

Food Delivery Sentiment Analysis Generative-AI

John Buczkowski
December 29th, 2025

Contents / Agenda

- Business Context, Problem Overview and Solution Approach
- Exploring Different Prompting Techniques
- Prompts and the structure of the Prompts
- Output from the Prompts
- Observations/Insights and conclusions
- Appendix

Project Goals and Objectives

Background:

In the food delivery industry, customer feedback drives service quality and brand loyalty. Companies like Uber Eats, DoorDash, and Grubhub rely on reviews to improve operational efficiency and customer satisfaction. With the high volume of reviews posted daily across multiple cities, manual analysis isn't feasible. Reviews in the form of unstructured text contain complaints, praise, and personal experience feedback. They require automated systems that can classify sentiment consistently and without bias.

Goal:

Automate sentiment classification of customer feedback for a food delivery platform using Generative AI and Prompt Engineering. Process customer reviews into Positive, Negative, or Neutral categories while ensuring factual consistency and interpretability. Map outputs with relevant tags (Delivery Time, Food Quality, Price, Packaging, Overall Experience) and suggested business actions.

Project Goals and Objectives

Objectives:

1. Design structured prompts enabling an LLM to perform sentiment classification on customer feedback
2. Implement four (4) prompting strategies:
 - Zero-Shot
 - Few-Shot
 - Zero-Shot Chain of Thought (CoT)
 - Few-Shot CoT
3. Compare prompting strategies in terms of accuracy, reasoning and contextual understanding
4. Determine best performing strategy
5. Generate suggested actions to guide business response, such as flagging delivery issues for operations or highlighting positive reviews for marketing.

Zero-Shot Prompt:

- Direct instructions with no examples. Model relies on its training to complete the task.

Few-Shot Prompt:

- Instructions with 3-5 labeled examples demonstrating the desired input → output pattern.

Zero-Shot Chain of Thought (CoT) Prompt:

- Instructions include reasoning steps but no examples. The model reasons through the problem step-by-step before producing output.

Few-Shot Chain of Thought (CoT) Prompt:

- Instructions along with examples that show both reasoning and output. Model learns the process through the provided examples.

Ground Truth:

- A subset of manually classified reviews (25-36) to serve as the benchmark to measure prompt model results agreement with **Ground Truth** labels.

1. Exploratory Data Analysis (EDA)

- Understand dataset structure and review content

2. Create Ground Truth

- Manually classify 25-36 reviews as benchmark for accuracy measurement

3. Develop Prompt Templates

- Zero-Shot
- Few-Shot
- Zero-Shot CoT
- Few-Shot CoT

4. Tune Each Prompt Type

- Test variations (temperature, examples, reasoning detail)
- Compare results against Ground Truth
- Select best-performing version of each type

5. Compare Across Prompt Types

- Evaluate all 4 optimized prompts
- Identify top 2 performers

6. Full Dataset Analysis

- Run top 2 prompts on complete dataset (100 reviews)
- Analyze sentiment distribution, tag frequency

7. Business Insights

- Sentiment by City
- Delivery Time patterns by sentiment
- Additional cross-analysis

8. Conclusions & Recommendations

- Summary of findings
- Suggested business actions

Prompt Engineering:

- Develop optimized prompts for sentiment classification using Zero-Shot, Few-Shot, and CoT
- Identify the best-performing prompt type through accuracy comparison
- Create reusable prompt templates for future analysis

Sentiment Insights:

- Classify customer reviews into **Positive**, **Negative**, and **Neutral** categories
- Identify common themes driving negative feedback (**delivery delays**, **food quality**, **pricing**)
- Map sentiment patterns across cities and delivery times

Process Improvement:

- Reduce manual effort in reviewing customer feedback
- Ensure consistent and unbiased sentiment classification
- Enable faster identification of issues requiring business attention

Increase Business Value:

- Generate actionable insights to relevant teams (operations, marketing, quality)
- Prioritize high-impact issues based on sentiment frequency and severity
- Increase scalability (process more reviews) while keeping costs low

Initial Exploratory Data Analysis (EDA)

Food Delivery Sentiment Analysis Description Statistics

Feature	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Customer_ID	100	100.00	C001	1	nan	nan	nan	nan	nan	nan	nan
City	100	100.00	New York	1	nan	nan	nan	nan	nan	nan	nan
Delivery_Time	100.00	nan	nan	nan	48.74	19.99	22.00	32.00	44.50	60.25	98.00
Review	100	98.00	The food was okay. The deli...	3	nan	nan	nan	nan	nan	nan	nan

Food Delivery Sentiment Analysis Basic Information

Feature	Non-Null Count	Null Count	Dtype
Customer_ID	100.00	0.00	object
City	100.00	0.00	object
Delivery_Time	100.00	0.00	int64
Review	100.00	0.00	object

Food Delivery Sentiment Analysis Dataset Shape

Rows	Columns
100	4

Food Delivery Sentiment Analysis Summary of Missing Feature Row Counts

Missing Features	Row Count
0	100

Summary statistics for features including count, mean, standard deviation, and distribution percentiles.

Data Shape — 4 columns, 100 rows of data

Missing values — There are no missing values, imputation is not required

High variance — **Delivery_Times** range from **22** → **98** minutes, with a mean of \approx **49** minutes

Potential outliers — There do not appear to be any outliers

Initial Exploratory Data Analysis (EDA)

Food Delivery Sentiment Analysis Random Data Sample

Customer_ID	City	Delivery_Time	Review
C057	Corpus Christi	66.00	The delivery was a bit slow. The food was okay, but nothing to write home about.
C044	Virginia Beach	61.00	My order was correct. The delivery was a bit slow, but the food was fine.
C038	Atlanta	60.00	The food was decent. The delivery was a little on the slow side.
C024	Nashville	64.00	The food was decent, but I had to wait a while for the delivery. It wasn't late accordi...
C074	Madison	34.00	A reliable choice for a tasty meal delivered right to your door. Highly recommend.
C096	Yonkers	98.00	This was the worst delivery experience I've ever had. The food was inedible and it took...
C062	Greensboro	25.00	Incredibly fast and friendly service. The food was delicious. I'll be back for more.
C075	Fort Wayne	50.00	The food was acceptable. The delivery time was within the estimated window.
C084	North Las Vegas	26.00	Super speedy delivery and the food was packed perfectly. Very satisfied with my order.
C048	Arlington	29.00	I'm impressed with the speed of the delivery. The food was fresh and tasty.
C060	St. Paul	33.00	The delivery was on time and the food was fresh and hot. A great experience.
C086	Boise	37.00	The food is consistently good and the delivery is reliable. A great option for dinner.
C070	Chandler	36.00	The food is consistently great and the delivery is always on time. I'm a happy camper.
C013	Fort Worth	30.00	I'm always happy with my orders from here. The food is delicious and the delivery is pr...
C071	Laredo	46.00	A standard food delivery. The food was decent and it arrived on time.
C090	Tacoma	29.00	Excellent food and speedy delivery. I couldn't ask for more. Highly recommended.
C059	Stockton	59.00	A very average experience. The food was fine and the delivery was on time.
C027	Oklahoma City	36.00	Great variety of options and the delivery is always dependable. A happy customer here.
C080	Scottsdale	35.00	I've ordered from here multiple times and it's always a great experience. Fast and tasty.
C077	Gilbert	73.00	The delivery took much longer than the estimate and the food was cold and unappetizing.

Table containing a random sampling of 20 rows from the imported raw dataset.

Ground Truth Dataset

Ground Truth Summary:

Ground Truth (GT) is a subset of the data with manually created verified, correct labels that represent the known right answers. It is used as a benchmark to measure how well the prompts perform. Without it, accuracy, precision, recall, or F1-scores cannot be calculated. For our purposes, we'll be using the GT for comparison by hand. It should be recognized that since it was created manually, it is subject to human biases.

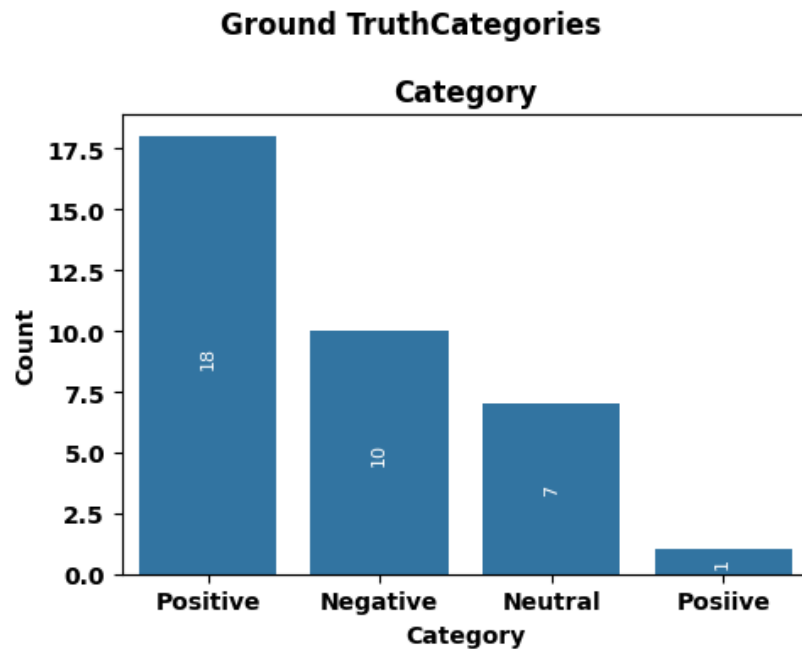
To create the **GT**, 36 reviews were selected from the dataset. Each one was examined, then manually classified as **Positive**, **Negative**, or **Neutral**, and assigned appropriate tags. This was saved as the reference dataset. This step must be completed before running any prompts to prevent bias.

To use, each prompt variation was run on the raw version of the rows in the GT. The models' raw data output was compared against the **GT**. This technique was used to select the best performer of each prompt type.

Ground Truth Dataset

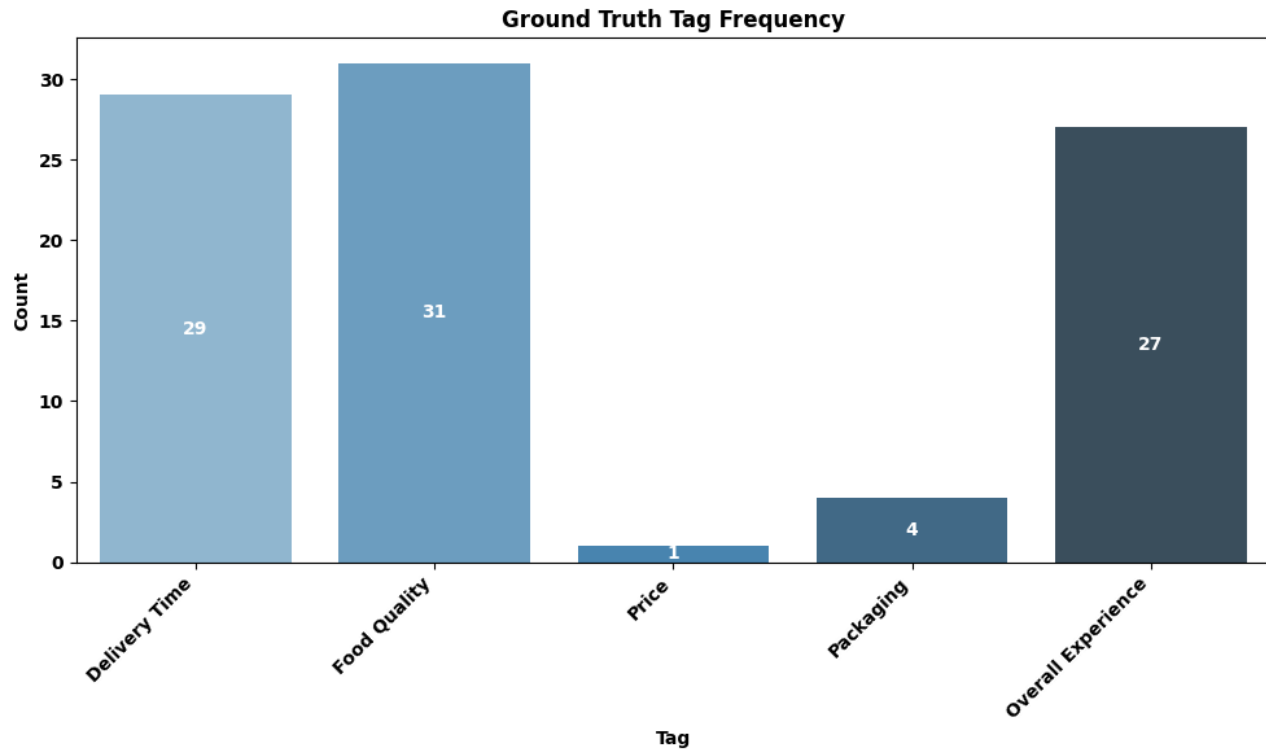
Customer_ID	City	Delivery_Time	Category	Tags	Review
C001	New York	28.00	Positive	Delivery Time,Food Quality,Overall Experience	I'm consistently impressed with this service. Tonight, I decided to try the salmon teri...
C002	Los Angeles	95.00	Negative	Delivery Time,Packaging,Overall Experience	This was an absolute disaster and I will not be ordering from here again. I placed my o...
C003	Chicago	58.00	Negative	Delivery Time,Food Quality,Overall Experience	The experience was a mixed bag. On the one hand, the food itself was quite good. I orde...
C004	Houston	33.00	Positive	Delivery Time,Food Quality,Overall Experience	This was my first time ordering, and it definitely won't be my last. I was hosting a sm...
C005	Phoenix	26.00	Positive	Delivery Time,Food Quality,Overall Experience	Incredibly fast delivery! The driver was courteous and the food was amazing. A+ experie...
C006	Philadelphia	62.00	Negative	Delivery Time,Food Quality,Price,Overall Experience	My order was correct and the driver was polite, so no issues there. The delivery time w...
C007	San Antonio	40.00	Positive	Delivery Time,Food Quality,Overall Experience	Great food and timely delivery. The portion sizes were generous. Very satisfied.
C008	San Diego	31.00	Positive	Food Quality,Overall Experience	I've ordered from here several times and the quality is always top-notch. The ingredien...
C009	Dallas	22.00	Positive	Delivery Time,Overall Experience	Lightning-fast delivery! The food was still steaming when it arrived. Impressive!
C010	San Jose	85.00	Negative	Delivery Time,Food Quality,Overall Experience	I wish I could give zero stars. The delivery took nearly two hours. When the driver fin...
C011	Austin	38.00	Neutral	Food Quality, Overall Experience	The quality of the food is consistently excellent. Delivery is reliable and the drivers...
C012	Jacksonville	55.00	Negative	Delivery Time,Food Quality	Food was okay, but the delivery was slow. It arrived later than the estimated time.
C013	Fort Worth	30.00	Positive	Delivery Time,Food Quality,Overall Experience	I'm always happy with my orders from here. The food is delicious and the delivery is pr...
C014	Columbus	48.00	Neutral	Delivery Time,Food Quality	The food was as expected. The delivery time was reasonable.
C015	Charlotte	42.00	Positive	Delivery Time,Food Quality,Overall Experience	A fantastic experience from start to finish. Easy ordering, fast delivery, and incredib...
C016	San Francisco	25.00	Positive	Delivery Time,Packaging,Overall Experience	Super quick delivery and the food was packed with care. Everything was perfect.
C017	Indianapolis	70.00	Negative	Delivery Time,Food Quality	The food was cold and soggy by the time it arrived. The delivery was extremely slow.
C018	Seattle	34.00	Positive	Food Quality,Overall Experience	The food is always fresh and flavorful. The delivery service is top-notch. I particular...
C019	Denver	53.00	Neutral	Delivery Time,Food Quality	The food was satisfactory. The delivery time was average. Nothing special to report.
C020	Washington	29.00	Positive	Delivery Time,Food Quality	I was pleasantly surprised by how fast the delivery was. The food was excellent as always.
C021	Boston	45.00	Neutral	Delivery Time,Food Quality,Overall Experience	An average experience. The food was okay and the delivery was on time.
C022	El Paso	41.00	Positive	Food Quality,Overall Experience	Reliable service and delicious meals. I've never had a bad experience with them.
C023	Detroit	32.00	Positive	Delivery Time,Food Quality,Packaging,Overall Experience	The delivery was even earlier than the estimated time. The food was hot and delicious. ...
C024	Nashville	64.00	Negative	Delivery Time,Food Quality,Overall Experience	The food was decent, but I had to wait a while for the delivery. It wasn't late accordi...
C025	Portland	27.00	Positive	Delivery Time,Food Quality,Overall Experience	Exceptional service! The food is always on point and the delivery is incredibly fast.
C026	Memphis	93.00	Negative	Delivery Time,Packaging,Overall Experience	I am beyond upset. Not only was my order an hour late, but they completely forgot one o...
C027	Oklahoma City	36.00	Positive	Overall Experience	Great variety of options and the delivery is always dependable. A happy customer here.
C028	Las Vegas	49.00	Neutral	Delivery Time,Food Quality	The food was fine. The delivery time was within the expected range.
C029	Louisville	24.00	Positive	Delivery Time,Food Quality,Overall Experience	Wow, that was fast! The food was as good as dining in. Will be ordering again soon.
C030	Baltimore	65.00	Negative	Delivery Time,Food Quality	The delivery took longer than I would have liked. The food was alright.
C031	Milwaukee	39.00	Positive	Food Quality,Overall Experience	Consistently great food and service. The delivery drivers are always professional and p...
C032	Albuquerque	56.00	Neutral	Delivery Time,Food Quality	The food was fine. The delivery was a bit slow, but not excessively so.
C033	Tucson	34.00	Positive	Food Quality,Overall Experience	My order arrived fresh and hot. The delivery was seamless. Very pleased.
C034	Fresno	78.00	Negative	Delivery Time,Food Quality,Overall Experience	The delivery driver got lost and my food was cold. Not a good experience.
C035	Sacramento	23.00	Positive	Delivery Time,Food Quality,Overall Experience	Incredibly efficient service. The food was delicious and the delivery was super fast.
C036	Kansas City	46.00	Neutral	Delivery Time,Food Quality	A pretty standard delivery experience. The food was okay and it arrived on time.

Ground Truth Dataset



Plot illustrating distribution **Ground Truth Categories**

Ground Truth Dataset



Plot illustrating distribution of Tags

Original Zero-Shot Prompt

Original Zero-Shot Prompt Template:

ROLE: You are a customer feedback analyst for a food delivery service.

TASK: Analyze the uploaded customer reviews and provide structured sentiment analysis.

Use Temperature=1

CLASSIFICATION GUIDELINES:

Positive: Customer expresses satisfaction, praise, or intent to return

Negative: Customer expresses dissatisfaction, complaints, or intent to not return

Neutral: Mixed feedback or neither clearly positive nor negative

TAGS (select ALL applicable from this list ONLY):

- Delivery Time, Food Quality, Price, Packaging
- Overall Experience: only when customer states "experience," "service," or uses general rating language

OUTPUT REQUIREMENTS:

- Category: Classify as exactly one of: Positive, Negative, or Neutral
- Tags: Select ALL applicable tags from the list above.
- Suggested_Action: Provide ONE specific, actionable business recommendation (10 words max)

Confidence: High / Medium / Low

OUTPUT FORMAT:

Return results as CSV format only. No explanations, no extra text.

Use these exact headers: Customer_ID, Category, Tags, Suggested_Action, Confidence

Rules:

- Wrap Tags in quotes if multiple: "Delivery Time, Food Quality"
- Wrap Suggested_Action in quotes
- Suggested_Action: 10 words max, no commas
- One row per customer

Run above Zero-Shot Prompt; Run numeric comparison against Ground Truth Results

Zero-Shot Prompt Definition Explanation

I. Core Purpose

- **Role:**
 - Frame of Reference provided to the AI
 - It tells the AI to use the relevant expertise, language, and thought process for that specific job or situation.
- **Task:**
 - Main instruction telling AI what to do. Core command that defines the objective.
 - It directs the AI to perform a specific action, like summarizing, classifying, generating, or analyzing.

II. Model Behavior Settings

- **Temperature:** 1 (More diverse/creative output)

III. Analysis Guidelines

- **Classification (Category):**
 - Overall category to be assigned to each review based on the review's textual contents
- **Tags (Select ALL applicable from list ONLY):**
 - Overall Experience: ONLY if customer states "experience," "service," or uses general rating language.
- **Confidence:**
 - Is the internal assessment of the AI's confidence in the model prediction (reliability)

IV. Output Structure & Format

- **Required Fields:** Fields to be contained in the result output
- **Format:** What format the output will have (in this case .csv)
- **CSV Headers:** Field/Column headers for the csv output
- **Specific Rules:** Specific rules to be applied to the output.

Original Zero-Shot Prompt Results

Customer_ID	Category	Tags	Suggested_Action	Confidence
C001	Positive	Delivery Time, Food Quality, Overall Experience	Highlight quality food, fast delivery, and friendly service.	High
C002	Negative	Delivery Time, Food Quality, Overall Experience	Investigate wrong order, late delivery, and customer service.	High
C003	Neutral	Delivery Time, Food Quality, Overall Experience	Improve delivery tracking and food temperature maintenance.	High
C004	Positive	Delivery Time, Food Quality, Overall Experience	Promote hot, fresh food and seamless ordering process.	High
C005	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize fast delivery, courteous driver, and amazing food.	High
C006	Neutral	Food Quality, Price	Review specialty sandwich quality, portion, and price alignment.	High
C007	Positive	Delivery Time, Food Quality	Highlight great food, timely delivery, and generous portions.	High
C008	Positive	Delivery Time, Food Quality, Overall Experience	Promote consistent top-notch quality and professional drivers.	High
C009	Positive	Delivery Time	Showcase lightning-fast delivery and hot food arrival.	High
C010	Negative	Delivery Time, Food Quality, Overall Experience	Address long delivery, rude driver, and poor food quality.	High
C011	Positive	Delivery Time, Food Quality	Emphasize consistent food quality and friendly drivers.	High
C012	Negative	Delivery Time, Food Quality	Improve delivery speed and ensure timely arrival.	High
C013	Positive	Delivery Time, Food Quality	Highlight delicious food and prompt delivery service.	High
C014	Neutral	Delivery Time, Food Quality	Seek feedback to elevate food and delivery experience.	High
C015	Positive	Delivery Time, Food Quality, Overall Experience	Promote easy ordering, fast delivery, and incredible food.	High
C016	Positive	Delivery Time, Packaging	Highlight quick delivery and careful food packaging.	High
C017	Negative	Delivery Time, Food Quality	Address cold, soggy food and extremely slow delivery.	High
C018	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize fresh, flavorful food and top-notch delivery service.	High
C019	Neutral	Delivery Time, Food Quality	Identify areas to improve satisfactory food and average delivery.	High
C020	Positive	Delivery Time, Food Quality	Highlight fast delivery and consistently excellent food.	High
C021	Neutral	Delivery Time, Food Quality, Overall Experience	Survey customer to identify specific areas for improvement.	High
C022	Positive	Food Quality, Overall Experience	Promote reliable service and delicious meals for customer retention.	High
C023	Positive	Delivery Time, Food Quality, Packaging, Overall Experience	Highlight early delivery, hot food, and fresh sushi quality.	High
C024	Neutral	Delivery Time, Food Quality, Overall Experience	Review initial delivery estimates and enhance meal excitement.	High
C025	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize exceptional service, on-point food, and fast delivery.	High
C026	Negative	Delivery Time, Food Quality, Overall Experience	Investigate late delivery, missing items, and unhelpful customer service.	High
C027	Positive	Overall Experience	Promote great variety and dependable delivery service.	High
C028	Neutral	Delivery Time, Food Quality	Gather feedback to improve fine food and expected delivery.	High
C029	Positive	Delivery Time, Food Quality, Overall Experience	Showcase fast delivery and restaurant-quality food.	High
C030	Negative	Delivery Time, Food Quality	Address slow delivery times and improve food quality.	High
C031	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize consistent food, professional drivers, and great service.	High
C032	Neutral	Delivery Time, Food Quality	Target slight delivery delays for improvement.	High
C033	Positive	Delivery Time, Food Quality	Highlight fresh, hot food and seamless delivery experience.	High
C034	Negative	Delivery Time, Food Quality, Overall Experience	Improve driver navigation and food temperature control.	High
C035	Positive	Delivery Time, Food Quality, Overall Experience	Promote efficient service, delicious food, and super fast delivery.	High
C036	Neutral	Delivery Time, Food Quality, Overall Experience	Identify opportunities to make "standard" experience stand out.	High

Original Zero-Shot Prompt Results

Numerical Comparison Summary (Zero-Shot Prompt, Original Tags, T=1 vs. Ground Truth):

1. Category Differences (vs. Ground Truth):

- C003: My Neutral vs. GT Negative - 1 difference
- C006: My Neutral vs. GT Negative - 1 difference
- C011: My Positive vs. GT Neutral - 1 difference
- C012: My Negative vs. GT Negative - Match
- C024: My Neutral vs. GT Negative - 1 difference
- C030: My Negative vs. GT Negative - Match

Total Category Differences: 4

2. Tags Differences (vs. Ground Truth):

- C002: My (Food Quality, Overall Experience) vs. GT (Packaging, Overall Experience) - 2 differences
- C006: My (no Delivery Time) vs. GT (Delivery Time) - 1 difference
- C007: My (no Overall Experience) vs. GT (Overall Experience) - 1 difference
- C008: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C009: My (no Overall Experience) vs. GT (Overall Experience) - 1 difference
- C011: My (Delivery Time, no Overall Experience) vs. GT (no Delivery Time, Overall Experience) - 2 differences
- C013: My (no Overall Experience) vs. GT (Overall Experience) - 1 difference
- C016: My (no Overall Experience) vs. GT (Overall Experience) - 1 difference
- C018: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C024: My (no Food Quality) vs. GT (Food Quality) - 1 difference
- C026: My (Food Quality, Overall Experience) vs. GT (Packaging, Overall Experience) - 2 differences
- C027: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C031: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C033: My (Delivery Time, no Overall Experience) vs. GT (no Delivery Time, Overall Experience) - 2 differences
- C036: My (Overall Experience) vs. GT (no Overall Experience) - 1 difference

Total Tags Differences: 20

3. Total Rows with at least one difference (Category or Tags): 15 out of 36 rows.

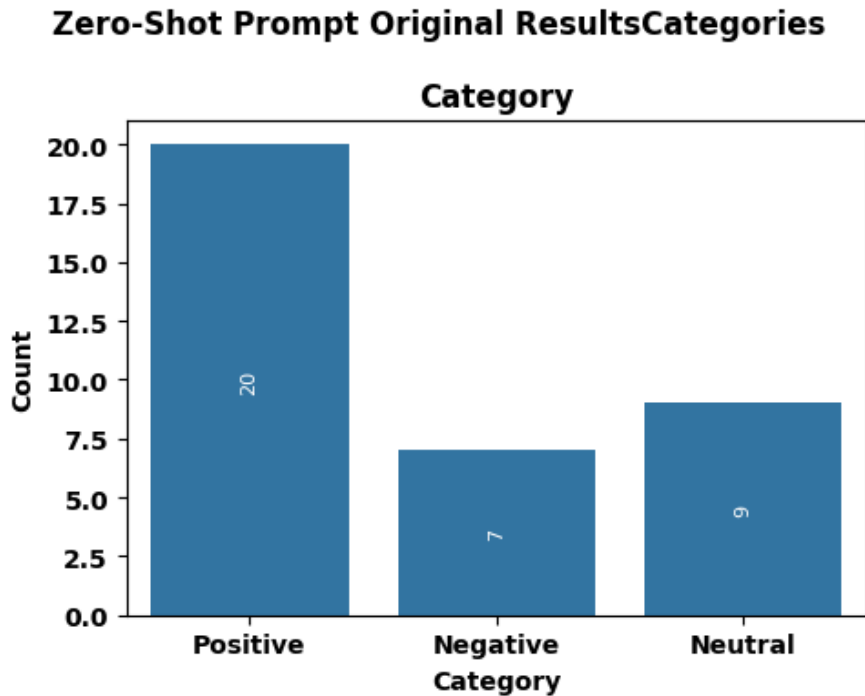
Original Zero-Shot Prompt Results

Conclusion:

- Zero-Shot prompt with the original settings, performed identically at temperatures = (0, 0.2, and 1).
- Category alignment: 4 discrepancies.
- Tag alignment: 20 discrepancies.
- Total rows with differences: 15.
- This consistent performance across various temperatures for the Zero-Shot prompt (without examples) strongly confirms that the model's interpretation of these specific, stricter textual guidelines is the sole determinant of its output in this scenario. The discrepancies are entirely rooted in the prompt's instructions themselves, not in any randomness or variability controlled by temperature.

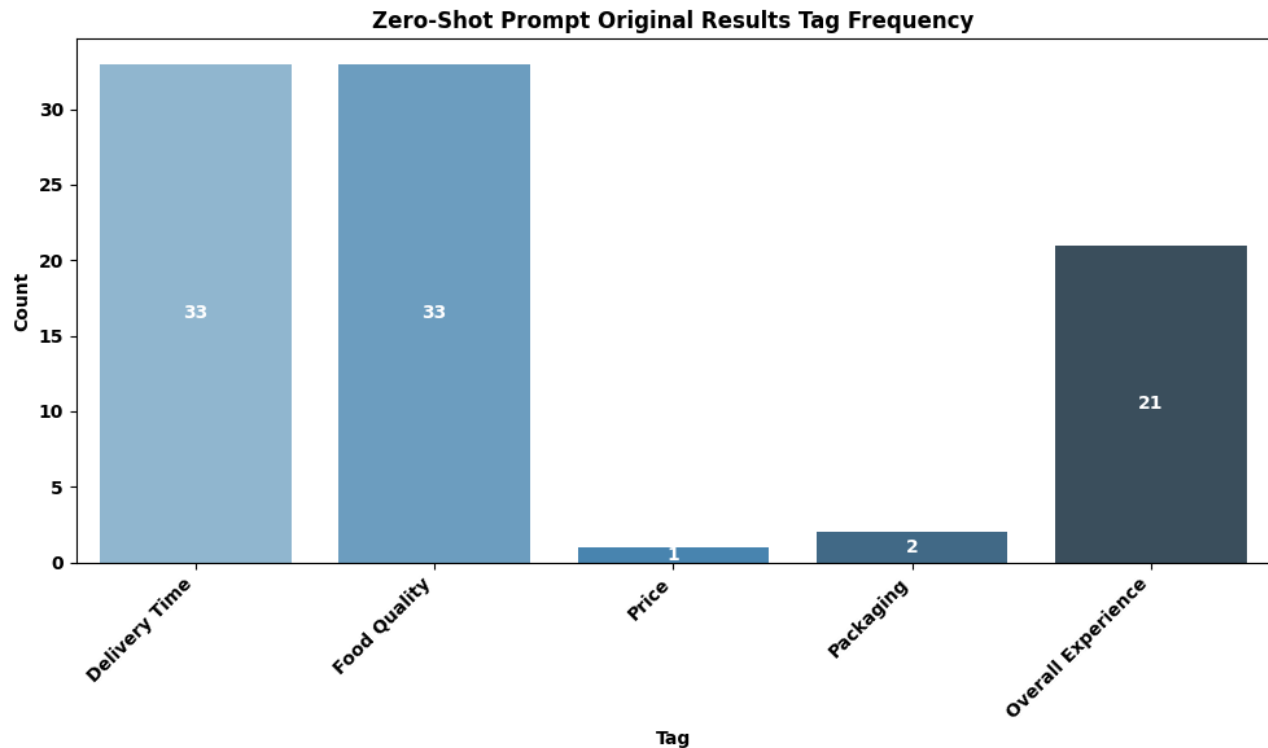
*** Result analysis and comparisons between **prompt** and **Ground Truth** were performed by AI Studio (Gemini) via the command prompt: “Run numeric comparison against Ground Truth Results”

Original Few-Shot Prompt Results



Plot illustrating distribution of **Categories**

Original Few-Shot Prompt Results



Plot illustrating distribution of **Tags**

Updated Zero-Shot Prompt

Updated Zero-Shot Prompt Template:

The “**Updated Zero-Shot Prompt Template**” is based on the original zero prompt with only the following difference:

- The Overall Experience tag definition was modified to:
 - Overall Experience: Apply when the customer explicitly states "experience," "service," "will/won't order again," or uses strong general rating language that clearly reflects their overall sentiment about the entire transaction (e.g., "very satisfied," "impressive," "always happy," "everything was perfect," "very pleased," "loved everything," "terrible meal")

Updated Zero-Shot Prompt Results

Customer_ID	Category	Tags	Suggested_Action	Confidence
C001	Positive	Delivery Time, Food Quality, Overall Experience	Highlight quality food, fast delivery, and friendly service.	High
C002	Negative	Delivery Time, Packaging, Food Quality, Overall Experience	Investigate wrong order, late delivery, and customer service.	High
C003	Negative	Delivery Time, Food Quality, Overall Experience	Improve delivery tracking and food temperature maintenance.	High
C004	Positive	Delivery Time, Food Quality, Overall Experience	Promote hot, fresh food and seamless ordering process.	High
C005	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize fast delivery, courteous driver, and amazing food.	High
C006	Negative	Delivery Time, Food Quality, Price, Overall Experience	Review specialty sandwich quality, portion, and price alignment.	High
C007	Positive	Delivery Time, Food Quality, Overall Experience	Highlight great food, timely delivery, and generous portions.	High
C008	Positive	Delivery Time, Food Quality, Overall Experience	Promote consistent top-notch quality and professional drivers.	High
C009	Positive	Delivery Time, Overall Experience	Showcase lightning-fast delivery and hot food arrival.	High
C010	Negative	Delivery Time, Food Quality, Overall Experience	Address long delivery, rude driver, and poor food quality.	High
C011	Neutral	Delivery Time, Food Quality, Overall Experience	Emphasize consistent food quality and friendly drivers.	High
C012	Negative	Delivery Time, Food Quality	Improve delivery speed and ensure timely arrival.	High
C013	Positive	Delivery Time, Food Quality, Overall Experience	Highlight delicious food and prompt delivery service.	High
C014	Neutral	Delivery Time, Food Quality	Seek feedback to elevate food and delivery experience.	High
C015	Positive	Delivery Time, Food Quality, Overall Experience	Promote easy ordering, fast delivery, and incredible food.	High
C016	Positive	Delivery Time, Packaging, Overall Experience	Highlight quick delivery and careful food packaging.	High
C017	Negative	Delivery Time, Food Quality	Address cold, soggy food and extremely slow delivery.	High
C018	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize fresh, flavorful food and top-notch delivery service.	High
C019	Neutral	Delivery Time, Food Quality	Identify areas to improve satisfactory food and average delivery.	High
C020	Positive	Delivery Time, Food Quality	Highlight fast delivery and consistently excellent food.	High
C021	Neutral	Delivery Time, Food Quality, Overall Experience	Survey customer to identify specific areas for improvement.	High
C022	Positive	Food Quality, Overall Experience	Promote reliable service and delicious meals for customer retention.	High
C023	Positive	Delivery Time, Food Quality, Packaging, Overall Experience	Highlight early delivery, hot food, and fresh sushi quality.	High
C024	Negative	Delivery Time, Food Quality, Overall Experience	Review initial delivery estimates and enhance meal excitement.	High
C025	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize exceptional service, on-point food, and fast delivery.	High
C026	Negative	Delivery Time, Packaging, Food Quality, Overall Experience	Investigate late delivery, missing items, and unhelpful customer service.	High
C027	Positive	Delivery Time, Overall Experience	Promote great variety and dependable delivery service.	High
C028	Neutral	Delivery Time, Food Quality	Gather feedback to improve fine food and expected delivery.	High
C029	Positive	Delivery Time, Food Quality, Overall Experience	Showcase fast delivery and restaurant-quality food.	High
C030	Negative	Delivery Time, Food Quality	Address slow delivery times and improve food quality.	High
C031	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize consistent food, professional drivers, and great service.	High
C032	Neutral	Delivery Time, Food Quality	Target slight delivery delays for improvement.	High
C033	Positive	Delivery Time, Food Quality, Overall Experience	Highlight fresh, hot food and seamless delivery experience.	High
C034	Negative	Delivery Time, Food Quality, Overall Experience	Improve driver navigation and food temperature control.	High
C035	Positive	Delivery Time, Food Quality, Overall Experience	Promote efficient service, delicious food, and super fast delivery.	High
C036	Neutral	Delivery Time, Food Quality, Overall Experience	Identify opportunities to make "standard" experience stand out.	High

Numerical Comparison Summary (Zero-Shot Prompt, Updated Overall Experience Tag, T=0.2 vs. Ground Truth):

1. Category Differences (vs. Ground Truth):

Total Tags Differences: 0

2. Tags Differences (vs. Ground Truth):

- **C002:** My (Delivery Time, Packaging, Food Quality, Overall Experience) vs. GT (Delivery Time, Packaging, Overall Experience) - 1 difference (My output has an extra "Food Quality" tag)
- **C006:** My (Delivery Time, Food Quality, Price, Overall Experience) vs. GT (Delivery Time, Food Quality, Price, Overall Experience) - Match
- **C007:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Delivery Time, Food Quality, Overall Experience) - Match
- **C00:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Food Quality, Overall Experience) - 1 difference (My output has an extra "Delivery Time" tag)
- **C009:** My (Delivery Time, Overall Experience) vs. GT (Delivery Time, Overall Experience) - Match
- **C011:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Food Quality, Overall Experience) - 1 difference (My output has an extra "Delivery Time" tag)
- **C013:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Delivery Time, Food Quality, Overall Experience) - Match
- **C016:** My (Delivery Time, Packaging, Overall Experience) vs. GT (Delivery Time, Packaging, Overall Experience) - Match
- **C018:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Food Quality, Overall Experience) - 1 difference (My output has an extra "Delivery Time" tag)
- **C024:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Delivery Time, Food Quality, Overall Experience) - Match
- **C026:** My (Delivery Time, Packaging, Food Quality, Overall Experience) vs. GT (Delivery Time, Packaging, Overall Experience) - 1 difference (My output has an extra "Food Quality" tag)
- **C027:** My (Delivery Time, Overall Experience) vs. GT (Overall Experience) - 1 difference (My output has an extra "Delivery Time" tag)
- **C031:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Food Quality, Overall Experience) - 1 difference (My output has an extra "Delivery Time" tag)
- **C033:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Food Quality, Overall Experience) - 1 difference (My output has an extra "Delivery Time" tag)
- **C036:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Delivery Time, Food Quality) - 1 difference (My output has an extra "Overall Experience" tag)

Total Tags Differences: 10

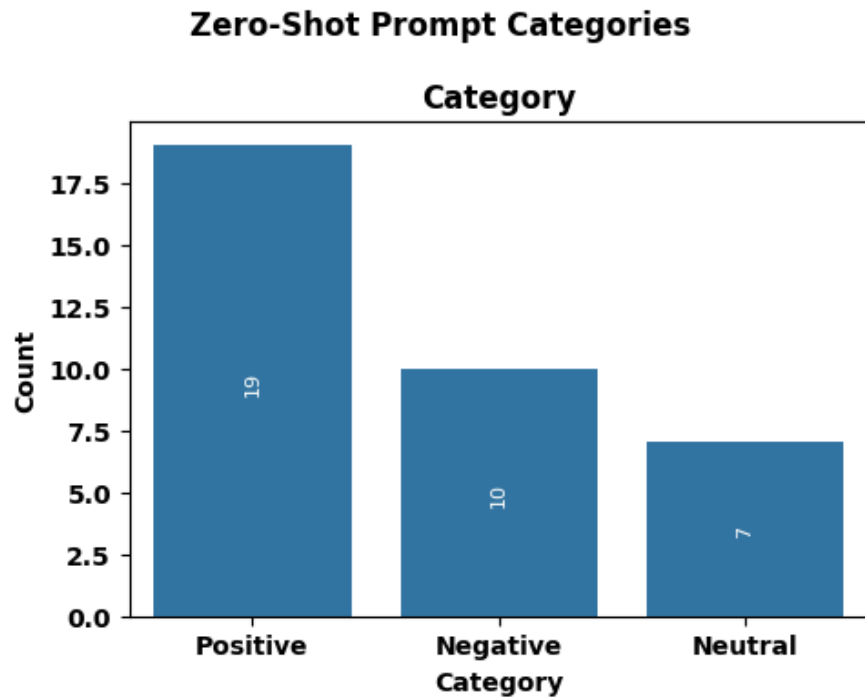
3. Total Rows with at least one difference (Category or Tags): 8 out of 36 rows.

Updated Zero-Shot Prompt Results

Conclusion:

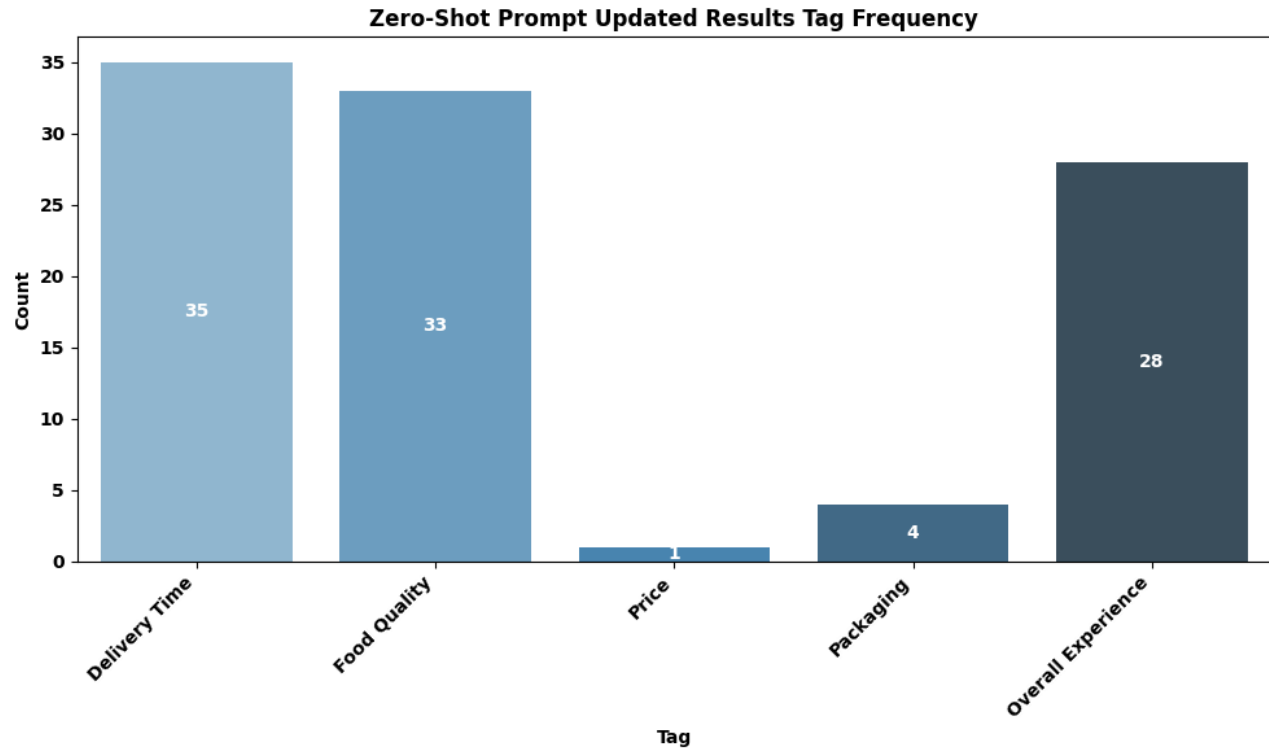
- Updated Zero-Shot prompt with the original settings, performed identically at temperatures = (0, 0.2, and 1).
- Category alignment: 0 discrepancies.
- Tag alignment: 10 discrepancies.
- Total rows with differences: 8.
- The refined textual guidelines are highly effective in teaching the model the interpretation for **Category classification**, achieving perfect alignment. The **10 Tag** discrepancies are not due to randomness or temperature variability. This occurs because the model's understanding of specific tag rules (like "Delivery Time" or "Packaging" vs. "Food Quality") doesn't perfectly match the **Ground Truth**, even with the current instructions.

Updated Zero-Shot Prompt Results



Plot illustrating distribution of **Categories**

Updated Zero-Shot Prompt Results



Plot illustrating distribution of **Tags**

Original Few-Shot Prompt

Original Few-Shot Prompt Template:

ROLE: You are a customer feedback analyst for a food delivery service.

TASK: Analyze the uploaded customer reviews and provide structured sentiment analysis.

Use Temperature=0.2

CLASSIFICATION GUIDELINES:

Positive: Customer expresses satisfaction, praise, or intent to return

Negative: Customer expresses dissatisfaction, complaints, or intent to not return

Neutral: Mixed feedback or neither clearly positive nor negative

TAGS (select ALL applicable from this list ONLY):

- Delivery Time, Food Quality, Price, Packaging
- Overall Experience: only when customer states "experience," "service," or uses general rating language

OUTPUT REQUIREMENTS:

- Category: Classify as exactly one of: Positive, Negative, or Neutral
- Tags: Select ALL applicable tags from the list above.
- Suggested_Action: Provide ONE specific, actionable business recommendation (10 words max)

Confidence: High / Medium / Low

OUTPUT FORMAT:

Return results as CSV format only. No explanations, no extra text.

Use these exact headers: Customer_ID, Category, Tags, Suggested_Action, Confidence

Rules:

- Wrap Tags in quotes if multiple: "Delivery Time, Food Quality"
- Wrap Suggested_Action in quotes
- Suggested_Action: 10 words max, no commas
- One row per customer

Original Few-Shot Prompt

EXAMPLES:

Review: "The food arrived 15 minutes late, which was annoying, but the pad thai was absolutely delicious and still hot. The driver was also very apologetic."

Category: Neutral

Tags: Delivery Time, Food Quality, Overall Experience

Suggested Action: Investigate delivery delay, maintain food quality.

Review: "I ordered the large family meal, and the portion sizes were surprisingly small for the price. The chicken was a bit dry, but the sides were fresh. Overall, it felt like a rip-off."

Category: Negative

Tags: Food Quality, Price, Overall Experience

Suggested Action: Review portion sizes and chicken preparation for value.

Review: "My smoothie spilled all over the bag because the lid wasn't properly sealed. The delivery was fast, but half my order was a sticky mess. Very disappointed."

Category: Negative

Tags: Delivery Time, Packaging, Overall Experience

Suggested Action: Improve packaging seal quality for liquid items.

Review: "The vegetarian burrito was exactly as described and tasted good. Delivery was on time. It was a perfectly fine meal, nothing to rave about."

Category: Neutral

Tags: Delivery Time, Food Quality

Suggested Action: Seek feedback to elevate "fine" to "memorable."

Review: "This is my go-to service. The sushi is always fresh, and the drivers are consistently friendly. I've never had a bad experience, which is why I keep coming back."

Category: Positive

Tags: Delivery Time, Food Quality, Overall Experience

Suggested Action: Highlight consistent quality and reliable service in marketing.

Run above Zero-Shot Prompt; Run numeric comparison against Ground Truth Results

Original Few-Shot Prompt Results

Customer_ID	Category	Tags	Suggested_Action	Confidence
C001	Positive	Delivery Time, Food Quality, Overall Experience	Highlight quality food, fast delivery, and friendly service.	High
C002	Negative	Delivery Time, Food Quality, Overall Experience	Investigate wrong order, late delivery, and customer service.	High
C003	Neutral	Delivery Time, Food Quality, Overall Experience	Improve delivery tracking and food temperature maintenance.	High
C004	Positive	Delivery Time, Food Quality, Overall Experience	Promote hot, fresh food and seamless ordering process.	High
C005	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize fast delivery, courteous driver, and amazing food.	High
C006	Neutral	Food Quality, Price	Review specialty sandwich quality, portion, and price alignment.	High
C007	Positive	Delivery Time, Food Quality	Highlight great food, timely delivery, and generous portions.	High
C008	Positive	Delivery Time, Food Quality, Overall Experience	Promote consistent top-notch quality and professional drivers.	High
C009	Positive	Delivery Time	Showcase lightning-fast delivery and hot food arrival.	High
C010	Negative	Delivery Time, Food Quality, Overall Experience	Address long delivery, rude driver, and poor food quality.	High
C011	Positive	Delivery Time, Food Quality	Emphasize consistent food quality and friendly drivers.	High
C012	Negative	Delivery Time, Food Quality	Improve delivery speed and ensure timely arrival.	High
C013	Positive	Delivery Time, Food Quality	Highlight delicious food and prompt delivery service.	High
C014	Neutral	Delivery Time, Food Quality	Seek feedback to elevate food and delivery experience.	High
C015	Positive	Delivery Time, Food Quality, Overall Experience	Promote easy ordering, fast delivery, and incredible food.	High
C016	Positive	Delivery Time, Packaging	Highlight quick delivery and careful food packaging.	High
C017	Negative	Delivery Time, Food Quality	Address cold, soggy food and extremely slow delivery.	High
C018	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize fresh, flavorful food and top-notch delivery service.	High
C019	Neutral	Delivery Time, Food Quality	Identify areas to improve satisfactory food and average delivery.	High
C020	Positive	Delivery Time, Food Quality	Highlight fast delivery and consistently excellent food.	High
C021	Neutral	Delivery Time, Food Quality, Overall Experience	Survey customer to identify specific areas for improvement.	High
C022	Positive	Food Quality, Overall Experience	Promote reliable service and delicious meals for customer retention.	High
C023	Positive	Delivery Time, Food Quality, Packaging, Overall Experience	Highlight early delivery, hot food, and fresh sushi quality.	High
C024	Neutral	Delivery Time, Food Quality, Overall Experience	Review initial delivery estimates and enhance meal excitement.	High
C025	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize exceptional service, on-point food, and fast delivery.	High
C026	Negative	Delivery Time, Food Quality, Overall Experience	Investigate late delivery, missing items, and unhelpful customer service.	High
C027	Positive	Overall Experience	Promote great variety and dependable delivery service.	High
C028	Neutral	Delivery Time, Food Quality	Gather feedback to improve fine food and expected delivery.	High
C029	Positive	Delivery Time, Food Quality, Overall Experience	Showcase fast delivery and restaurant-quality food.	High
C030	Negative	Delivery Time, Food Quality	Address slow delivery times and improve food quality.	High
C031	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize consistent food, professional drivers, and great service.	High
C032	Neutral	Delivery Time, Food Quality	Target slight delivery delays for improvement.	High
C033	Positive	Delivery Time, Food Quality	Highlight fresh, hot food and seamless delivery experience.	High
C034	Negative	Delivery Time, Food Quality, Overall Experience	Improve driver navigation and food temperature control.	High
C035	Positive	Delivery Time, Food Quality, Overall Experience	Promote efficient service, delicious food, and super fast delivery.	High
C036	Neutral	Delivery Time, Food Quality, Overall Experience	Identify opportunities to make "standard" experience stand out.	High

Numerical Comparison Summary (Zero-Shot Prompt, Original Tags, T=0.2 vs. Ground Truth):

1. Category Differences (vs. Ground Truth):

Total Category Differences: 0

2. Tags Differences (vs. Ground Truth):

- **C002:** My (Delivery Time, Packaging, Food Quality, Overall Experience) vs. GT (Delivery Time, Packaging, Overall Experience) - 1 difference (My output has an extra "Food Quality" tag)
- **C006:** My (Delivery Time, Food Quality, Price, Overall Experience) vs. GT (Delivery Time, Food Quality, Price, Overall Experience) - Match
- **C007:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Delivery Time, Food Quality, Overall Experience) - Match
- **C008:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Food Quality, Overall Experience) - 1 difference (My output has an extra "Delivery Time" tag)
- **C009:** My (Delivery Time, Overall Experience) vs. GT (Delivery Time, Overall Experience) - Match
- **C011:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Food Quality, Overall Experience) - 1 difference (My output has an extra "Delivery Time" tag)
- **C013:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Delivery Time, Food Quality, Overall Experience) - Match
- **C016:** My (Delivery Time, Packaging, Overall Experience) vs. GT (Delivery Time, Packaging, Overall Experience) - Match
- **C018:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Food Quality, Overall Experience) - 1 difference (My output has an extra "Delivery Time" tag)
- **C024:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Delivery Time, Food Quality, Overall Experience) - Match
- **C026:** My (Delivery Time, Packaging, Food Quality, Overall Experience) vs. GT (Delivery Time, Packaging, Overall Experience) - 1 difference (My output has an extra "Food Quality" tag)
- **C027:** My (Delivery Time, Overall Experience) vs. GT (Overall Experience) - 1 difference (My output has an extra "Delivery Time" tag)
- **C031:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Food Quality, Overall Experience) - 1 difference (My output has an extra "Delivery Time" tag)
- **C033:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Food Quality, Overall Experience) - 1 difference (My output has an extra "Delivery Time" tag)
- **C036:** My (Delivery Time, Food Quality, Overall Experience) vs. GT (Delivery Time, Food Quality) - 1 difference (My output has an extra "Overall Experience" tag)

Total Tags Differences: 10

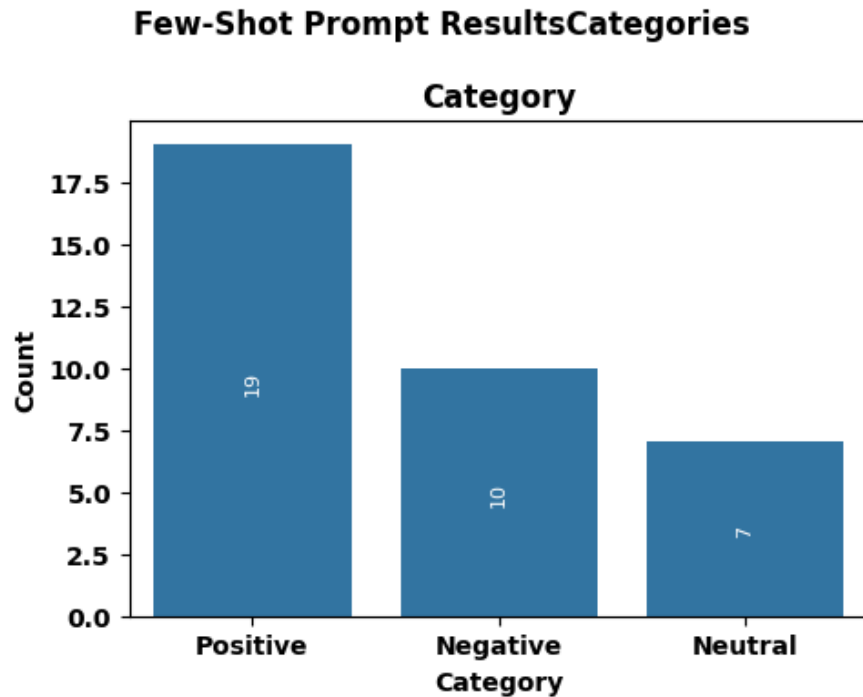
3. Total Rows with at least one difference (Category or Tags): 8 out of 36 rows.

Original Few-Shot Prompt Results

Conclusion:

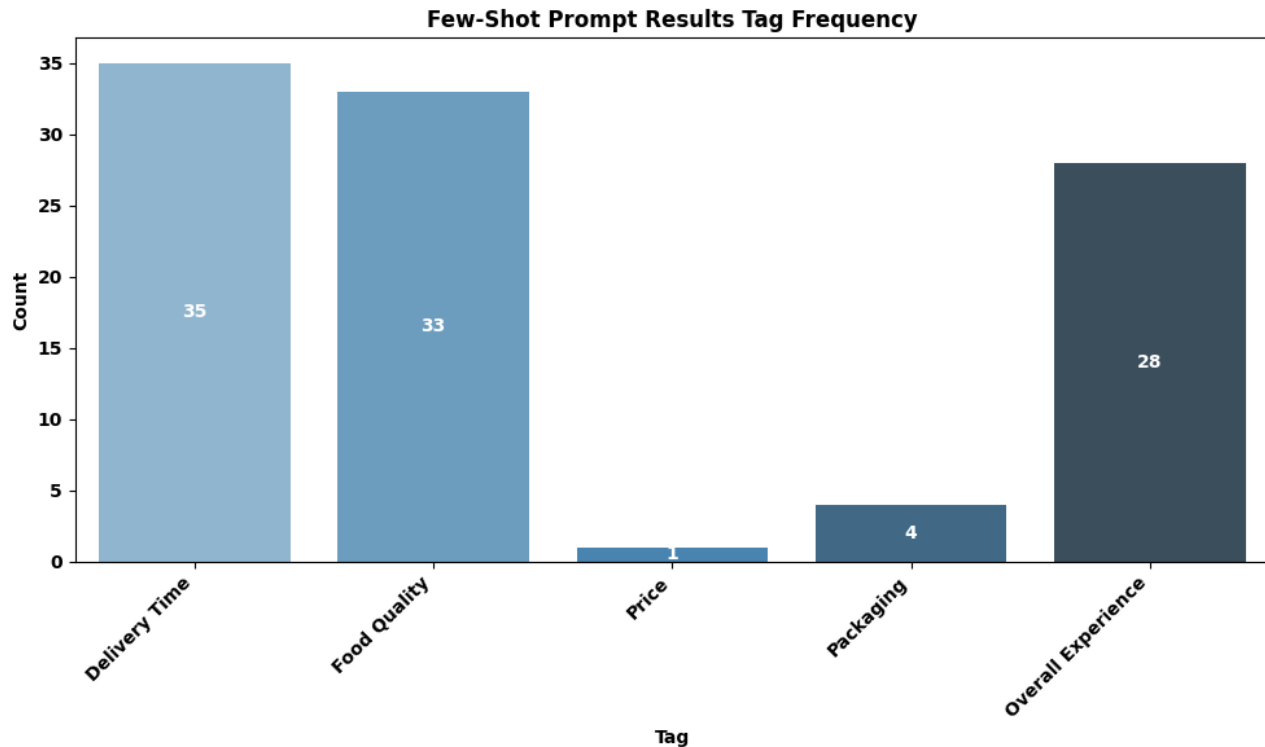
- Few-Shot prompt with the original definition settings, performed identically at temperatures = (0, 0.2, and 1).
- Category alignment: 0 discrepancies.
- Tag alignment: 10 discrepancies.
- Total rows with differences: 8.
- The 5 custom examples are highly effective in teaching the model the interpretation for **Category** classification, achieving perfect alignment. This consistent performance across various temperatures for the Few-Shot prompt (without examples) strongly confirms that the model's interpretation of these specific, stricter textual guidelines is the sole determinant of its output in this scenario. The **Tag** discrepancies are entirely rooted in the prompt's instructions themselves, not in any randomness or variability controlled by temperature.

Original Few-Shot Prompt Results



Plot illustrating distribution of **Categories**

Original Few-Shot Prompt Results



Plot illustrating distribution of **Tags**

Updated Few-Shot Prompt

Updated Few-Shot Prompt Template:

ROLE: You are a customer feedback analyst for a food delivery service.

TASK: Analyze the uploaded customer reviews and provide structured sentiment analysis.

Use Temperature=0.2

CLASSIFICATION GUIDELINES:

Positive: Customer expresses satisfaction, praise, or intent to return

Negative: Customer expresses dissatisfaction, complaints, or intent to not return. If a review contains both positive/neutral and significant negative feedback (e.g., major delays, cold food, wrong order, poor value for money), classify as Negative.

Neutral: Mixed feedback where positive and negative aspects are balanced, or neither clearly positive nor negative.

TAGS (select ALL applicable from this list ONLY):

- Delivery Time
- Food Quality: Includes taste, temperature (if affecting quality), freshness, accuracy of order content (e.g., wrong dish, missing items).
- Price
- Packaging: Includes physical integrity of packaging, spillage, or incorrect items due to packing error.
- Overall Experience: Apply when the customer explicitly states "experience," "service," "will/won't order again," or uses strong general rating language that clearly reflects their overall sentiment about the entire transaction (e.g., "very satisfied," "impressive," "always happy," "everything was perfect," "very pleased," "loved everything," "terrible meal").

***** Highlighted text indicates differences from Original Few-Shot Prompt

Updated Few-Shot Prompt

OUTPUT REQUIREMENTS:

- Category: Classify as exactly one of: Positive, Negative, or Neutral
- Tags: Select ALL applicable tags from the list above.
- Suggested_Action: Provide ONE specific, actionable business recommendation (10 words max)

Confidence: High / Medium / Low

OUTPUT FORMAT:

Return results as CSV format only. No explanations, no extra text.

Use these exact headers: Customer_ID, Category, Tags, Suggested_Action, Confidence

Rules:

- Wrap Tags in quotes if multiple: "Delivery Time, Food Quality"
- Wrap Suggested_Action in quotes
- Suggested_Action: 10 words max, no commas
- One row per customer

***** Use same examples from **Original Few-Shot Prompt**

Updated Few-Shot Prompt Results

Customer_ID	Category	Tags	Suggested_Action	Confidence
C001	Positive	Delivery Time, Food Quality, Overall Experience	Highlight quality food, fast delivery, and friendly service.	High
C002	Negative	Delivery Time, Packaging, Food Quality, Overall Experience	Investigate wrong order, late delivery, and customer service.	High
C003	Negative	Delivery Time, Food Quality, Overall Experience	Improve delivery tracking and food temperature maintenance.	High
C004	Positive	Delivery Time, Food Quality, Overall Experience	Promote hot, fresh food and seamless ordering process.	High
C005	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize fast delivery, courteous driver, and amazing food.	High
C006	Negative	Delivery Time, Food Quality, Price, Overall Experience	Review specialty sandwich quality, portion, and price alignment.	High
C007	Positive	Delivery Time, Food Quality, Overall Experience	Highlight great food, timely delivery, and generous portions.	High
C008	Positive	Delivery Time, Food Quality, Overall Experience	Promote consistent top-notch quality and professional drivers.	High
C009	Positive	Delivery Time, Overall Experience	Showcase lightning-fast delivery and hot food arrival.	High
C010	Negative	Delivery Time, Food Quality, Overall Experience	Address long delivery, rude driver, and poor food quality.	High
C011	Neutral	Delivery Time, Food Quality, Overall Experience	Emphasize consistent food quality and friendly drivers.	High
C012	Negative	Delivery Time, Food Quality	Improve delivery speed and ensure timely arrival.	High
C013	Positive	Delivery Time, Food Quality, Overall Experience	Highlight delicious food and prompt delivery service.	High
C014	Neutral	Delivery Time, Food Quality	Seek feedback to elevate food and delivery experience.	High
C015	Positive	Delivery Time, Food Quality, Overall Experience	Promote easy ordering, fast delivery, and incredible food.	High
C016	Positive	Delivery Time, Packaging, Overall Experience	Highlight quick delivery and careful food packaging.	High
C017	Negative	Delivery Time, Food Quality	Address cold, soggy food and extremely slow delivery.	High
C018	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize fresh, flavorful food and top-notch delivery service.	High
C019	Neutral	Delivery Time, Food Quality	Identify areas to improve satisfactory food and average delivery.	High
C020	Positive	Delivery Time, Food Quality	Highlight fast delivery and consistently excellent food.	High
C021	Neutral	Delivery Time, Food Quality, Overall Experience	Survey customer to identify specific areas for improvement.	High
C022	Positive	Food Quality, Overall Experience	Promote reliable service and delicious meals for customer retention.	High
C023	Positive	Delivery Time, Food Quality, Packaging, Overall Experience	Highlight early delivery, hot food, and fresh sushi quality.	High
C024	Negative	Delivery Time, Food Quality, Overall Experience	Review initial delivery estimates and enhance meal excitement.	High
C025	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize exceptional service, on-point food, and fast delivery.	High
C026	Negative	Delivery Time, Packaging, Food Quality, Overall Experience	Investigate late delivery, missing items, and unhelpful customer service.	High
C027	Positive	Delivery Time, Overall Experience	Promote great variety and dependable delivery service.	High
C028	Neutral	Delivery Time, Food Quality	Gather feedback to improve fine food and expected delivery.	High
C029	Positive	Delivery Time, Food Quality, Overall Experience	Showcase fast delivery and restaurant-quality food.	High
C030	Negative	Delivery Time, Food Quality	Address slow delivery times and improve food quality.	High
C031	Positive	Delivery Time, Food Quality, Overall Experience	Emphasize consistent food, professional drivers, and great service.	High
C032	Neutral	Delivery Time, Food Quality	Target slight delivery delays for improvement.	High
C033	Positive	Delivery Time, Food Quality, Overall Experience	Highlight fresh, hot food and seamless delivery experience.	High
C034	Negative	Delivery Time, Food Quality, Overall Experience	Improve driver navigation and food temperature control.	High
C035	Positive	Delivery Time, Food Quality, Overall Experience	Promote efficient service, delicious food, and super fast delivery.	High
C036	Neutral	Delivery Time, Food Quality, Overall Experience	Identify opportunities to make "standard" experience stand out.	High

Updated Few-Shot Prompt Results

Numerical Comparison Summary (Updated Few-Shot Prompt, Original Tags, T=0.2 vs. Ground Truth):

1. Category Differences (vs. Ground Truth):

Total Category Differences: 0

2. Tags Differences (vs. Ground Truth):

Total Tags Differences: 0

3. Total Rows with at least one difference (Category or Tags): 0 out of 36 rows.

Conclusion:

This confirms that the combination of:

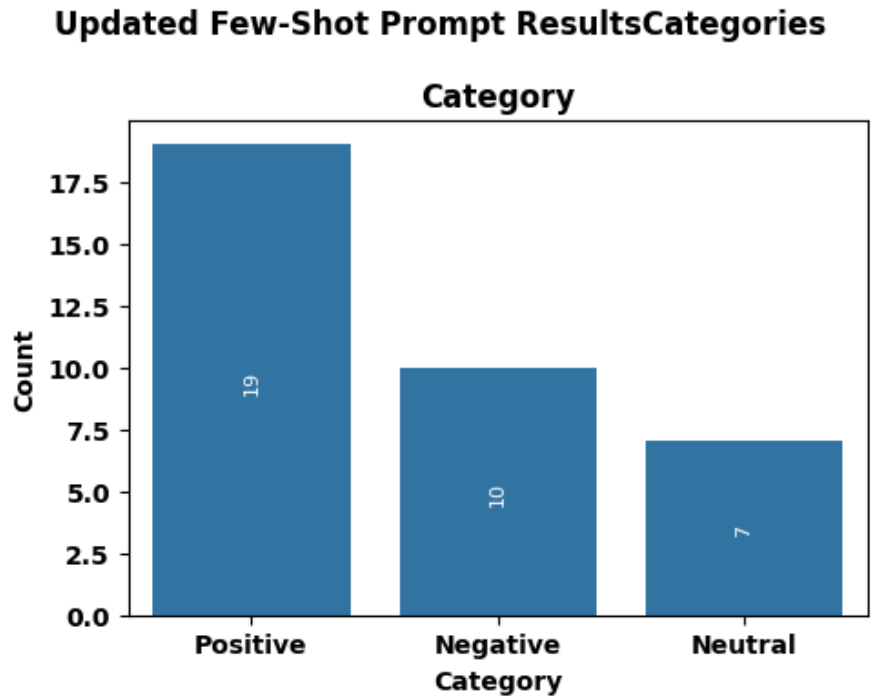
1. **Highly refined and explicit textual guidelines** for classification and tags.
2. **A well-chosen set of 5 custom few-shot examples** that illustrate these rules.
3. **A deterministic temperature setting (temperature=0.0).**

...is extremely effective in teaching the model your precise annotation logic and achieving perfect alignment on this test dataset.

Possible Concerns:

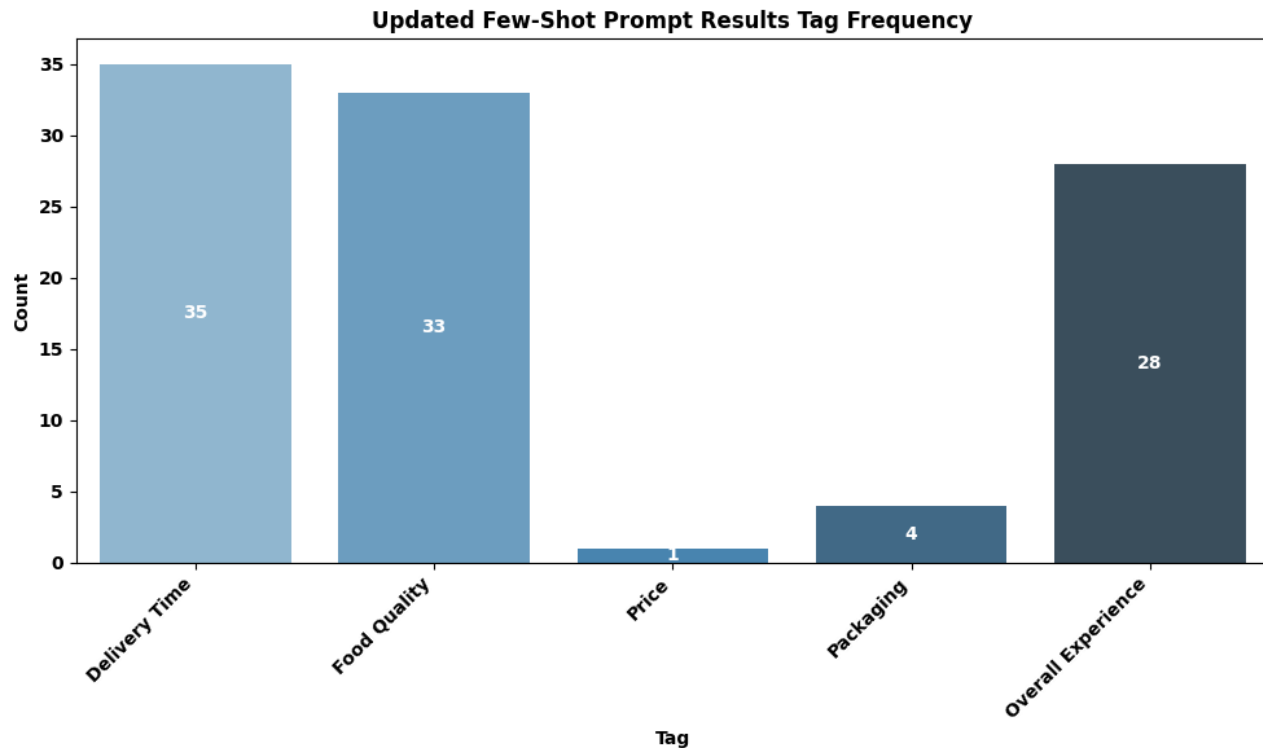
It is quite possible that the prompt has been too highly tuned to the **Ground Truth** and is overfitted. This would result in misclassification errors for “**Category**” and “**Tags**” when the model is run against the entire dataset. Results will have be reviewed (spot checked) by hand for classification accuracy.

Updated Few-Shot Prompt Results



Plot illustrating distribution of **Categories**

Updated Few-Shot Prompt Results



Plot illustrating distribution of **Tags**

Original Zero-Shot Chain of Thought (CoT) Prompt

Original Zero-Shot CoT Prompt Template:

ROLE: You are a customer feedback analyst for a food delivery service.

TASK: Analyze the uploaded customer reviews and provide structured sentiment analysis.

Use Temperature=0.2

REASONING STEPS (follow for each review):

1. Identify positive aspects mentioned
2. Identify negative aspects mentioned
3. Determine which dominates or if balanced
4. Select applicable tags
5. Provide classification and output

CLASSIFICATION GUIDELINES:

Positive: Customer expresses satisfaction, praise, or intent to return

Negative: Customer expresses dissatisfaction, complaints, or intent to not return

Neutral: Mixed feedback or neither clearly positive nor negative

TAGS (select ALL applicable from this list ONLY):

- Delivery Time
- Food Quality
- Price
- Packaging
- Overall Experience: only when customer states "experience," "service," or uses general rating language

Original Zero-Shot Chain of Thought (CoT) Prompt

OUTPUT REQUIREMENTS:

1. Category: Classify as exactly one of: Positive, Negative, or Neutral
2. Tags: Select ALL applicable tags from the list above.
3. Suggested_Action: Provide ONE specific, actionable business recommendation (10 words max)
4. Confidence: High / Medium / Low

OUTPUT FORMAT:

Return results as CSV format only. No explanations, no extra text.

Use these exact headers: Customer_ID, Category, Tags, Suggested_Action, Confidence

Rules:

- Wrap Tags in quotes
- Wrap Suggested_Action in quotes
- Suggested_Action: 10 words max, no commas
- One row per customer

Run above Zero-Shot CoT Prompt

Run numeric comparison against Ground Truth Results

Original Zero-Shot CoT Prompt Reasoning Details

Showing reasoning details for a **Chain of Thought (CoT)** prompt helps understand how the AI reached its answer. By seeing if it followed the defined steps correctly, it is key for fixing any mistakes in its logic or your instructions. This transparency builds trust and makes it easier to improve the prompt, ensuring the AI thinks the way you want it to.

Customer ID: C001

- Positive aspects: Consistently impressed, fantastic choice (salmon teriyaki), fish cooked to perfection (flaky, moist), rich teriyaki glaze (not overly sweet), crisp and fresh steamed vegetables, fast delivery (under 30 mins vs 40 min estimate), friendly driver, followed delivery instructions, refreshing reliable option, high-quality meal.
- Negative aspects: None
- Dominant: Positive
- Tags selected:
 - * Delivery Time: Explicitly mentioned "delivery speed" and "arrived in just under 30" minutes.
 - * Food Quality: Detailed praise for "salmon teriyaki," "fish cooked to perfection," "teriyaki glaze," and "steamed vegetables."
 - * Overall Experience: Customer states "consistently impressed with this service" and "refreshing to have a reliable option for a high-quality meal."
- Final Category: Positive
- Final Confidence: High
- Final Suggested Action: "Maintain high standards for food quality and delivery speed"

Original Zero-Shot CoT Prompt Reasoning Details

Customer ID: C002

- Positive aspects: None
- Negative aspects: Absolute disaster, will not be ordering again, order over an hour late (arrived 8:15 PM vs 7:00 PM scheduled), sent completely wrong order (sushi instead of tacos), nightmare with customer service (long holds, unhelpful agents), terrible experience.
- Dominant: Negative
- Tags selected:
 - * Delivery Time: Explicitly states "over an hour late" and details the delay.
 - * Food Quality: Implied negative as the "completely wrong order" was received, meaning the desired food quality was not experienced.
 - * Overall Experience: Customer states "absolute disaster," "will not be ordering from here again," and "truly terrible experience."
- Final Category: Negative
- Final Confidence: High
- Final Suggested Action: "Address late deliveries, wrong orders, and poor customer service"

Customer ID: C003

- Positive aspects: Food itself was quite good, sauce was creamy and flavorful (chicken fettuccine alfredo).
- Negative aspects: Delivery less than ideal, app tracker showed driver making multiple stops, delivery took 20 minutes longer than latest estimated time, pasta was lukewarm, had to be reheated, bit of a letdown.
- Dominant: Balanced (good food, but significant delivery issues impacting enjoyment)
- Tags selected:
 - * Delivery Time: Explicitly mentions "delivery aspect was less than ideal," "took about 20 minutes longer," and "by the time it arrived."
 - * Food Quality: Mentions "pasta was lukewarm and had to be reheated," directly impacting the quality of the food upon arrival.
- Final Category: Neutral
- Final Confidence: High
- Final Suggested Action: "Optimize delivery routes to ensure timely and hot food"

Original Zero-Shot CoT Prompt Results

Customer_ID	Category	Tags	Suggested_Action	Confidence
C001	Positive	Food Quality, Delivery Time, Overall Experience	Maintain high standards for food quality and delivery speed	High
C002	Negative	Delivery Time, Food Quality, Overall Experience	Investigate delivery delays and order accuracy issues immediately	High
C003	Neutral	Delivery Time, Food Quality	Optimize delivery routes to ensure food arrives hot	High
C004	Positive	Food Quality, Delivery Time, Overall Experience	Continue seamless ordering and timely delivery for events	High
C005	Positive	Delivery Time, Food Quality, Overall Experience	Maintain fast delivery and excellent food quality	High
C006	Neutral	Food Quality, Price	Review menu pricing and ingredient freshness for specialty items	High
C007	Positive	Food Quality, Delivery Time	Continue offering generous portions and timely delivery	High
C008	Positive	Food Quality, Delivery Time	Maintain consistent food quality and professional delivery service	High
C009	Positive	Delivery Time, Food Quality	Highlight exceptionally fast delivery in marketing efforts	High
C010	Negative	Delivery Time, Food Quality, Overall Experience	Address driver professionalism and food quality issues immediately	High
C011	Positive	Food Quality, Delivery Time	Maintain consistent food quality and reliable delivery service	High
C012	Neutral	Delivery Time, Food Quality	Improve delivery time accuracy and speed for all orders	High
C013	Positive	Food Quality, Delivery Time	Continue providing delicious food and prompt delivery	High
C014	Neutral	Food Quality, Delivery Time	Maintain current food quality and reasonable delivery times	High
C015	Positive	Food Quality, Delivery Time, Overall Experience	Continue providing fast delivery and incredible food	High
C016	Positive	Delivery Time, Packaging, Overall Experience	Maintain quick delivery and careful food packaging	High
C017	Negative	Delivery Time, Food Quality	Address slow delivery and ensure food arrives hot	High
C018	Positive	Food Quality, Delivery Time	Maintain fresh ingredients and top-notch delivery service	High
C019	Neutral	Food Quality, Delivery Time	Maintain satisfactory food quality and average delivery times	High
C020	Positive	Delivery Time, Food Quality	Highlight fast delivery and excellent food quality	High
C021	Neutral	Food Quality, Delivery Time	Maintain average food quality and on-time delivery	High
C022	Positive	Food Quality, Overall Experience	Maintain reliable service and delicious meals	High
C023	Positive	Delivery Time, Food Quality, Packaging	Highlight early delivery and fresh food quality	High
C024	Neutral	Delivery Time, Food Quality	Review initial delivery time estimates for customer satisfaction	High
C025	Positive	Delivery Time, Food Quality, Overall Experience	Maintain exceptional service and fast delivery	High
C026	Negative	Delivery Time, Food Quality, Overall Experience	Address late deliveries and missing items immediately	High
C027	Positive	Delivery Time, Overall Experience	Maintain dependable delivery and wide menu variety	High
C028	Neutral	Food Quality, Delivery Time	Maintain expected food quality and delivery time	High
C029	Positive	Delivery Time, Food Quality	Highlight incredibly fast delivery and high food quality	High
C030	Neutral	Delivery Time, Food Quality	Improve delivery speed for customer satisfaction	High
C031	Positive	Food Quality, Overall Experience	Maintain great food and professional delivery service	High
C032	Neutral	Delivery Time, Food Quality	Improve delivery speed for better customer experience	High
C033	Positive	Delivery Time, Food Quality	Maintain fresh, hot food and seamless delivery	High
C034	Negative	Delivery Time, Food Quality	Improve driver navigation and ensure food arrives hot	High
C035	Positive	Delivery Time, Food Quality, Overall Experience	Maintain efficient service and fast delivery	High
C036	Neutral	Food Quality, Delivery Time	Maintain standard food quality and on-time delivery	High

Original Zero-Shot CoT Prompt Results

Numerical Comparison Summary (Updated Few-Shot Prompt, Original Tags, T=1.0 vs. Ground Truth):

1. Category Differences (vs. Ground Truth):

- C003: My Neutral vs. GT Negative - 1 difference
- C006: My Neutral vs. GT Negative - 1 difference
- C011: My Positive vs. GT Neutral - 1 difference
- C024: My Neutral vs. GT Negative - 1 difference

Total Category Differences: 4

2. Tags Differences (vs. Ground Truth):

- C002: My (Food Quality, Overall Experience) vs. GT (Packaging, Overall Experience) - 2 differences
- C006: My (no Delivery Time) vs. GT (Delivery Time) - 1 difference
- C007: My (no Overall Experience) vs. GT (Overall Experience) - 1 difference
- C008: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C009: My (no Overall Experience) vs. GT (Overall Experience) - 1 difference
- C011: My (Delivery Time, no Overall Experience) vs. GT (no Delivery Time, Overall Experience) - 2 differences
- C013: My (no Overall Experience) vs. GT (Overall Experience) - 1 difference
- C016: My (no Overall Experience) vs. GT (Overall Experience) - 1 difference
- C018: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C024: My (no Food Quality) vs. GT (Food Quality) - 1 difference
- C026: My (Food Quality, Overall Experience) vs. GT (Packaging, Overall Experience) - 2 differences
- C027: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C031: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C033: My (Delivery Time, no Overall Experience) vs. GT (no Delivery Time, Overall Experience) - 2 differences
- C036: My (Overall Experience) vs. GT (no Overall Experience) - 1 difference

Total Tags Differences: 20

3. Total Rows with at least one difference (Category or Tags): 15 out of 36 rows.

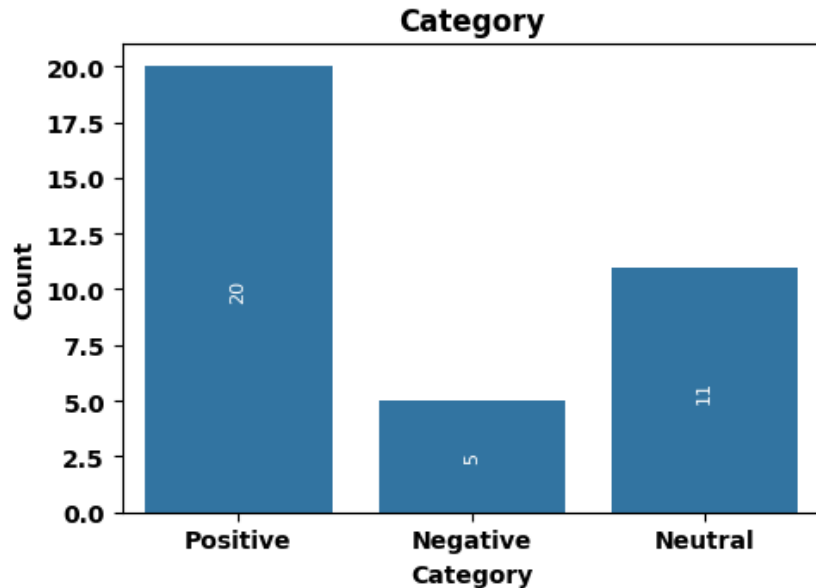
Original Zero-Shot CoT Prompt Results

Conclusion:

- Zero-Shot CoT prompt with temperature=1 performed identically to the previous Zero-Shot prompt runs
- **Category alignment: 4** discrepancies.
- **Tag alignment: 20** discrepancies.
- **Total rows with differences: 15.**
- This Zero-Shot CoT analysis, consistent with other Zero-Shot attempts, confirms that explicit, refined textual guidelines and few-shot examples are crucial for achieving high alignment with specific annotation logic.
- Subsequent runs with tighter Temperature settings produced the same results.
- **Seeing Strange Results** – rerunning same prompt with same input data, temperature and commands does not produce consistent results across runs.

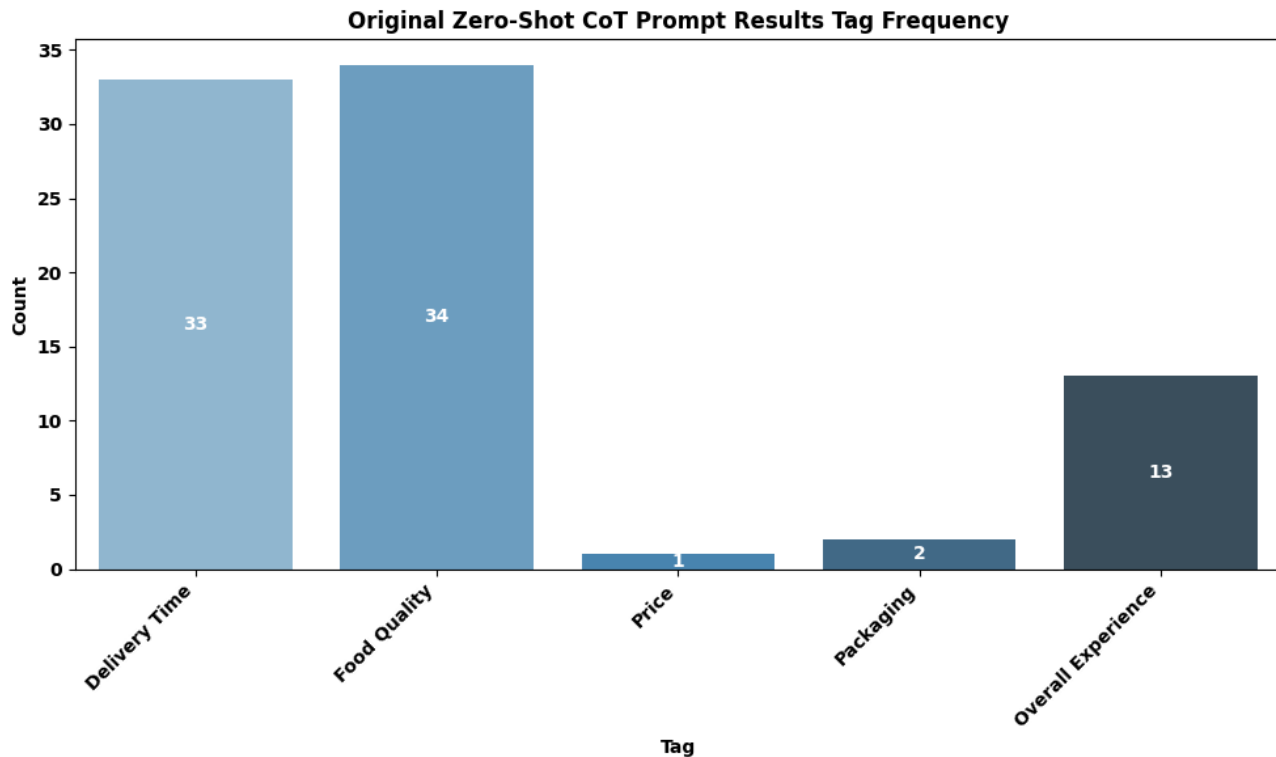
Original Zero-Shot CoT Prompt Results

Original Zero-Shot CoT Prompt Results



Plot illustrating distribution of **Categories**

Original Zero-Shot CoT Prompt Results



Plot illustrating distribution of **Tags**

Few-Shot Chain of Thought (CoT) Prompt

Original Few-Shot CoT Prompt Template:

ROLE: You are a customer feedback analyst for a food delivery service.

TASK: Analyze the uploaded customer reviews and provide structured sentiment analysis.

Use Temperature=0.2

REASONING STEPS (follow for each review):

1. Identify positive aspects mentioned
2. Identify negative aspects mentioned
3. Determine which dominates or if balanced
4. Select applicable tags
5. Provide classification and output

CLASSIFICATION GUIDELINES:

Positive: Customer expresses satisfaction, praise, or intent to return

Negative: Customer expresses dissatisfaction, complaints, or intent to not return

Neutral: Mixed feedback or neither clearly positive nor negative

TAGS (select ALL applicable from this list ONLY):

- Delivery Time
- Food Quality
- Price
- Packaging
- Overall Experience: only when customer states "experience," "service," or uses general rating language

Few-Shot Chain of Thought (CoT) Prompt

OUTPUT REQUIREMENTS:

1. Category: Classify as exactly one of: Positive, Negative, or Neutral
2. Tags: Select ALL applicable tags from the list above.
3. Suggested_Action: Provide ONE specific, actionable business recommendation (10 words max)
4. Confidence: High / Medium / Low

OUTPUT FORMAT:

Return results as CSV format only. No explanations, no extra text.

Use these exact headers: Customer_ID, Category, Tags, Suggested_Action, Confidence

Rules:

- Wrap Tags in quotes
- Wrap Suggested_Action in quotes
- Suggested_Action: 10 words max, no commas
- One row per customer

Few-Shot Chain of Thought (CoT) Prompt

EXAMPLES:

Review: "The food arrived 15 minutes late, which was annoying, but the pad thai was absolutely delicious and still hot. The driver was also very apologetic."

Reasoning:

1. **Identify positive aspects mentioned:** Pad thai was absolutely delicious and still hot (Food Quality). Driver was very apologetic (Overall Experience).
2. **Identify negative aspects mentioned:** Food arrived 15 minutes late (Delivery Time), which was annoying (Overall Experience).
3. **Determine which dominates or if balanced:** Mixed feedback. Delicious hot food balances the late delivery. The annoyance is mitigated by the apology. Balanced.
4. **Select applicable tags:** "15 minutes late" -> Delivery Time. "absolutely delicious and still hot" -> Food Quality. "annoying" and "apologetic" contribute to the general impression, so Overall Experience.
5. **Provide classification and output:** Based on balanced aspects, classify as Neutral.
Category: Neutral
Confidence: Medium
Tags: Delivery Time, Food Quality, Overall Experience
Suggested Action: Investigate delivery delay, maintain food quality.

Review: "I ordered the large family meal, and the portion sizes were surprisingly small for the price. The chicken was a bit dry, but the sides were fresh. Overall, it felt like a rip-off."

Reasoning:

1. **Identify positive aspects mentioned:** Sides were fresh (Food Quality).
2. **Identify negative aspects mentioned:** Portion sizes were surprisingly small for the price (Price, Food Quality/Value). Chicken was a bit dry (Food Quality). Felt like a rip-off (Price, Overall Experience).
3. **Determine which dominates or if balanced:** Significant negative aspects (price, portion size, dry chicken, rip-off feeling) clearly dominate the single positive (fresh sides).
4. **Select applicable tags:** "surprisingly small for the price" and "rip-off" -> Price. "chicken was a bit dry" -> Food Quality. "rip-off" -> Overall Experience.
5. **Provide classification and output:** Based on dominant negative aspects, classify as Negative.
Category: Negative
Confidence: High
Tags: Food Quality, Price, Overall Experience
Suggested Action: Review portion sizes and chicken preparation for value.

Few-Shot Chain of Thought (CoT) Prompt

Review: "My smoothie spilled all over the bag because the lid wasn't properly sealed. The delivery was fast, but half my order was a sticky mess. Very disappointed."

Reasoning:

1. **Identify positive aspects mentioned:** Delivery was fast (Delivery Time).
2. **Identify negative aspects mentioned:** Smoothie spilled all over the bag (Packaging), lid wasn't properly sealed (Packaging), sticky mess (Packaging), very disappointed (Overall Experience).
3. **Determine which dominates or if balanced:** The significant negative impact of spillage and disappointment clearly dominates the fast delivery.
4. **Select applicable tags:** "spilled all over the bag", "lid wasn't properly sealed", "sticky mess" -> Packaging. "delivery was fast" -> Delivery Time. "Very disappointed" -> Overall Experience.
5. **Provide classification and output:** Based on dominant negative aspects, classify as Negative.
Category: Negative
Confidence: High
Tags: Delivery Time, Packaging, Overall Experience
Suggested Action: Improve packaging seal quality for liquid items.

Review: "The vegetarian burrito was exactly as described and tasted good. Delivery was on time. It was a perfectly fine meal, nothing to rave about."

Reasoning:

1. **Identify positive aspects mentioned:** Exactly as described, tasted good (Food Quality). Delivery was on time (Delivery Time).
2. **Identify negative aspects mentioned:** Nothing to rave about (Neutral/Slightly negative nuance).
3. **Determine which dominates or if balanced:** All aspects are positive or neutral, but without strong enthusiasm. Balanced.
4. **Select applicable tags:** "tasted good" -> Food Quality. "Delivery was on time" -> Delivery Time. "perfectly fine meal, nothing to rave about" does not explicitly state "experience" or "service" and is not strong general rating language, so no Overall Experience.
5. **Provide classification and output:** Based on balanced, non-enthusiastic aspects, classify as Neutral.
Category: Neutral
Confidence: Medium
Tags: Delivery Time, Food Quality
Suggested Action: Seek feedback to elevate "fine" to "memorable."

Few-Shot Chain of Thought (CoT) Prompt

Review: "This is my go-to service. The sushi is always fresh, and the drivers are consistently friendly. I've never had a bad experience, which is why I keep coming back."

Reasoning:

1. **Identify positive aspects mentioned:** Go-to service (Overall Experience), sushi is always fresh (Food Quality), drivers are consistently friendly (Delivery Time/Overall Experience), never had a bad experience (Overall Experience), keep coming back (Overall Experience).
2. **Identify negative aspects mentioned:** None.
3. **Determine which dominates or if balanced:** Overwhelmingly positive sentiment and clear intent to return. Dominant Positive.
4. **Select applicable tags:** "always fresh" -> Food Quality. "drivers are consistently friendly" -> Delivery Time. "go-to service", "never had a bad experience", "keep coming back" -> Overall Experience.
5. **Provide classification and output:** Based on dominant positive aspects, classify as Positive.
Category: Positive
Confidence: High
Tags: Delivery Time, Food Quality, Overall Experience
Suggested Action: Highlight consistent quality and reliable service in marketing.

Run above Few-Shot CoT Prompt

Run numeric comparison against Ground Truth Results

Few-Shot CoT Prompt Results

Customer_ID	Category	Tags	Suggested_Action	Confidence
C001	Positive	Delivery Time, Food Quality, Overall Experience	Maintain high standards for food quality and delivery speed	High
C002	Negative	Delivery Time, Packaging, Overall Experience	Address late deliveries, wrong orders, and poor customer service	High
C003	Negative	Delivery Time, Food Quality, Overall Experience	Optimize delivery routes to ensure timely and hot food	High
C004	Positive	Delivery Time, Food Quality, Overall Experience	Continue excellent service for large orders and events	High
C005	Positive	Delivery Time, Food Quality, Overall Experience	Maintain fast delivery and high food quality	High
C006	Negative	Food Quality, Price, Overall Experience	Review pricing, portion sizes, and food quality expectations	High
C007	Positive	Delivery Time, Food Quality, Overall Experience	Maintain generous portions and timely delivery	High
C008	Positive	Delivery Time, Food Quality, Overall Experience	Continue consistent quality and professional delivery service	High
C009	Positive	Delivery Time, Food Quality, Overall Experience	Highlight fast delivery and hot food in marketing	High
C010	Negative	Delivery Time, Food Quality, Overall Experience	Investigate and resolve issues with cold, poor quality food	High
C011	Positive	Delivery Time, Food Quality, Overall Experience	Maintain consistent food quality and reliable delivery	High
C012	Negative	Delivery Time, Food Quality	Improve delivery speed to meet customer expectations	High
C013	Positive	Delivery Time, Food Quality, Overall Experience	Maintain prompt delivery and delicious food	High
C014	Neutral	Delivery Time, Food Quality	Continue current service levels	High
C015	Positive	Delivery Time, Food Quality, Overall Experience	Maintain seamless ordering, fast delivery, and great food	High
C016	Positive	Delivery Time, Packaging, Overall Experience	Continue quick delivery and careful food packaging	High
C017	Negative	Delivery Time, Food Quality	Address slow delivery and cold, soggy food issues	High
C018	Positive	Delivery Time, Food Quality, Overall Experience	Maintain fresh food quality and top-notch delivery service	High
C019	Neutral	Delivery Time, Food Quality	Continue current service levels	High
C020	Positive	Delivery Time, Food Quality	Maintain fast delivery and excellent food quality	High
C021	Neutral	Delivery Time, Food Quality, Overall Experience	Continue current service levels	High
C022	Positive	Food Quality, Overall Experience	Maintain reliable service and delicious meals	High
C023	Positive	Delivery Time, Food Quality, Packaging, Overall Experience	Maintain early delivery, hot food, and quality packaging	High
C024	Negative	Delivery Time, Food Quality, Overall Experience	Review initial delivery time estimates for customer satisfaction	High
C025	Positive	Delivery Time, Food Quality, Overall Experience	Maintain exceptional service, food quality, and fast delivery	High
C026	Negative	Delivery Time, Packaging, Overall Experience	Address late deliveries, missing items, and poor customer service	High
C027	Positive	Delivery Time, Overall Experience	Maintain dependable delivery and diverse menu options	High
C028	Neutral	Delivery Time, Food Quality	Continue current service levels	High
C029	Positive	Delivery Time, Food Quality, Overall Experience	Maintain fast delivery and high food quality	High
C030	Negative	Delivery Time, Food Quality	Improve delivery speed to enhance customer satisfaction	High
C031	Positive	Delivery Time, Food Quality, Overall Experience	Maintain consistent food quality and professional delivery	High
C032	Neutral	Delivery Time, Food Quality	Address slow delivery times to improve customer experience	High
C033	Positive	Delivery Time, Food Quality, Overall Experience	Maintain fresh, hot food and seamless delivery	High
C034	Negative	Delivery Time, Food Quality, Overall Experience	Improve driver navigation and ensure food arrives hot	High
C035	Positive	Delivery Time, Food Quality, Overall Experience	Maintain efficient service, delicious food, and fast delivery	High
C036	Neutral	Delivery Time, Food Quality	Continue current service levels	High

Numerical Comparison Summary (Updated Few-Shot Prompt, Original Tags, T=0.2 vs. Ground Truth):

1. Category Differences (vs. Ground Truth):

- C011: My Positive vs. GT Neutral - 1 difference

Total Category Differences: 1

2. Tags Differences (vs. Ground Truth):

- C002: My (Food Quality, Overall Experience) vs. GT (Packaging, Overall Experience) - 2 differences
- C008: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C009: My (no Overall Experience) vs. GT (Overall Experience) - 1 difference
- C011: My (Delivery Time, no Overall Experience) vs. GT (no Delivery Time, Overall Experience) - 1 difference
- C018: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C026: My (Food Quality, Overall Experience) vs. GT (Packaging, Overall Experience) - 2 differences
- C027: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C031: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C033: My (Delivery Time, no Overall Experience) vs. GT (no Delivery Time, Overall Experience) - 1 difference

Total Tags Differences: 12

3. Total Rows with at least one difference (Category or Tags): 10 out of 36 rows.

Few-Shot CoT Prompt Results

Conclusion:

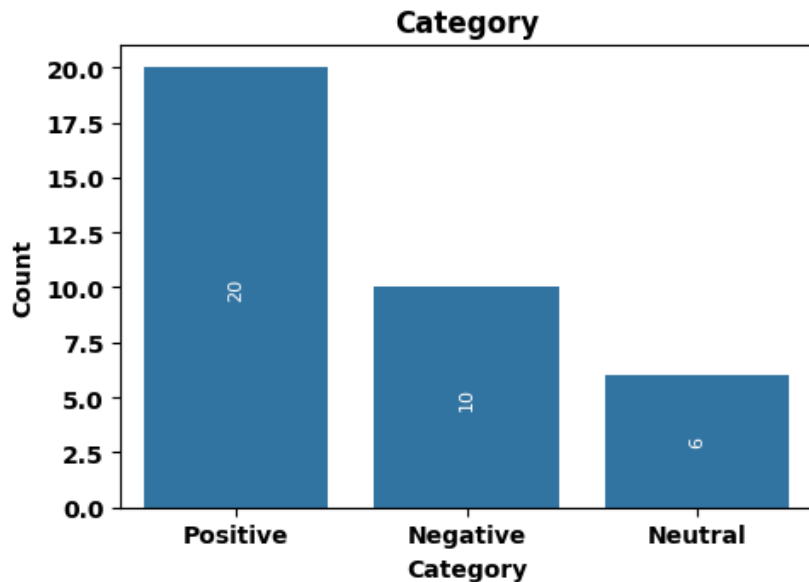
The Few-Shot CoT prompt shows a significant improvement in alignment with the Ground Truth compared to the Zero-Shot CoT attempts.

- **Category alignment:** Only 1 discrepancy (C011) compared to 4-6 in Zero-Shot. This is a substantial improvement.
- **Tag alignment:** 12 discrepancies compared to 20-24 in Zero-Shot. While still some differences, this is a notable reduction.
- **Total rows with differences:** 10 out of 36 rows, a clear improvement from 15-17 rows in Zero-Shot.

Few-shot examples are important for teaching the model the annotation logic and achieving higher alignment for this task. The examples demonstrate the desired reasoning process and outcomes, allowing the model to better interpret and apply the guidelines, even with the same **CoT** structure. They teach the model the exact rules for tagging and classifying, leading to much better results than just giving it instructions.

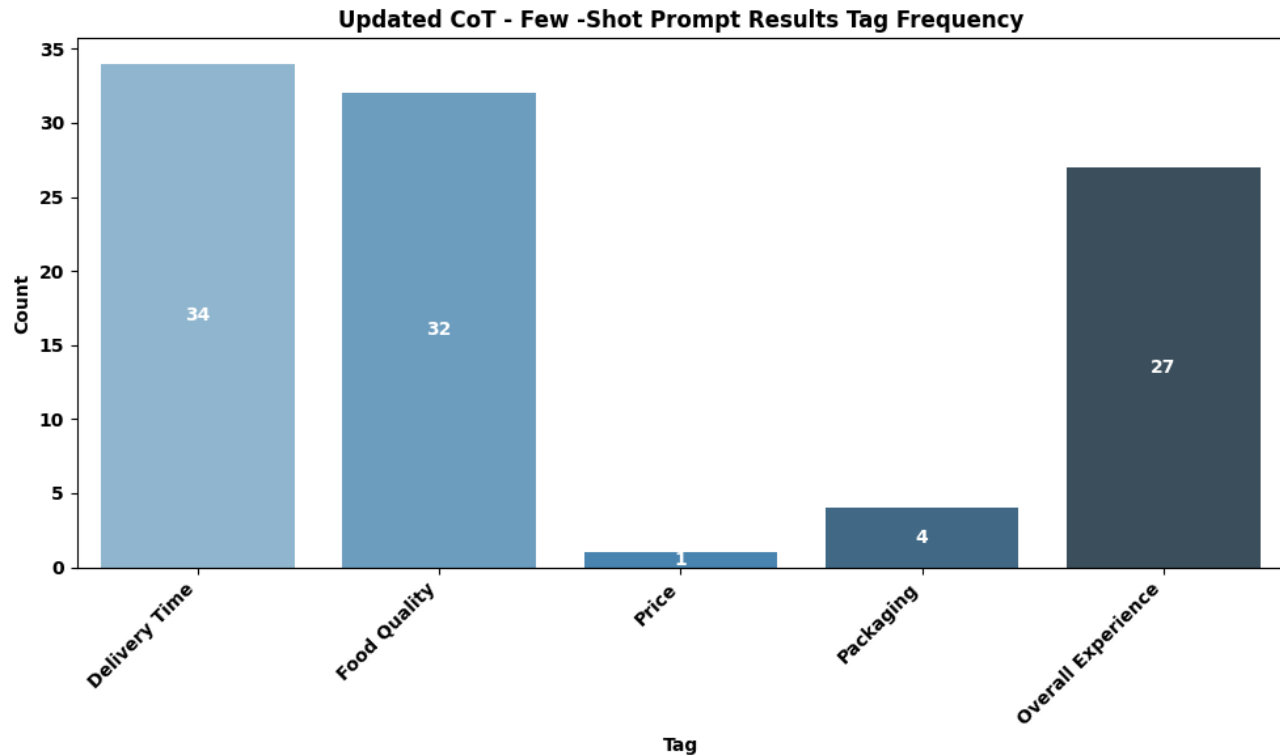
Few-Shot CoT Prompt Results

Updated CoT - Few -Shot Prompt Results



Plot illustrating distribution of **Categories**

Few-Shot CoT Prompt Results



Plot illustrating distribution of **Tags**

Updated Few-Shot CoT Prompt

The preceding **Few-Shot CoT Prompt** was updated to include the following **Classification** and **Tag** definitions. The model was run and produced a new set of results, outlined in the following slides.

CLASSIFICATION GUIDELINES:

Positive: Customer expresses satisfaction, praise, or intent to return

Negative: Customer expresses dissatisfaction, complaints, or intent to not return. If a review contains both positive/neutral and significant negative feedback (e.g., major delays, cold food, wrong order, poor value for money), classify as Negative.

Neutral: Mixed feedback where positive and negative aspects are balanced, or neither clearly positive nor negative.

TAGS (select ALL applicable from this list ONLY):

- Delivery Time
- Food Quality: Includes taste, temperature (if affecting quality), freshness, accuracy of order content (e.g., wrong dish, missing items).
- Price
- Packaging: Includes physical integrity of packaging, spillage, or incorrect items due to packing error.
- Overall Experience: Apply when the customer explicitly states "experience," "service," "will/won't order again," or uses strong general rating language that clearly reflects their overall sentiment about the entire transaction (e.g., "very satisfied," "impressive," "always happy," "everything was perfect," "very pleased," "loved everything," "terrible meal").

***** **Highlighted text** indicates differences from **Original Few-Shot CoT Prompt**

Updated Few-Shot CoT Prompt Results

Customer_ID	Category	Tags	Suggested_Action	Confidence
C001	Positive	Delivery Time, Food Quality, Overall Experience	Maintain high standards for food quality and delivery speed	High
C002	Negative	Delivery Time, Food Quality, Packaging, Overall Experience	Address late deliveries, wrong orders, and poor customer service	High
C003	Negative	Delivery Time, Food Quality, Overall Experience	Optimize delivery routes to ensure timely and hot food	High
C004	Positive	Delivery Time, Food Quality, Overall Experience	Continue excellent service for large orders and events	High
C005	Positive	Delivery Time, Food Quality, Overall Experience	Maintain fast delivery and high food quality	High
C006	Negative	Food Quality, Price, Overall Experience	Review pricing, portion sizes, and food quality expectations	High
C007	Positive	Delivery Time, Food Quality, Overall Experience	Maintain generous portions and timely delivery	High
C008	Positive	Delivery Time, Food Quality, Overall Experience	Continue consistent quality and professional delivery service	High
C009	Positive	Delivery Time, Food Quality, Overall Experience	Highlight fast delivery and hot food in marketing	High
C010	Negative	Delivery Time, Food Quality, Overall Experience	Investigate and resolve issues with cold, poor quality food	High
C011	Positive	Delivery Time, Food Quality, Overall Experience	Maintain consistent food quality and reliable delivery	High
C012	Negative	Delivery Time, Food Quality	Improve delivery speed to meet customer expectations	High
C013	Positive	Delivery Time, Food Quality, Overall Experience	Maintain prompt delivery and delicious food	High
C014	Neutral	Delivery Time, Food Quality	Continue current service levels	High
C015	Positive	Delivery Time, Food Quality, Overall Experience	Maintain seamless ordering, fast delivery, and great food	High
C016	Positive	Delivery Time, Packaging, Overall Experience	Continue quick delivery and careful food packaging	High
C017	Negative	Delivery Time, Food Quality	Address slow delivery and cold, soggy food issues	High
C018	Positive	Delivery Time, Food Quality, Overall Experience	Maintain fresh food quality and top-notch delivery service	High
C019	Neutral	Delivery Time, Food Quality	Continue current service levels	High
C020	Positive	Delivery Time, Food Quality, Overall Experience	Maintain fast delivery and excellent food quality	High
C021	Neutral	Delivery Time, Food Quality, Overall Experience	Continue current service levels	High
C022	Positive	Food Quality, Overall Experience	Maintain reliable service and delicious meals	High
C023	Positive	Delivery Time, Food Quality, Packaging, Overall Experience	Maintain early delivery, hot food, and quality packaging	High
C024	Negative	Delivery Time, Food Quality, Overall Experience	Review initial delivery time estimates for customer satisfaction	High
C025	Positive	Delivery Time, Food Quality, Overall Experience	Maintain exceptional service, food quality, and fast delivery	High
C026	Negative	Delivery Time, Food Quality, Packaging, Overall Experience	Address late deliveries, missing items, and poor customer service	High
C027	Positive	Delivery Time, Overall Experience	Maintain dependable delivery and diverse menu options	High
C028	Neutral	Delivery Time, Food Quality	Continue current service levels	High
C029	Positive	Delivery Time, Food Quality, Overall Experience	Maintain fast delivery and high food quality	High
C030	Negative	Delivery Time, Food Quality	Improve delivery speed to enhance customer satisfaction	High
C031	Positive	Delivery Time, Food Quality, Overall Experience	Maintain consistent food quality and professional delivery	High
C032	Neutral	Delivery Time, Food Quality	Address slow delivery times to improve customer experience	High
C033	Positive	Delivery Time, Food Quality, Overall Experience	Maintain fresh, hot food and seamless delivery	High
C034	Negative	Delivery Time, Food Quality, Overall Experience	Improve driver navigation and ensure food arrives hot	High
C035	Positive	Delivery Time, Food Quality, Overall Experience	Maintain efficient service, delicious food, and fast delivery	High
C036	Neutral	Delivery Time, Food Quality	Continue current service levels	High

Updated Few-Shot CoT Prompt Results

Numerical Comparison Summary (Updated Few-Shot CoT Prompt, Original Tags, T=0.2 vs. Ground Truth):

1. Category Differences (vs. Ground Truth):

- C011: My Positive vs. GT Neutral - 1 difference

Total Category Differences: 1

2. Tags Differences (vs. Ground Truth):

- C002: My (Food Quality, Overall Experience) vs. GT (Packaging, Overall Experience) - 2 differences
- C008: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C009: My (no Overall Experience) vs. GT (Overall Experience) - 1 difference
- C011: My (Delivery Time, no Overall Experience) vs. GT (no Delivery Time, Overall Experience) - 1 difference
- C018: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C026: My (Food Quality, Overall Experience) vs. GT (Packaging, Overall Experience) - 1 difference
- C027: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C031: My (Delivery Time) vs. GT (no Delivery Time) - 1 difference
- C033: My (Delivery Time, no Overall Experience) vs. GT (no Delivery Time, Overall Experience) - 1 difference

Total Tags Differences: 10

3. Total Rows with at least one difference (Category or Tags): 9 out of 36 rows.

Updated Few-Shot CoT Prompt Results

Conclusion:

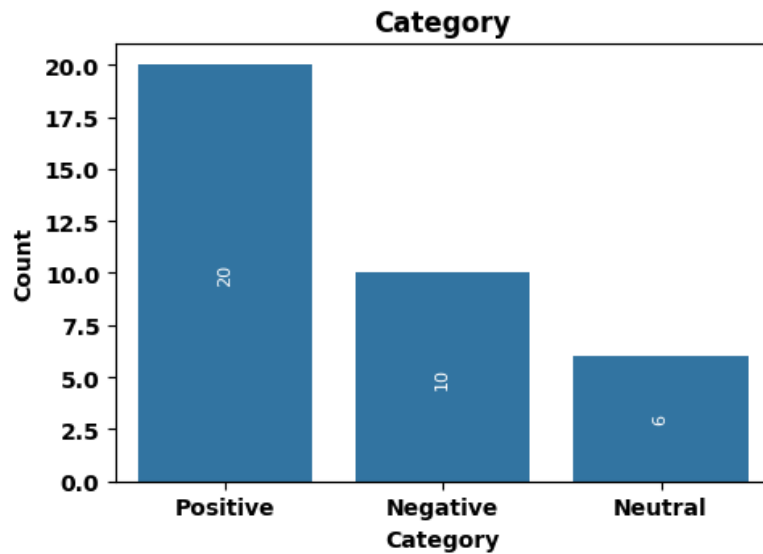
The updated Few-Shot CoT prompt shows a small improvement when compared against the Ground Truth Results dataset.

- **Category alignment:** Only 1 discrepancy (C011)
- **Tag alignment:** 10 discrepancies, a minor improvement
- **Total rows with differences:** 9 out of 36 rows, again a slight improvement

Few-shot examples combined with clearer guidelines significantly improved the model's performance. We saw better tag accuracy and very consistent sentiment categories, proving this approach is more effective.

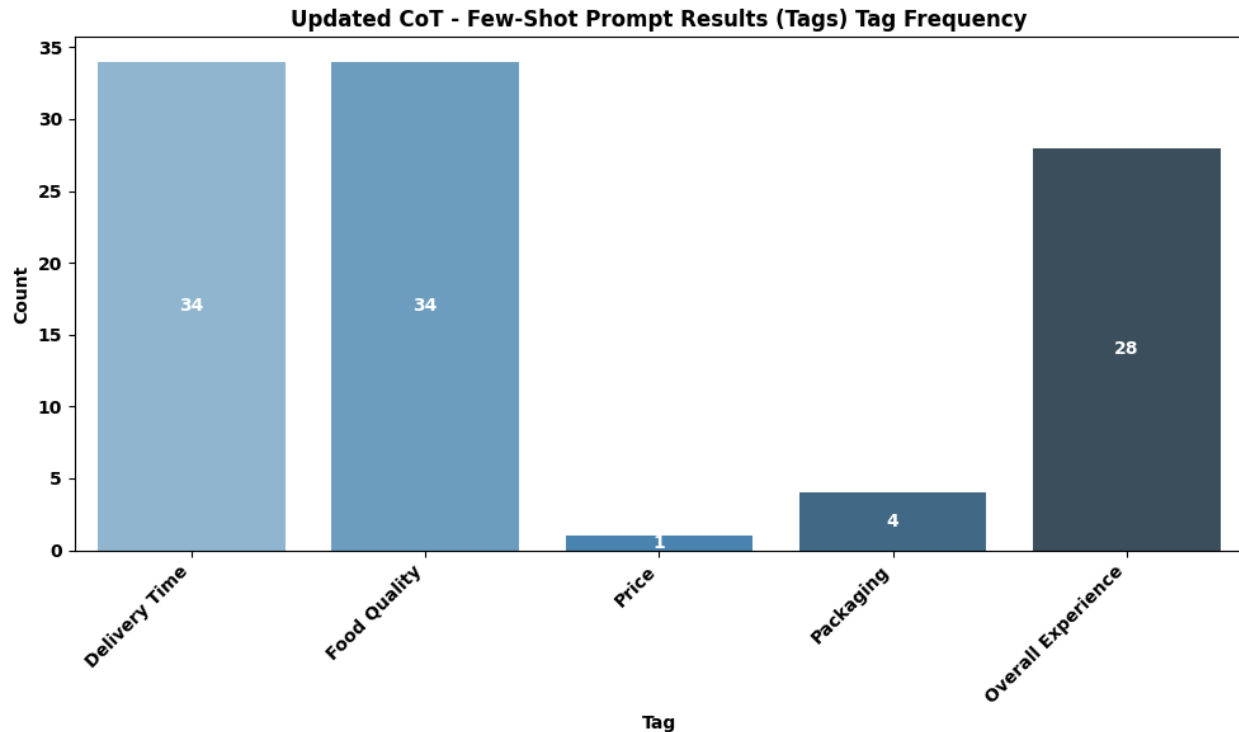
Updated Few-Shot CoT Prompt Results

Updated CoT - Few-Shot Prompt Results (Tags)Categories



Plot illustrating distribution of **Categories**

Updated Few-Shot CoT Prompt Results



Plot illustrating distribution of Tags

Few-Shot CoT Prompt Results

Numerical Comparison Summary (Updated Few-Shot Prompt, Original Tags, T=0.2 vs. Ground Truth):

1. Category Differences (vs. Ground Truth):

Total Category Differences: 0

2. Tags Differences (vs. Ground Truth):

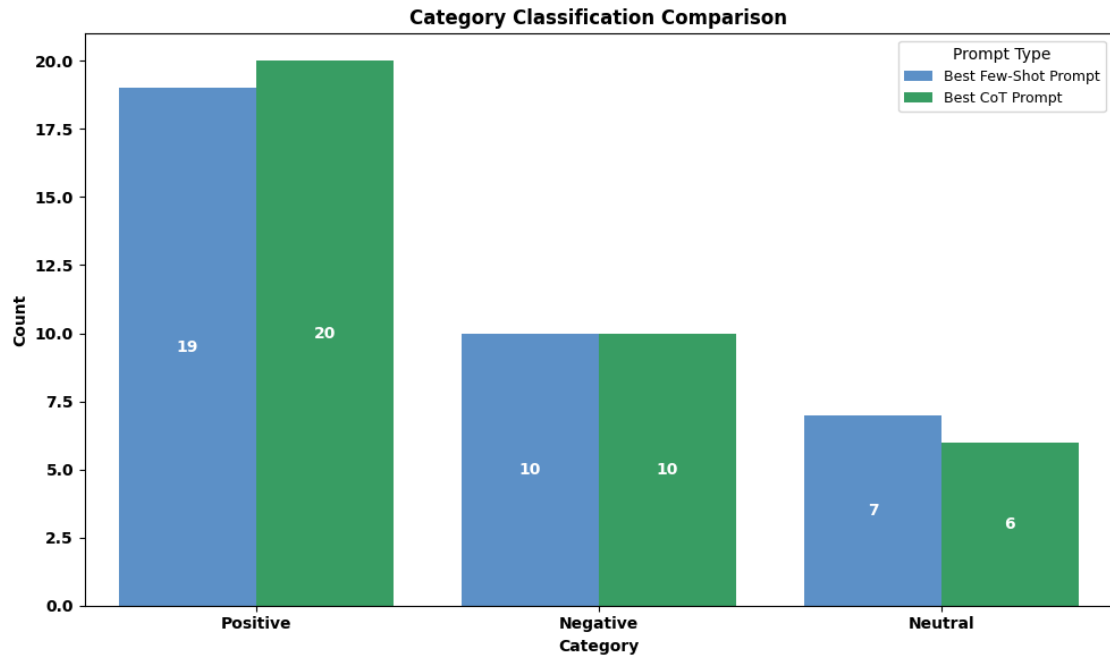
Total Tags Differences: 10

3. Total Rows with at least one difference (Category or Tags): 8 out of 36 rows.

Conclusion:

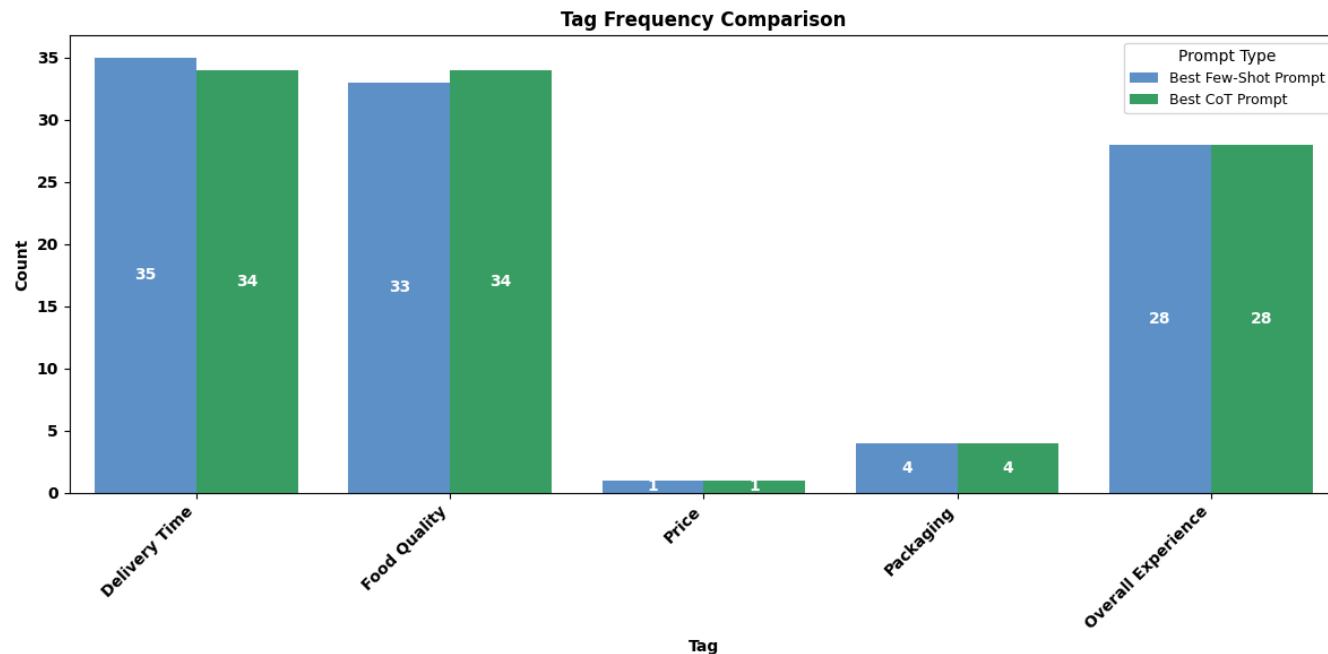
The results for Few-Shot CoT prompt are exactly the same as those for Few-Shot CoT.

Best Performing Models Comparison



Plot comparing distribution of **Category** for best **Few-Shot** and **Few-Shot CoT** prompts.

Best Performing Models Comparison



Plot comparing distribution of **Tags** for best **Few-Shot** and **Few-Shot CoT** prompts.

Best Performing Models Comparison

The "**Best Few-Shot Prompt**" achieved perfect accuracy (**0** discrepancies) in both category and tag classification. The "**Few-Shot CoT Prompt**" was very close, with only **1** category and **10** tag discrepancies.

I found it surprising that the **Few-Shot Prompt** would beat out **Few-Shot CoT** in terms of classification accuracy

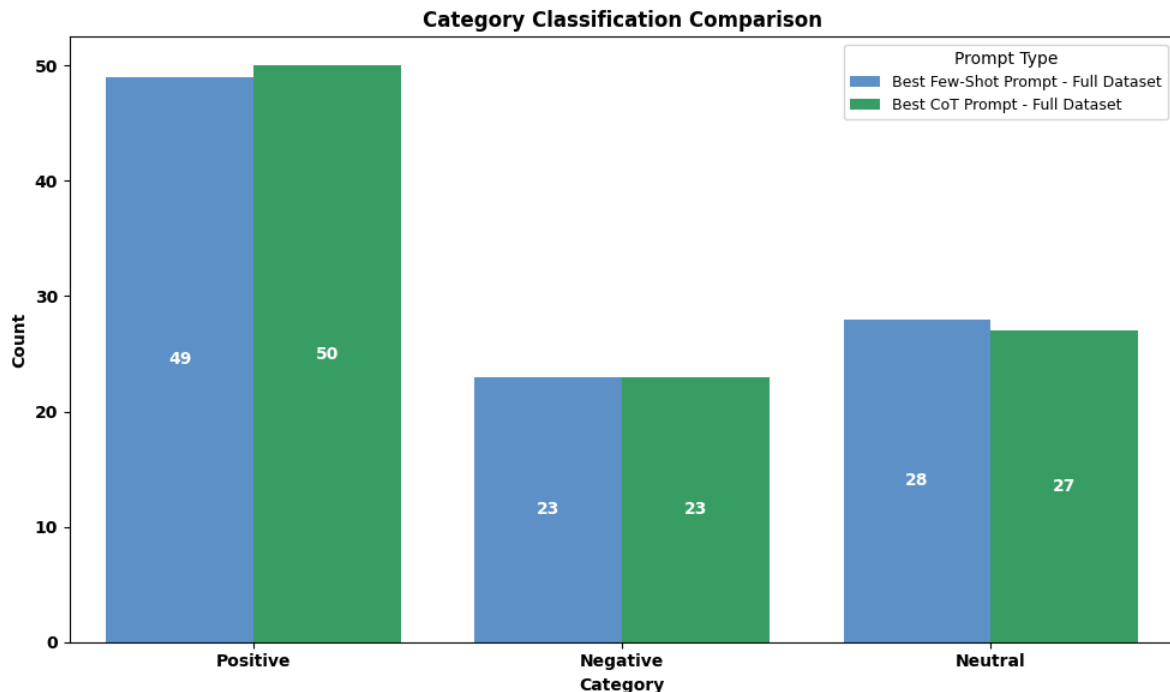
Possible Reasons for This?

- **Directness vs. Steps:** The **Few-Shot Prompt** might have directly learned from examples without extra reasoning steps, leading to flawless results for this specific task.
- **Subtle Interpretation:** While generally CoT helps, its explicit steps might have caused slight confusion in a few trickier cases compared to the example-driven **Few-Shot Prompt**.
- **Task Fit:** For well-defined tasks with good examples, adding **CoT** might not always improve accuracy and can sometimes introduce minor deviations.
- **Sometimes keeping it simpler**, lean and clean outperforms overthought/analyzed/complicated solutions
- The perfect accuracy is an **alarm bell** for possible overfitting, which might result in worse performance of the Few-Shot Prompt model when applied to the entire dataset.

Next Steps?

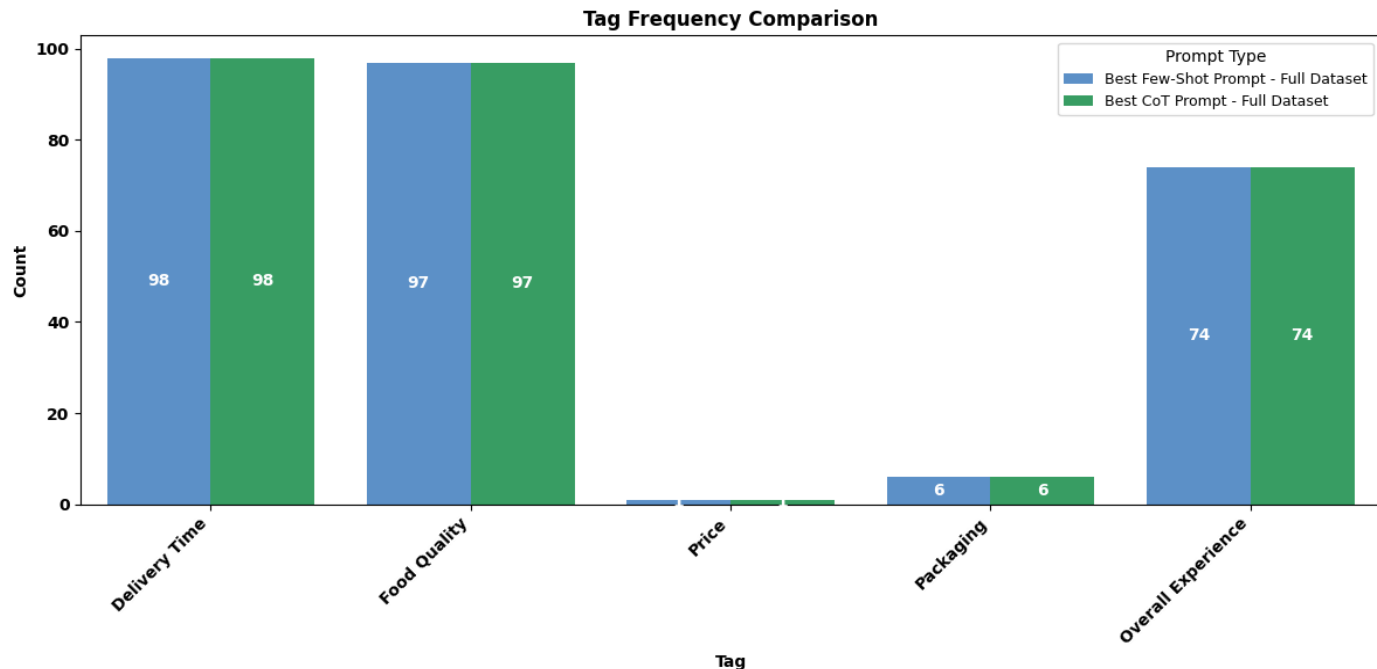
- Run both these "Best Performing" models against the entire dataset and compare the results.
- Since this is unstructured data with no known/provided target label features, we'll have to randomly manually select and spot-compare some data rows

Best Performing Models Comparison (Full Dataset)



Plot comparing distribution of **Category** for best **Few-Shot** and **Few-Shot CoT** prompts. The models have a very similar **Category** frequency distribution.

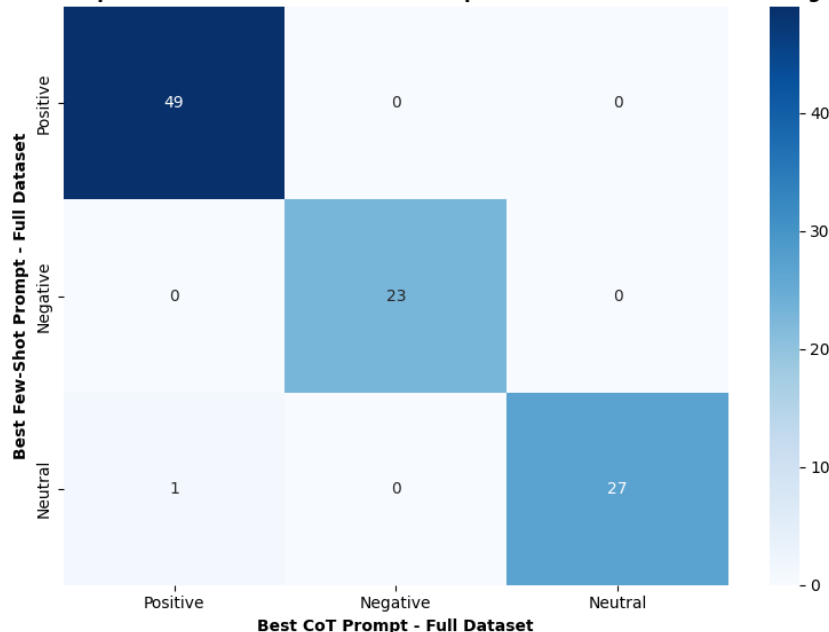
Best Performing Models Comparison (Full Dataset)



Plot comparing distribution of **Tags** for best **Few-Shot** and **Few-Shot CoT** prompts. The models have exactly the same **Tag** frequency distribution.

Best Performing Models – Category Comparison

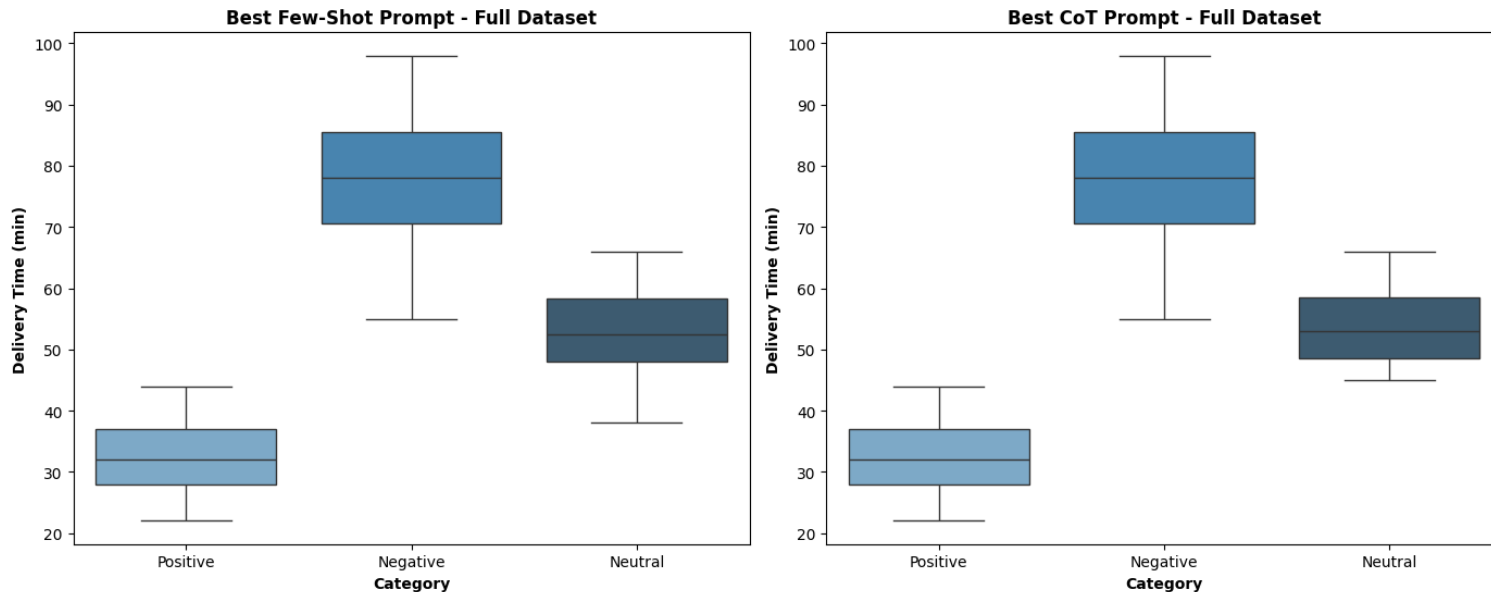
Best Few-Shot Prompt - Full Dataset vs Best CoT Prompt - Full Dataset Classification Agreement



Both prompt types agree on 99 out of 100 classifications, with only 1 review classified differently (**Few-Shot**: Neutral, **CoT**: Positive). This indicates both methods perform nearly identically on this dataset.

Best Performing Models – Category Comparison

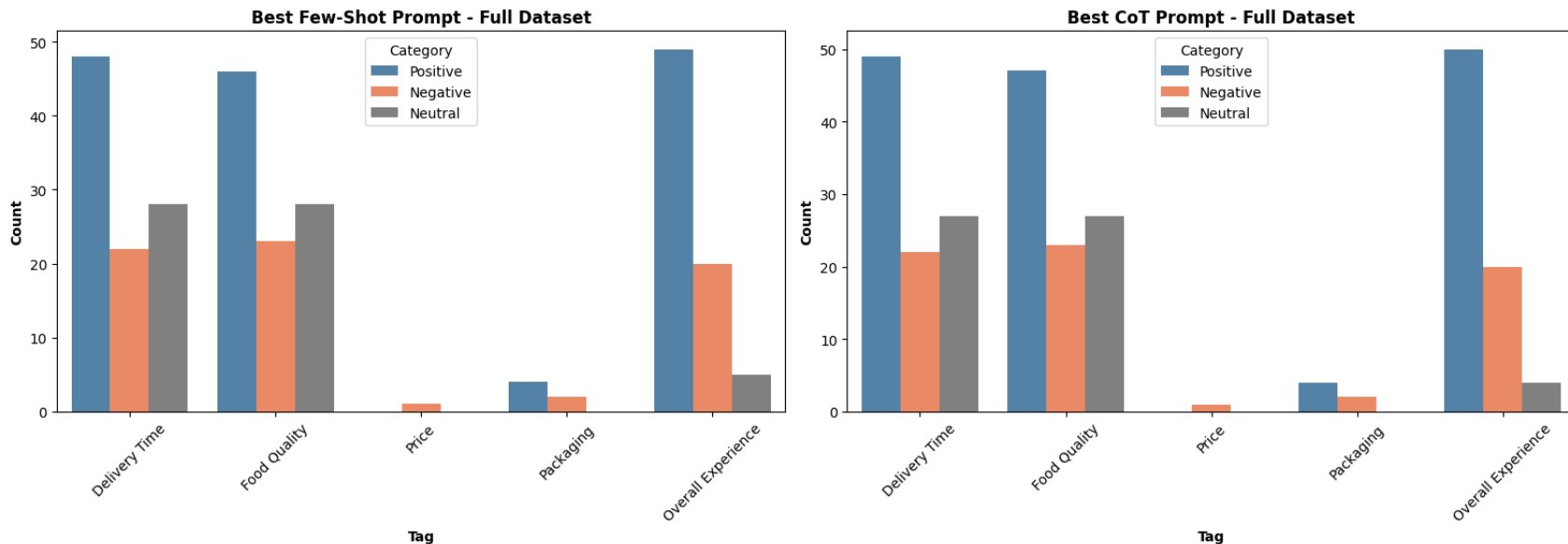
Delivery Time by Category Comparison



Both prompt types show the same pattern: negative reviews have significantly longer delivery times (median ~78 min), while positive reviews have the shortest (median ~30 min). This confirms that delivery time is a strong driver of customer sentiment, and both prompts capture this relationship consistently.

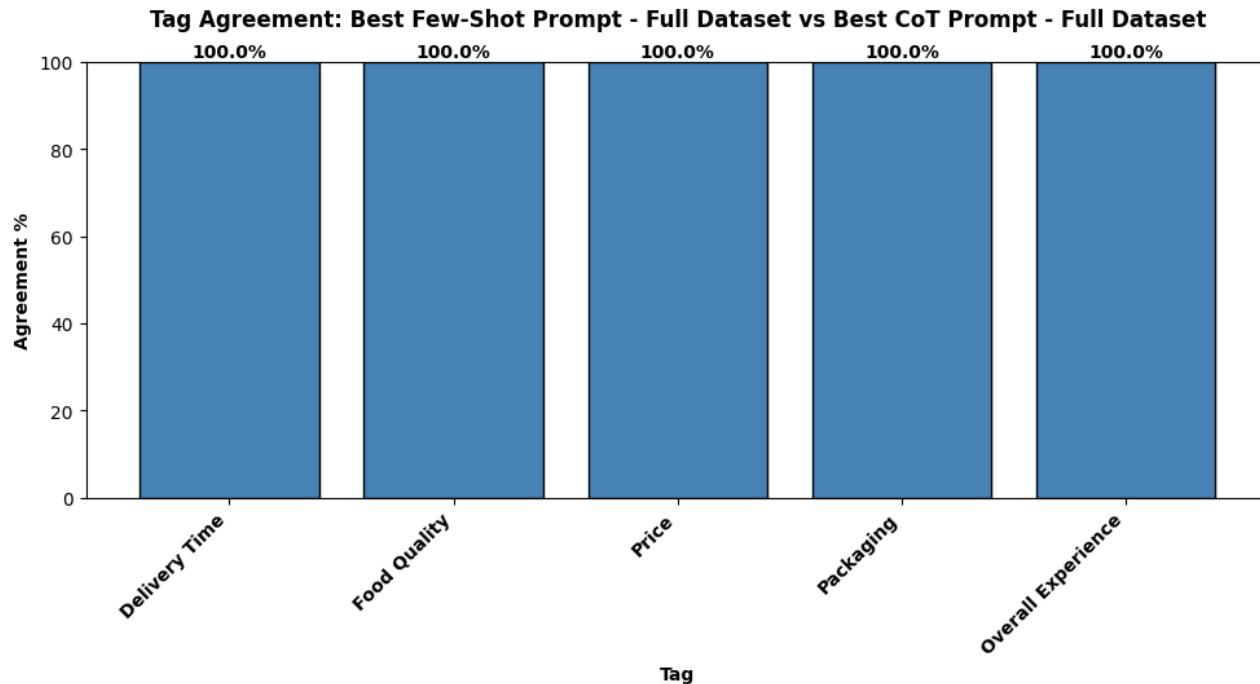
Best Performing Models – Category Comparison

Tags by Category Comparison



Tag assignment in both plot times is almost identical, with Delivery Time, Food Quality, and Overall Experience being the most common across all sentiment categories. Price and Packaging are rarely tagged, suggesting these factors are mentioned infrequently in customer reviews.

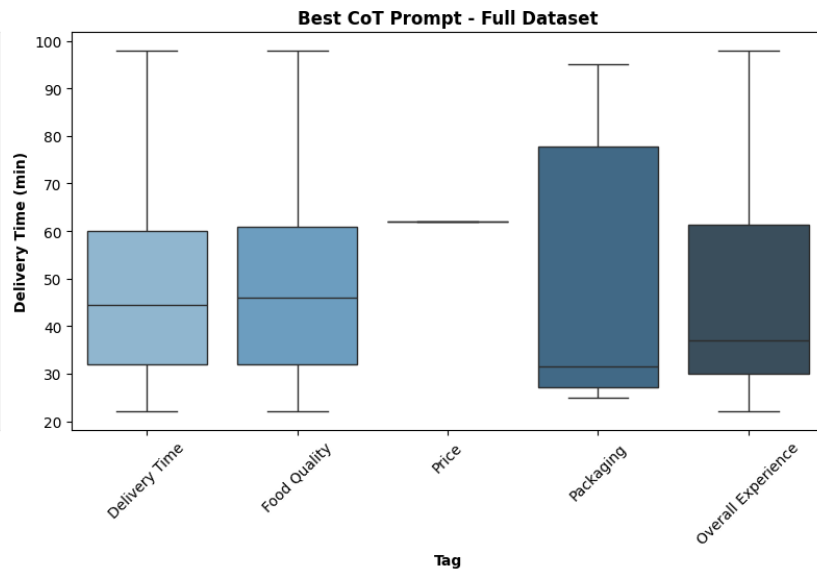
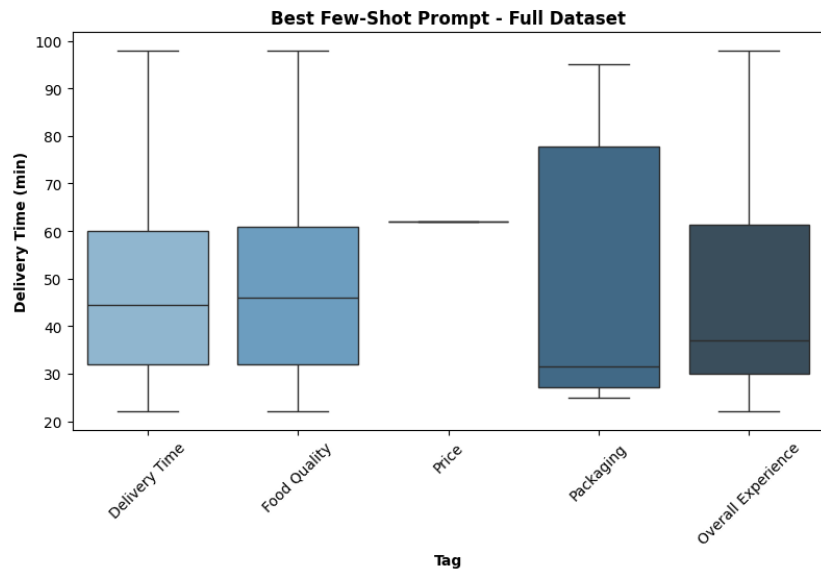
Best Performing Models – Tags Comparison



Both prompt types agree 100% on tag assignments across all five categories. This indicates that **Few-Shot** and **CoT** prompts identify the same themes in customer reviews with no variation.

Best Performing Models – Tags Comparison

Delivery Time by Tags Comparison



Both prompts show the same delivery time patterns across all tags. Packaging has the widest spread of delivery times, while Price was tagged in only a single review—indicating these factors are rarely the focus of customer feedback compared to Delivery Time, Food Quality, and Overall Experience.

Few-Shot ↔ Few-Shot Cot Comparison Results

Additional Observations:

Category Consistency: Both models demonstrate very high consistency in category classification, with only a single disagreement (**C011**). This suggests that the few-shot examples and refined guidelines are highly effective in establishing a robust sentiment classification logic for both approaches.

Tag Discrepancies: For the tag differences, the "**Best CoT Few-Shot Prompt**" tends to include more tags than the "**Best Few-Shot Prompt**" (e.g., adding 'Delivery Time' or 'Food Quality' where the other did not). This suggests the CoT reasoning might be slightly more exhaustive in identifying all potentially applicable tags based on the review text. The simpler Few-Shot model is more selective, tagging only the main points.

"Overall Experience" Tag: Both models show strong agreement on the "Overall Experience" tag, indicating the guidelines for this tag are well-understood by both.

"Price" and "Packaging" Tags: These tags appear infrequently in the reviews, and both models show perfect agreement on their application in the observed differences.

Few-Shot ↔ Few-Shot Cot Comparison Results

Numerical Comparison Summary (Updated Few-Shot Prompt, Original Tags, T=1.0 vs. Ground Truth):

1. Category Differences:

- C011: CoT Few-Shot (Positive) vs. Few-Shot (Neutral) - 1 difference

Total Category Differences: 1

2. Tags Differences (vs. Ground Truth):

- C008: CoT Few-Shot (Delivery Time, Food Quality, Overall Experience)
vs. Few-Shot (Food Quality, Overall Experience) - 1 difference (CoT added 'Delivery Time')
- C009: CoT Few-Shot (Delivery Time, Food Quality, Overall Experience)
vs. Few-Shot (Delivery Time, Overall Experience) - 1 difference (CoT added 'Food Quality')
- C018: CoT Few-Shot (Delivery Time, Food Quality, Overall Experience)
vs. Few-Shot (Food Quality, Overall Experience) - 1 difference (CoT added 'Delivery Time')
- C031: CoT Few-Shot (Delivery Time, Food Quality, Overall Experience)
vs. Few-Shot (Food Quality, Overall Experience) - 1 difference (CoT added 'Delivery Time')
- C033: CoT Few-Shot (Delivery Time, Food Quality, Overall Experience)
vs. Few-Shot (Food Quality, Overall Experience) - 1 difference (CoT added 'Delivery Time')

Total Tags Differences: 20

3. Total Rows with at least one difference (Category or Tags): 6 out of 100 rows.

Few-Shot ↔ Few-Shot Cot Comparison Results

General Summary:

It is important to remember that we are working on a small manufactured sample dataset that isn't really representative of real-world data (or maybe it is?); regardless we are basing all our conclusions and insights on this small cultivated dataset.

Both the "**Best CoT Few-Shot Prompt**" and the "**Best Few-Shot Prompt**" perform exceptionally well on this dataset, demonstrating near-perfect alignment in category classification. The "**Best Few-Shot Prompt**" achieves a slightly higher overall alignment with the Ground Truth (0 discrepancies vs. 6 for **CoT Few-Shot**), particularly in tag selection.

The minor differences in tagging suggest that while **CoT** reasoning can lead to a more comprehensive identification of aspects, the simpler **Few-Shot** approach, when perfectly tuned with examples, can achieve slightly more precise or parsimonious tag assignments that align perfectly with the Ground Truth's specific annotation philosophy. For this particular task, the added complexity of explicit CoT steps did not translate into noticeable performance improvements over a well-optimized Few-Shot approach.

Few-Shot ↔ Few-Shot CoT Comparison Results

Conclusion:

- The Zero-Shot CoT analysis, consistent with other Zero-Shot attempts, confirms that explicit, refined textual guidelines and few-shot examples are crucial for achieving high alignment with specific annotation logic.
- **Few-Shot** with clear examples works better than step-by-step reasoning for rule-based labelling. A few well constructed examples go further in improving the model than explaining it's reasoning.
- **CoT** added more complexity, but not necessarily better results. More steps don't necessarily mean better results.
- **Prompt Model performance improvements** through tighter defined prompt structures far outweigh benefits of temperature adjustment.
- Subsequent runs with tighter Temperature settings produced the same results.
- **Observed Inconsistent Results** – rerunning same prompt with same input data, temperature and commands does not produce consistent results across runs. This was mainly with running **CoT** prompts.

Few-Shot ↔ Few-Shot Cot Comparison Results

Final Recommendation:

The final recommendation is to use **Few-Shot Prompt** over Chain-of-Thought prompts.

- **It gives more consistent labels.**
Few-Shot followed the labeling rules more closely and avoided extra, unnecessary tags.
- **It's easier to maintain.**
You can change or fix behavior by updating examples instead of rewriting reasoning steps.
- **It's faster and cheaper.**
Few-Shot prompts are shorter and don't require extra reasoning, which reduces cost and response time.
- **It's easier to debug.**
When something looks wrong, you can compare the output directly to the examples.
- **It matched the ground truth best.**
In this analysis, Few-Shot aligned most closely with the benchmark labels.

Additional Observations

1. Chain-of-Thought Increased Output Variability

- When using CoT prompts, repeated executions with the same input and settings occasionally produced different outputs. This made it difficult to tract consistent results at times.

2. Issues with consistent formatted output results.

- Despite providing concise/exact instructions, Gemini still occasionally produced improperly formatted CSV output

3. Incorrect Prompt Parsing.

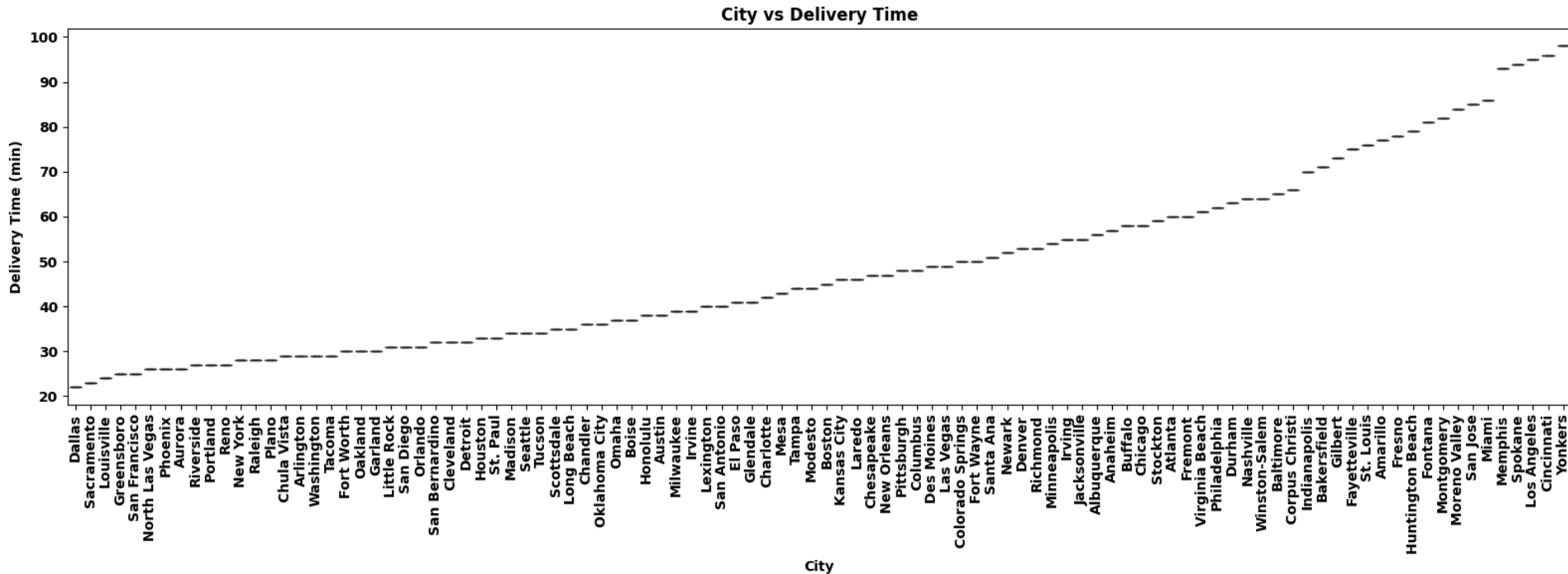
- Gemini would sometimes indicate there were prompt structure/definition errors when there in fact were none.

4. Gemini is still very much a work in progress platform

- Given the instabilities and inconsistancies I experienced Gemini is still an AI platform in development.

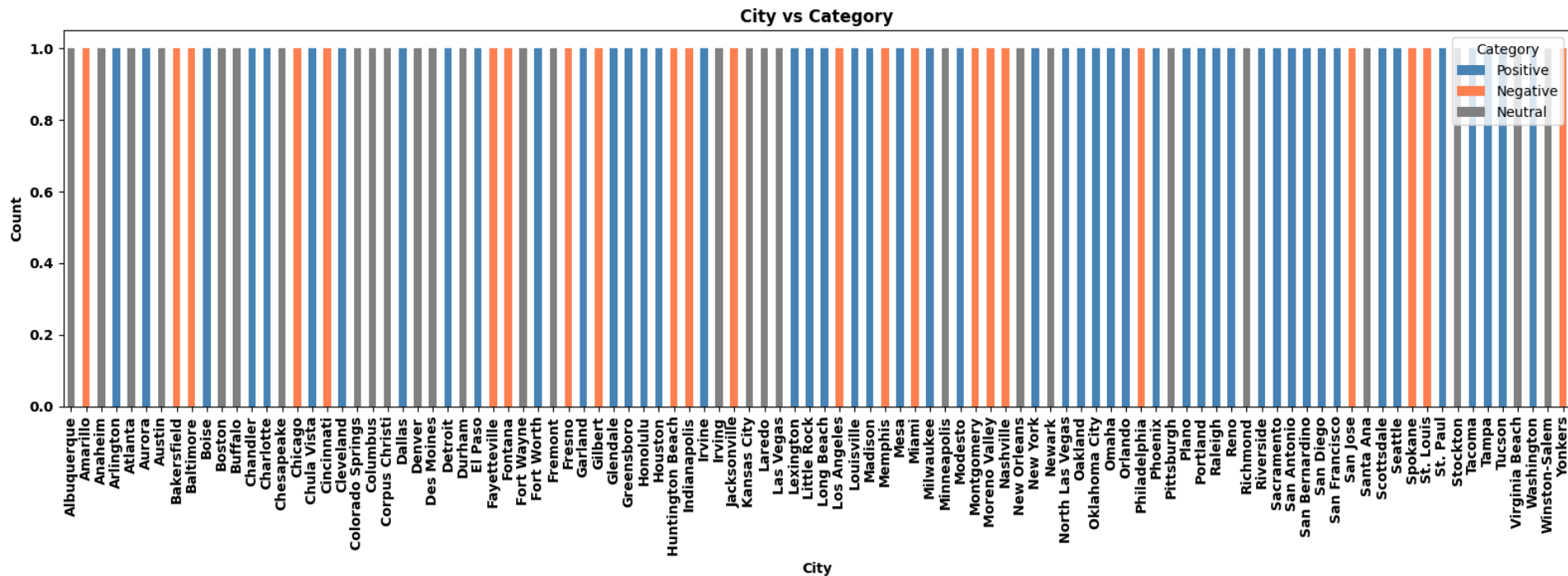
APPENDIX

Additional EDA – Feature Distributions & Relationships



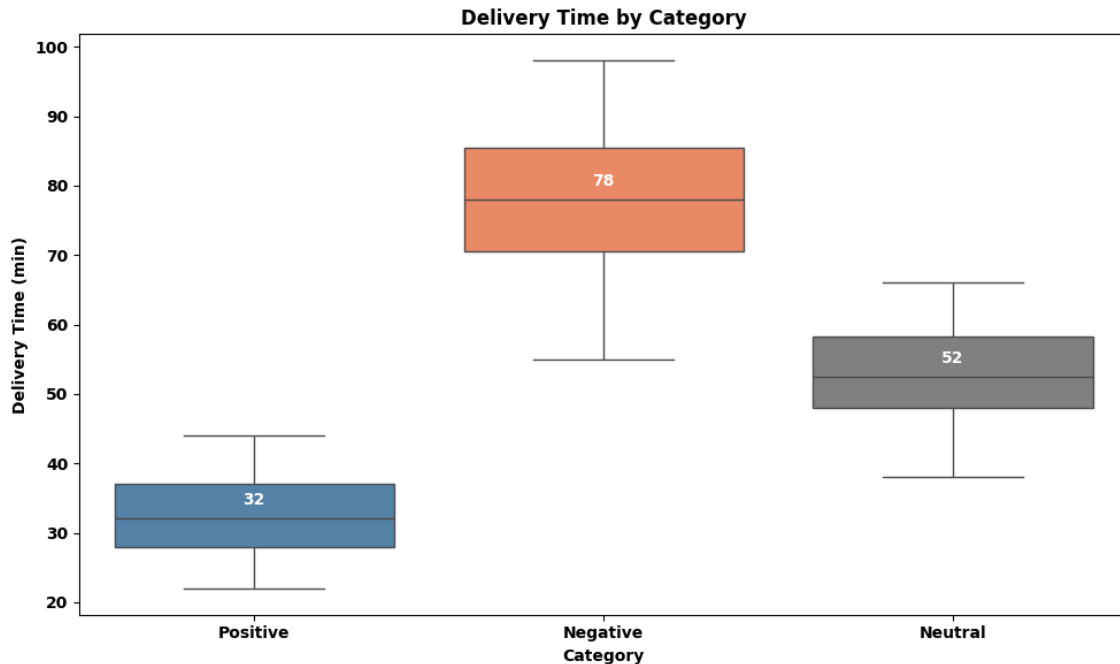
Delivery times range from 22 minutes (Dallas) to 98 minutes (Yonkers) across the dataset. This is not a meaningful plot, since there is only one review per city. More data per city would produce more meaningful patterns

Additional EDA – Feature Distributions & Relationships



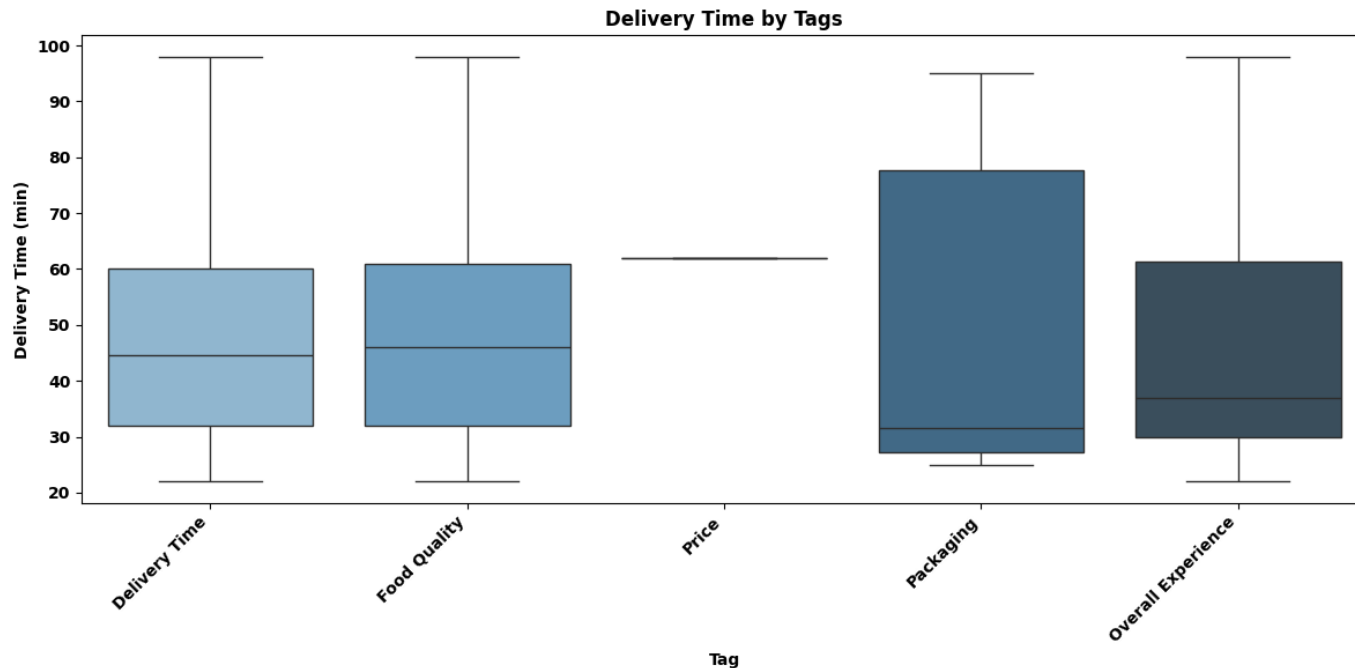
Each city has exactly one review, with sentiment evenly distributed across Positive (blue), Negative (coral), and Neutral (gray). This is not a meaningful plot, since there is only one data point per city. More data per city would produce more meaningful patterns.

Additional EDA – Feature Distributions & Relationships



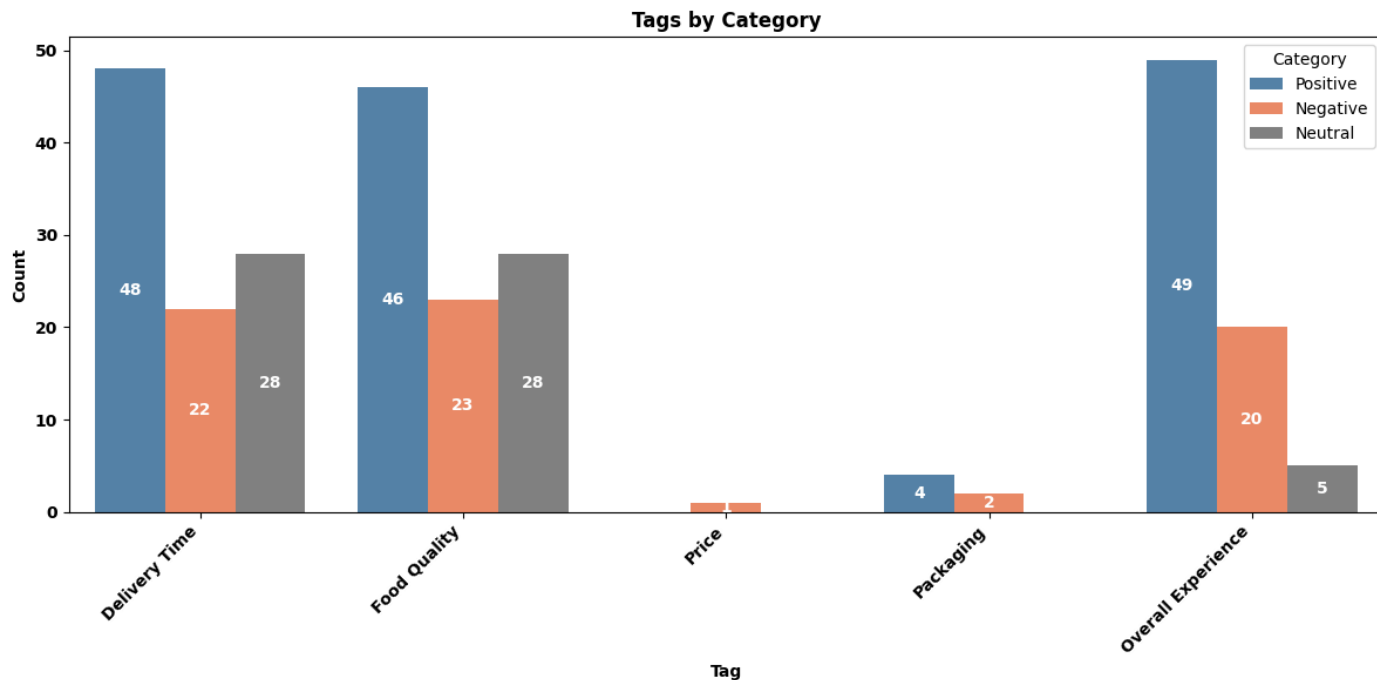
Negative reviews have significantly longer delivery times (median 78 min) compared to Positive reviews (median 32 min), with Neutral reviews falling in between (median 52 min). Delivery time is a strong driver of customer sentiment; longer waits lead to negative feedback.

Additional EDA – Feature Distributions & Relationships



Delivery Time, Food Quality, and Overall Experience tags show similar delivery time distributions, indicating these are mentioned across all delivery speeds. Price was tagged in only one review (~62 min), while Packaging shows a wide range with a higher median, suggesting packaging issues are more noticeable during longer deliveries (People more critical when they are already annoyed about a late delivery).

Additional EDA – Feature Distributions & Relationships



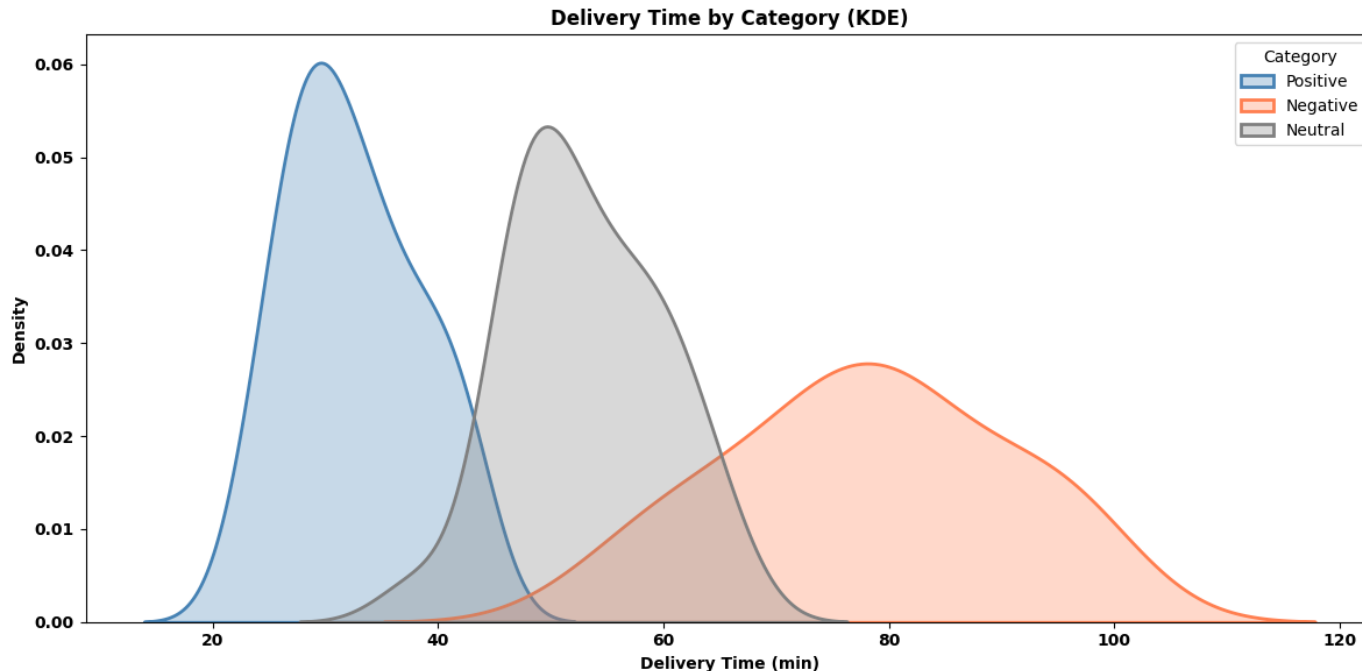
Delivery Time, Food Quality, and Overall Experience are the most frequently tagged, with Positive reviews having the highest counts across all three. Price and Packaging are rarely tagged, with only 1 and 6 total mentions respectively

Additional EDA – Feature Distributions & Relationships



Delivery times are right-skewed with most orders arriving between 25-40 minutes, and a median of 44.5 minutes. The mean (48.7 min) is higher than the median, indicating some longer deliveries are pulling the average up.

Additional EDA – Feature Distributions & Relationships



The three sentiment categories show clearly separated delivery time distributions—Positive reviews peak around 30 minutes, Neutral around 50 minutes, and Negative around 75-80 minutes. This demonstrates a strong relationship between delivery time and customer sentiment, with minimal overlap between categories

Entire Few-Shot Prompt with Zero (0) Discrepancies

The following is the the Few-Shot Prompt definition that produced results that had zero discrepancies with the Ground Truth dataset. **This is the best performing prompt model that I chose.**

Few-Shot Prompt Template:

ROLE: You are a customer feedback analyst for a food delivery service.

TASK: Analyze the uploaded customer reviews and provide structured sentiment analysis.

Use Temperature=0.2

CLASSIFICATION GUIDELINES:

Positive: Customer expresses satisfaction, praise, or intent to return

Negative: Customer expresses dissatisfaction, complaints, or intent to not return. If a review contains both positive/neutral and significant negative feedback (e.g., major delays, cold food, wrong order, poor value for money), classify as Negative.

Neutral: Mixed feedback where positive and negative aspects are balanced, or neither clearly positive nor negative.

TAGS (select ALL applicable from this list ONLY):

- **Delivery Time**
- **Food Quality:** Includes taste, temperature (if affecting quality), freshness, accuracy of order content (e.g., wrong dish, missing items).
- **Price**
- **Packaging:** Includes physical integrity of packaging, spillage, or incorrect items due to packing error.
- **Overall Experience:** Apply when the customer explicitly states "experience," "service," "will/won't order again," or uses strong general rating language that clearly reflects their overall sentiment about the entire transaction (e.g., "very satisfied," "impressive," "always happy," "everything was perfect," "very pleased," "loved everything," "terrible meal").

Entire Few-Shot Prompt with Zero (0) Discrepancies

OUTPUT REQUIREMENTS

- Category: Classify as exactly one of: Positive, Negative, or Neutral
- Tags: Select ALL applicable tags from the list above.
- Suggested_Action: Provide ONE specific, actionable business recommendation (10 words max)

Confidence: High / Medium / Low

OUTPUT FORMAT:

Return results as CSV format only. No explanations, no extra text.

Use these exact headers: Customer_ID, Category, Tags, Suggested_Action, Confidence

Rules:

- Wrap Tags in quotes if multiple: "Delivery Time, Food Quality"
- Wrap Suggested_Action in quotes
- Suggested_Action: 10 words max, no commas
- One row per customer

Entire Few-Shot Prompt with Zero (0) Discrepancies

EXAMPLES:

Review: "The food arrived 15 minutes late, which was annoying, but the pad thai was absolutely delicious and still hot. The driver was also very apologetic."

Category: Neutral

Confidence: Medium

Tags: Delivery Time, Food Quality, Overall Experience

Suggested Action: Investigate delivery delay, maintain food quality.

Review: "I ordered the large family meal, and the portion sizes were surprisingly small for the price. The chicken was a bit dry, but the sides were fresh. Overall, it felt like a rip-off."

Category: Negative

Confidence: High

Tags: Food Quality, Price, Overall Experience

Suggested Action: Review portion sizes and chicken preparation for value.

Review: "My smoothie spilled all over the bag because the lid wasn't properly sealed. The delivery was fast, but half my order was a sticky mess. Very disappointed."

Category: Negative

Confidence: High

Tags: Delivery Time, Packaging, Overall Experience

Suggested Action: Improve packaging seal quality for liquid items.

Review: "The vegetarian burrito was exactly as described and tasted good. Delivery was on time. It was a perfectly fine meal, nothing to rave about."

Category: Neutral

Confidence: Medium

Tags: Delivery Time, Food Quality

Suggested Action: Seek feedback to elevate "fine" to "memorable."

Review: "This is my go-to service. The sushi is always fresh, and the drivers are consistently friendly. I've never had a bad experience, which is why I keep coming back."

Category: Positive

Confidence: High

Tags: Delivery Time, Food Quality, Overall Experience

Suggested Action: Highlight consistent quality and reliable service in marketing.



Happy Learning !

