

Would You Eat Off This Data Table? Answer Key

1. Spot the Problems - Common issues students should find:

- Inconsistent capitalization ("Choco Bites" vs "choco bites")
- Duplicates or near-duplicates of the same review
- Inconsistent labeling ("Positive," "POS," "neg," "Negative?" etc.)
- Spelling errors and inconsistent phrasing ("To sour" instead of "Too sour")
- Missing data (e.g., Snack "?" with blank label)
- Unclear or duplicate punctuation ("Too salty salty salty" vs "TOO SALTY SALTY SALTY")
- Category confusion (same phrase labeled multiple different ways)

Teacher Note: Encourage students to think about *why* each inconsistency causes problems - for example, AI might treat "neg" and "Negative" as different categories, or think "Trail mix" and "Trail Mix" are two separate snacks.

2. Duplicates - Likely duplicates or near-duplicates:

- "Choco Bites" entries (#1, #2, and #10) - same review repeated, slight formatting differences.
- "Salties" entries (#3, #6, #12, #13) - same review written multiple ways ("Too salty salty salty," "TOO SALTY SALTY SALTY," "Tooo salty!!").
- "Fruit Crunch" entries (#5, #7, #9) - same review repeated with spelling and capitalization differences.
- "Trail Mix" entries (#8 and #11) - same review, capitalization and spelling errors.

Teacher Note: AI can treat these as separate data points unless duplicates are merged or removed. This causes bias toward overrepresented examples.

3. Missing or Incorrect Labels

Examples:

- Snack "?" - Missing name and missing label. Should likely be removed.
- "Fruit Crunch" entry labeled "N/A" - not a valid category.
- "Salties" labeled "Negative?" - uncertain label. Needs clarification or removal.
- "NUETRAL" - misspelled; should be standardized to "Neutral."

Teacher Note: AI doesn't know that "NUETRAL" or "neg." mean "Neutral" or "Negative." Data cleaning means choosing one consistent label style and applying it everywhere.

4. Standardizing the Data - Model student answers might include:

- Use consistent capitalization (e.g., "Choco Bites," "Fruit Crunch," "Trail Mix," "Salties").
- Standardize label categories to exactly three options: **Positive, Neutral, Negative.**
- Fix spelling and punctuation errors in reviews ("Too sour" instead of "To sour").
- Remove duplicates and unclear entries.
- Replace blank or invalid labels with correct ones or remove them.
- Use the same text format for all reviews (capitalize first letter, consistent punctuation).

Teacher Note: Standardization helps AI treat each item equally and avoids noise from formatting errors.