

## Summary Report

For Task 1, I was given a dataset containing descriptions of Complaint, Cause, and Correction fields, along with a taxonomy sheet that listed standardized terms for Root Cause, Symptom Condition, Symptom Component, Fix Condition, and Fix Component. The objective was to tag each record in the dataset by identifying and mapping these standardized taxonomy values based on the unstructured Complaint-Cause-Correction text fields.

### Approach:

#### Data Preparation:

First, I cleaned the taxonomy data to remove nulls and extracted unique values for each taxonomy field into separate lists. This ensured that I was working with only valid, non-duplicate taxonomy terms.

#### Matching Logic:

I defined a matching function that searched each row's Complaint, Cause, and Correction text for the presence of any taxonomy terms using a simple substring check. This function was case-insensitive and returned matched values as comma-separated strings if multiple terms were found.

#### Field-wise Mapping:

- **Root Cause:** I searched across all three fields (Complaint, Cause, Correction) for any root cause term and stored the match in a single "Root Cause" column.
- **Symptom Condition / Component and Fix Condition / Component:** Each of these was matched separately against Complaint, Cause, and Correction fields, creating three columns per type (e.g., Symptom Condition 1, 2, 3), to capture context-specific mapping.

#### Insights (Bonus Thought):

**Pattern:** The tagged data can now be used to identify the most frequently occurring root causes or symptom conditions, which can help prioritize engineering or support resources.

**Automation:** If high-frequency complaint keywords map consistently to certain conditions or fixes, rule-based or ML-based auto-tagging systems can be trained.