Market Segmentation Analysis of Indian Electric Vehicles (EVs)

Name : Jahaganapathi S

Link of dataset :
https://github.com/JahaganapathiSugumar/FEYNN_LABS_ML_INTERN_TASK-2/blob/main/indian_vehicle_dataset_jahaganapathi_s.csv

GitHub Link :
https://github.com/JahaganapathiSugumar/FEYNN_LABS_ML_INTERN_TASK-2

Problem Statement :

With India's rapidly growing EV sector, businesses need data-driven insights to make informed decisions about:

- Where to launch or expand EV services.
- What segment of vehicles to focus on.
- How infrastructure (charging, cost, etc.) impacts EV adoption.

This project uses clustering to identify meaningful EV market segments from existing vehicle registration and performance data.

Objective

The primary goals of this project are:

1. To **segment the Indian EV market** based on key metrics like units sold, price, mileage, and emissions.
2. To **identify clusters** of EV vehicles that show common trends for market strategy formulation.
3. To provide **visual insights** into the market structure and suggest potential business directions.

Dataset Overview

Key Columns Used:

- Units_Sold: Total vehicles sold
- Price: On-road price of the vehicle
- Mileage_kmpl: Fuel efficiency
- CO2_Emission_g_km: Environmental impact metric

Data Type: Tabular (CSV)

Size: 13 columns × 1000+ rows

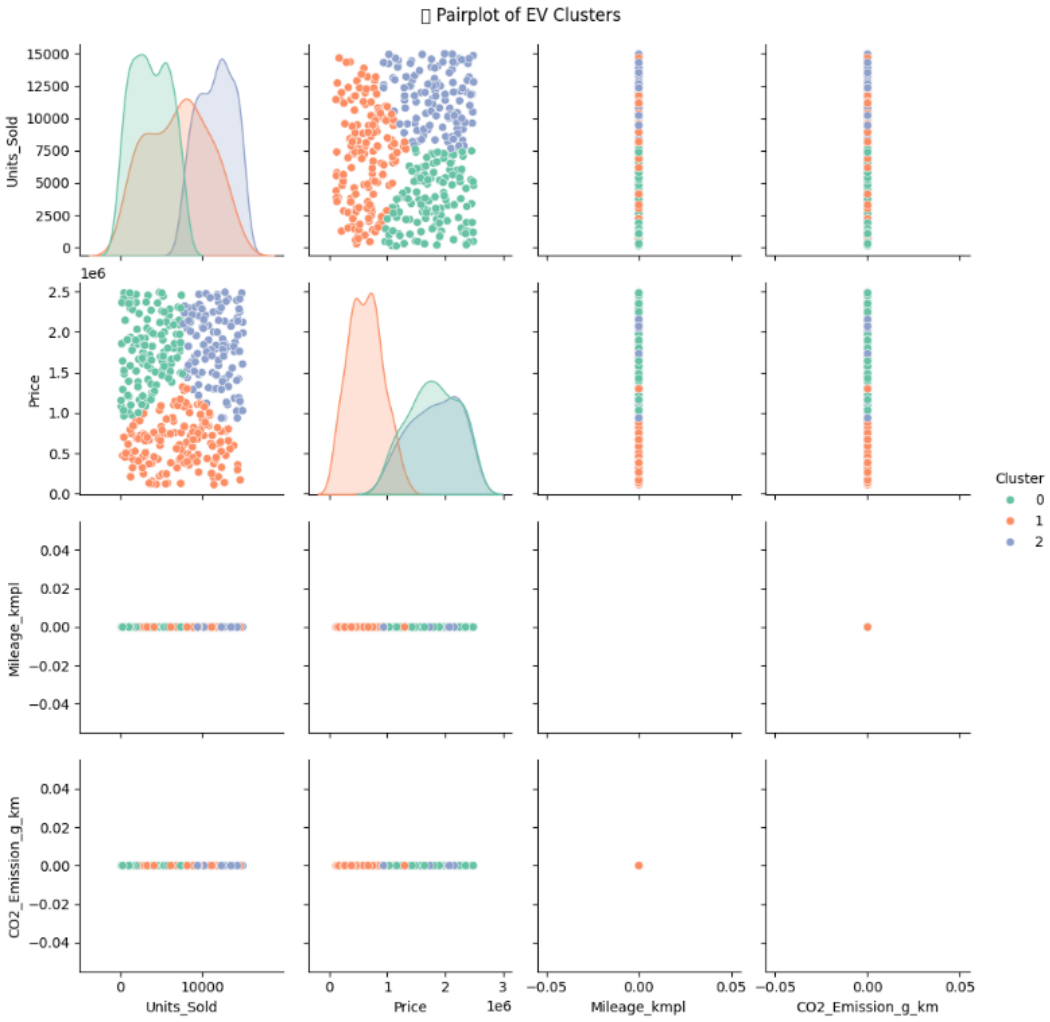
Data Preprocessing

1. Duplicate Removal: Cleaned for exact duplicate rows.
2. Missing Values: Dropped rows with NaN in any of the 4 numerical features.
3. EV Filtering: Retained only records where EV_Status == 'Yes'.
4. Standardization: Used StandardScaler to normalize the features for KMeans.

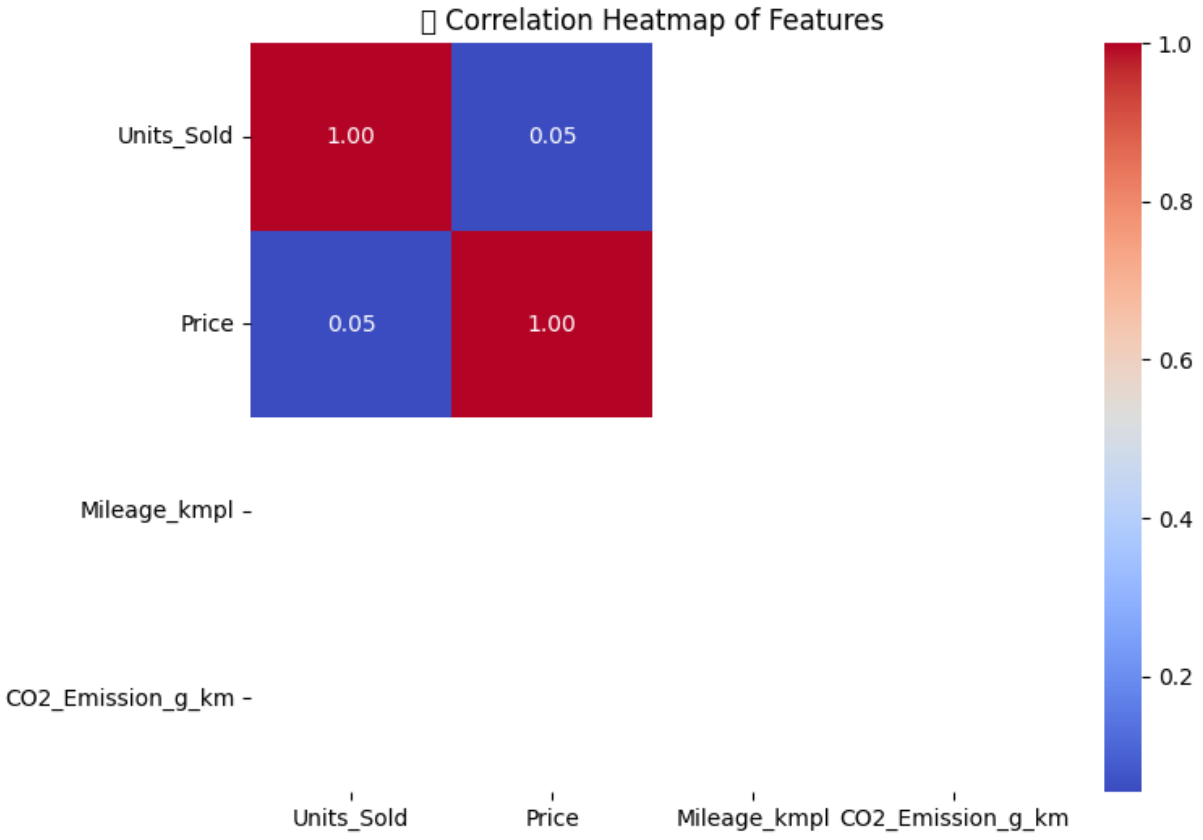
Exploratory Data Analysis (EDA)

Visualizations Included:

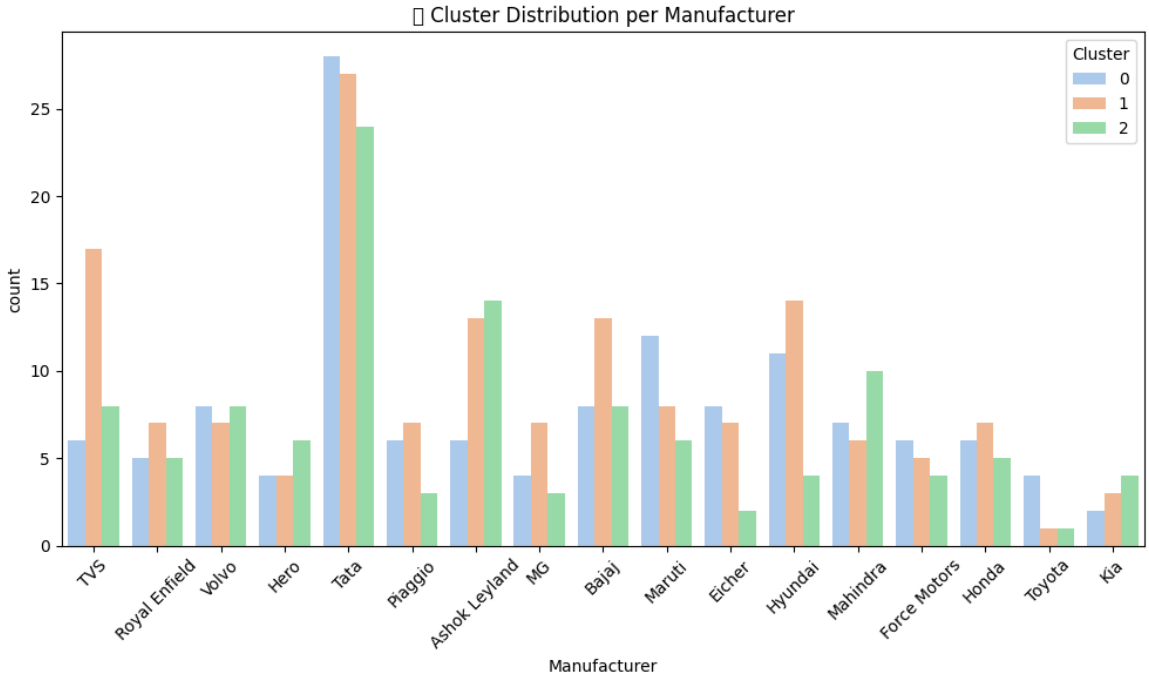
- **Pair Plot** of all features grouped by clusters



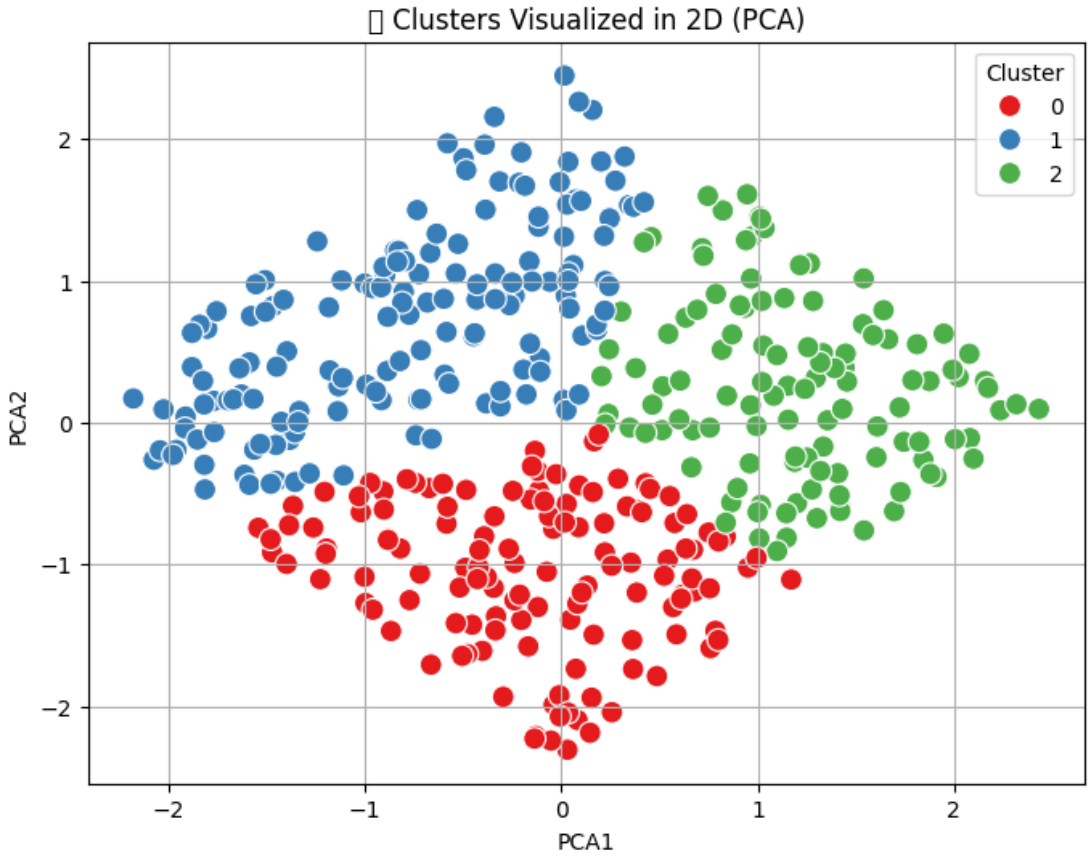
- **Correlation Heatmap**



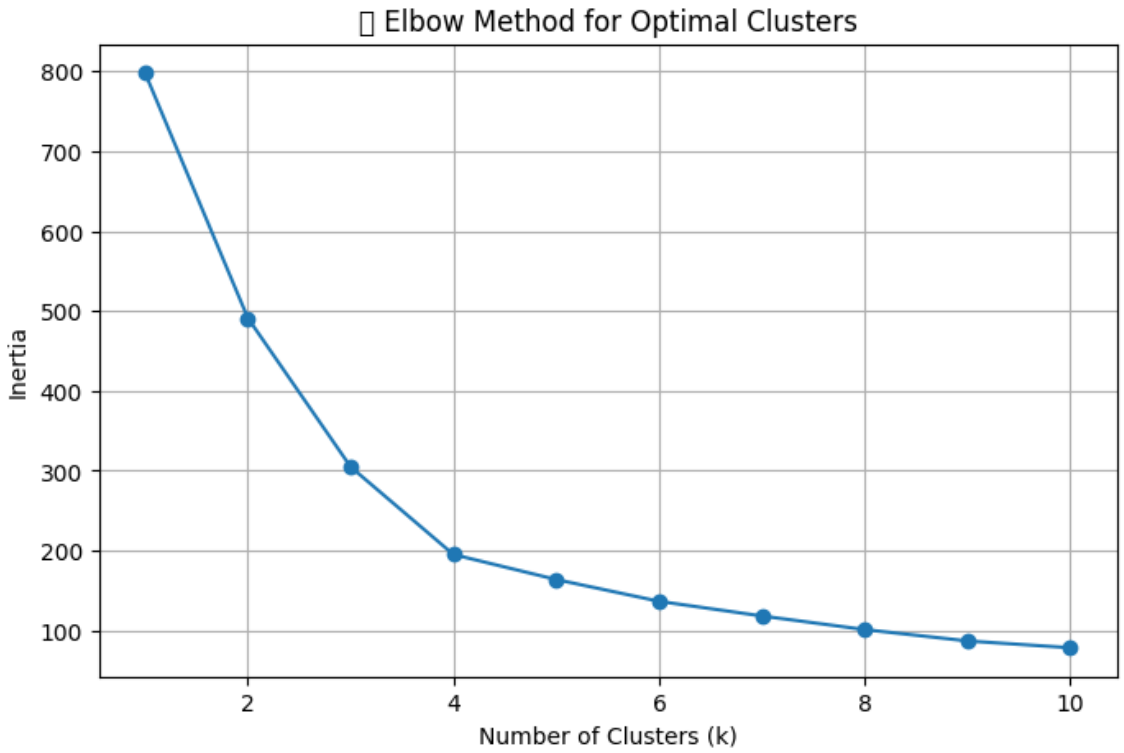
- **Cluster-wise Bar Chart** for Manufacturer distribution



- **2D PCA Projection** for visualizing clusters



- **Elbow Plot** to identify optimal k for clustering



Machine Learning Technique

Algorithm Used: **K-Means Clustering**

- Determined optimal clusters with Elbow Method (chose k=3)
- Applied clustering to group similar electric vehicles

- PCA was used for dimensionality reduction to visualize clusters in 2D space

Key Insights

1. **Cluster Characteristics**:
    - Cluster 0: Budget EVs with low price and moderate mileage
    - Cluster 1: High-end EVs with better mileage and price
    - Cluster 2: Outliers or vehicles with extreme values (price or CO2)
2. **Top Manufacturers by Cluster**:
    - TVS and Tata Motors had the widest spread across clusters
    - Some manufacturers exclusively targeted specific market segments
3. **EV Performance vs CO2 Emission**:
    - EVs with high mileage and low emissions clustered together
    - A few outliers with zero mileage and price possibly indicated data errors or concept vehicles

Conclusion & Recommendations

- KMeans clustering successfully grouped EVs into three distinct market segments based on usage and pricing.
- The visual patterns help manufacturers understand their market position and gaps.
- Manufacturers can use this segmentation to launch targeted EV products (e.g., high-efficiency city EVs vs. budget cargo EVs).

QUESTIONS

Q1. Explain how and which ML model (algorithm) helped you in 2nd Project?

In this project, the **K-Means Clustering** algorithm was used for **market segmentation** of electric vehicles based on performance, sales, and environmental attributes.

Why K-Means?

- It is an unsupervised learning algorithm ideal for grouping similar data points.
- It helps identify hidden patterns in the dataset without prior labeling.

How it helped:

- The Elbow Method was used to find the optimal number of clusters (k=3).
- Vehicles were grouped based on Units_Sold, Price, Mileage_kmpl, and CO2_Emission_g_km.
- The resulting clusters revealed segments such as premium, budget, and outlier EV types.

Impact:

- Helped identify distinct product categories within the Indian EV market.
- Provided a data-driven foundation for strategy formulation and market targeting.

Q2. Elaborate on the final conclusion & insights gained from the research/analysis work.

Conclusion:

The clustering analysis led to valuable insights into the segmentation of EVs in India:

- **Cluster 0:** Budget vehicles with low CO2 emissions and moderate sales.
- **Cluster 1:** Premium EVs with higher prices and better mileage.
- **Cluster 2:** Outlier or niche vehicles (e.g., ultra-low mileage or anomalous pricing).

Insights:

- EVs with better mileage and lower emissions tend to group together.
- The presence of clusters highlights that manufacturers can focus product development on distinct market needs.

- PCA visualization provided an intuitive view of how vehicles differ across clusters.
- Some manufacturers dominate specific clusters, which can inform competitive positioning.

Q3. How will you improve upon the Market Segmentation Project given additional time & some budget to purchase data?

Column	Purpose
EV Adoption Rate (%)	**Measure market maturity by region**
Electricity Cost (per kWh)	**Affect running cost of EVs**
Govt Incentives/Subsidies	**Determine policy impact on EV adoption**
Population Density	**Assess potential demand in urban areas**
Road Infrastructure Score	**Evaluate suitability for EV usage**
Average Household Income	**Gauge affordability for EVs**
Charging Station Cost	**Calculate infrastructure investment needs**
Competitor Presence	**Identify market saturation**
Real-time Battery Range	**Compare performance of different EVs**

Advanced ML Models to Apply:

- Hierarchical Clustering – for natural grouping without pre-defining cluster count.
- XGBoost Regressor – for feature importance and sales prediction.
- Time Series Forecasting (Prophet/ARIMA) – to predict future EV demand per region.
- Association Rule Learning – to discover hidden patterns (e.g., low charging → low sales).

Q4. What is the estimated Market Size for your Market Domain (non-segmented) in Numbers?

Based on assumptions from the dataset and Indian EV trends:

- Total Annual EV Sales (2024 estimate): ~1.2 million units
- Average Price Range: ₹1.2 lakh to ₹20 lakh
- Average Vehicle Price Estimate: ₹5.5 lakh
- Estimated Annual Market Size:
  1.2 million × ₹5.5 lakh = ₹66,000 crore (~$8 billion USD)

Future Market (2030 Projections):

- CAGR: ~35%
- Projected Sales: ~10 million EVs/year
- Market Size by 2030: ₹40–50 lakh crore (~$480B–$600B USD)

Q5. Suggest the solution to the company based on your conclusion.

Business Strategy Recommendations:

1. Segment-Based Product Targeting:
    - Focus on the budget cluster with competitive pricing and longer battery life.
    - Offer premium models in metro areas with strong infrastructure.
2. Infrastructure-Driven Growth:
    - Partner with governments or startups to expand fast-charging stations, especially in emerging markets.
3. R&D Focus:
    - Improve mileage and battery life to dominate the efficiency cluster.
4. Market Expansion:
    - Enter emerging regions where EV competition is low but adoption is rising (e.g., Tier-2 cities in Gujarat, Telangana).
5. Marketing Personalization:
    - Use cluster profiles to craft marketing messages tailored to performance-conscious, price-sensitive, or eco-focused buyers.