#### **Axion Ray Task 1 Summary**

#### **A. Approach to Tagging Each Field**

The tagging process involved mapping free-text data from the fields Complaint, Cause, and Correction to predefined categories in a taxonomy. The key steps were:

1. **Data Preparation:**
   1. Extracted predefined categories for Root Cause, Symptom Condition, Symptom Component, Fix Condition, and Fix Component from the taxonomy sheet.
   2. Ensured that any missing or incorrect values were handled gracefully by dropping NaN values and ensuring consistent data types.
2. **Text Matching Logic:**
   1. Created a matching function (find\_all\_matches) to identify occurrences of taxonomy categories within the free-text data.
   2. This function converted both the free-text and taxonomy categories to lowercase to perform case-insensitive matching.
   3. Applied this function to each field (Complaint, Cause, Correction) to detect relevant keywords and tag them to corresponding columns.
3. **Handling Multiple Matches:**
   1. Since multiple conditions or components could be present in a single text field, the code was designed to fill up to three corresponding columns (Symptom Condition 1, Symptom Condition 2, etc.). This ensured comprehensive tagging without losing data.
   2. If no matches were found, None was explicitly assigned to maintain data integrity and consistency across the dataset.

#### **B. Potential Insights with a Larger Dataset (10k Rows)**

Analyzing a larger dataset of similar nature (e.g., 10,000 rows) could unlock valuable operational and strategic insights:

1. **Root Cause Analysis:**
   1. Identify the most common root causes of failures or complaints. This could help prioritize process improvements, training, or supplier quality initiatives.
2. **Pattern Recognition:**
   1. Discover recurring patterns in symptoms and components, revealing systemic issues within specific product lines or manufacturing processes.
   2. Use these insights to implement predictive maintenance strategies or enhance product design.
3. **Efficiency and Cost Savings:**
   1. Understanding the most frequent corrective actions (Fix Condition) can help optimize repair processes, reduce downtime, and manage spare part inventory more effectively.
   2. Identify common failure points to reduce warranty costs and improve customer satisfaction.
4. **Text Analytics and NLP:**
   1. With a large dataset, implementing advanced Natural Language Processing (NLP) techniques could further improve tagging accuracy.
   2. Clustering or topic modeling might reveal hidden trends or suggest new categories for future taxonomy updates.
5. **Benchmarking and Comparison:**
   1. Analyze data across different time periods, product categories, or geographies to detect trends in performance or quality issues.
   2. Benchmark performance metrics (like defect rates) to industry standards, driving continuous improvement.