Failure Description Tagging and Analysis Report

1. Column Analysis

The dataset contains two columns:

- FailureDescription: Describes the issue or repair activity in free text.

- Tags: Extracted keywords representing actions such as 'replace', 'inspect', 'adjust'.

The 'Tags' column is generated programmatically from FailureDescription using a rule-based keyword matching and NLP.

2. Data Cleaning Summary

- Null values in 'FailureDescription' were removed as they do not contribute to tagging.

- Text was standardized to lowercase for uniformity.

- Duplicate entries were checked and removed if any.

- Applied regular expressions and spaCy for named entity recognition and verb extraction.

- Mapped verbs like 'replaced', 'replacing' to base verb 'replace' to ensure consistent tagging.

3. Visualizations

We created:

- A word cloud showing frequent words in the FailureDescription.

- A bar chart displaying frequency of generated tags.

These visualizations helped identify dominant maintenance actions such as 'replace', 'inspect', and 'adjust'.

4. Generated Tags & Key Takeaways

- Most frequent tags include: 'replace', 'inspect', 'found', 'adjust'.

Failure Description Tagging and Analysis Report

- Actions related to the steering wheel, control arm, and suspension system appeared often.

- High frequency of 'replace' indicates major maintenance trends.

Recommendations:

- Investigate recurring issues tied to frequent replacements.

- Proactively stock spare parts for components frequently tagged.

- Use tagging to automate future maintenance categorization.

Discrepancies addressed:

- Null or missing descriptions were filtered.

- Columns were typecast properly before transformation.