❖ Data Task 2 Report: Steering Wheel Repair Dataset Analysis

A. Column Analysis

The dataset contains over 60 columns spanning vehicle identifiers, repair details, customer complaints, dealer information, and diagnostic outcomes. Key column types include:

Identifiers: VIN, TRANSACTION_ID, REPAIRING_DEALER_CODE
Repair Details: REPAIR_DATE, GLOBAL_LABOR_CODE_DESCRIPTION, TOTALCOST, LBRCOST
Customer Feedback: CUSTOMER_VERBATIM, CORRECTION_VERBATIM
Vehicle Specs: ENGINE_DESC, TRANSMISSION_DESC, BODY_STYLE
Dealer Info: DEALER_NAME, REPAIR_DLR_CITY, STATE

Every column was evaluated for:
 Data type (datetime, float, or object)
Uniqueness (e.g., labor codes are repeated, VINs are unique)
missing values (more than 50% of some columns were null).
Relevance to business (for example, TOTALCOST and CUSTOMER_VERBATIM are essential for sentiment and cost analysis)

B. Data Cleaning Summary

Missing Values: Columns that were missing more than 50% were removed. The mode (categorical) and median (numerical) were used to impute the remaining nulls.
Text Standardization: In order to address discrepancies, categorical fields were capitalized and cleared of whitespace.
Date Parsing: Month was obtained by converting REPAIR_DATE to datetime format.
Outlier Removal: To eliminate extreme values, IQR filtering was applied to TOTALCOST and LBRCOST.

Free Text Processing: CORRECTION and CUSTOMER_VERBATIMTokenizing VERBATIM allowed for the extraction of tags.

C. Visualizations

1. Distribution of Repair Costs: The majority of repairs cost less than ₹500, but a small number cost more than ₹1000, according to a TOTALCOST histogram that has a right-skewed distribution.

2. Most Popular Repair Types:"STEERING WHEEL REPLACEMENT" is the most common repair type, according to the bar plot of GLOBAL_LABOR_CODE_DESCRIPTION.

3. The Monthly Repair Volume: Time-series bar chart indicates that January–February 2024 saw the highest repair activity, which may indicate batch-related or seasonal problems.

D. Generated Tags & Key Takeaways

Tags Extracted

From free-text fields, we generated tags like:

- **Failure Tags**: HEATED, INOP, BUTTON, STITCHING, PEELING
- **Component Tags**: STEERING, MODULE, TRIM, COVER
- **Action Tags**: REPLACED, CHECKED, VERIFIED, ORDERED
- **Bonus Tags**: PRA, PREAUTH, GM AUTHORIZATION, VIDEO, TRANSIT DAMAGE, REPROGRAMMED

These tags were added to each record for quick filtering and summarization.

* Bonus Insights & Recommendations

* Unique & Meaningful Tags

1. Procedural Tags

   REPLACED, INSTALLED, TORQUED, CLEARED, RETESTED, VERIFIED, DIAGNOSIS, CHECKED, REMOVE, ORDERED

These indicate technician workflows and repair validation steps, useful for process optimization.

2.Warranty & Authorization Tags

   PRA, PREAUTH, GM AUTHORIZATION, SOP, RO, CLAIM APPROVED

These reflect warranty approval processes and service order tracking, critical for compliance and cost recovery.

3.Customer Behavior Tags

VIDEO, ADVISOR, INSPECT, REPORT, CONFIRMED, REQUESTED

* Insights

    1.The most common problem is heated steering wheel failures.

    2.All versions have common material flaws like peeling and stitching.

    3.More than 60% of records have warranty tags (PRA, PREAUTH).

* Recommendations

1.Enhance Component Quality: Pay attention to stitching durability and hot modules.

  2.Automate Warranty Workflows: To expedite approvals, use PRA tags.

  3.Standardize Dealer Repairs: Use procedural tags to reinforce SOPs.