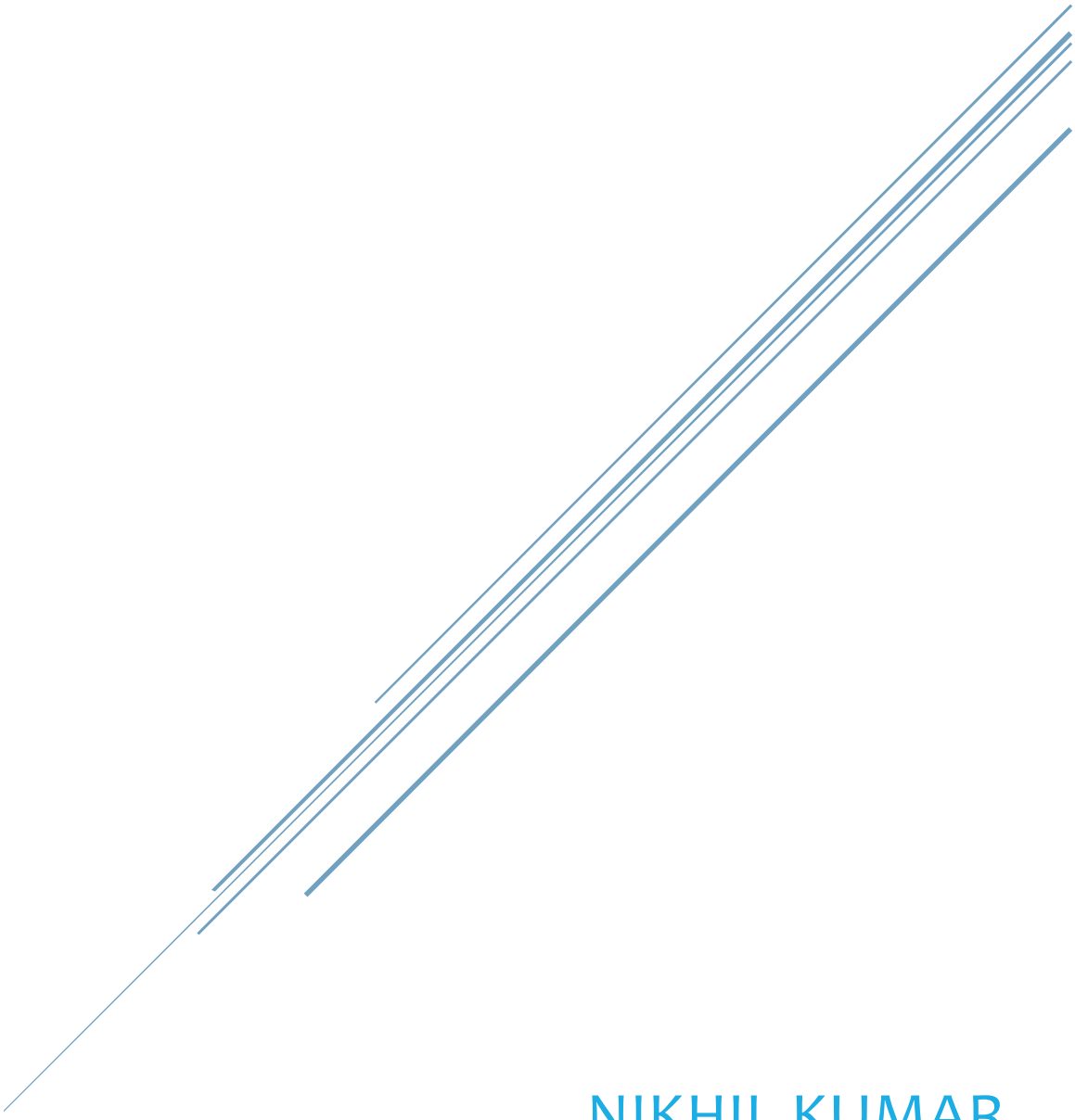
Data Analysis and Insights Generation using Python

AxionRay assignment – Task 2



NIKHIL KUMAR

Nikhild888@gmail.com

Summary Report:

Column analysis:

** Data types, Unique values and distribution of the data has been clearly reported in the **Jupyter notebook** attached. Following points provides the brief overview of the dataset.

1. This data provides information on repairs carried out on different types of vehicles over a period of two months (January and February of 2024) across dealerships. The data seems to be prepared for the purpose of insurance claims or general reporting for vehicles built in three countries: USA, Canada, and Mexico.
2. Most of the repair work is related to the steering wheel assembly.
3. It seems useful to consider only the columns with fewer unique objects, as they help in identifying relationships. Columns such as GLOBAL_LABOR_CODE_DESCRIPTION, BODY_STYLE, BUILD_COUNTRY, etc. provide actionable data.
4. Most of the columns contain unique identifiers, such as vehicle identification numbers, part numbers, region codes, and dealership numbers. These are being ignored, as finding relationships within them seems out of scope.

Data cleaning:

Dropped a few columns based on their relevance to the analysis:

1. **Purely acted as identifiers and did not offer any relationships or patterns**:
   Examples include TRANSACTION_ID, LAST_KNOWN_DLR_NAME, PLANT, LAST_KNOWN_DLR_CITY, REPAIRING_DEALER_CODE, DEALER_NAME, REPAIR_DLR_CITY, REPAIR_DLR_POSTAL_CD, COMPLAINT_CD, ENGINE_SOURCE_PLANT, ENGINE_TRACE_NBR, VIN_MODL_DESGTR, TRANSMISSION_SOURCE_PLANT, TRANSMISSION_TRACE_NBR, SRC_TXN_ID, SRC_VER_NBR.
2. **Similar values represented in different forms**:
   Examples include DEALER_NAME, STATE, ENGINE_DESC, TRANSMISSION_DESC, TRANSACTION_CNTR, SALES_REGION_COD, REPORTING_COST, GLOBAL_LABOR_CODE, COUNTRY_SALE_ISO, PLATFORM.
3. **Lack of clarity or relevance to the analysis and decision-making**:
   Some columns had labels or keywords that were not well-defined or directly applicable to the scope of the analysis, making them less useful. Examples include REPAIR_AGE, VEH_TEST_GRP, ORD_SELLING_SRC_CD, OPTN_FAMLY_CERTIFICATION, OPTF_FAMLY_EMISSIOF_SYSTEM, TRANSACTION_CATEGORY, LAST_KNOWN_DELVRY_TYPE_CD.

After dropping a few columns, the dataset still had discrepancies like null values, inconsistent capitalization, incorrect data types, and outliers in cost-related columns. Nulls needed to be handled, text fields standardized, and data types converted. Outliers in cost values removed using plot and find method to minimize the skewness. All methods used to minimize discrepancies are provided in jupyter notebook.

Top 5 Critical Columns and Reasons for Selection:

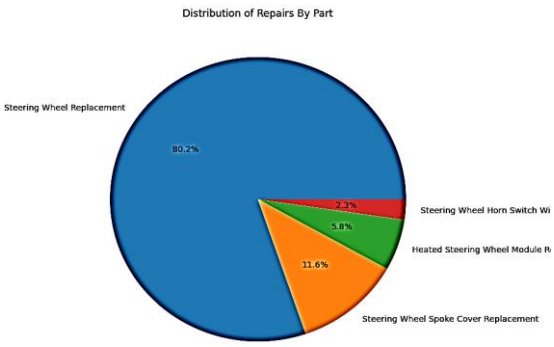
1. **CORRECTION_VERBATIM**: This column is key as it helps identify the root cause and the actions taken. Unlike using VIN numbers or other identifiers, CORRECTION_VERBATIM was chosen because most of its values were unique, allowing us to categorize repairs, tag and identify which categories had the most issues.
2. **BODY_STYLE**: Had to choose between PLATFORM and BODY_TYPE, and BODY_TYPE provided a broader perspective, which was more useful for analysis. It also served as a filter for stakeholders to easily identify associated issues.
3. **GLOBAL_LABOR_CODE_DESCRIPTION**: This column was essential for understanding the frequency of issues in specific parts and replacements. It provided a clear view of the most affected parts, helping to focus on critical areas.
4. **BUILD_COUNTRY**: This helped us identify geographical patterns in reported complaints and understand where vehicles were manufactured. It also offered insights into the quality checks conducted in different regions.
5. **TOTAL_COST**: As a major factor in the analysis, TOTAL_COST helped to calculate the overall cost of repairs and replacements across various BODY_STYLES and parts. It was crucial for comparing costs across regions and assessing part-specific costs.

Insights and Recommendations:

1. Pie chart revealed that **80%** repair activity was associated with **Steering Wheel Replacement**, indicating frequent issues with these components.
    - Focus on improving the **quality and material of steering wheel components** to reduce the frequency of these repairs.
2. Average total cost by labor code shows that certain repairs, such as the **Steering Wheel Horn Switch Wiring Harness Replacement** and **Steering Wheel Replacement** incur higher average costs compared to other repairs. Average labor cost was also high for these repairs indicating that these repairs are time consuming and complex in nature.
    - Need for investigating the Root Cause of high cost of repair and find ways to reduce costs, component quality and reliability needs to improved. Efficient repair methods to minimize the labor cost. As these are both time consuming and costly processes.
3. The **Steering Wheel Horn Switch Wiring Harness Replacement** was also more frequent, with occurrences having an average mileage of less than 5,000 km and has **highest labor cost**.
    - Strict quality checks and improvements in design of product needs to be done as it is essential to mitigate the high-cost repairs and reduced labor costs.
4. Among all the build countries, the **USA** accounted for the majority of complaints, registering approximately **70%** of the total. Suggesting potential areas for targeted quality improvements.
    - Conduct a thorough root cause analysis to identify specific quality issues prevalent in USA-built vehicles. Analyzing manufacturing process, from parts sourcing to final assembly including suppliers and OEMs responsible.
5. Crew cab and 4-door utility vehicles recorded a significantly higher number of repairs (**87%**) compared to other vehicle types.
    - This suggests these models might encounter more frequent issues, potentially linked to their usage patterns. Design modifications could enhance durability and quality.

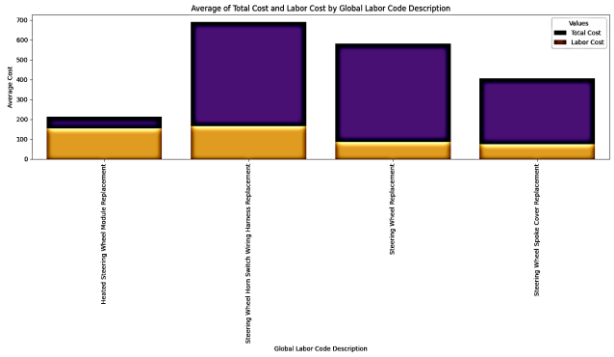
Visualization:

1. Distribution of Repairs By Labor code (Part):



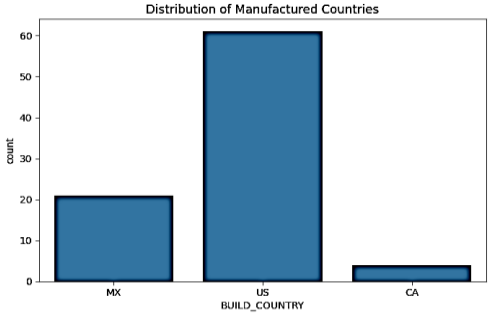
This pie chart shows the distribution of repairs by part. The majority of repairs involve Steering Wheel Replacement (80.2%), followed by Steering Wheel Spoke Cover Replacement (11.6%), Heated Steering Wheel Module Replacement (5.8%), and Steering Wheel Horn Switch Wiring Harness Replacement (2.3%).

2. Average of Total Cost and Labor Cost by Global Labor Code Description:



This bar chart displays the average total cost and labor cost for different types of repairs categorized by global labor code descriptions I,e. Kind of part repaired or replaced. The costs are broken down to show the proportion of labor costs within the total repair costs.

3. Distribution of Complaint with respect to Manufactured Countries:



This bar chart depicts the distribution of manufactured countries for the parts involved in the repairs. The majority of complaints are from vehicles that are manufactured in the US, followed by Mexico (MX) and Canada (CA).

Food for thought:

1. **Dominant Repair Type**: The majority of repairs involved **Steering Wheel Replacement**, accounting for 80.2% of all repairs. This indicates a potential recurring issue or a high wear-and-tear part. It also calls for Root Cause analysis as quality seems to be compromised and an opportunity to investigate and improve the durability of this part.
2. **Highest Costs:** Total cost and labour cost associated with Steering wheel replacement and Steering Wheel Horn Switch Wiring Harness Replacement respectively emphasizes its financial impact and potentially complex repair process.
3. **Build Country and Quality:** Vehicles manufactured in the USA account for nearly 80% of the complaints. Although this may be due to the larger sales in the region, it still highlights the need for stringent quality checks and accountability for all parts.
4. **Failure rate:** Electrical components (Horn switch harness) tend fail at much earlier than mechanical parts in relation with distance travelled indicating high wear rate or other factors. Suggesting the need for enhanced component protection or improved product engineering.

All the above highlights the need to focus on three key areas, implementing cost reduction strategies, improving designs to enhance durability by analyzing counterparts and diversification of part supply, and boosting customer satisfaction through timely maintenance.

----------*----------

** Attached Jupyter notebook (Python scripts) and Excel workbook in mail. Can be accessed from HERE too.