# **TASK 1-DATA ANALYSIS**

# Neha Kumari

# **Approach to Tagging**

The dataset consists of free-text fields (Complaint, Cause, and Correction) that need to be categorized based on predefined categories from the Taxonomy sheet. To ensure logical and accurate tagging, we implemented the following approach:

#### **1. Preprocessing of Text Data:**

* Converted all text to lowercase for uniformity.
* Removed special characters and extra spaces to standardize inputs.
* Tokenized text to facilitate pattern recognition and matching.

#### **2. Mapping Free Text to Taxonomy Categories:**

* Used fuzzy matching techniques (FuzzyWuzzy) to compare text with predefined taxonomy categories.
* Set an 80% similarity threshold for assigning a category.
* If a match was below the threshold, the entry was labeled as **"Needs Review"** for manual verification.

#### **3. Logical Assignment of Categories:**

#### **Root Cause:** Mapped descriptions of the underlying issue to known root causes.

#### **Symptom\_Condition & Symptom\_Component:** Extracted symptom-related keywords from the "Complaint" field and linked them to known conditions and components.

* **Fix\_Condition & Fix\_Component:** Analyzed corrective actions in the "Correction" field and associated them with repair conditions and components.

#### **4. Handling Ambiguous Cases:**

* Low-confidence matches were stored separately in a **"Low\_Confidence\_Matches.xlsx"** file for manual review.
* This ensures transparency and allows for iterative improvement in tagging accuracy.

### **Potential Insights & Observations**

* + **Frequent Root Causes:** Identifying the most common failure patterns can help prioritize preventive measures and design improvements.
  + **Symptom Patterns & Fix Efficiency:** By analyzing symptom-to-fix relationships, we can optimize troubleshooting guides and enhance service efficiency.
  + **Error Reduction:** A structured taxonomy-based approach reduces ambiguity, leading to better decision-making in maintenance and quality assurance.
  + **Automation Enhancement:** Future improvements can integrate machine learning models (e.g., NLP-based classifiers) to enhance contextual understanding beyond simple fuzzy matching.
  + **Trend Analysis for Predictive Maintenance:** Identifying recurring patterns in symptoms and fixes can help forecast potential failures, reducing downtime and improving operational efficiency.
  + This structured approach ensures **efficient and accurate tagging**, providing valuable insights for **process optimization, predictive maintenance, and quality improvement.**