Data Tagging Summary for Task 1

1. Approach

To systematically tag the dataset, I applied a keyword-based matching approach using predefined taxonomy categories. The process involved:

- Root Cause Tagging: Extracted keywords from the "Cause" column and matched them with predefined taxonomy categories.

- Symptom Condition & Component Tagging: Matched words in "Complaint" with known symptom categories.

- Fix Condition & Component Tagging: Mapped repair actions in "Correction" to fix categories.

2. Challenges

- Ambiguous Cases: Some complaints lacked enough detail to determine an exact cause.

- Multiple Symptoms/Fixes: Some records contained multiple issues in a single entry.

- Data Formatting Issues: Inconsistent terminology required preprocessing for better matching.

3. Insights

- The most frequent root causes were 'Not Installed' and 'Loose,' highlighting common assembly issues.

- Most complaints related to leaks, missing parts, and faulty connections.

- Around 20% of the data was incomplete or ambiguous, suggesting the need for better data collection.

This analysis can help manufacturers improve defect tracking and optimize quality control measures.