VED ML Data Modelling

ICE, HEV, EV, and PHEV Analysis | Supervised Learning | Unsupervised Learning

Unsupervised Learning - Clustering

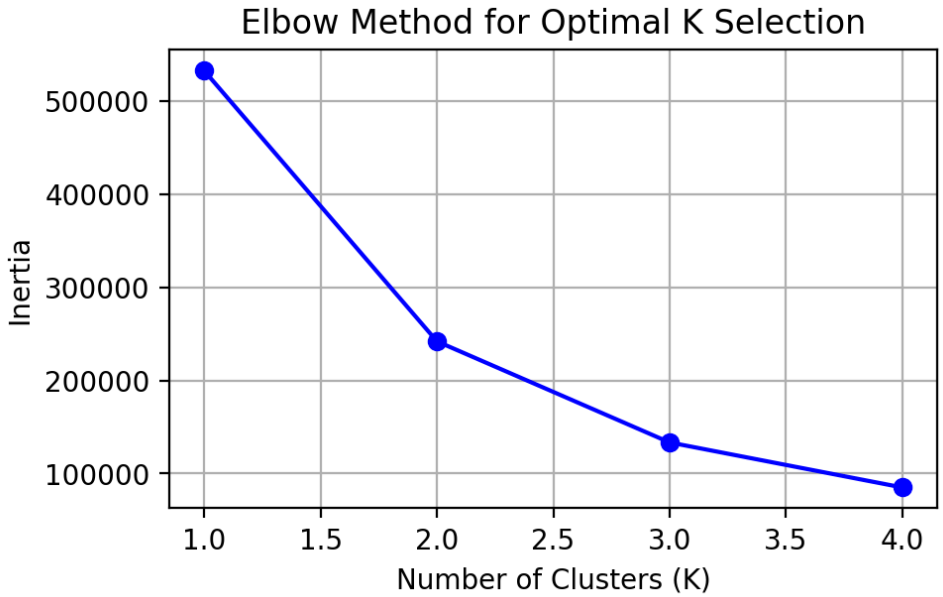
- **Model:** K-means
- **Inertia:** Calculates the difference in inertia to suggest an optimal number of clusters (elbow point).
- **Clusters:** Predicts cluster labels for the input data.

K-Means Clustering: Elbow Method and Cluster Assignment

This section demonstrates unsupervised learning using K-Means clustering:

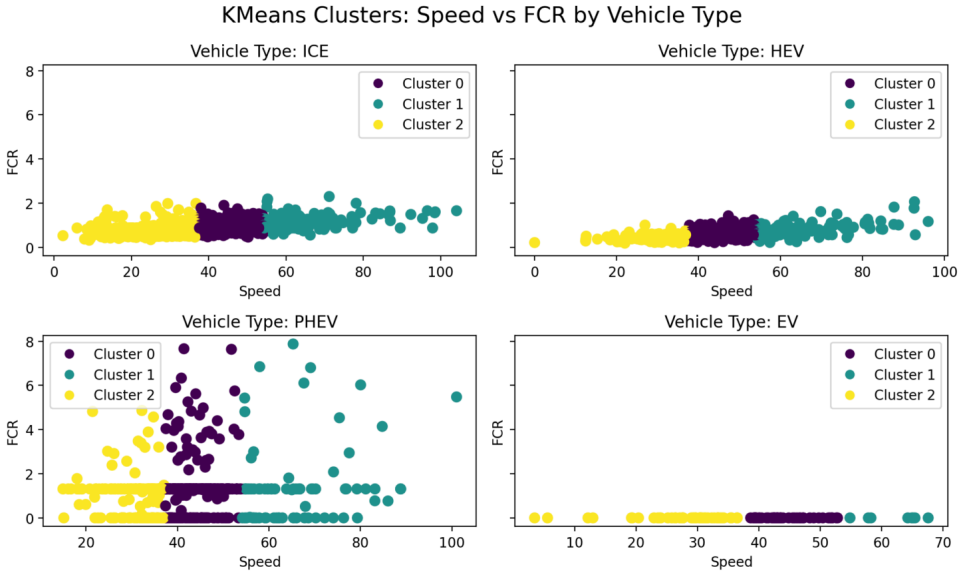
- **Elbow Method:** Visualizes the inertia (within-cluster sum of squares) for different cluster counts to help select the optimal number of clusters.
- **Cluster Assignment:** Assigns cluster labels to each data point and displays the resulting cluster assignments.

The elbow plot helps determine the best value for K, and the resulting clusters are shown in a table.



K-Means Clustering (Speed vs FCR by Vehicle Type)

- Loads the combined dataset and selects relevant features.
- Maps vehicle type strings to numeric codes for clustering.
- Runs the elbow method to suggest optimal cluster count.
- Fits KMeans (K=3) and assigns cluster labels.
- Visualizes clusters for each vehicle type (Speed vs FCR).
- Shows the cluster assignment table.



Outcome & Understanding:

- The plots show KMeans clustering results for Speed vs FCR across four vehicle types (ICE, HEV, PHEV, EV).
- Each subplot represents a vehicle type, with points colored by cluster.
- **Distinct Clusters:** Each vehicle type forms 2–3 clusters, indicating different driving or consumption patterns within each type.
- **ICE & HEV:** Most data points are at lower FCR values, with clusters separated by speed.
- **PHEV:** Shows a wider spread in FCR, with clusters capturing both low and high FCR at various speeds.
- **EV:** FCR is near zero (as expected for electric vehicles), with clusters mainly reflecting speed differences.

Conclusion:
KMeans clustering effectively segments vehicle operational patterns by type, revealing characteristic speed and fuel/energy consumption behaviors for each vehicle class.

Cluster Assignment Table (first 10 rows):

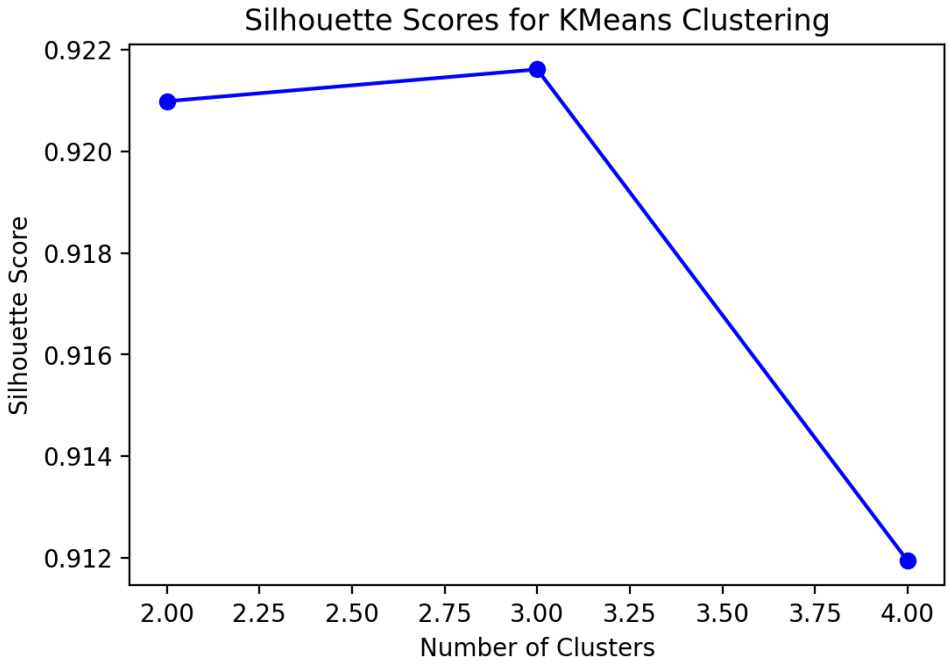
	Vehicle Type	Vehicle Speed[km/h]	FCR	Cluster
0	0	71.1996	2.2984	1
1	0	55.2804	2.1885	1
2	0	55.1316	2.076	1
3	0	78.1669	1.9946	1
4	0	36.7285	1.9842	2
5	0	29.4306	1.9812	2
6	0	43.9556	1.8991	0
7	0	26.4415	1.8488	2
8	0	55.066	1.8242	1
9	0	38.0215	1.7772	0

Silhouette Score for KMeans Clustering

This visualization helps identify the optimal number of clusters using the **Silhouette Score** metric.

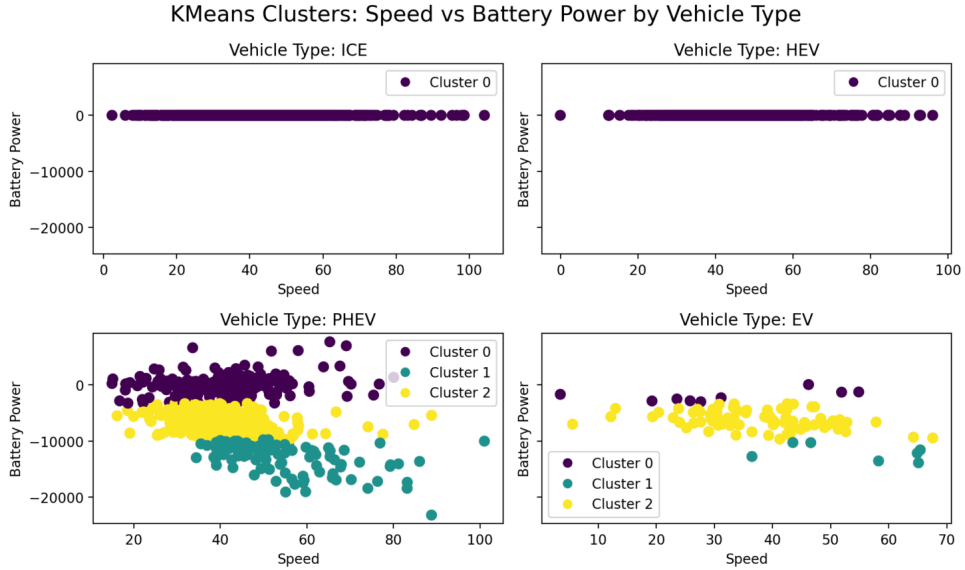
- **Data Used**: 'Vehicle Speed[km/h]' and 'HV Battery Power[Watts]'
- **Range Tested:** Clusters from 2 to 4
- The **higher** the silhouette score, the **better** the clustering quality.
- Iterates over cluster numbers 2 to 4.
- Applies KMeans clustering on selected features.
- Computes **silhouette score** for each model to evaluate clustering quality.
- Displays both a **line chart** of scores and a **data table** for reference.

	Number of Clusters	Silhouette Score
0	2	0.921
1	3	0.9216
2	4	0.9119



K-Means Clustering (Speed vs Battery Power by Vehicle Type)

- 2x2 grid of scatter plots, one for each vehicle type, showing the relationship between 'Vehicle Speed[km/h]' and 'HV Battery Power[Watts]' for each type..
- It iterates over unique vehicle types, selects the corresponding subset of data, and plots the points colored by their KMeans cluster assignment.
- Visualizes clusters for each vehicle type (Speed vs Battery Power).
- Shows the cluster assignment table.



Outcome & Understanding (Battery Power Clustering):

- **ICE & HEV:** All data points are assigned to a single cluster (Cluster 0), indicating almost no variation in battery power—expected, as these vehicles have little or no high-voltage battery activity.
- **PHEV & EV:** Multiple clusters are identified, reflecting more diverse battery power usage patterns. PHEVs show a wide range of battery power (including negative values, likely due to regenerative braking or charging), while EVs also display clustering based on battery power and speed.

Conclusion:

Clustering reveals that battery power is only a meaningful differentiator for PHEV and EV types. ICE and HEV vehicles show negligible battery power variation, while PHEV and EV vehicles exhibit distinct operational patterns based on speed and battery power.

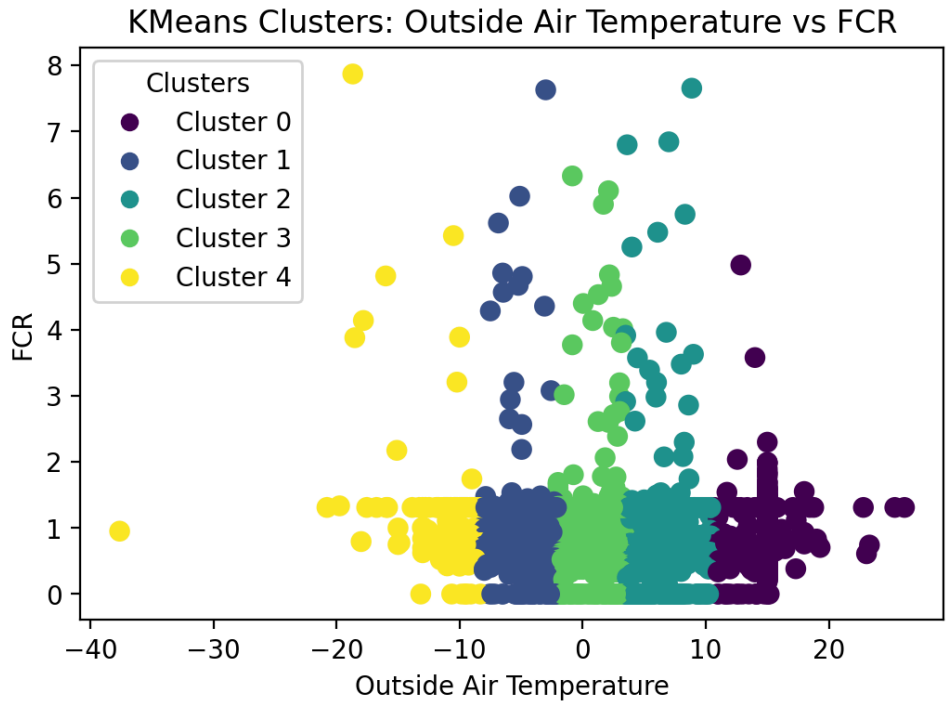
Cluster Assignment Table (first 10 rows):

	Vehicle Type	Vehicle Speed[km/h]	HV Battery Power[Watts]	Cluster
0	0	71.1996	0	0
1	0	55.2804	0	0
2	0	55.1316	0	0
3	0	78.1669	0	0
4	0	36.7285	0	0
5	0	29.4306	0	0
6	0	43.9556	0	0
7	0	26.4415	0	0
8	0	55.066	0	0
9	0	38.0215	0	0

KMeans Clustering: Outside Air Temperature vs FCR

This scatter plot visualizes KMeans clustering on:

- **X-axis**: Outside Air Temperature (`OAT[DegC]`)
- **Y-axis**: Fuel Consumption Rate (`FCR`)
- Points are **colored by cluster assignment** using the 'viridis' colormap.
- Useful for identifying patterns in vehicle behavior based on environmental conditions.



Cluster Analysis: Outside Air Temperature vs FCR

The scatter plot visualizes **KMeans clustering** results for the relationship between:

- **Outside Air Temperature (OAT[°C])**
- **Fuel Consumption Rate (FCR)**

Key Observations:

- The data is grouped into **5 distinct clusters (Cluster 0–4)**.
- **Cluster separation** appears to reflect patterns in how vehicles consume fuel under different outside temperatures.
- **Cluster 4** (yellow) tends to appear at **very low temperatures** (below -10°C), showing **higher FCR variability**.
- **Cluster 0 and 1** (dark colors) dominate the **moderate-to-warm temperature** range, where FCR is generally low.
- **Cluster 3** shows more spread across moderate temperatures and seems to include more varied FCR.

Insights:

- Fuel consumption behavior varies with **environmental conditions**, especially temperature.
- The model successfully segments driving patterns that may correspond to **cold-weather inefficiencies**, **efficient warm-weather driving**, or **transitional performance zones**.

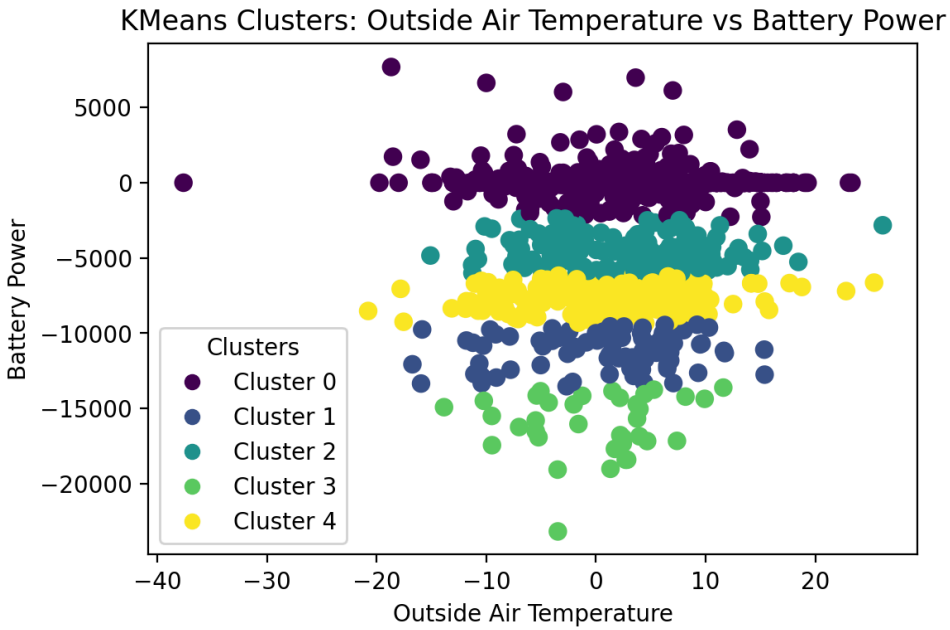
Understanding these clusters can help:

- Optimize driving or energy usage under specific temperature conditions.
- Inform EV or hybrid battery management strategies.

KMeans Clustering: Outside Air Temperature vs Battery Power

This scatter plot visualizes KMeans clustering on:

- **X-axis**: Outside Air Temperature (`OAT[DegC]`)
- **Y-axis**: HV Battery Power (`Watts`)
- Points are **colored by cluster assignment** using the 'viridis' colormap.
- Useful for identifying patterns in vehicle behavior based on environmental conditions.



Cluster Analysis: Outside Air Temperature vs Battery Power

This scatter plot visualizes **KMeans clustering** applied to:

- **X-axis**: Outside Air Temperature (`OAT[DegC]`)
- **Y-axis**: HV Battery Power (`[Watts]`)
- Points are **grouped into 5 clusters (Cluster 0–4)** and **colored using the 'viridis' colormap**.

Key Observations:

- **Cluster 0** (purple) dominates the **upper power range**, indicating **higher battery output** at varying temperatures.
- **Cluster 1 and 3** (dark blue and light green) represent data with **significant negative battery power**, suggesting **regenerative braking or power intake scenarios**, especially in **moderate temperatures**.
- **Cluster 4** (yellow) appears consistent across low to moderate temperatures with a narrow range of battery power usage.
- The separation between clusters shows **how battery power usage varies across temperature bands** and operating modes.

Insights:

- Clusters effectively segment **vehicle operating conditions** — such as energy consumption, regeneration, or idle states — influenced by **external temperature**.
- This analysis can help improve:
- **Battery management algorithms**
- **Climate-aware energy optimization**
- **Driving behavior insights across seasons**