

## **Task 1 - Data Tagging**

### **Data for task 1**

#### **Guidelines:**

The dataset consists of the following:

- Free-text data (Columns: Complaint, Cause, Correction) that needs to be tagged.
- Taxonomy Sheet: A reference list with predefined categories for Root Cause, Symptom\_Condition, Symptom\_Component, Fix\_Condition, and Fix\_Component.

**Your task is to tag the data by applying logical reasoning and aligning it with the categories provided in the taxonomy.**

**Note:** Two Examples of how to tag the data are already provided in Task 1. Please refer to those examples.

#### **Deliverables for Task 1:**

1. Submit the tagged dataset in CSV or Excel format.
2. A summary report (200-400 words) explaining –
  - a. How you approached the tagging of each given field (Root Cause, Symptom\_Condition, Symptom\_Component, Fix\_Condition, and Fix\_Component.)
  - b. Any potential insights that can be generated (Food for thought and has bonus marks)

## **Task 2 - Data Analysis and Insights Generation using Python**

### **Data for task 2**

#### **Guidelines**

1. Column-Wise Analysis:
  - Perform a column-wise analysis of the provided dataset.
  - Describe each column in terms of its data type, unique values, distribution, and overall significance for stakeholders
2. Data Cleaning:
  - Handle missing or invalid values using appropriate methods (e.g., imputation, deletion).
  - Address inconsistencies in categorical columns (e.g., typos, inconsistent capitalization).
  - Ensure numerical columns are in the correct format and free from outliers, where applicable.
3. Identifying Critical Columns:
  - Select the top 5 critical columns that might be most insightful for stakeholders according to your data understanding.
  - Provide reasoning for your selection.
  - Generate visualizations (e.g., bar plots etc) using Python to represent these insights effectively. (atleast 3)
4. Generating tags/features from free text available :
  - Generate meaningful tags from the free text fields to summarize information, example - failure conditions and components etc..
5. Summary and Insights (Food for thought and has bonus marks)
  - Write a summary of the tags generated, including potential insights derived from the dataset.
  - Provide actionable recommendations for stakeholders based on your analysis.
  - Highlight discrepancies in the dataset (e.g., null values, missing primary keys) and how did you approach.

#### **Deliverables for Task 2:**

1. Submit the cleaned and tagged records in CSV or Excel format.
2. Submit a detailed report covering the above pointers, including (Maximum 2 page):
  - a. Column analysis
  - b. Data cleaning summary and
  - c. Visualizations d. Generated tags & Key takeaways
3. Attach Python scripts used for the analysis.