

VED ML Data Loading & Preprocessing

Data Loading and Preprocessing

1. Data Loading

This section is responsible for importing both static and dynamic datasets required for further analysis. Static data is loaded from Excel files, while dynamic data is aggregated from multiple CSV files within a specified directory.

ICE & HEV Records

357

PHEV & EV Records

27

2. Data Cleaning

The data cleaning process involves replacing placeholder values, correcting data types, renaming columns for consistency, and merging static datasets. Duplicate records are also identified and reported.

Data Cleaning Summary

Total Static Records

384

Dynamic Data Sample Size

5,131,987

No duplicate records were found in the static dataset.

No duplicate records were found in the dynamic sample dataset.

3. Data Joining

In this step, the sampled dynamic data is merged with the consolidated static dataset using the 'VehId' key. The results of the join operation, including the number of matched and unmatched records, are displayed below.

Total Records After Join

5,131,987

Matched Records

5,131,987

Unmatched Records

0

4. Data Transformation

This section performs several data transformations, including:

- Categorization of Outside Air Temperature (OAT)
 - Conversion of day numbers to datetime objects
 - Calculation of distance traveled
 - Computation of Fuel Consumption Rate (FCR) using available sensor data
- The results of these transformations are summarized and visualized below.

Transformation Results

Distribution of Outside Air Temperature (OAT) Categories:

Count	count
Cool	928330
Cold	655929
Mild	137473
Warm	4340
Extremely Cold	2419
Hot	2230

Fuel Consumption Rate (FCR) Statistical Summary:

Statistic	Value
count	3770025
mean	0.7839
std	0.9327
min	0
25%	0.2305
50%	0.4363
75%	1.1027
max	19.9242

HV Battery Power[Watts]:

The column 'HV Battery Power[Watts]' is calculated as the product of 'HV Battery Voltage[V]' and 'HV Battery Current[A]' for each record in the dataset. This represents the instantaneous electrical power output (in Watts) of the high-voltage battery at each timestamp.

Formula:

$$\text{HV Battery Power [Watts]} = \text{HV Battery Voltage [V]} \times \text{HV Battery Current [A]}$$

Final Dataset Overview:

DayNum	VehId	Trip	Timestamp(ms)	Latitude[deg]	Longitude[deg]	Vehicle Speed[km/h]	MAF[g/sec]
1.5867	8	706	0	42.2776	-83.6987	40	22.13
1.5867	8	706	200	42.2776	-83.6987	40	22.13
1.5867	8	706	1100	42.2776	-83.6987	45	22.13
1.5867	8	706	2100	42.2776	-83.6987	47	6.15
1.5867	8	706	4200	42.2776	-83.6987	48	21.44
1.5867	8	706	5200	42.2783	-83.6988	52	21.44
1.5867	8	706	6300	42.2783	-83.6988	55	26.51
1.5867	8	706	7400	42.2783	-83.6988	59	26.51
1.5867	8	706	8400	42.2783	-83.6988	59	11.66
1.5867	8	706	10600	42.279	-83.6989	60	13.39

Final Cleaned Data Info:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5131987 entries, 0 to 5131986
Data columns (total 37 columns):
#   Column                                Dtype
---  -----
0   DayNum                                float64
1   VehId                                int64
2   Trip                                  int64
3   Timestamp(ms)                        int64
4   Latitude[deg]                        float64
5   Longitude[deg]                       float64
6   Vehicle Speed[km/h]                  float64
7   MAF[g/sec]                           float64
8   Engine RPM[RPM]                      float64
9   Absolute Load[%]                    float64
10  OAT[DegC]                            float64
11  Fuel Rate[L/hr]                      float64
12  Air Conditioning Power[kW]           float64
13  Air Conditioning Power[Watts]        float64
14  Heater Power[Watts]                  float64
15  HV Battery Current[A]                float64
16  HV Battery SOC[%]                   float64
17  HV Battery Voltage[V]                float64
18  Short Term Fuel Trim Bank 1[%]        float64
19  Short Term Fuel Trim Bank 2[%]        float64
20  Long Term Fuel Trim Bank 1[%]         float64
21  Long Term Fuel Trim Bank 2[%]         float64
22  Vehicle Type                          object
23  Vehicle Class                        object
24  Engine Configuration & Displacement  object
25  Transmission                         object
26  Drive Wheels                         object
27  Generalized_Weight                   float64
28  OAT_Category                        object
29  DateTime                            datetime64[ns]
30  Date                                datetime64[ns]
31  Time                                object
32  Distance[km]                        float64
33  Displacement_L                      float64
34  correction                          float64
35  FCR                                 float64
```

36 HV Battery Power[Watts] float64
dtypes: datetime64[ns](2), float64(25), int64(3), object(7)
memory usage: 1.4+ GB