CAF Energy Efficiency Analysis

https://github.com/Carloseidd/CAF_Capstone_project

Master in Big Data and Business Analytics

Corporate Project

Executive Summary

Group 07

July 2024

Thomas Russell

Mario Garcia

Rodrigo Reyes

Carlos Eid

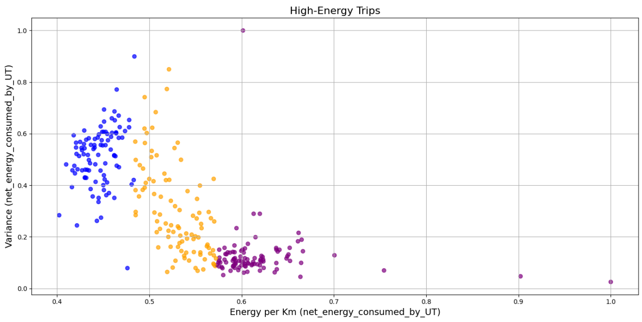
Yazeed Yabroudi

“We hereby certify that this executive summary t is our own original work in its entirety, unless where indicated and referenced”

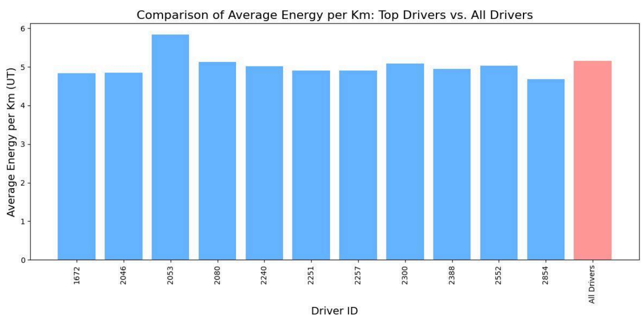
This report details an analysis of CAF train driver journey dataset to assess the energy efficiency of train drivers and anomalous energy trends over the time frame of 2022-2024 Data preprocessing, optimization, evaluation, and data visualisation are some of the techniques used to find consensus for the two assigned questions.

Our mandate from CAF for question 1, was **which train drivers carry out the most optimal driving from the point of view of energy without altering schedules.** Using Python software we meticulously cleaned the dataset of null values and train journeys without driver ID. We also filtered the drivers with significant trip counts between 1,000 and 33,000 to have a statistically sound set of data with variances that can be observed. The dataset included parameters such as net energy consumption, auxiliary energy, and traction energy categorised by trip and driver.

We conducted a dual analysis focussing on the mean energy consumption per kilometre, and the variance in energy consumption across trips of the drivers. Energy metrics were normalised to ensure comparability across different scales and conditions. We also categorised trips into low, medium, and high based on net energy consumption energy metrics. Our comparative study allowed us to visually inspect the three graphs, and find the top performers across all types of trips. For each type of trip, we used unique values to identify the top drivers based on the clustering who were most efficient in variance and energy per km. Using this method we identified 11 drivers who performed within our parameters. However, we further identified the top 3 drivers whose performance was even better and provided further savings.



Significant energy savings are possible when comparing the best drivers to the overall average. Efficient driving practices employed by the top drivers can result in financial savings and an incentive to further improve energy management practices across CAF. Our findings for the energy consumption are as follows which indicate The total energy savings over the three years is 1,651, 498 kWh and the percentage of energy savings over the three years will be 6.73%. Using the cost of electricity in Spain of €0.13 per kWh as per https://www.globalpetrolprices.com/Spain/electricity_prices/, a saving of approximately 200k euros can be achieved. If every driver drove similarly to the top 3 drivers based on the total kilometres driven from 2022 to 2024 the amount of savings would have been much higher.



This exploration into driver efficiency highlights the potential for a scalable approach to energy management that can be applied to the train transportation industry. By leveraging data-driven insights, CAF can improve its energy efficiency and economic performance.

Our mandate from CAF for question #2 was **to identify equipment/trains with anomalous energy consumption**, which could indicate equipment faults and inefficient operational practices. If left unchecked, these anomalies could increase operational costs, impact sustainability efforts, and potentially affect safety.

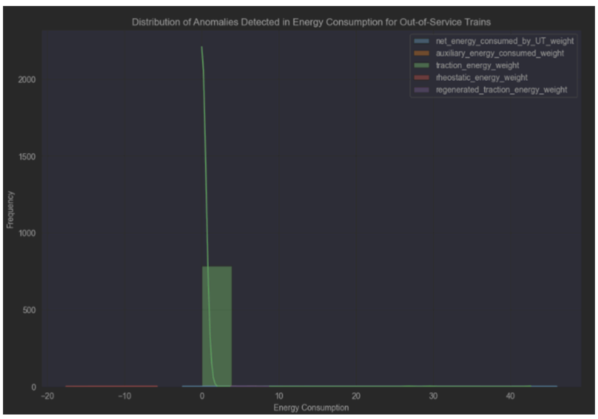
Using Python and libraries such as pandas, numpy, scikit-learn, and matplotlib, we applied the Isolation Forest algorithm to detect anomalies in energy consumption patterns. This approach was chosen for its efficiency in handling large datasets and its ability to analyse data without making assumptions about distribution.

Our analysis focused on detecting outliers in several key areas:

1. Overall energy consumption
2. Out-of-service trains
3. Energy consumption per kilometre
4. Route-specific energy consumption
5. Seasonal energy consumption patterns

Key findings from our analysis include:

- Significant anomalies in energy consumption were detected across various trains and trips.
- Out-of-service trains showed unexpected energy usage patterns.
- Certain routes exhibited unusual energy consumption, suggesting route-specific factors.
- Seasonal variations in energy consumption were identified.




```
Number of seasonal energy outliers: 1

Seasonal energy consumption:
   season  season_energy  anomaly_season_energy
0  Autumn       7.869306                      1
1  Spring       7.975565                      1
2  Summer       7.805433                      1
3  Winter       8.068962                     -1

Details of seasonal energy outliers:
   season  season_energy
3  Winter       8.068962

Overall average energy consumption: 7.93
Season with highest energy consumption: Winter
Season with lowest energy consumption: Summer
```

Calculations indicate significant potential for energy and cost savings:

- Total potential energy savings: 3131.98 MWh, equivalent to 12.49% of total energy consumption.

- Estimated monetary savings for 2022: €639,863.54
- Estimated monetary savings for 2023: €311,945.22
- Total potential savings for both years: €951,808.76

These findings highlight the importance of addressing anomalous energy consumption in CAF's train operations. By identifying and rectifying these issues, CAF can significantly improve energy efficiency, reduce operational costs, and enhance sustainability efforts. Our recommendations include conducting detailed inspections of anomalous trains, implementing real-time monitoring, and developing targeted energy efficiency programs based on our findings.