Exploratory Data Analysis and Feature Engineering Report

Vehicle Dataset Analysis for Taabi

Manjit Singh

14/08/2024

Introduction

Objective:

The goal of this analysis is to perform exploratory data analysis (EDA) on the provided vehicle dataset to gain insights into vehicle attributes and trends. This analysis will help identify key patterns and areas for further investigation, with the ultimate aim of improving vehicle performance, fuel efficiency, safety, and emissions.

Scope:

The dataset includes various vehicle performance metrics collected through IoT devices, such as fuel consumption, engine load, vehicle speed, and driver behaviour. We aim to understand these metrics and their relationships through detailed analysis and feature engineering.

Dataset Overview

The dataset comprises several vehicle performance metrics, including but not limited to vehicle speed, engine load, fuel consumption, and various sensor readings. The data is loaded from the provided source and inspected for initial understanding.

Data Types and Size:

- **Number of Rows**: 18434

- **Number of Columns**: 141

- **Variable Types**:

  - Numerical: Vehicle speed, engine load, fuel consumption, etc.

  - Categorical: Brake switch status, clutch switch status, etc.

Exploratory Data Analysis (EDA)

Univariate Analysis:

- **Numerical Features**:

  - Engine load and vehicle speed show high variance and skewness.

  - Fuel consumption varies significantly, indicating different driving conditions.

- **Categorical Features**:

  - Brake switch status predominantly shows "Released" indicating minimal braking activity.

  - Clutch switch status is mostly "Pressed," reflecting frequent engagement.

Bivariate Analysis:

- **Fuel Level vs FL Level**: Positive correlation indicates that higher fuel levels are associated with higher FL levels.

- **RPM vs Fuel Consumption**: Positive trend showing higher RPM results in increased fuel consumption.

Visualizations:

- **Scatter Plots**: Used to visualize relationships between engine torque percent and driver's demand engine torque percent, and other key metrics.

- **Histograms**: Illustrated the distribution of variables like engine temperature and coolant levels.

Outliers and Unusual Patterns:

- Identified outliers in engine temperature and fuel consumption.

- Noted unusual patterns in coolant levels showing constant engine temperature.

Feature Engineering

- **Fuel Efficiency Ratio**: Calculated as fuel consumption divided by vehicle speed.

- **Torque Utilization**: Ratio of engine torque level to driver’s demand engine torque percentage.

- **Temperature Range**: Deviation of engine temperature from a baseline value.

- **Braking Intensity**: Computed based on brake switch status and vehicle speed.

- **Clutch Engagement Time**: Percentage of time the clutch switch is pressed.

- **Normalized Features**: Scaled features to a range of [0, 1] for comparability.

Summary of Key Insights:

- **Performance Correlations**: Strong correlation between vehicle speed and engine load. Increased vehicle speed typically results in higher engine load.

- **Fuel Efficiency**: Higher fuel consumption is linked to lower fuel levels, suggesting a need for improved fuel management strategies.

- **Driving Behaviour**: High clutch engagement and brake switch status indicate frequent vehicle operation and driving conditions.

Recommendations:

- **Optimize Speed Management**: Implement systems to manage vehicle speed for better performance and fuel efficiency.

- **Improve Fuel Management**: Develop strategies to monitor and manage fuel levels effectively.

- **Monitor Driving Behaviour**: Regularly check braking and clutch systems for proper maintenance and operational efficiency.

- **Enhance Temperature Management**: Address engine temperature stability and coolant management for optimal performance.

Conclusion

This analysis provided a comprehensive understanding of the vehicle dataset, highlighting key performance metrics and their relationships. The insights and recommendations derived from this study aim to enhance vehicle performance, fuel efficiency, and overall operational effectiveness.