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

#### **a. How I Approached the Tagging of Each Given Field**

The tagging process involved filling empty cells in the 'Task' sheet using predefined categories (tags) from the 'Taxonomy' sheet. Here’s how I approached tagging each field:

* **Root Cause**: For the Root Cause column, I matched the empty cells with the predefined tags related to issues such as "Not Tightened", "Not Installed", or "Leaking". These tags were chosen based on a direct mapping from the tags available in the 'Taxonomy' sheet. In practice, this involved filling empty cells with a default tag such as "Not Tightened" until a more specific rule or additional context could be determined.
* **Symptom Condition**: For the Symptom Condition columns (e.g., Symptom Condition 1, Symptom Condition 2), I used the predefined list of symptoms (e.g., "Loose", "Won't stay open", "Crushed") to fill in missing values. These symptoms are associated with different mechanical issues or failures and provided a useful categorization for incomplete data.
* **Symptom Component**: Similarly, for Symptom Component columns, such as Symptom Component 1, I used the list of potential components (e.g., "Fuel Door", "Cab P Clip", "Boom") to fill empty cells. These components typically represent the parts of the machinery involved in the failure or malfunction described in the task.
* **Fix Condition**: In the case of Fix Condition columns, I used tags like "Retightened", "Installed", and "Repaired" to fill missing values. These terms relate to the actions taken to fix the identified issues, and I populated these fields based on available context in other columns of the dataset.
* **Fix Component**: For Fix Component columns, I chose tags like "Gas Strut", "O-Ring", and "Bulkhead Connector" to fill in the empty cells. These represent the specific parts that were repaired or replaced, aligning with the issues identified in the Symptom Component field.

The approach was fairly straightforward: for each empty cell in these fields, I replaced the missing value with the most common or relevant tag from the taxonomy, using predefined lists. In real-world applications, a more context-sensitive approach would be needed, where the filling would be based on understanding the nature of each failure and the likely corrective actions.

#### **b. Potential Insights from a Larger Dataset for 10k Rows**

With a dataset of 10,000 rows, the analysis could yield several insightful trends and patterns. Some potential insights include:

* **Root Cause Analysis**: By analyzing the frequency of specific root causes (e.g., "Not Tightened", "Not Installed"), it would be possible to identify the most common mechanical failures across a larger set of data. This could help identify areas where quality control or factory processes need to be improved, such as tightening procedures or part installation.
* **Common Symptoms and Components**: With a larger dataset, we could identify which components are most frequently associated with specific symptoms. For example, if "Fuel Door" appears often as the symptom component for issues where the fuel door "Won't stay open", this could suggest a recurring issue with that specific component, prompting more focused maintenance or part replacement strategies.
* **Cost Optimization in Repairs**: By analyzing the Fix Condition and Fix Component columns, we could estimate the most common fixes (e.g., "Retightened", "Installed") and which parts are frequently replaced (e.g., "Gas Strut", "Bulkhead Connector"). This could provide insights into cost optimization—identifying which repairs are most frequent and where maintenance resources should be allocated.
* **Predictive Maintenance**: With a larger dataset, it would be possible to use machine learning algorithms to predict failures based on past data. For example, by correlating the Symptom Condition and Root Cause columns, predictive models could forecast which components are most likely to fail, allowing for proactive maintenance and reducing downtime.
* **Anomaly Detection**: Outliers or rare occurrences in root causes or fixes could be flagged using statistical analysis or machine learning, potentially uncovering unknown issues that require immediate attention.

In summary, a larger dataset would enable more granular analysis and the extraction of actionable insights, helping to optimize maintenance schedules, reduce costs, and improve machine reliability.