REPORT:

1] DATA IMPORT AND PREPARATION

The dataset was imported from an Excel file and loaded into a DataFrame for analysis.

2] DATA CLEANING AND PREPROCESSING

Handling Missing Values: The dataset contained several columns with null values, and three columns had a few null values.

To maintain data integrity and avoid complexity and redundancy, these columns and the respective records with null values were removed.

Duplicate Check: The dataset was checked for duplicate records, and none were found. But added the code for removing any duplicates found.

4] OUTLIER DETECTION AND TREATMENT

Outlier Detection: Outliers were detected in several variables using box plots.

Outlier Treatment: The outliers were treated using the capping technique, adjusting extreme values to reduce their impact on the analysis.

5] CORRELATION ANALYSIS

Correlation Between Variables: A correlation heatmap was plotted to identify relationships among variables. Significant correlations were found

between certain variables, such as:'Fuel consumption' and 'runtime','Engine torque mode' and 'engine load'

6]VARIABLE ANALYSIS

Univariate Analysis:

Numerical Variables: Histograms were used to visualize the distribution of numerical variables, providing insights into their behavior.

Categorical Variables: Count plots were used to display the frequency of categories within categorical variables, highlighting the distribution of values.

Bivariate Analysis:

Scatter Plots: Scatter plots were utilized to explore relationships between pairs of variables. For example:

The relationship between 'engine load' and 'vehicle speed',

The relationship between 'engine load' and 'RPM'

7] FEATURE ENGINEERING

New Feature Creation: New features were engineered to enhance model performance and improve the interaction among variables. This step is crucial for uncovering hidden patterns and relationships in the data.

8] STANDARDIZATION

Numerical Variables: Standardization was applied to numerical variables to ensure they have a mean of zero and a standard deviation of one. This step is essential for improving the performance and convergence of machine learning models.

RECOMMENDATIONS:

1] Ensure Consistent and Complete Data:
It is essential to handle null values properly to avoid complexity. This can be achieved by either removing records with null values or imputing missing values. Ensure that you identify how many variables have null values and the extent of these missing values within the dataset.

2] Minimize Outliers:
Ensure that the dataset does not have an excessive number of outliers, as this can lead to biased analysis. Implement techniques to detect and appropriately treat outliers to maintain data integrity.

3] Use Clear and Understandable Feature Names:
Assign clear and understandable names to features. This will facilitate better understanding and analysis of the dataset, helping to draw more accurate insights.

4] Conduct Time Series Analysis:
If the dataset includes timestamps, perform time series analysis to identify patterns and trends over time. This is crucial for understanding temporal dynamics and making informed predictions.

5] Create Power BI Visualization Reports:
Develop comprehensive Power BI visualization reports to understand various scenarios and support decision-making. Visualizations can provide deeper insights and highlight key aspects of the data.

6] Improve Coolant Usage:
According to the univariate analysis, there is a need to improve the usage of coolant. Investigate and implement measures to optimize coolant consumption.

7]Enhance Fuel Economy:
Focus on improving fuel economy based on the insights gained from the data analysis. Identifying factors affecting fuel consumption can lead to strategies for better fuel efficiency.