Summary Report: Data Tagging Approach

Regarding task 1,

(A) summary report

I had to take a careful approach through the data tagging process and ensure that every field aligned to the taxonomy provided to and to ensure accuracy and consistency throughout the dataset.

Root Cause: I determined the key reason for each problem by reviewing the "Cause" column and if necessary, interpreting "Correction" descriptions. For example, if a hose was leaking due to not being tightened at the factory, I tagged the root cause as "Not Tightened". If the cause was unclear or not provided at all, I tagged it as "Not Mentioned". Taking this step helped me standardize repetitions in failure reasons and common issues identified at the manufacturing or assembly point.

Symptom Condition: This field is meant to capture what the customer saw or experienced. I extracted symptom conditions from the "Complaint" column for taxonomy terms: "Leak", "Not Working", "Fault/Error Codes" or "Broke". If the symptom condition was unclear or was not provided, I went with "Not Mentioned". This allowed us to understand which symptoms were reported most often and how the failure ultimately manifested itself in the equipment.

Symptom Component: I documented the component that showed symptoms, like "Fuel Door," "Auto Boom Sensor," "Bolts," "Braided Steel," etc. If the component was not mentioned, I indicated "Not Mentioned." This is useful for seeing what components are more fragile or prone to failure.

Fix Condition and Fix Component: These fields capture the correction made and the component. I mapped corresponding actions (i.e. "Installed," "Replaced," or "Retightened") from the column "Correction." Only completed fixes were referenced; simply inspections or incomplete attempts were coded as "Not Mentioned." This maintains the integrity of the data representing actual correction rather than a planned/attempted correction.

(B) Potential Insights

The things that I noticed is that there are frequent similarities in root cause like “not tightened” or “no Oring” appears multiple times in the dataset. This shows this type of issue occurs more often than others.

The analysis reveals symptom and fix patterns. “Leak,” “Not Working,” and “Error Codes” are the most commonly reported symptoms. “Installed,” “Retightened,” and “Replaced” are the most often used fixes. Several instances have multiple symptom and fixes in a single complaint. One fault can affect multiple components, which could complicate maintenance.

So by tagging the dataset according to the given taxonomy, we can able to systematically classify into complaint, cause, and corrective action. Overall, the tagging facilitated the understanding of the dataset and provided valuable information that could guide future preventive initiatives as well as improvements to maintenance practices.