**Axion Ray Task 2 :** Data Analysis and Insights Generation using Python

#### **Column Analysis**

The dataset provides comprehensive insights into vehicle production, costs, and reported issues. Here is a summary of the key columns:

1. **Platform**: Identifies the production platform used for manufacturing vehicles. This column is crucial for understanding the allocation of production resources and optimizing platform usage. Over-reliance on specific platforms can lead to inefficiencies.
2. **Body Style**: Describes the type of vehicle (e.g., SUV, Sedan, Truck). This helps analyze cost efficiency across different vehicle types and identify high-cost segments.
3. **Total Cost** and **Reporting Cost**: These columns provide financial insights into manufacturing and reported issue costs. Identifying high-cost areas helps improve cost management.
4. **State**: Indicates the geographic location of manufacturing or issue reporting. This column is useful for pinpointing regional disparities in costs or failure reports.
5. **Free Text Feedback**: Captures detailed descriptions of reported issues. This column is valuable for identifying recurring problems and generating actionable tags.

Each column contributes to a holistic view of manufacturing processes, providing valuable information for stakeholders to make data-driven decisions.

#### **Data Cleaning Summary**

The dataset underwent several preprocessing steps to ensure it was ready for analysis:

1. **Missing or Invalid Values**: The dataset did not contain significant missing data, but ambiguities in the **Free Text Feedback** column were addressed. Terms with typos or unclear meanings, such as "pra," were standardized for consistency.
2. **Categorical Columns**: Inconsistencies in formatting, such as capitalization variations (e.g., "SUV" vs. "suv"), were corrected to ensure uniformity.
3. **Numerical Columns**: Columns such as **Total Cost** and **Reporting Cost** were checked for errors or extreme outliers. Outliers were retained where relevant, as they often indicated inefficiencies or noteworthy instances.

Through these cleaning steps, the dataset was standardized and prepared for accurate analysis, improving its overall reliability.

#### **Visualizations and Key Insights**

1. **Platform Distribution**:  
   A bar chart revealed that certain platforms dominate vehicle production. This indicates an imbalance in resource allocation, with overutilized platforms facing potential strain. A balanced approach across platforms can enhance efficiency and reduce risks.
2. **Cost by Body Style**:  
   A bar chart comparing **Total Costs** across body styles highlighted that **SUVs** are the most expensive to produce. Sedans, on the other hand, are relatively cost-efficient. Stakeholders should focus on optimizing SUV production processes to lower costs.
3. **Reporting Cost vs. Total Cost**:  
   A scatter plot of **Reporting Cost** against **Total Cost** displayed a strong positive correlation. Outliers in the data indicate instances where reporting costs are disproportionately high compared to total costs. Investigating these cases could lead to streamlining reporting processes and cutting unnecessary expenses.
4. **State-Wise Analysis**:  
   A map visualization showed significant cost disparities across different states. States with higher costs or frequent issues may require additional quality control measures or improved manufacturing strategies to bring costs in line with benchmarks.

#### **Generated Tags and Insights**

From the **Free Text Feedback** column, several meaningful tags were generated, including:

* **‘steering’, ‘heated’, ‘replace’, ‘removed’, ‘coming off’, and ‘wheel’**.

**Key Insights**:

1. **Mechanical Failures**: Tags such as **‘steering’** and **‘heated’** highlight recurring issues with steering mechanisms and overheating components.
2. **Replacement Trends**: High frequencies of tags like **‘replace’** and **‘removed’** indicate recurring part replacements, suggesting issues with component durability.
3. **Assembly Issues**: Tags like **‘coming off’** suggest problems with parts detaching during use, indicating potential flaws in assembly or design.
4. **Customer Feedback Patterns**: Terms like **‘customer states’** reflect recurring language used in feedback, emphasizing the need to analyze customer concerns systematically.

#### **Conclusion and Recommendations**

**Insights**:

1. Certain platforms are overused, indicating a need for better resource allocation.
2. SUVs incur the highest production costs, while Sedans are more cost-efficient.
3. States with higher costs highlight regional inefficiencies that require attention.
4. Recurring mechanical failures and replacements point to quality issues in key components.

**Recommendations**:

1. **Platform Optimization**: Balance production workloads across all platforms to enhance resource efficiency.
2. **Cost Management**: Investigate and streamline SUV production processes to reduce overall costs.
3. **Regional Quality Control**: Focus on regions with higher manufacturing costs to address inefficiencies and improve consistency.
4. **Product Quality Improvements**: Address recurring issues with steering and heated components to reduce failures and replacement costs.
5. **Customer Feedback Analysis**: Develop a systematic approach to analyzing feedback, using generated tags to prioritize key areas of improvement.