:51:39<17:59, 215.85s/it]Topics: 7, Coherence Score: 0.4586
96%	95/99 [8:55:11<14:18, 214.66s/it]Topics: 6, Coherence Score: 0.4435
97%	96/99 [8:58:50<10:48, 216.03s/it]Topics: 5, Coherence Score: 0.4326
98%	97/99 [9:02:41<07:20, 220.48s/it]Topics: 4, Coherence Score: 0.4033
99%	98/99 [9:06:38<03:45, 225.39s/it]Topics: 3, Coherence Score: 0.3925
100%	99/99 [9:10:40<00:00, 333.74s/it]Topics: 2, Coherence Score: 0.3633

optimal number of topics is 23

Probability Distributions by Topic

Dominant Topic

Topic with highest probability per review

Sentiment Distribution

- Positive, Neutral, Negative percentages per topic
- Grouped reviews by dominant topic

Insights

Identify satisfaction drivers and problem areas

	lemmatized_text	Dominant_Topic	Topic_Name	Positive_%	Neutral_%	Negative_%
0	buy several vitality can dog food product find...	8	Pets	92.139103	1.330650	6.530247
1	product arrive label jumbo salt peanutsthe pea...	12	Reviews	74.457831	4.698795	20.843373
2	confection around century light pillowy citrus...	21	Protein	90.320122	2.896341	6.783537
3	look secret ingredient robitussin believe find...	2	Beverage	93.104184	2.049530	4.846285
4	great taffy great price wide assortment yummy ...	11	Tea	93.914131	1.417257	4.668612
5	get wild hair taffy order five pound bag taffy...	11	Tea	93.914131	1.417257	4.668612
6	saltwater taffy great flavor soft chewy candy ...	15	Cooking	97.566520	0.655168	1.778313
7	taffy good soft chewy flavor amazing would def...	14	Chocolate	95.802771	0.896496	3.300733
8	right I m mostly sprout cat eat grass love rot...	16	Spicy	92.096944	3.371970	4.531085
9	healthy dog food good digestion also good smal...	8	Pets	92.139103	1.330650	6.530247

The screenshot shows a Jupyter Notebook interface with the following details:

- File Bar:** CO Final_Result.ipynb, File, Edit, View, Insert, Runtime, Tools, Help.
- Toolbar:** Share, Gemini.
- Code Cell:** Review 140, 201, 249, 252, 813, 1004, 1084, 1087, 1245, 1275, 1414, 2212, 2259, 2379, 2534, 2599, 2873, 3753, 3863, 4392, 4423, 4429, 4691, 4698, 4866, 4977, Topic Name: Texture, Sentiment: Positive: 88.58%, Neutral: 3.09%, Negative: 8.33%.
- Code Cell:** Review 14, 70, 101, 144, 169, 170, 192, 200, 203, 216, 258, 266, 311, 337, 344, 362, 385, 387, 401, 417, 434, 439, 450, 470, 513, 515, 535, 541, 561, 566, 567, Topic Name: Packaging, Sentiment: Positive: 81.47%, Neutral: 4.38%, Negative: 14.16%.
- Code Cell:** Review 3, 78, 174, 176, 178, 196, 207, 233, 250, 265, 267, 268, 269, 287, 298, 299, 320, 356, 652, 658, 659, 660, 662, 663, 729, 746, 783, 1066, 1088, 1161, Topic Name: Beverage, Sentiment: Positive: 93.1%, Neutral: 2.05%, Negative: 4.85%.
- Code Cell:** Review 10, 29, 318, 338, 751, 755, 778, 988, 995, 1043, 1071, 1238, 1371, 1410, 1419, 1479, 1480, 1481, 1919, 1934, 2219, 2551, 2561, 2570, 3031, 3748, 3750, Topic Name: Sauces, Sentiment: Positive: 91.36%, Neutral: 2.38%, Negative: 6.26%.
- Code Cell:** Review 326, 649, 650, 1939, 2265, 2564, 2695, 2704, 2957, 3209, 3224, 3227, 3232, 3242, 3260, 3263, 3270, 3292, 3310, 3343, 3345, 3354, 3360, 3423, 3457, 3477, Topic Name: Coffee, Sentiment: Positive: 85.55%, Neutral: 3.44%, Negative: 11.01%.
- Code Cell:** Review 34, 225, 246, 334, 374, 386, 393, 409, 744, 902, 905, 907, 911, 918, 926, 1065, 1130, 1170, 1246, 1312, 1440, 1443, 1445, 1447, 1451, 1452, 1453, 1459, Topic Name: Meals, Sentiment: Positive: 91.52%, Neutral: 2.71%, Negative: 5.78%.

Activate Windows
Go to Settings to activate Windows

Factors Impacting Customer Satisfaction

Key Factors

- Taste
- Packaging
- Delivery
- Pricing

Sentiment Analysis

Measured positive, neutral, negative per factor

Business Use

Target improvements and leverage strengths



Observations & Sentiment Patterns



Highly Positive

Cooking, Chocolate, Tea, Beverage topics



Negative-Leaning

Reviews, Complaints, Delivery topics



Neutral Sentiment

Generally low across topics



Implication

Focus on improving weak areas

Topic_No	Topic_Name	Positive%	Neutral%	Negative%
0	Texture	88.58%	3.09%	8.33%
1	Packaging	81.47%	4.38%	14.16%
2	Beverage	93.1%	2.05%	4.85%
3	Sauces	91.36%	2.38%	6.26%
4	Coffee	85.55%	3.44%	11.01%
5	Meals	91.52%	2.71%	5.78%
6	Kids	92.94%	1.97%	5.09%
7	Snacks	95.32%	1.19%	3.49%
8	Pets	92.14%	1.33%	6.53%
9	Delivery	82.5%	4.09%	13.41%
10	Complaints	81.14%	2.53%	16.33%
11	Tea	93.91%	1.42%	4.67%
12	Reviews	74.46%	4.7%	20.84%
13	Gluten	95.64%	2.85%	1.52%
14	Chocolate	95.8%	0.9%	3.3%
15	Cooking	97.57%	0.66%	1.78%
16	Spicy	92.1%	3.37%	4.53%
17	Energy	86.12%	2.84%	11.04%
18	Organic	87.3%	5.29%	7.41%
19	Taste	90.52%	1.77%	7.71%
20	K-Cups	93.76%	1.51%	4.73%
21	Protein	90.32%	2.9%	6.78%
22	Pricing	93.91%	2.04%	4.04%

Conclusion & Future work

Summary

Positive sentiment dominates.

Negatives clustered in reviews.

Final Thought

Data-driven insights empower better business decisions.

Future Work

Use advanced models.

Balance the review data.



A large, abstract background image on the left side of the slide, featuring swirling patterns of pastel-colored ink or smoke in shades of blue, purple, pink, and white against a light blue gradient.

Thank You!!

We appreciate your time and interest in our presentation.