Task 2 – Data Analysis and Insights Generation using Python

Dataset Summary

- File name: DA - Task 2..xlsx
- Total records (rows): 100
- Total columns: 52

Data types:

- Object (Text): 33
- Integer: 12
- Float: 6
- Date: 1

The dataset contains vehicle repair details, including information about parts replaced, repair costs, customer complaints, and service locations. It helps in understanding which components fail often, cost patterns, and customer concerns.

Column-Wise Analysis (Summary)

Column Name	DataType	ExampleValue	Importance
Vin	Object	MA1TA2	Vehicle unique ID (Primary Key)
Transaction_id	Int	12345	Unique service transaction number
Repair_date	Date	2024-07-14	Date when repair happened
Causal_part_nm	Object	Brake Pad	Main part responsible for the failure
Totalcost	Float	5400.50	Total cost of repair (labor + parts)
Platform	Object	Scorpio-N	Vehicle platform/model
Repair_age	Int	36	Vehicle age in months at time of repair
Customer_verbatim	Object	"Brake noise when driving"	Customer complaint text
Correction_verbatim	Object	"Replaced brake pads"	Technician's correction details
Engine_desc / transmission_desc	Object	Diesel 2.2L	Engine/Transmission info

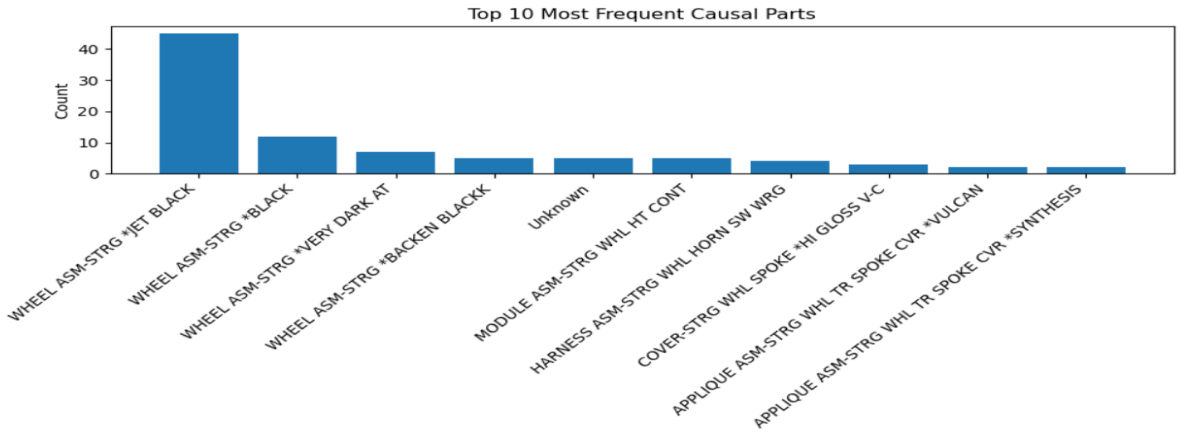
Data Cleaning Summary

Step	Action Taken	Result
Remove empty columns	Dropped CAMPAIGN_NBR (all null)	Removed
Handle missing values	Filled missing text with "Unknown"; numeric with median	All nulls fixed
Standardize text	Converted to title case, removed spaces	Clean
Fix data types	Converted date and numeric fields properly	Correct formats
Remove duplicates	Checked & removed if any	Done
Outlier handling	Removed top 1% extreme values in cost columns	Clean numeric data

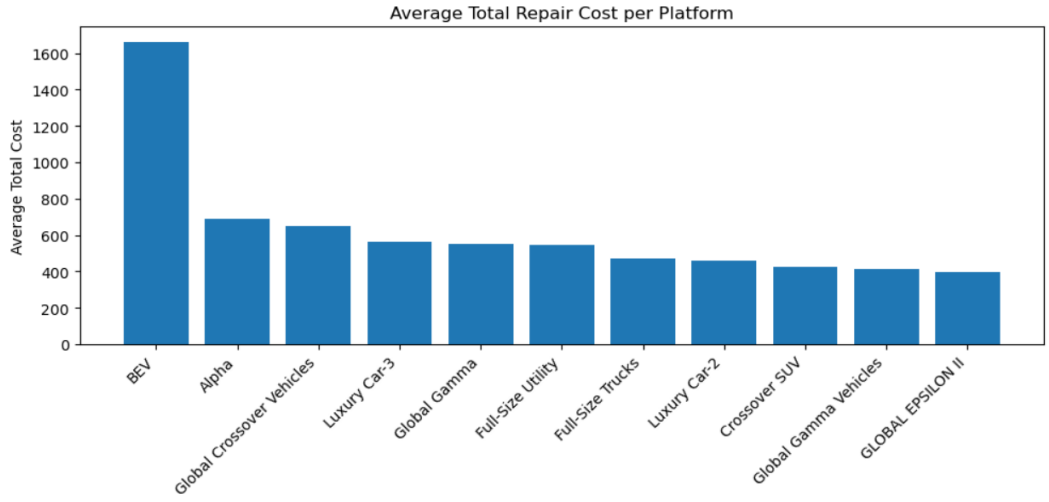
Top 5 Critical Columns

Column	Important because it
Causal_part_nm	Help identify which part fails most often
Platform	Show which vehicle model faces more issues
Totalcost	Help control warranty and service costs
Repair_age	Show when failures usually happen
Customer_verbatim	Gives direct customer feedback for issue trends

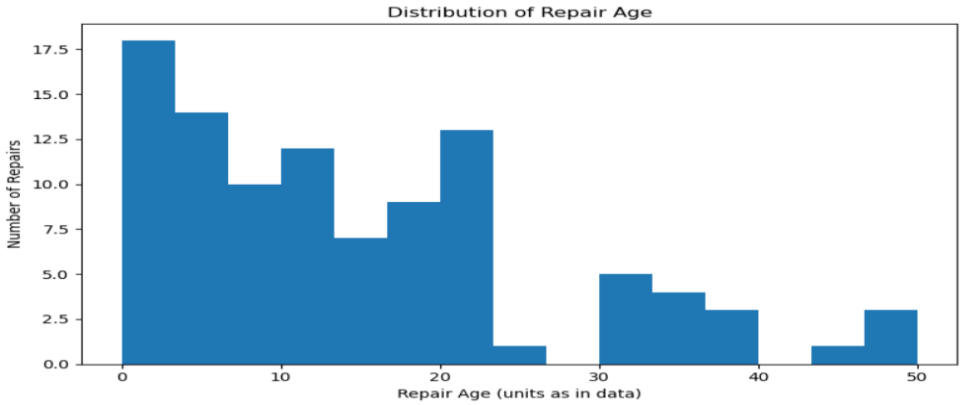
Visualizations: Top 10 Causal Parts – Bar chart showing which components fail most often



Average Repair Cost by Platform – Bar chart comparing average cost per vehicle model



Distribution of Repair Age – Histogram showing at what age most repairs happen



Findings from Visualizations : Some parts (like brakes or engine mounts) appear frequently, Certain platforms show higher repair costs and Most issues occur around 2–3 years after purchase.

Tag Generation : Two columns **-** CUSTOMER_VERBATIM and CORRECTION_VERBATIM are analyzed to create summary tags.
Tagging Methodology: Used simple keyword search in text (like engine, brake, noise, leak)
Created new columns: SYMPTOM_TAG (issue type, e.g., Noise, Leakage) and COMPONENT_TAG (part name, e.g. Engine, "Brake)

Summary & Insights
Main Insights

1. Most repairs are related to brake and engine parts.
2. Repair cost increases with vehicle age.
3. Some platforms have higher-than-average repair costs, indicating possible design or quality issues.
4. Customer complaints often include noise, vibration, or leakage — indicating recurring failure types.
5. Majority of repairs occur after 24–36 months, suggesting post-warranty concerns

Recommendations

- Focus on high-failure parts for design improvement.
- Use repair age data to plan preventive maintenance programs.
- Address costly repairs through supplier quality checks.
- Regularly monitor customer text feedback to identify new issues early.