

Task 2 - Data Analysis & Insights Report

1. Column Analysis

The dataset contained **52 columns**, covering various aspects such as transaction details, customer complaints, repair actions, costs, and vehicle specifications. Key observations:

- **Data Types:** Mix of categorical, numerical, text, and date fields.
- **Missing Values:** Some columns, like `CAMPAIGN_NBR`, had 100% missing values and were dropped, while others like `ENGINE_TRACE_NBR` had partial missing data.
- **Unique Identifiers:** `VIN` and `TRANSACTION_ID` ensured record uniqueness.
- **Significant Columns for Analysis:** `CUSTOMER_VERBATIM`, `CORRECTION_VERBATIM`, `TOTALCOST`, `PLATFORM`, and `REPAIR_DATE` were selected for deeper insights.

2. Data Cleaning Summary

- **Dropped Columns:** Removed columns with excessive missing values (e.g., `CAMPAIGN_NBR`).
- **Filled Missing Values:**
 - Categorical values were replaced with the most frequent category.
 - Numerical values (e.g., `TOTALCOST`) were imputed using the median.
- **Standardization:** Converted text fields (e.g., `PLATFORM`, `BODY_STYLE`) to uppercase.
- **Outlier Handling:** Applied **IQR (Interquartile Range)** to remove extreme values in numerical fields.
- **Duplicate Removal:** Ensured no redundant records.

3. Key Visualizations & Insights

- **Repair Cost Distribution:** Most repairs cost between **\$200-\$500**, but some exceed **\$3,000**, suggesting high-cost repair cases.
- **Most Repaired Vehicle Platforms:** **Full-Size Trucks** had the highest number of repairs, indicating potential reliability issues.
- **Repair Trends Over Time:** Spikes in repair costs suggest **seasonal effects or recall-related repairs**.

4. Generated Tags from Free-Text Data

From the free-text fields (`CUSTOMER_VERBATIM`, `CORRECTION_VERBATIM`, `CAUSAL_PART_NM`), we generated structured tags under two broad categories:

- **Failure Conditions** — e.g., Inoperative, Heating Failure, Physical Damage, Surface Defect, Noise Issue, Loose Fitment, Material Wear, Electrical Failure.
- **Affected Components** — e.g., Steering Wheel, Heated Steering Wheel Module, Wiring Harness, Spoke Cover / Trim, Control Module, Airbag Assembly.

The tagging process used keyword mapping and contextual matching to ensure that symptoms, causes, and fixes were logically aligned. For example, “heated steering wheel inop” was tagged as *Heating Failure* (condition) and *Heated Steering Wheel Module* (component).

5. Actionable Recommendations

- **Engineering & Design** — Review supplier quality for heated steering wheel modules; consider redesigning connectors to reduce intermittent faults.
- **Materials & Manufacturing** — Upgrade leather/adhesive specifications to reduce peeling and fraying under heat and wear.
- **Service Process** — Provide technicians with targeted diagnostic flowcharts for heating failures to reduce unnecessary full-wheel replacements.
- **Customer Communication** — Proactively inform customers about known cosmetic issues and offer early interventions to improve satisfaction.

6. Discrepancies & Approach

- **Null / Missing Values** — Some records had blank CUSTOMER_VERBATIM or CORRECTION_VERBATIM fields; these were tagged as “Unspecified” to maintain dataset integrity.
- **Inconsistent Component Naming** — Variations like “WHEEL ASM-STRG” vs. “Steering Wheel” were normalized via a synonym mapping table.
- **Potential Missing Primary Keys** — While VIN appears unique, some transactions share the same VIN with different TRANSACTION_IDs; uniqueness was validated using a composite key (VIN + TRANSACTION_ID).

Conclusion

This analysis provides valuable insights into repair trends, cost patterns, and failure conditions. These findings can help optimize maintenance, reduce costs, and enhance vehicle reliability. Key takeaways include:

- **Frequent high-cost repairs** should be investigated further to understand root causes and mitigate unnecessary expenses.
- **Full-Size Trucks require additional quality control measures** due to their high repair rates.
- **Automated issue tagging** can streamline fault detection and improve predictive maintenance.