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.

Summary Report for Task 1 — Tagging Free-Text Complaint Data

a. Tagging Approach

To tag the dataset accurately, I used a hybrid NLP strategy combining **spaCy** for linguistic parsing and **fuzzy matching** for semantic alignment with the taxonomy. The process involved:

- **Text Consolidation**: Merged the **Complaint, Cause, and Correction** fields into a single narrative per record to capture full context.
- **Keyword Extraction**: Used spaCy to extract noun phrases, verbs, and adjectives that represent symptoms, causes, and fixes.
- **Taxonomy Matching**: Applied fuzzy string matching to align extracted keywords with predefined categories in the taxonomy sheet. This allowed for flexible tagging even when terminology varied (e.g., “not tightened” vs “loose”).
- **Tag Completion**: Filled in missing values for Root Cause, Symptom Condition, Symptom Component, Fix Condition, and Fix Component based on the best match from the taxonomy.
- **Python Strategy: Hierarchical Tagging (Top 3 Matches)**
  We’ll extract multiple relevant keywords from each record and match the **top 3** to fill:
  Symptom Condition 1, 2, 3
  Symptom Component 1, 2, 3
  Fix Condition 1, 2, 3
  Fix Component 1, 2, 3

This approach balances automation with interpretability, ensuring that tags are contextually relevant and semantically accurate.

b. Insights Generated

The tagged dataset reveals several operational patterns:

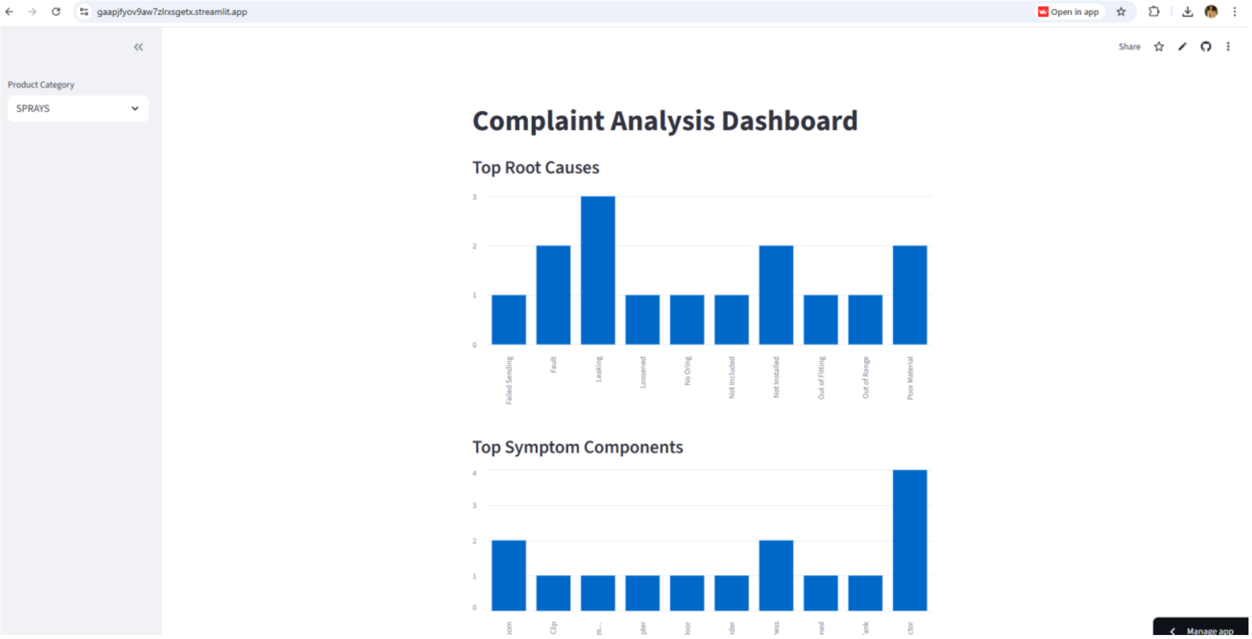
- **Recurring Root Causes**: **"Not Tightened," "Poor Material," and "Missing Components"** dominate, indicating potential gaps in factory QA and assembly protocols.
- **Frequent Symptom Components**: "**Cab P Clip," "Fuel Door," and "Bulkhead Connector"** appear repeatedly, suggesting these parts may require design review or better installation procedures.
- **Common Fixes**: **"Retightened," "Replaced," and "Installed"** are the most frequent corrective actions, which can inform technician training and inventory stocking.
- **Product Category Trends**: Most issues are concentrated in the **"SPRAYS"** category, hinting at either higher usage or more complex assembly compared to **"BALER."**

These insights can guide preventive maintenance, quality control audits, and targeted technician upskilling. With location and time data added, this could evolve into a predictive service dashboard.

I also deployed a piece of code to github and made a dashboard using streamlit.

Link: https://gaapjfyov9aw7zlrxsgetx.streamlit.app/

Github link: https://github.com/tsahil007/assignment/tree/main



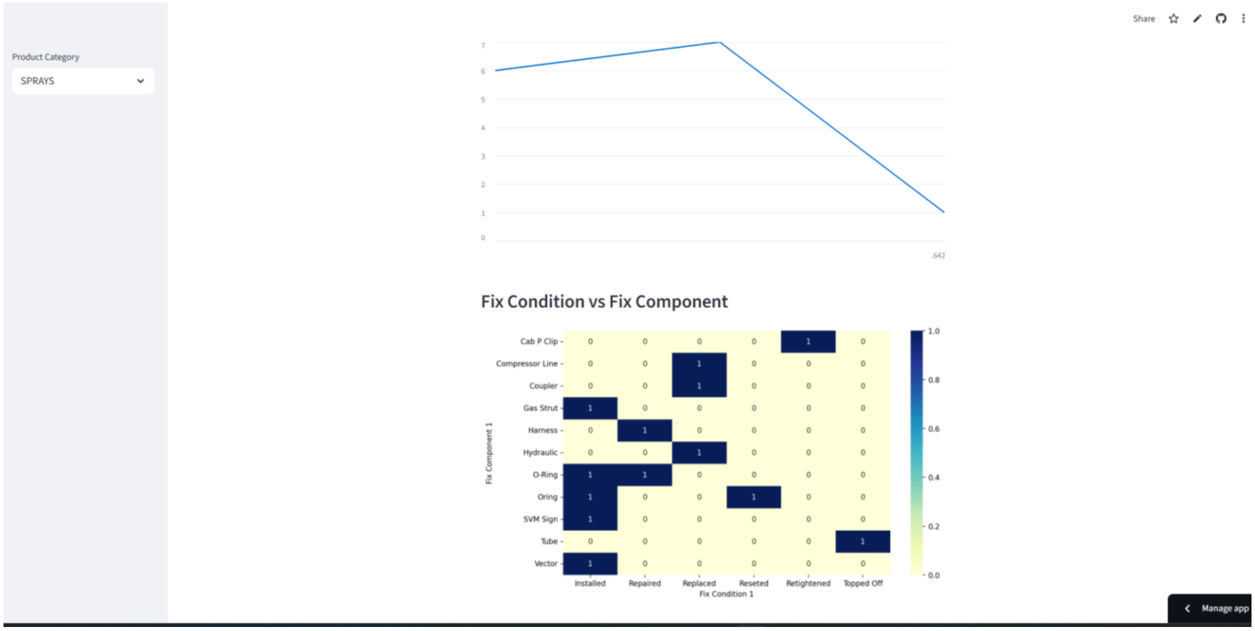
Product Category

SPRAYS

Complaint Analysis Dashboard

Top Root Causes

Top Symptom Components



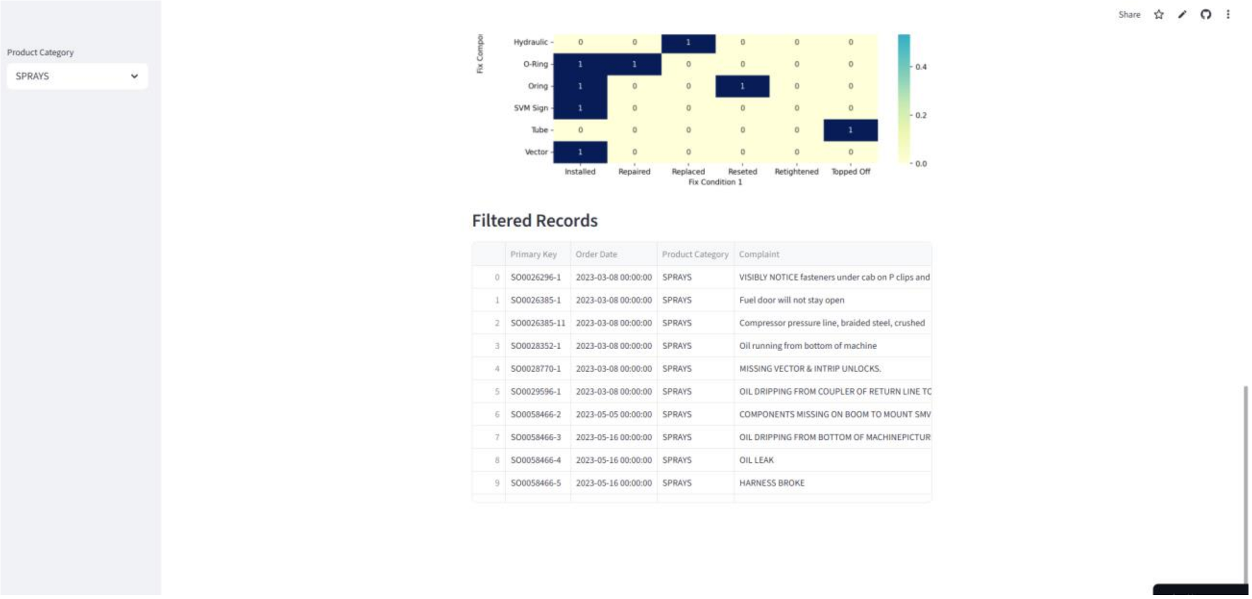
Product Category

SPRAYS

Fix Condition vs Fix Component

Fix Component 1	Installed	Repaired	Replaced	Reseted	Retightened	Topped Off
Cab P Clip	0	0	0	0	1	0
Compressor Line	0	0	1	0	0	0
Coupler	0	0	1	0	0	0
Gas Strut	1	0	0	0	0	0
Harness	0	1	0	0	0	0
Hydraulic	0	0	1	0	0	0
O-Ring	1	1	0	0	0	0
Oring	1	0	0	1	0	0
SVM Sign	1	0	0	0	0	0
Tube	0	0	0	0	0	1
Vector	1	0	0	0	0	0

Filtered records based upon product Category:



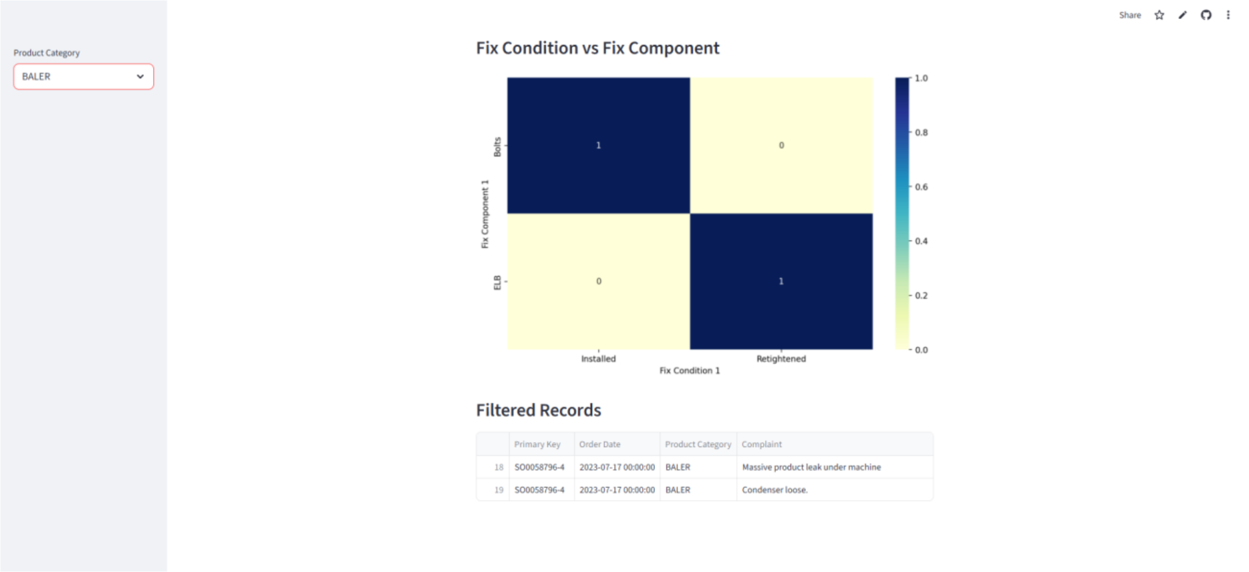
Product Category

SPRAYS

	Installed	Repaired	Replaced	Reseted	Retightened	Topped Off
Hydraulic	0	0	1	0	0	0
O-Ring	1	1	0	0	0	0
Oring	1	0	0	1	0	0
SVM Sign	1	0	0	0	0	0
Tube	0	0	0	0	0	1
Vector	1	0	0	0	0	0

Filtered Records

	Primary Key	Order Date	Product Category	Complaint
0	SO0026296-1	2023-03-08 00:00:00	SPRAYS	VISIBLY NOTICE fasteners under cab on P clips and
1	SO0026385-1	2023-03-08 00:00:00	SPRAYS	Fuel door will not stay open
2	SO0026385-11	2023-03-08 00:00:00	SPRAYS	Compressor pressure line, braided steel, crushed
3	SO0028352-1	2023-03-08 00:00:00	SPRAYS	Oil running from bottom of machine
4	SO0028770-1	2023-03-08 00:00:00	SPRAYS	MISSING VECTOR & INTRIP UNLOCKS.
5	SO0029596-1	2023-03-08 00:00:00	SPRAYS	OIL DRIPPING FROM COUPLER OF RETURN LINE TC
6	SO0058466-2	2023-05-05 00:00:00	SPRAYS	COMPONENTS MISSING ON BOOM TO MOUNT SMV
7	SO0058466-3	2023-05-16 00:00:00	SPRAYS	OIL DRIPPING FROM BOTTOM OF MACHINEPICTUR
8	SO0058466-4	2023-05-16 00:00:00	SPRAYS	OIL LEAK
9	SO0058466-5	2023-05-16 00:00:00	SPRAYS	HARNESS BROKE



Product Category

BALER

Fix Condition vs Fix Component

	Installed	Retightened
Bolts	1	0
ELB	0	1

Filtered Records

	Primary Key	Order Date	Product Category	Complaint
18	SO0058796-4	2023-07-17 00:00:00	BALER	Massive product leak under machine
19	SO0058796-4	2023-07-17 00:00:00	BALER	Condenser loose.