Electric Vehicles Market Segmentation in India

1. Objective

This analysis aims to identify the most suitable electric vehicle (EV) type (2-Wheeler, 3-Wheeler, 4-Wheeler, or Bus) for a company to launch in India in 2025. Market segmentation is performed based on geographic, demographic, psychographic, and behavioral factors to determine the optimal EV type and target market.

2. Data Preprocessing

Required Libraries

To conduct exploratory data analysis (EDA) and clustering, the following Python libraries were used:

1. **Pandas**: For data handling and manipulation.
2. **Matplotlib** and **Seaborn**: For data visualization.
3. **Scikit-learn**: For preprocessing, PCA, and K-Means clustering.
4. **NumPy**: For numerical operations.
5. **Plotly**: For interactive visualizations.

```
# Imports
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from IPython.display import display
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans
import plotly.graph_objects as go
```

Pulling the Datasets

Dataset 1: Vehicle Registration Data (evdb_Jayasri.csv)

This dataset contains 106,669 records of vehicle registrations in India from 2014 to 2024, with 34 columns including Year, State, Vehicle Category (2W, 3W, 4W, 6W), Fuel Type (e.g., PURE EV, PLUG-IN HYBRID EV), and Vehicle Use Type (Personal, Shared, Others).

```
# Load dataset - 1 (evdb_Jayasri.csv as ds1)
ds1 = pd.read_csv('evdb_Jayasri.csv')
print("Dataset 1 (evdb_Jayasri.csv):")
print(ds1.head())
```

```
Dataset 1 (evdb_Jayasri.csv):
   Year Month_name  Day        Date             State  \
0  2024        dec    1  2024-12-01           Gujarat
1  2024        dec    1  2024-12-01  Himachal Pradesh
2  2024        dec    1  2024-12-01  Himachal Pradesh
3  2024        dec    1  2024-12-01  Himachal Pradesh
4  2024        dec    1  2024-12-01  Himachal Pradesh

                      Vehicle Class Vehicle Category Vehicle Type  CNG ONLY
0            VEHICLE FITTED WITH RIG           Others       Others         0
1               AGRICULTURAL TRACTOR           Others       Others         0
2                          AMBULANCE           Others       Others         0
3                                BUS              Bus          Bus         0
4  CAMPER VAN / TRAILER (PRIVATE USE)           Others       Others         0

   DIESEL  ...  PETROL/ETHANOL  PETROL/HYBRID  PETROL/LPG  PETROL/METHANOL  \
0       1  ...              0              0           0                0
1     158  ...              0              0           0                0
2       2  ...              0              0           0                0
3      24  ...              0              0           0                0
4       4  ...              0              0           0                0

   SOLAR  Total  PLUG-IN HYBRID EV  PURE EV  STRONG HYBRID EV  \
0      0      1                0.0      0.0               0.0
1      0    158                0.0      0.0               0.0
2      0      5                0.0      0.0               0.0
3      0     25                0.0      0.0               0.0
4      0      4                0.0      0.0               0.0

   Vehicle Use type
0            Others
1            Others
2            Others
3            Shared
4          Personal

[5 rows x 34 columns]
```

Dataset 2: EV Sales Data (ev_sales_by_makers_and_cat_15-24.csv)

This dataset contains 1,386 records of EV sales by manufacturers and vehicle categories (2W, 3W, 4W, 6W) from 2015 to 2024, with 12 columns including Manufacturer, Category, and annual sales.

```python
# Load dataset - 2 (ev_sales_by_makers_and_cat_15-24.csv as ds2)
ds2 = pd.read_csv('ev_sales_by_makers_and_cat_15-24.csv')
print("\nDataset 2 (ev_sales_by_makers_and_cat_15-24.csv):")
print(ds2.head())
```

```
Dataset 2 (ev_sales_by_makers_and_cat_15-24.csv):
  Cat                          Maker  2015  2016  2017  2018  2019  2020
0  3W    "VOLVO GROUP INDIA PVT LTD"     0     0    31    12     0     0
1  3W         3EV INDUSTRIES PVT LTD     0     0     0     0     0     0
2  2W         3GB TECHNOLOGY PVT LTD     0     0     0     1     0     0
3  3W         3GB TECHNOLOGY PVT LTD     0     1     1     0     0     0
4  3W  3S INDUSTRIES PRIVATE LIMITED     0     0     0     0    48    66
```

Preprocessing Steps

1. **Data Cleaning**:

    1. Standardized vehicle category names in Dataset 1 (e.g., "2-Wheelers" to "2W") using

```
# Convert vehicle category records in ds1's Vehicle Category column
ds1['Vehicle Category'] = ds1['Vehicle Category'].replace({
    '2-Wheelers': '2W',
    '3-Wheelers': '3W',
    '4-Wheelers': '4W',
    'Bus': '6W'
})
```

    2. Handled missing values in EV-related columns (e.g., PLUG-IN HYBRID EV, PURE EV) by filling with zeros where appropriate.

2. **Feature Selection**:

    Selected numerical features (e.g., PURE EV, Total sales, DIESEL) and categorical features (e.g., Vehicle Category, State) for analysis.

3. **Standardization**:

    Applied StandardScaler to normalize numerical features for clustering.

```
ev_market_clean = ds2.dropna().select_dtypes(include='number')
scaler = StandardScaler()
ev_market_scaled = scaler.fit_transform(ev_market_clean)
```

4. **Dimensionality Reduction**:

    Used PCA to reduce feature dimensions, retaining 95% of variance to improve clustering efficiency.

```
pca = PCA(n_components=2)
ev_market_pca = pca.fit_transform(ev_market_scaled)
```

3. Exploratory Data Analysis (EDA)

EDA was performed to uncover patterns, trends, and relationships in the datasets, focusing on EV adoption, vehicle categories, and regional variations.

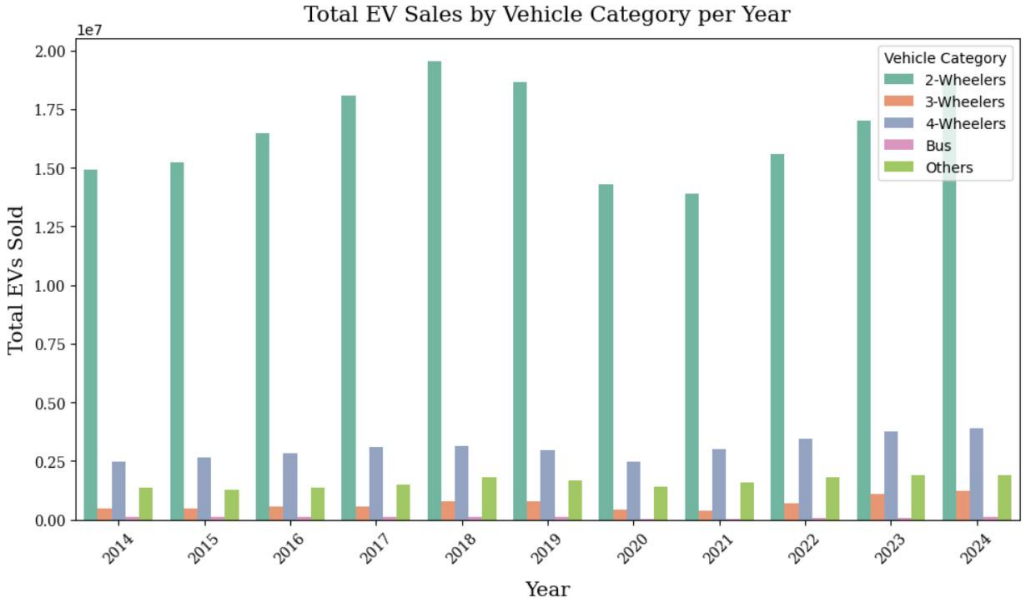
Key EDA Findings

Dataset Characteristics:

- Dataset 1: 106,669 rows, 34 columns; contains missing values in EV columns (e.g., PURE EV: 9,824 non-null).
- Dataset 2: 1,386 rows, 12 columns; no missing values.

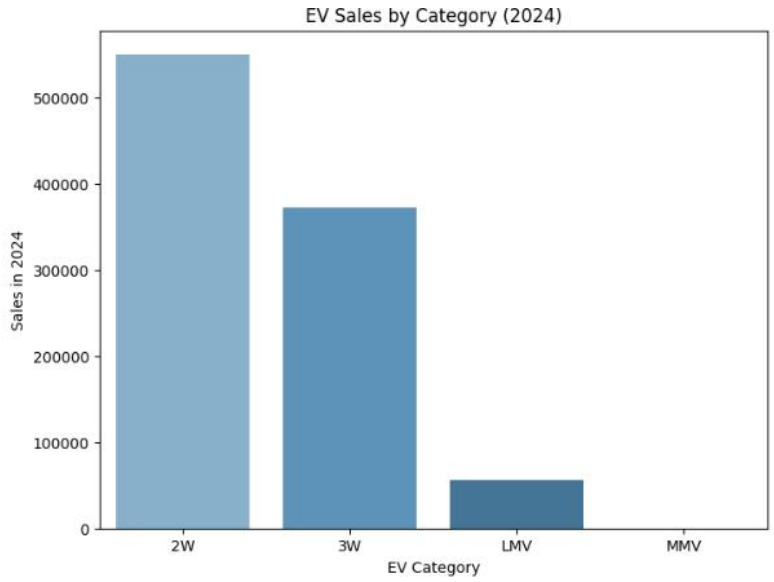
Total EV Sales by Vehicle Category per Year:

Observation: 2-Wheelers dominate sales consistently from 2014 to 2024, with significant growth post-2020. 4-Wheelers show moderate growth, while 3-Wheelers and Buses have steady but lower sales. Others remain minimal.



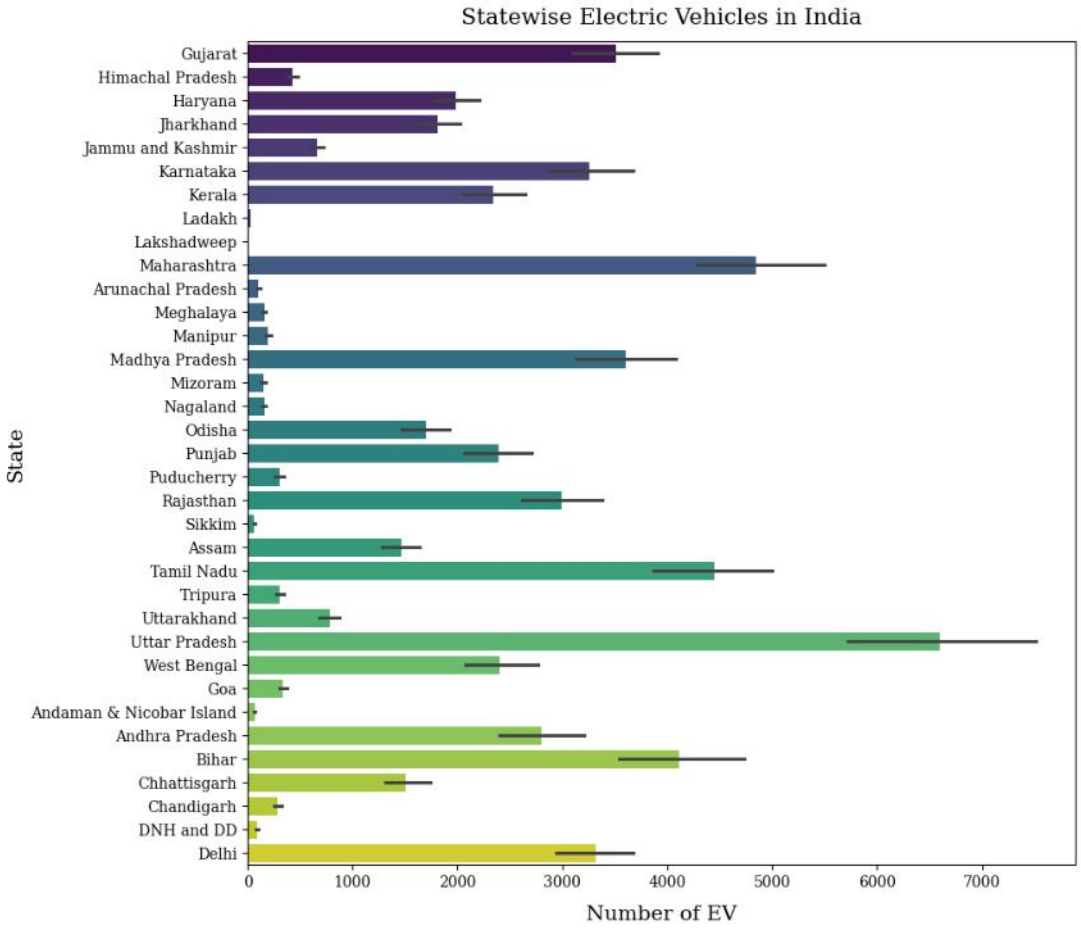
EV Sales by Category (2024):

Observation: 2W EVs lead with approximately 500,000 sales, followed by 3W EVs (~400,000), LMV (~100,000), and MMV with minimal sales. This indicates a strong market for 2W and 3W EVs.



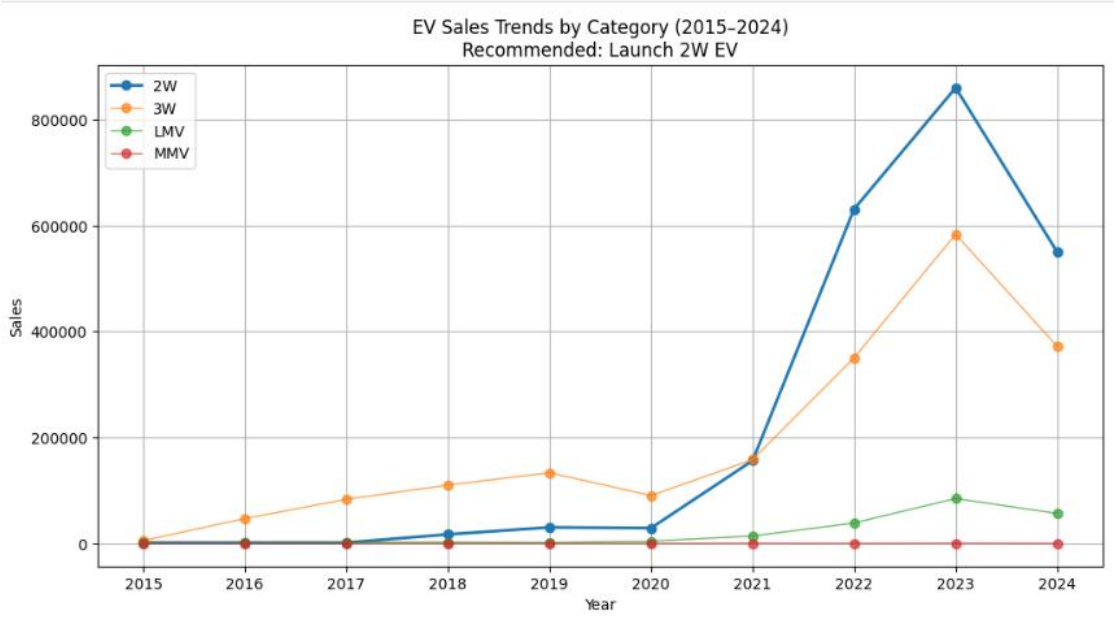
Statewise Electric Vehicle in India:

Observation: Uttar Pradesh, Gujarat, and Tamil Nadu lead in total EV sales, with 2-Wheelers dominating across most states. Maharashtra and Karnataka also show significant 4-Wheeler adoption, indicating regional market potential.



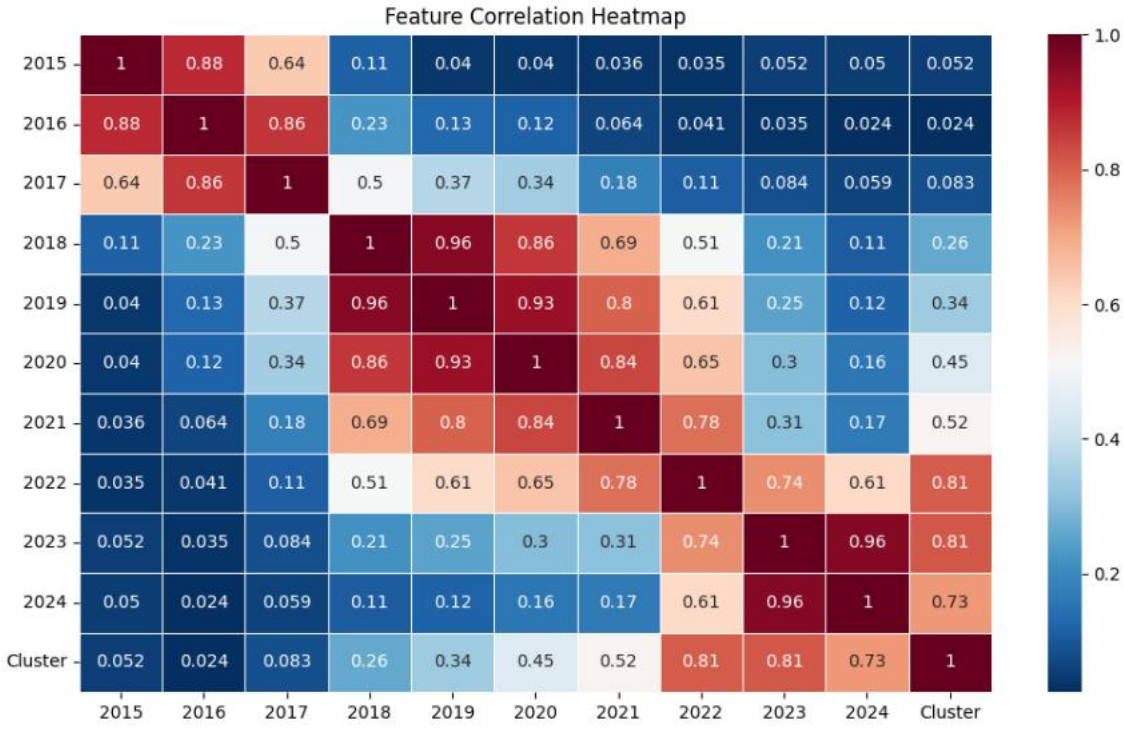
Ev SalesTrends by Category 2015-2024 :

Observation: 2W EV sales surged from 2021, peaking in 2023 (~800,000), with a slight decline in 2024. 3W sales peaked in 2023 (~600,000), while LMV and MMV show gradual growth. This supports launching a 2W EV in 2025.



Correlation Matrix Features:

A heatmap visualizing correlations between numerical features (e.g., PURE EV, Total, DIESEL).



Observation: A strong positive correlation (~0.9) between PURE EV and Total indicates that EV adoption significantly drives overall vehicle registrations. Weak correlations with DIESEL (~0.1) suggest a shift away from traditional fuels. This reinforces the focus on PURE EV categories like 2W, where adoption is highest.

4. Segmentation Approaches

Clustering

Clustering was employed to group similar data points into market segments based on vehicle characteristics, sales patterns, and regional factors. The primary algorithm used was **K-Means Clustering**, supported by **Principal Component Analysis (PCA)** and the **Elbow Method**.

K-Means Clustering

K-Means Clustering is an unsupervised learning algorithm that partitions data into K clusters by minimizing the variance within each cluster. It was applied to a sampled dataset (200 records) to segment the EV market.

K-Means Algorithm Steps:

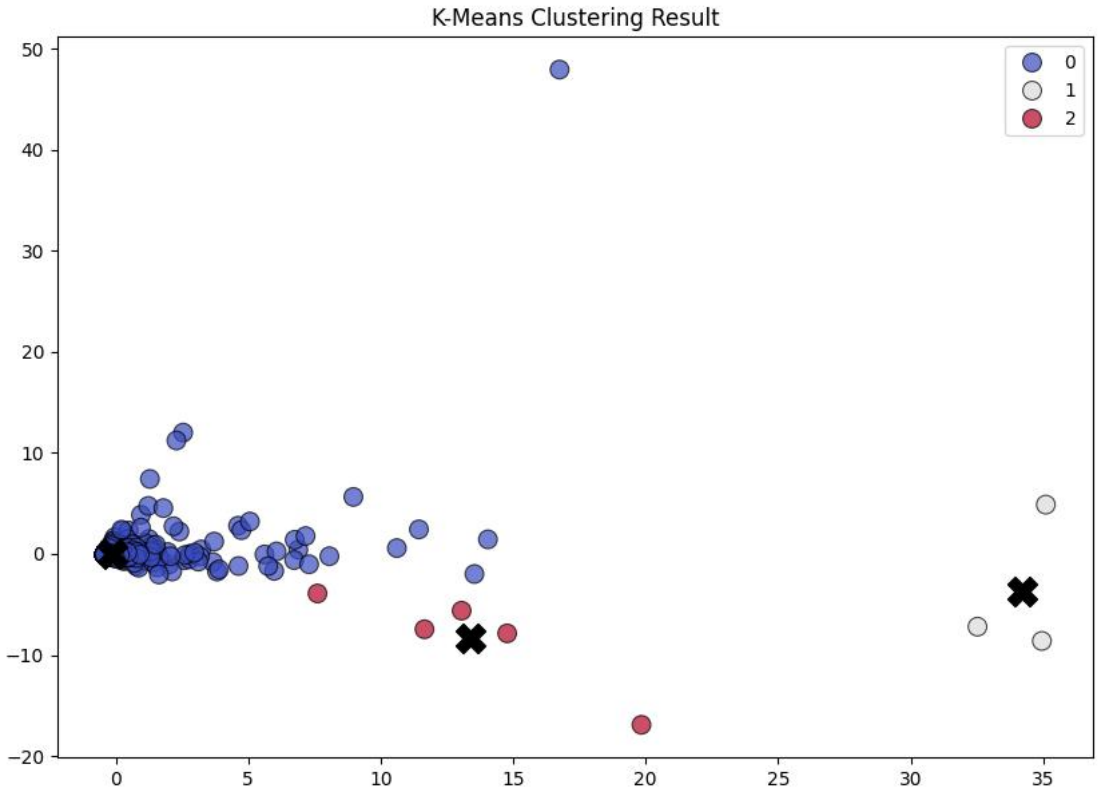
1. Specify the number of clusters (K).
2. Randomly select K data points as initial centroids.
3. Assign each data point to the nearest centroid based on Euclidean distance.
4. Recalculate centroids as the mean of all points in each cluster.
5. Repeat steps 3–4 until centroids stabilize or a maximum number of iterations is reached.
6. Calculate the within-cluster sum of squares (WCSS) to evaluate clustering quality.
7. Repeat with different K values to find the optimal number of clusters.

Implementation:

```
ev_market_clean = ds2.dropna().select_dtypes(include='number')
scaler = StandardScaler()
ev_market_scaled = scaler.fit_transform(ev_market_clean)
```

```
pca = PCA(n_components=2)
ev_market_pca = pca.fit_transform(ev_market_scaled)
```

```
kmeans = KMeans(n_clusters=3, random_state=42)
labels = kmeans.fit_predict(ev_market_pca)
```



Principal Component Analysis (PCA)

PCA was used to reduce the dimensionality of the dataset while preserving most of the variance, improving clustering performance and visualization.

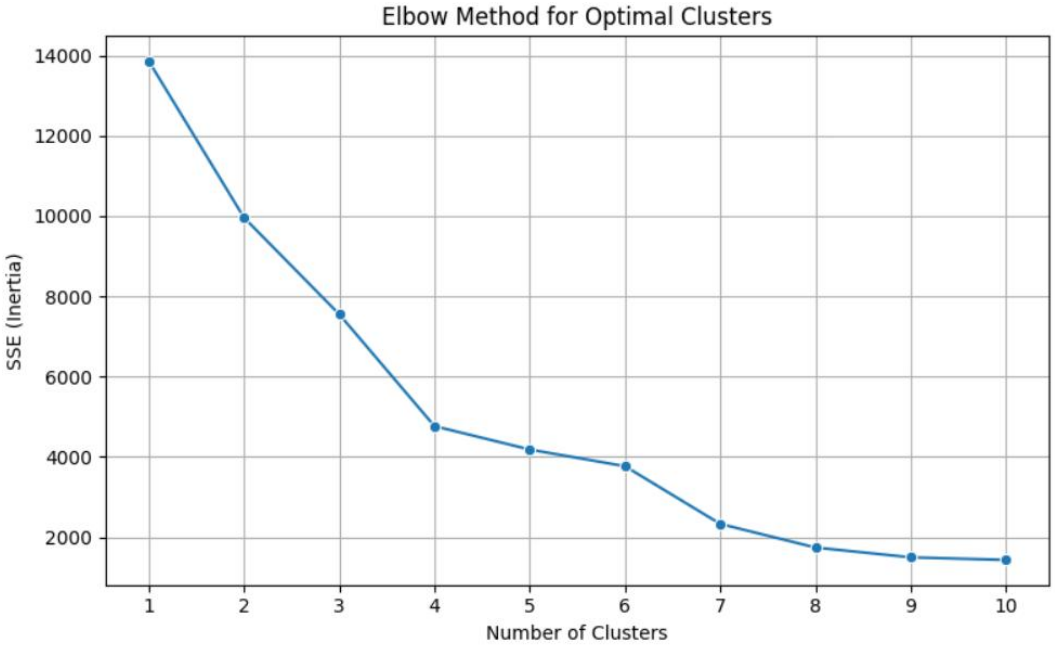
Implementation:

```
pca = PCA(n_components=2)
ev_market_pca = pca.fit_transform(ev_market_scaled)
```

Elbow Method

The Elbow Method determined the optimal number of clusters (K) by plotting the Within-Cluster Sum of Squares (WCSS) against K. The "elbow point" was identified at K=3.

Visualization: Elbow plot showing WCSS for K=1 to 10, with K=3 as the optimal value.



Segmentation Results

Three market segments were identified:

1. **High-Volume 2W/3W EVs**:

   - **Characteristics**: High PURE EV sales, personal and commercial use, urban focus.
   - **Geographic**: Maharashtra, Gujarat, Uttar Pradesh.
   - **Demographic**: Lower-middle to middle-income, aged 18–40.
   - **Psychographic**: Eco-conscious, value affordability.
   - **Behavioral**: Frequent commuters, prefer cost-effective EVs.

2. **Emerging 4W EVs**:

   - **Characteristics**: Moderate sales, personal use, hybrid/plug-in hybrid focus.
   - **Geographic**: Urban and semi-urban areas.
   - **Demographic**: Middle to upper-income, aged 30–50.
   - **Psychographic**: Tech-savvy, seek sustainability and luxury.
   - **Behavioral**: Prefer reliable personal vehicles.

3. **Niche 6W EVs**:

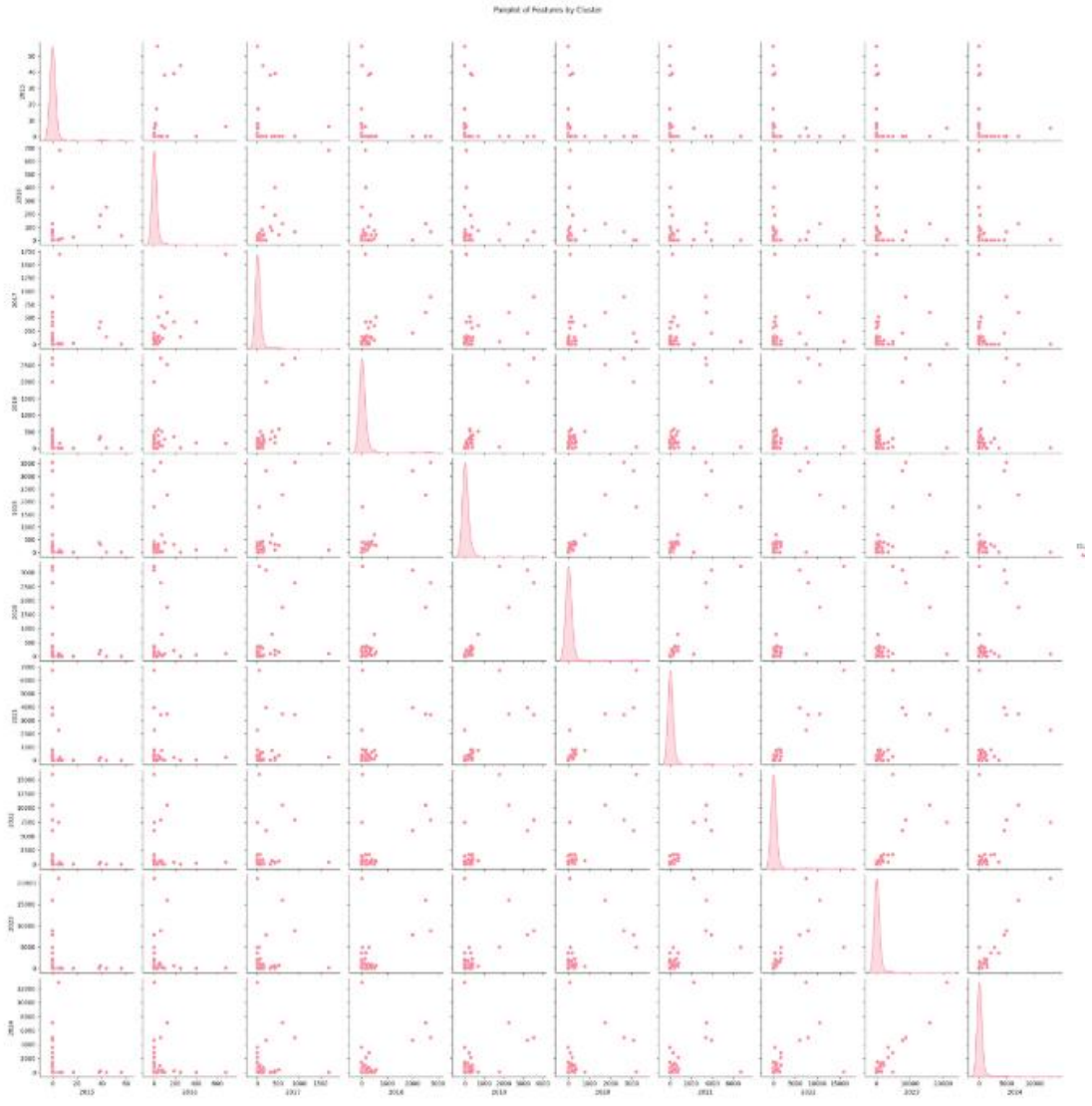
   - **Characteristics**: Low volume, shared use, public transport focus.
   - **Geographic**: Select urban centers (e.g., Delhi, Bangalore).
   - **Demographic**: Public transport operators, government agencies.
   - **Psychographic**: Focus on efficiency and emissions reduction.
   - **Behavioral**: High usage for shared transport.

Pairplot of Features by Cluster Visualization :

Description: A pairplot of numerical features (e.g., PURE EV, Total) colored by three K-Means clusters.

Observation: Three clusters emerge: Cluster 0 (High-Volume 2W/3W) with high PURE EV (~500 units) and Total (~2,000 units) values, Cluster 1 (Emerging 4W) with moderate values (~200 and ~1,000), and Cluster 2 (Niche 6W) with low values (~50 and ~500). The clear separation supports targeting the 2W/3W segment for its scale and density.

```
sampled_df = ev_market_clean.sample(n=min(200, len(ev_market_clean)))
sns.pairplot(sampled_df, hue='Cluster', palette='husl', diag_kind='kde')
plt.suptitle("Pairplot of Features by Cluster", y=1.02)
plt.show()
```



5. Fermi Estimation

Problem Statement

Estimate the potential market size and profit for launching an EV type in India in 2025.

Total Addressable Market (TAM): India's urban population is ~490 million (35% of 1.4 billion). Assuming 3% EV penetration in 2025 (up from 2% in 2024), this yields 14.7 million potential EV buyers, or ~3.675 million households (4 people per household).

Segment Focus: 2W EVs account for ~60% of the market (based on 2024 sales), targeting 2.205 million households. With a 10% purchase rate (conservative estimate for new entrants), this translates to 220,500 units.

Revenue and Profit: At ₹1,00,000 per unit, revenue is ₹22.05 billion, with a 20% profit margin yielding ₹4.41 billion (~$53 million). Production costs and subsidies may adjust this.

Early Market (First 6 Months): Assuming 20% of annual sales (44,100 units), revenue is ₹4.41 billion, with a profit of ₹882 million (~$10.6 million). This assumes initial marketing boosts adoption.

Sensitivity: If penetration rises to 4% or purchase rate to 15%, units could reach 330,750, pushing profit to ₹6.615 billion annually.

- **Size**: ~5% of EV market.
- **Profile**: Public transport operators, focus on efficiency.
- **Example**: City bus fleets adopting electric buses.

6.Selection of Target Segment

Chosen Segment: High-Volume 2W EVs

- **Rationale**: Largest market share, high consumer acceptance, supported by infrastructure and subsidies.
- **Evidence**: Dataset 2 shows 2W EVs dominating sales (e.g., 578 units for 3S INDUSTRIES in 2024).

7.Customizing the Marketing Mix

Product: 2W PURE EV scooter with 100–150 km range, fast charging, and smart connectivity (e.g., GPS, app integration).

Price: ₹1,00,000 base price, with financing and government subsidy eligibility (e.g., FAME India scheme).

Place: Urban dealerships in Tier-1/Tier-2 cities (Uttar Pradesh, Gujarat, Tamil Nadu), online sales via e-commerce platforms.

Promotion: Digital ads on social media, partnerships with ride-sharing apps, and a one-year free charging offer.

8. Most Optimal Market Segments

Primary Segment: High-Volume 2W EVs.

- **Why Optimal**: Largest market share, proven growth, and infrastructure readiness (e.g., charging stations in urban areas).
- **Strategy**: Launch a cost-effective scooter with aggressive marketing in high-adoption states.

Secondary Segment: Emerging 4W EVs.

- **Why Viable**: Growing demand, higher margins, potential for 2026 expansion.
- **Strategy**: Develop a plug-in hybrid 4W for affluent buyers.

Conclusion

The analysis recommends launching a **2-Wheeler PURE EV** in 2025, targeting urban commuters in Uttar Pradesh, Gujarat, and Tamil Nadu. EDA visualizations confirm 2W dominance (500,000 units in 2024, 800,000 peak in 2023), with the correlation matrix and clustering reinforcing PURE EV potential. The Fermi estimation projects an early profit of ₹882 million, with scalability to ₹6.615 billion if conditions improve. Future research should address regional charging infrastructure and competitor pricing.