assignment summary

In this assignment, I learned how to clean, combine, and analyze two datasets containing telemetry information. This involved handling datetime formats, identifying overlaps, performing outer joins, and cleaning the resulting dataset. For exploratory data analysis (EDA), we began by examining basic metrics such as speed, odometer readings, and fuel levels. Through histograms, box plots, and correlations, I observed a wide variation in vehicle usage, with odometer readings revealing a mix of heavily-used and less-utilized vehicles. Most vehicles maintained a high average fuel level, but a few displayed data integrity issues, such as fuel levels exceeding 100%. Speed data showed many vehicles idle or at rest, while others were used on highways, contributing to significant variation in the dataset.

We applied normalization to account for the differences in scale between the speed, fuel level, and odometer readings. This allowed us to make fair comparisons across these variables. Vehicle-specific observations, such as vehicle 11's high speed and fuel level, suggested regular long-distance trips, while vehicles like vehicle 5 indicated shorter trips despite maintaining high fuel levels. Furthermore, outlier detection techniques, including interquartile range (IQR) and Z-score methods, revealed variations in vehicle mileage, further highlighting differences in vehicle usage patterns.

Approach to Calculate Fuel Economy

- **Step 1: Sort the data chronologically**

  The telemetry data for each vehicle is sorted by the timestamp to ensure all data is in chronological order for accurate distance and fuel calculations.

- **Step 2: Calculate Distance Traveled**

  Rows with missing odometer values are removed. The **distance traveled** is calculated as the difference between consecutive odometer readings. This provides the total distance covered by each vehicle.

- **Step 3: Identify Fuel Consumption Events**

Rows with missing fuel level values are dropped, and the **fuel level difference** is computed. We identify fuel consumption by checking where the fuel level drops, ignoring refueling events.

- **Step 4: Calculate Fuel Consumed**

  Using the vehicle's tank capacity and the percentage drop in fuel level, the **fuel consumed** is calculated for each fuel drop event. This is accumulated to determine the total fuel consumed by the vehicle.

- **Step 5: Compute Fuel Economy (MPG)**

  The fuel economy for each vehicle is calculated as the ratio of the **total distance traveled** to the **total fuel consumed**. The results for each vehicle are stored for further analysis.

- **Step 6: Calculate Average Speed**

  For each vehicle, the **average speed** is also calculated. Vehicles with moderate and consistent speeds (e.g., between **30-50 mph**) typically achieve better fuel economy compared to those with high-speed variations.