**Detailed Report**

**Column Analysis**

CUSTOMER\_VERBATIM: Free description of vehicle components by customer.   
  
CLEANED\_VERBATIM: Preprocessed text of customer invoice.   
  
ENTITIES: Named entities extracted from cleaned text.   
  
TAGS: A combination of keywords and entities that summarize error conditions and components.   
  
**Data Cleansing Overview**

Preprocessing: Text is converted to lower case, punctuation is removed, text is tokenized, and stop words are removed. Keyword extraction: TF-IDF was used to identify important keywords.   
  
Entity extraction: SpaCy was used to extract named entities.   
  
Tag generation: Keywords and entities were combined to generate meaningful tags.   
  
**Visualization**   
  
Visualizations can help provide insight into the data and understand the distribution and frequency of tags.  
Visualization provides insight into the distribution and frequency of tags.   
  
**Generated tags and key insights**

Generated tags: Add terms such as "steering wheel", "heated steering wheel", "stitching", and "cruise bar". Key Findings:   
  
Common Failure Conditions: Common issues include "loose steering wheel", "steering wheel heater not working", and "loose stitching".   
  
Component Specific Issues: Most issues are steering wheel related.   
  
Customer Concerns: Customers report both functional and aesthetic issues.   
 **Actionable Recommendations**

**Quality Control:** We implement more stringent quality control measures for steering wheel parts.  
We implement stricter quality control measures on steering wheel parts, focusing on common failure points such as seams and heating functions. Regular inspection and testing helps us identify and fix problems before the product reaches the customer**.**   
  
**Customer Communication**: Proactively address common issues with steering wheels.

**Technician Training:** Provide additional training to diagnose and repair steering wheel issues. Product Improvement: Revised design and materials of steering wheel parts to improve durability.

**Discrepancies and Approach**

**Null Values**

Null values in the CUSTOMER\_VERBATIM column were identified and handled to ensure data quality. Rows with null values were removed to prevent incomplete or inaccurate analysis.

**Approach**: The dropna method was used to remove rows with null values in the CUSTOMER\_VERBATIM column.

**Missing Primary Keys**

Each record in the dataset was assigned a unique identifier to maintain data integrity and avoid data duplication. Ensuring unique identifiers for each record is crucial for accurate analysis and reporting.

**Approach**: A unique identification column was added to the DataFrame using the index value. Improve customer communication and proactively address common steering wheel component issues. Providing clear instructions for reporting issues and timely support can improve customer satisfaction and loyalty.

### **Inconsistent Data** Inconsistent data entries were identified and standardized using string manipulation techniques. This step ensures consistency and accuracy of the analysis. Approach: Data entries were standardized using regular expressions and string manipulation methods. B. Convert text to lower case and remove special characters.