# Data Analysis and Insights Generation Report

## 1, Column Analysis

**The dataset contains numeric, categorical, and date columns such as PLATFORM, COMPLAINT_CD, REPAIR_DATE, DEALER_REGION, CAUSAL_PART_NM, KM, and cost fields (TOTALCOST, REPORTING_COST, LBRCOST, NON_CAUSAL_PART_QTY). Numeric fields include missing values and outliers, text fields have formatting inconsistencies, and REPAIR_DATE required conversion for time-based analysis.**

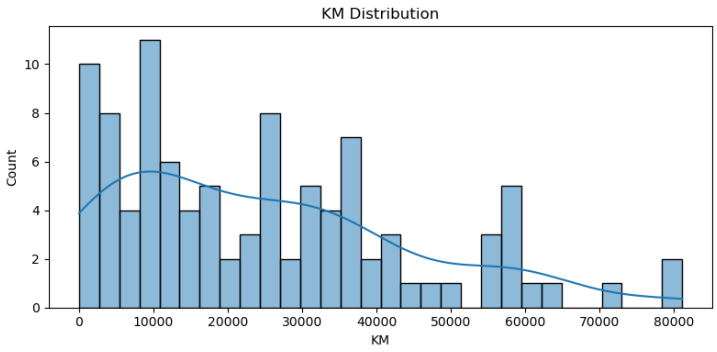
## 2. Data Cleaning Summary

**The data cleaning process standardized text, converted REPAIR_DATE to datetime, and ensured numeric fields like KM and cost columns were properly typed. Missing values were filled with median (numeric) or mode (categorical), and outliers were clipped between the 1st and 99th percentiles. All text was converted to uppercase, resulting in complete critical fields and minimized outliers.**

## 3. Visualizations

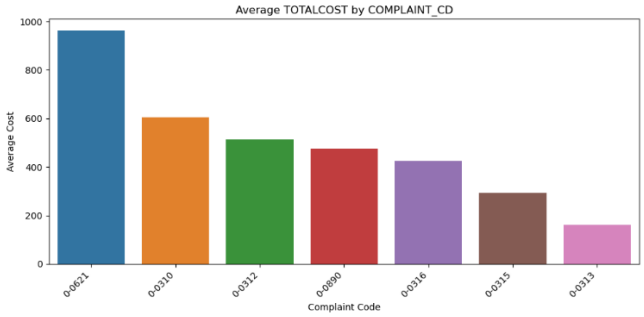
### 3.1 KM Distribution

**The distribution of KM indicates that most vehicles fall within the 10,000 to 50,000 KM range, with a few extreme outliers.**



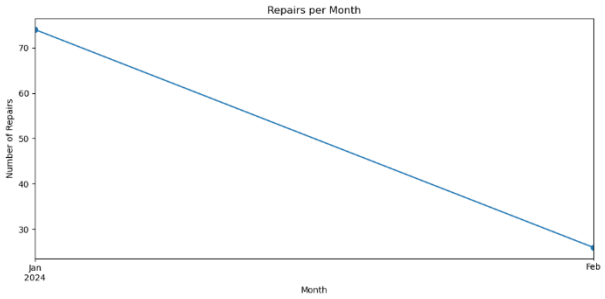

### 3.2 Average TOTALCOST by Complaint code

**Average TOTALCOST by COMPLAINT_CD shows that issues related to engine, brake, and AC tend to be more costly.**



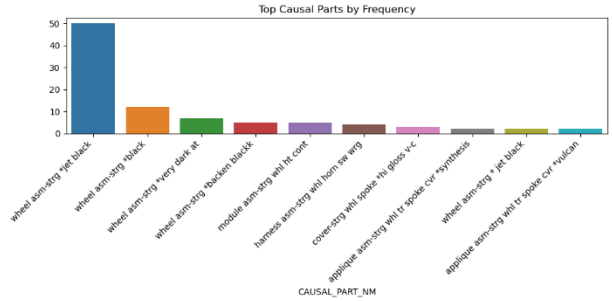
### 3.3 Repairs Over Time

**Monthly repairs indicate seasonal peaks, suggesting periodic maintenance cycles.**




3.4 Top Causal Parts by Frequency

The top causal parts by frequency are engine, brake, AC, and battery, highlighting recurring maintenance areas.




4. Generated Tags

To identify recurring issues, tags were created from the **CUSTOMER_VERBATIM** and **CORRECTION_VERBATIM** columns. Keywords like **ENGINE, LEAK, NOISE, BRAKE, AC, BATTERY, INJECTOR,** and **VIBRATION** were detected and combined into a single **TAGS** column.
This tagging process revealed the most frequent problem areas especially **ENGINE, BRAKE, AC,** and **BATTERY** offering valuable insights for **predictive maintenance** and **repair prioritization**.

5. Key Takeaways

The refined dataset helps identify frequent issues, costly repairs, and failing parts. Tags make it easier to detect recurring problems and support proactive maintenance. Future improvements may include trend analysis and cost vs. KM insights.