Market Segmentation Analysis on EV Charging Patterns using Machine Learning

By Riya Maurya

1. Introduction

With the rapid growth of electric vehicle (EV) adoption, understanding user charging behavior is crucial for optimizing infrastructure, energy distribution, and policy planning. The goal of this project is to analyze EV charging patterns and identify distinct user segments using unsupervised machine learning techniques, particularly KMeans clustering.

2. Dataset Description

The dataset used in this project contains detailed information on EV charging sessions. The key features include:

- Battery Capacity (kWh)
- Energy Consumed (kWh)
- Charging Duration (hours)
- Charging Rate (kW)
- Charging Cost (USD)
- State of Charge (Start and End %)
- Distance Driven since last charge (km)
- Temperature (°C)
- Vehicle Age (years)

The dataset has been cleaned and preprocessed for analysis.

3. Methodology

3.1 Data Preprocessing

- Dropped irrelevant or identifier columns
- Checked and confirmed no null values
- Scaled numerical features using `StandardScaler`
- Encoded categorical variables (if present) using one-hot encoding

3.2 Dimensionality Reduction

- Principal Component Analysis (PCA) was applied for visualization purposes
- However, since the first two components explained limited variance, PCA was not used for clustering

3.3 Clustering

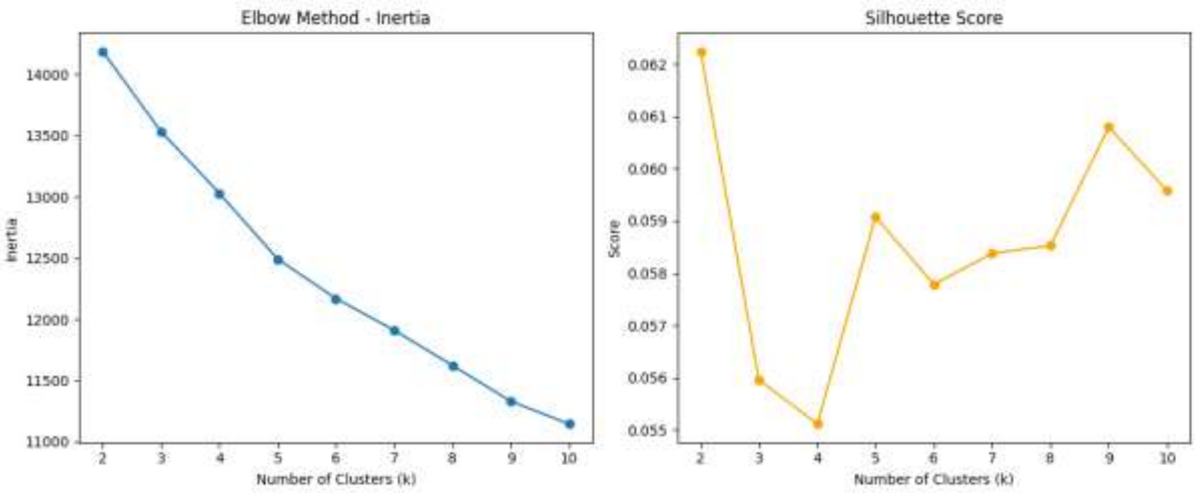
- KMeans clustering was used to segment the dataset

- Elbow Method and Silhouette Score were used to determine the optimal number of clusters
- 5 clusters were selected based on results

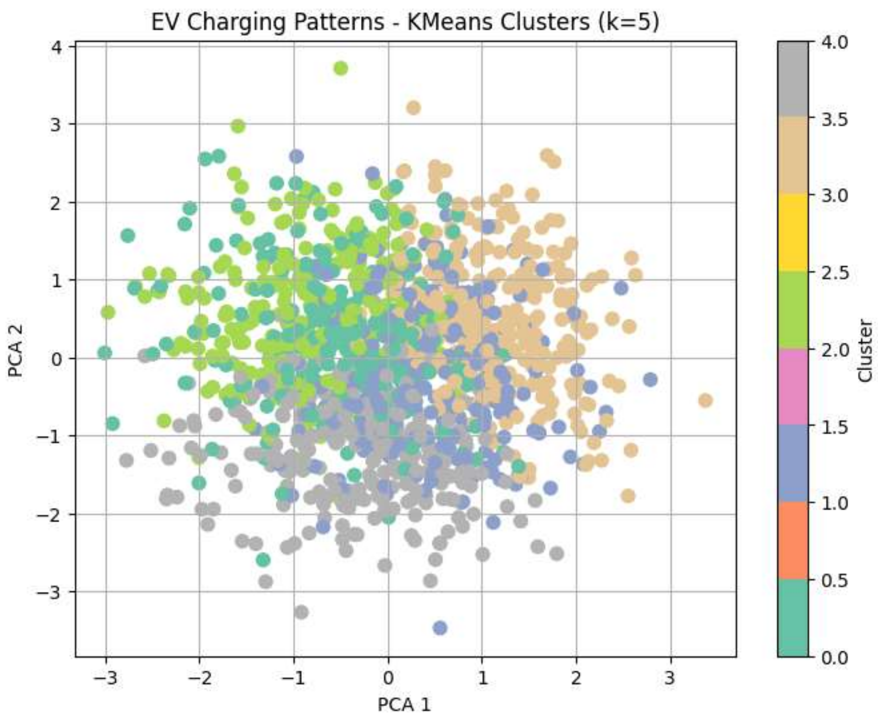
4. Results & Visualizations

Clustering visualizations include:

- Elbow plot to find optimal k and Silhouette score comparison



- PCA scatter plot (for visualization only)



- Cluster-wise average values for all features

Cluster	Battery Capacity (kWh)	Energy Consumed (kWh)	Charging Duration (hours)	Charging Rate (kW)	Charging Cost (USD)	State of Charge (Start %)	State of Charge (End %)	Distance Driven (since last charge) (km)	Temperature (°C)	Vehicle Age (years)
0	89.080660	43.082492	2.068051	32.712746	19.571353	31.109614	75.070703	89.718297	13.676615	3.161247
1	70.240492	38.898326	1.808183	25.226734	17.324348	72.524147	69.896470	143.503221	9.125011	3.004304
2	71.254030	52.396039	3.115557	19.077695	18.481908	42.941744	83.729647	176.133459	7.191841	4.861444
3	71.608572	43.123646	2.082964	23.201145	29.576523	57.326416	75.540930	118.183944	28.318257	4.974694
4	71.226795	36.875683	2.366933	29.369577	26.700722	39.488720	72.448059	240.465648	16.199018	2.107896

5. Insights

Cluster 0 – Fast Chargers, Short Rides

- High charging rate (~33 kW), short duration (~2 hrs)
- Low start SOC, high battery capacity
- Newer vehicles, shorter trips (~90 km)

Cluster 1 – Efficient Chargers

- Medium charging rate and duration
- Balanced SOC and moderate costs
- Used in cooler environments

Cluster 2 – Long-Distance Users

- Longest distances (~176 km), slow charging (~19 kW)
- High SOC increase
- Possibly intercity travelers

Cluster 3 – High Cost in Hot Regions

- Highest charging cost, hot environments (~28°C)
- Moderate distance and duration

Cluster 4 – Maximum Distance Drivers

- Longest distance (~240 km) and efficient charging
- Likely commercial/rideshare vehicles

6. Conclusion

The analysis successfully identified five distinct EV charging user segments. These segments highlight how different user needs (short vs. long trips, fast vs. slow charging, cost sensitivity, etc.) influence charging behavior.

This segmentation is valuable for:

- Charging station placement and planning
- Designing custom EV energy plans
- Developing EV features tailored to user types

In conclusion, machine learning-based segmentation provides actionable insights for stakeholders in the EV ecosystem.

7. References

- Python Libraries: pandas, matplotlib, seaborn, scikit-learn
- Research on EV infrastructure and adoption trends
- Dataset provided for educational analysis