

Detailed Report

1.Column Analysis

Key Findings (Summary):

- Data Types: Mix of strings, numerics, dates, and integers.
- Unique Text Fields: 'VIN', 'CORRECTION_VERBATIM', and 'CUSTOMER_VERBATIM' show high uniqueness, indicating detailed free-text inputs.
- Numerical Insights: Columns like 'REPAIR_AGE', 'KM', and cost metrics ('REPORTING_COST', 'TOTALCOST', 'LBRCOST') reflect vehicle age, mileage, and financials.
- Categorical Insights: Key groupings by parts ('CAUSAL_PART_NM'), platform, region ('BUILD_COUNTRY', 'STATE'), and transaction types.
- Identifiers: Multiple columns track vehicles, transactions, dealers, engines, and transmissions for detailed traceability.
- Temporal Analysis: Date fields enable trend and seasonal pattern analysis.
- Cost Metrics: Financial columns help assess warranty claims and repair costs.

2.Data cleaning summary

The dataset underwent essential cleaning to ensure consistency, accuracy, and readiness for analysis. Key steps included:

1. Handling Missing Values:

- Dropped rows with fewer than 3 non-null values.
- Removed the completely empty column 'CAMPAIGN_NBR'.
- Filled missing values in categorical columns with 'Unknown'.
- No missing values remained in numerical columns; however, mean imputation and IQR-based capping were prepared as fallback strategies.

2. Resolving Categorical Inconsistencies:

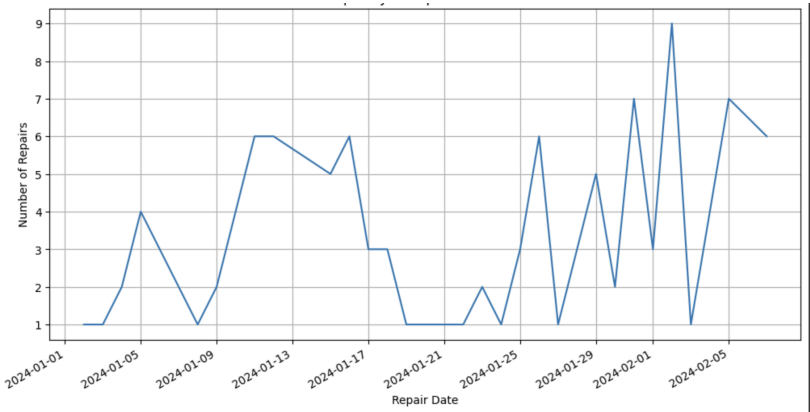
- Standardized text formatting in object columns by applying `.str.strip().str.title()` to fix capitalization and spacing inconsistencies.

3. Validating Numerical Columns & Outlier Treatment:

- Converted all numerical fields to float64 for consistency.
- Applied Interquartile Range (IQR) method to cap extreme outliers.

After cleaning, the dataset was well-structured and suitable for in-depth analysis and visualization.

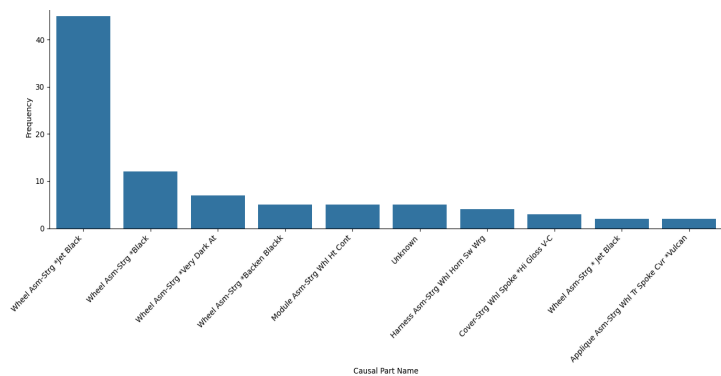
3. Data Visualization - 1.Number of Repairs over Time:



Key Insights:

1. Repair Spike: Repairs peaked on Feb 2, 2024 with 9 cases, indicating a surge in activity.
2. Mid-January Dip: A noticeable drop in repairs occurred between Jan 17–24, possibly due to operational slowdowns.
3. High Variability: Daily repair counts fluctuated significantly, suggesting inconsistent repair patterns.

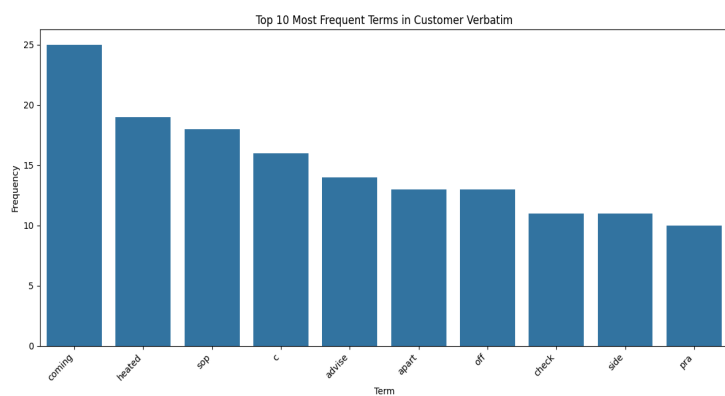
2. Top 10 Most Frequent Causal Parts:



Key Insights:

1. Frequent Issue: Wheel Asm-Strg *Jet Black* is the top failure point with highest frequency.
2. Color Variants: Several steering wheel parts with color labels show repeated issues.
3. Inconsistencies: Typos and "Unknown" entries suggest a need for better data standardization.

3. Top 10 Most Frequent Terms in Customer Verbatim chart:



Key Insights:

1. Frequent Complaints: Terms like "coming" and "heated" suggest recurring issues with heated components or features not activating.
2. Escalation Language: Words like "advise" and "check" indicate customers are actively seeking service resolution or clarification.
3. Data Noise: Terms like "c" and "sop" may be artifacts or abbreviations, suggesting the need for text preprocessing or abbreviation mapping.

4. Generated tags :Component-Specific

- Steering Wheel
- Wheel Assembly
- Heated Steering
- Air Bag
- Module
- Light.

5. Key Takeaways from Data Analysis:

- Manufacturing/Engineering: Investigate and improve the quality of 'Wheel Asm-Strg *Jet Black' and 'Wheel Asm-Strg *Black', focusing on "coming apart" and "peeling" issues identified in text data.
- Customer Service: Train staff on common customer complaint terms like 'coming apart,' 'heated steering inop,' and 'loose' for better communication and diagnosis.
- Warranty/Finance: Analyze costs for repairs with 'TOTALCOST' above ~\$570 to identify key cost drivers and reduction opportunities.
- Quality Control: Investigate VINs 3Hrs9Eed0Lh255650 and 1Hrffhel8Rz133325 with multiple repairs to understand recurring issues and repair effectiveness.
- Inventory Management: Ensure adequate stock of 'Wheel Asm-Strg *Jet Black', 'Wheel Asm-Strg *Black', and 'Module Asm-Strg Whl Ht Cont' based on their high frequency as causal parts.