Task 2 – Data Analysis Summary

Dataset Overview

The dataset provides detailed information on vehicle repairs, customer-reported issues, and associated service costs. It consists of multiple attributes covering technical details, dealer/service information, and textual complaints. This dataset was analyzed to detect recurring issues, identify cost drivers, and highlight reliability trends that can inform quality control and customer support strategies.

1. Data Preparation & Cleaning

Before analysis, several steps were taken to ensure data accuracy and usability:

- Missing Values: Certain columns (e.g., CAUSAL_PART_NM, CUSTOMER_VERBATIM, CORRECTION_VERBATIM) contained gaps. These were treated using placeholders or imputation methods to preserve consistency.
- Text Normalization: Complaint and correction narratives were standardized (case formatting, trimming spaces) to improve keyword extraction and tag generation.
- Handling Outliers: Abnormal values in cost-related fields (e.g., extremely high labor costs) were flagged. Unrealistic values were either replaced with nulls or excluded from trend calculations.
- Duplicate Records: A small percentage of repeated transaction IDs were identified, indicating a need for stronger validation in data collection systems.

2. Insights & Observations

Frequent Failure Patterns

- Steering Wheel & Electronics: A large share of complaints were linked to steering wheel replacement, cosmetic issues (stitching, fading), and heated steering malfunctions.
- Driver Assistance Features: Alerts and failures tied to electronic driver-assist modules were reported, showing possible gaps in software or sensor calibration.

Trends by Vehicle Characteristics

- Body Style Impact: Crew Cab trucks and 4-Door Utility vehicles reported the highest complaint volumes.
- Vehicle Age & Usage: Older vehicles and higher-mileage categories showed higher repair costs and greater likelihood of electronic part failures.

Cost Structure

- Average repair costs were heavily skewed toward labor expenses (~55–65%), suggesting that technician involvement is a major driver of costs.
- Commonly replaced components included steering assemblies, sensors, wiring harnesses, and cosmetic parts.

Data Quality Concerns

- ~10–15% of complaint narratives were incomplete, reducing clarity for root cause analysis.
- A few categorical fields (PLANT, STATE, CERTIFICATION) contained missing or inconsistent values.

3. Recommendations

1. Quality Improvements:
   - Investigate frequent steering-related and cosmetic failures to strengthen supplier quality checks.
   - Enhance durability testing of driver-assist and electronic systems.

2. Technician & Dealer Support:
   - Provide refresher training to reduce assembly-related errors and improve diagnostic accuracy.
   - Standardize repair documentation to minimize incomplete customer/technician inputs.

3. Data Governance:
   - Introduce stricter validation for transaction IDs to avoid duplication.
   - Improve complaint data entry systems with mandatory fields for key details.

4. Targeted Maintenance:
   - Offer special service recommendations for aging/high-mileage vehicles to reduce long-term failures.
   - Use predictive analytics for seasonal or regional issue monitoring.

Conclusion

This analysis highlights recurrent steering and electronic issues as the primary customer pain points, with labor costs forming the largest share of repair expenses. Strengthening quality assurance, technician training, and data collection practices will help lower costs and improve customer satisfaction. Longer-term, integrating predictive monitoring and refining maintenance

strategies for specific vehicle categories can provide a proactive edge in managing failures and optimizing reliability.