

Task 1 Summary Report – Data Tagging

In this task, I worked on converting unstructured free-text entries from the Complaint, Cause, and Correction columns into structured tags based on a predefined taxonomy. The goal was to assign accurate labels for five key fields: Root Cause, Symptom Condition, Symptom Component, Fix Condition, and Fix Component.

Approach to Tagging

To start, I imported the taxonomy into Excel and applied data validation drop-downs to the tagging columns. This helped ensure I could only select values from the approved list, reducing the risk of tagging errors. Before diving into tagging, I took time to understand the dataset, reviewing common phrases like “not installed” or “leaking”—and studied the taxonomy to distinguish between similar categories like “Not Installed” and “Not Included.”

How I Tagged Each Field

- **Root Cause:** I mainly looked at the Cause column, but also cross-checked the Complaint and Correction fields to validate the tag. For example, if the Cause mentioned "Leaking" and the Complaint said "Dripping," I chose a label like "Out of Fitting" to reflect the issue's origin.
- **Symptom Condition & Component:** These were mostly identified from the Complaint column. I extracted up to three condition–component pairs, such as "Getting Fault Code" + "Condenser" or "No Power" + "Control Board," aligning them to the closest match in the taxonomy.
- **Fix Condition & Component:** Based on the Correction column, I tagged the action taken and the component involved—for example, "Installed" + "Gas Strut" or "Cleaned Out" + "Coupler."

Handling Ambiguity

When a clear taxonomy match wasn’t available, I chose the closest reasonable option. If I couldn’t find a good fit, I left the tag blank and added a note for review. This helped maintain accuracy while avoiding forced matches.

Insights Gained

Tagging the data revealed recurring issues like installation errors, component misalignment, and leaks. These patterns suggest areas where preventive training or process improvements could make a big difference. It also showed the potential to use this structured data for automated tagging and better root-cause analytics in the future.

"Overall, this structured tagging process laid the foundation for better analytics and operational improvement."