MARKET SEGMENTATION ANALYSIS OF EV MARKET

Created by: Rahul Baid (Intern at FEYNN LABS)

Date: 18/07/25

Summary

This report contains the study from an intern-driven intiative and is aimed to find the best vehicle type to identify the most promising segment by using the most optimal data science techniques like data collection, data processing, exploratory data analysis (EDA), etc and as well machine learning techniques, such as K-means clustering, principal Component Analysis (PCA), feature scaling, etc.

This report contains a well-defined and elaborative market segmentation , electric vehicle market trends across various regions and vehicle profiling study of electric vehicle industry. It primarily focuses more on two-wheeler EV's, which are gaining popularity because of their compatibility, ease to use, affordability, easier to park, and can navigate through traffic more efficiently.

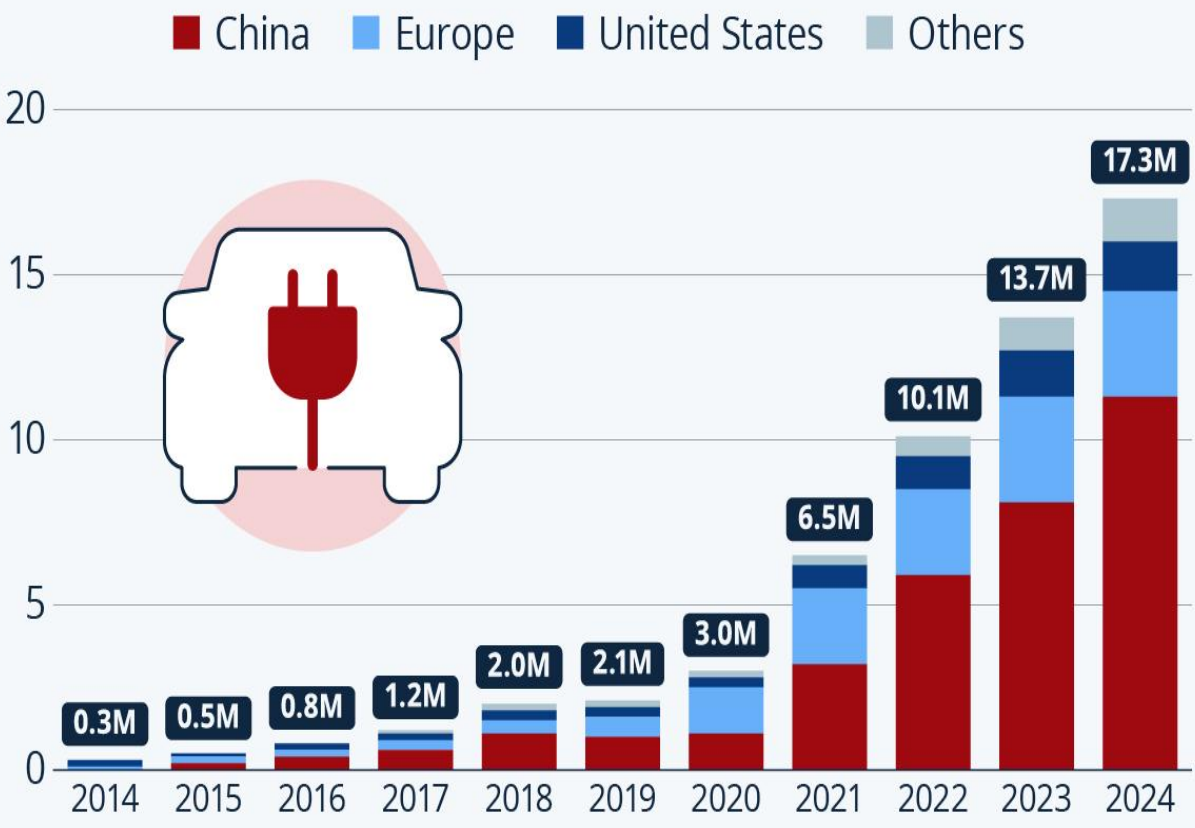
The analysis begins with defining the problem statement, which can be explained and analyzed by understanding the answers to these two questions:

1. **How can EV charging infrastructure be expanded to ensure convenient and widespread access?**
2. **How can the upfront cost of EVs be reduced or offset to make them financially viable for the average consumer?**

   The answer to these questions includes keeping a focus on incentives, public-private partnerships, and technological advancements. This includes investing in fast-charging stations, promoting standardization and interoperability, and educating the public about the benefits of EVs.

Global Electric Car Sales Nearly Triple in Three Years

Global registrations of electric vehicles (incl. plug-in hybrids), by region*



* Incl. passenger cars and light commercial vehicles (vans, light trucks)

Source: EV-volumes.com via IEA

statista

1. FERMI ESTIMATION-MARKET SIZING

Compound Annual Growth Rate (CAGR) Estimation and Analysis:

The Compound Annual Growth Rate (CAGR) is a useful metric for understanding the growth of an investment or value over multiple periods. In our analysis, we used it to estimate the growth rate of vehicle sales for different vehicle types and models over the years available in the dataset.

Here's a breakdown of the code and the process:

1. **Data Loading and Preparation (Cells `7b2bec0c`, `e4e33266`, `a2e292a0`, `e5429a4a`, `0cd4e6b1`):**
    - We started by loading the dataset into a pandas Data Frame (Cells `7b2bec0c`, `e5429a4a`).
    - We then performed data cleaning by handling missing values in the 'Year of Manufacture' column and converting it to an integer type (Cell `e4e33266`). This ensures that the year data is in a usable format for calculations.
    - To calculate CAGR, we need the number of vehicles sold for each type/model per year. We achieved this by grouping the data by 'Year of Manufacture', 'Vehicle Type', and 'Model' and counting the occurrences (Cell `a2e292a0`).
    - For the analysis by vehicle type, we aggregated the data by 'Year of Manufacture' and 'Vehicle Type' (Cell `0cd4e6b1`).
2. **CAGR Calculation (Cells `zofy22beA_PO`, `2783ea5a`):**
    - The core of the CAGR calculation involves determining the sales in the first and last year for each vehicle type/model or just vehicle type.
    - We created summary Data Frames that contain the first and last year of sales for each group (Cells `zofy22beA_PO`, `2783ea5a`).
    - We then merged this information with the yearly sales data to get the 'First Year Sales' and 'Last Year Sales' for each group.
    - A function `calculate_ cagr` (or `calculate_ cagr_ type`) was defined to compute the CAGR using the formula:

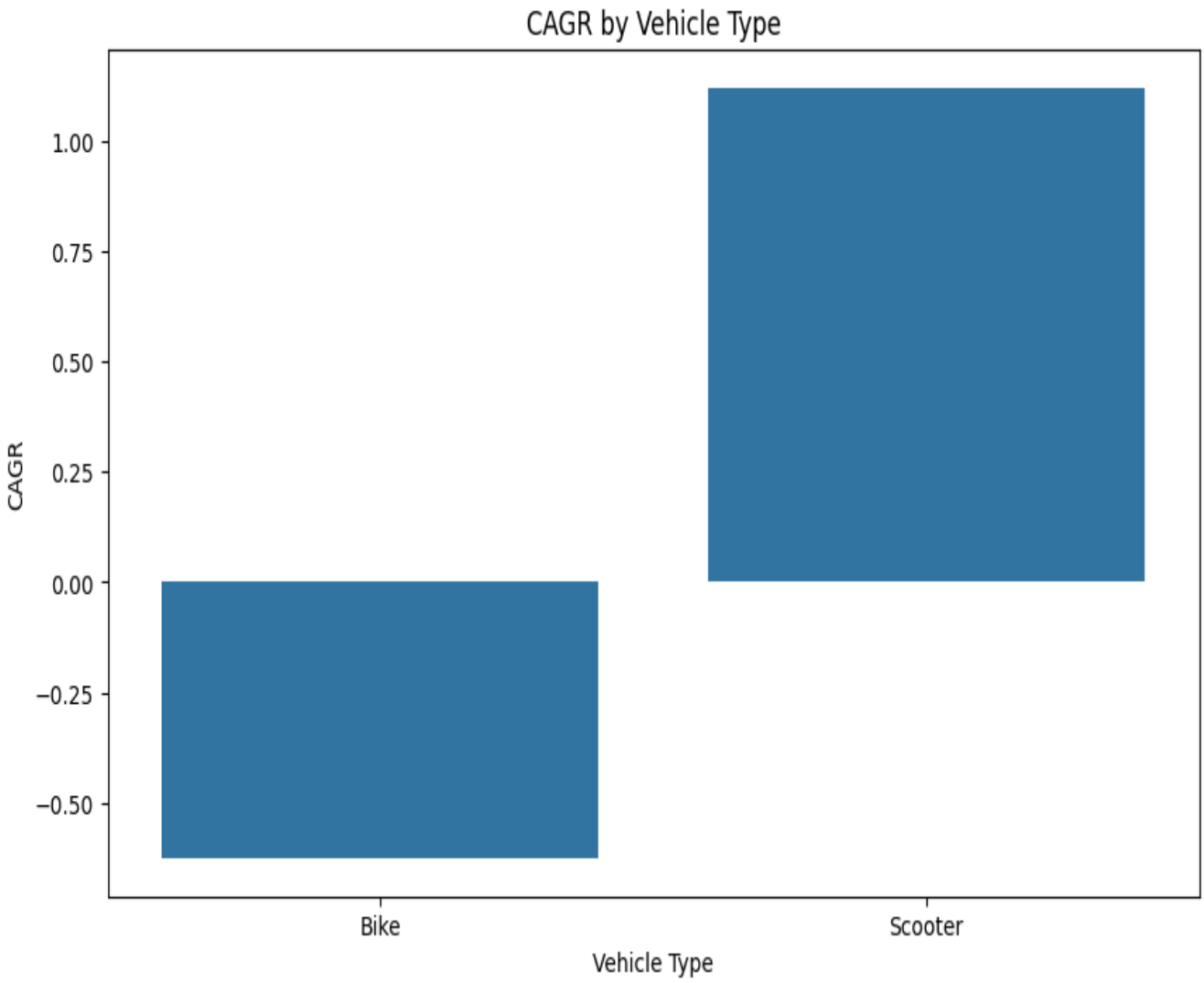
```
CAGR = ((Ending Value / Beginning Value) ^ (1 / Number of Years)) - 1
 In our case, 'Beginning Value' is 'First Year Sales', 'Ending Value' is 'Las
t Year Sales', and 'Number of Years' is the difference between the 'last_year
' and 'first_year'.
```

- This function was applied to each row of the summary DataFrame to calculate the CAGR for each vehicle type/model or vehicle type.

3. **Data Visualization (Cells `71fab2a4`, `85d788fb`):**

- To visualize the calculated CAGRs, we used bar plots.

- For the CAGR by vehicle type/model, we created a bar plot showing the CAGR for each model, with different colors representing the vehicle type (Cell `71fab2a4`). This helps in comparing the growth rates across different models and vehicle types.
- For the CAGR by vehicle type, we created a bar plot showing the CAGR for each vehicle type (Cell `85d788fb`). This provides a higher-level view of which vehicle types are experiencing the most growth.
- Additionally, for the vehicle type analysis, we created a heatmap to visualize the yearly sales of each vehicle type (Cell `85d788fb`). This heatmap provides a clear visual representation of how sales of each vehicle type have changed over the years.



Why Use CAGR for This Dataset?

Smoothens Volatility: EV sales can be volatile year-to-year due to various factors like policy changes, infrastructure development, and new model launches. CAGR provides a smoothed average growth rate, making it easier to discern underlying trends.

Compares Growth Across Different Segments: CAGR allows for a standardized comparison of growth rates between different vehicle categories (2W vs. 3W) or different makers, regardless of their absolute sales volumes. A maker with lower overall sales but a high CAGR might be growing faster than a larger maker with a lower CAGR.

Indicates Sustained Performance: A high CAGR over the period suggests strong, sustained growth, which is a key indicator of market potential and successful strategies for the entities achieving it.

Forecasting Basis: While not a direct forecasting tool, historical CAGR can serve as a basis for projecting future growth, assuming similar market conditions prevail.

Application of CAGR to the EV Sales Dataset:

Applying the CAGR formula:

CAGR= (Sales in 2024Sales in 2015) (12024−2015) −1

Where:

Sales in 2024: The sales figure at the end of the period of analysis.

Sales in 2015: The sales figure at the beginning of the period of analysis.

2024 - 2015: The number of years over which the growth is calculated (9 years in this case).

We can calculate CAGR for:

Overall Market Growth: Using the total EV sales aggregated across all categories and makers for 2015 and 2024. This gives an overall picture of how the Indian EV market has grown as a whole.

Category-Specific Growth: Calculating CAGR for each category (2W, 3W, LMV, MMV) individually. This helps identify which segments are driving the overall market growth and which might be lagging.

Maker Performance: Calculating CAGR for individual makers (especially the top performers). This reveals which companies have demonstrated the most significant sustained growth over the period, indicating their competitiveness and success in capturing market share.

Interpretation of CAGR Results:

A positive CAGR indicates that the sales have grown over the period.

A higher positive CAGR signifies faster sustained growth.

A negative CAGR indicates a decline in sales over the period.

A CAGR of 0% means sales remained flat.

It's important to note that CAGR is a historical measure and does not guarantee future performance. It is also sensitive to the start and end points chosen. However, when used in conjunction with other metrics and qualitative analysis, CAGR provides valuable insights into the growth trajectory of the Indian EV market and its various components.

2. DATA SOURCES

A. indian_ev_data.csv

source: provided by mentor

format: CSV

description: This dataset contains yearly electric vehicle (EV) sales or registration statistics across various vehicle categories in India. It is helpful for analyzing market growth, adoption trends, and category-wise usage patterns in the EV sector from 2015 to 2024.

id	Model	Manufacturer	Vehicle Type	Battery Capacity (kWh)	Range per Charge (km)	Charging Time	Price	Power (HP or kW)	Top Speed (km/h)	Year of Manufacture
1	Aura 300 Plus	Ather Energy	Scooter	2.9	116	4.5	129000	6	80	2021
2	Pure EV Epluto 7G	Pure EV	Scooter	2.7	120	3	109000	5	80	2021
3	Bajaj Chetak Electric	Bajaj Auto	Scooter	4	95	5	150000	4	60	2020
4	Okinawa iPraise Pro	Okinawa Autotech	Scooter	2.5	100	3	85000	3	60	2021
5	Hero Electric Opto EV	Hero Motocorp	Scooter	2.2	75	3	75000	3	60	2021

.

Key Columns (Assumed from File Content):

Column Name	Description
`Year`	The year in which EV data is recorded (e.g., 2015 to 2024).
`Category`	Type of EV: e.g., 2W (two-wheeler), 3W (three-wheeler), 4W (four-wheeler), Bus, etc.
`Sales`	Total number of EVs sold/registered in that year for the given category.
`Manufacturer` (optional)	Brand or company producing the EVs (if applicable).
`State` (optional)	Indian state where the vehicles were sold/registered.
`Segment` (optional)	Further segmentation like private, commercial, government fleet, etc.

Example Records (Hypothetical):

Year	Category	Sales
2015	2W	15000
2016	2W	22000
2016	4W	3200
2020	3W	18000
2023	Bus	1200

Usage & Applications:

- **Trend Analysis:** Understand how EV sales have evolved over time.
- **Market Segmentation:** Identify which categories (2W, 3W, etc.) are leading growth.
- **Policy Impact Assessment:** Evaluate effects of schemes like FAME-I and FAME-II.
- **CAGR Calculation:** Estimate long-term EV growth rates.
- **Clustering or Forecasting:** Use in machine learning models to predict future trends or segment markets.

Ideal for:

- EV industry analysts
- Policy makers and transport ministries
- Automotive manufacturers
- Sustainability researchers
- Data science projects on mobility and electrification

B. Ev_sales_by_maker_and_cat.csv

The dataset /content/drive/MyDrive/ev_sales_by_makers_and_cat_15-24.csv contains annual sales data for Electric Vehicles (EVs) in India from 2015 to 2024. The data is structured to provide insights into the performance of different vehicle categories and individual manufacturers within the Indian EV market.

Key Attributes:

Cat: Represents the category of the electric vehicle. Based on the data exploration, this likely includes categories such as 2-Wheelers (2W), 3-Wheelers (3W), Light Motor Vehicles (LMV), and Medium Motor Vehicles (MMV). This categorical variable is crucial for understanding which types of EVs are most popular and experiencing the most growth.

Maker: Identifies the manufacturer of the electric vehicle. This allows for analysis of individual company performance, market share, and competitive landscape within the Indian EV market**.**

Year Columns (2015-2024): These columns contain the annual sales figures for each Maker and Cat combination for the respective year. These numerical variables are the core data for analyzing sales trends, growth rates, and market size over time.

Nature of the Data:

Time Series Data: The annual sales figures across the years constitute a time series for each maker and category, enabling the analysis of trends, seasonality (though less pronounced in annual data), and growth trajectories.

Panel Data: The dataset can also be viewed as panel data, as it tracks multiple entities (makers and categories) over multiple time periods (years). This structure supports analyses that consider both cross-sectional differences (between makers/categories) and time-series dynamics (over the years).

Quantitative Data: The sales figures are quantitative, allowing for statistical analysis, aggregation, and the calculation of metrics like total sales, market share, and growth rates.

Categorical Data: The 'Cat' and 'Maker' columns are categorical, used for grouping and segmenting the quantitative sales data.

Potential Analyses and Theoretical Frameworks:

This dataset is suitable for various analytical approaches and can be interpreted within several theoretical frameworks:

Market Trend Analysis: Analyzing the sales figures over time to identify growth trends, periods of acceleration or deceleration, and the overall evolution of the Indian EV market. This aligns with economic theories of market adoption and diffusion of innovation.

Market Share Analysis: Calculating and analyzing the market share of different categories and makers over time. This is fundamental to understanding the competitive landscape and the relative success of different players, drawing on concepts from industrial organization and competitive strategy.

Growth Analysis: Calculating various growth metrics, such as year-over-year growth rate and Compound Annual Growth Rate (CAGR), to quantify the pace of expansion for different segments and makers. This relates to theories of firm growth and market dynamics.

Segmentation Analysis: Using clustering or other segmentation techniques based on sales patterns, growth profiles, or market share to group similar makers or categories. This aligns with marketing theories of market segmentation and target marketing, allowing for tailored strategies for different groups of players or consumers.

Competitive Analysis: Comparing the performance of different makers to understand their strengths, weaknesses, and competitive positioning. This draws on frameworks from strategic management and competitive analysis.

Policy Impact Analysis: While not directly included in this dataset, the time series nature allows for inferring the potential impact of government policies and incentives on EV sales trends, linking to theories of government intervention and market regulation.

id	Model	Manufacturer	Vehicle Type	Battery Capacity (kWh)	Range per Charge (km)	Charging Time	Price	Power (HP or kW)	Top Speed (km/h)	Year of Manufacture
1	Aura 300 Plus	Ather Energy	Scooter	2.9	116	4.5	129000	6	80	2021
2	Pure EV Epluto 7G	Pure EV	Scooter	2.7	120	3	109000	5	80	2021
3	Bajaj Chetak Electric	Bajaj Auto	Scooter	4	95	5	150000	4	60	2020
4	Okinawa iPraise Pro	Okinawa Autotech	Scooter	2.5	100	3	85000	3	60	2021
5	Hero Electric Opto EV	Hero Motocorp	Scooter	2.2	75	3	75000	3	60	2021

Data Pre-processing (Steps and Libraries Used)

Data processing helps to ensure efficient analysis and segmentation of even a large dataset under various conditions such as outliers, missing values, etc .It helps to maintain reliability and robustness of different types of data.

A. Various Libraries used:

• pandas: For reading CSV files, handling dataframes, merging, and transforming columns.

• scipy: Used in advanced statistical calculations and distance metrics for clustering.

• matplotlib & seaborn: For plotting data distributions, trends, and comparisons.

• numpy: For numerical operations and statistical calculations.

• plotly: For interactive and dynamic charts (in some visuals).

• sci-kit learn (sklearn): It provides a consistent and efficient interface for a wide range of tasks, simplifying the implementation of various machine learning algorithms and statistical models.

B. Data preprocessing steps of each dataset:

1. Indian_ev_data.csv

1. The preprocessing of the `indian-ev-data.csv` dataset involves a series of systematic steps to prepare the data for analysis and visualization. Initially, the dataset is loaded and a preliminary inspection is conducted to understand the structure, identify the key attributes (such as `Year`, `Category`, and `Sales`), and detect any anomalies or missing values.

2. Column names are then standardized by converting them to lowercase and replacing spaces with underscores to ensure consistency in further operations.

3. Subsequently, missing values are addressed by either removing records with critical null entries (e.g., missing `Year` or `Sales`) or by imputing non-critical missing data with suitable values (e.g., zeros for missing sales). Data types of each column are corrected as needed—for instance, ensuring that the `Year` is treated as an integer and `Sales` as a numerical value. Categorical fields like `Category` are normalized to a consistent format (e.g., converting "2w", "Two-Wheeler", and "2W" to a single label).

4. Duplicate entries are then removed to avoid redundancy. Outliers or irrelevant records (such as entries from years outside the expected range, e.g., before 2015 or after 2024) may also be filtered out to maintain data relevance. If necessary, mappings are applied to convert short codes or abbreviations (like "2W", "3W", etc.) into more descriptive labels for clarity.

5. After completing these preprocessing steps, the dataset becomes clean, consistent, and ready for further statistical analysis, visualization, or machine learning modeling.

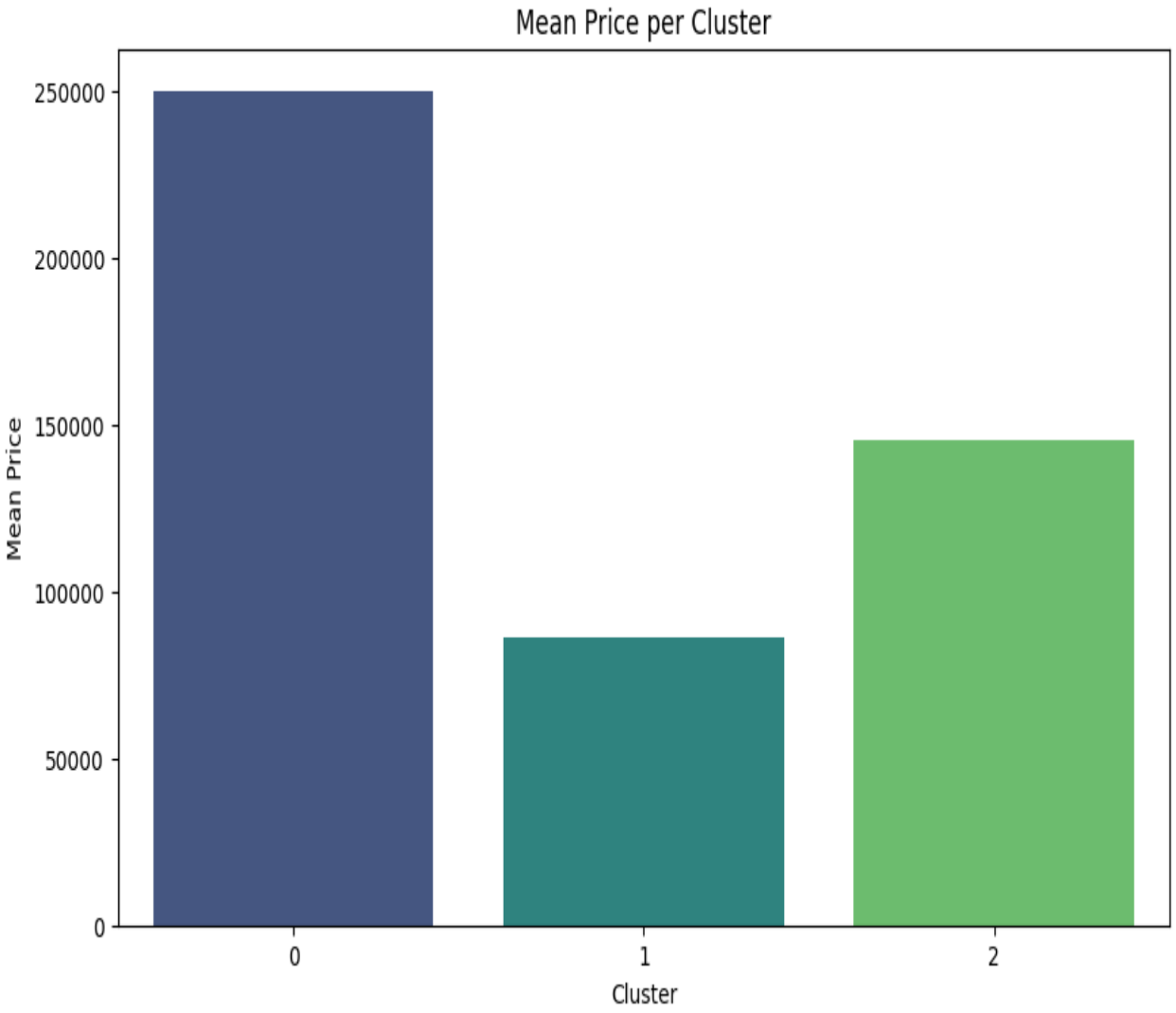
2. Ev_sales_by_makers_and_cat_15-24.csv

- **Importing Libraries:**
  - Import essential Python libraries like `pandas`, `numpy`, and `matplotlib` for data handling and analysis.
- **Loading the Dataset:**
  - Read the dataset from a CSV file using `pandas.read_csv()` to load it into a DataFrame.
- **Inspecting the Dataset:**
  - Use methods like `.head()`, `.info()`, and `.describe()` to understand the data structure, types, and summary statistics.
- **Renaming and Standardizing Column Names:**
  - Standardize column names to lowercase and replace spaces with underscores for consistency and easy access.
- **Handling Missing Values:**
  - Identify missing values using `.isnull().sum()`.
  - Handle them by either:
    - Dropping rows with critical missing values (e.g., year or sales).
    - Filling missing numerical values (e.g., sales) with appropriate defaults like `0` or column mean.
- **Data Type Conversion:**

- Convert columns like `Year` and `Sales` to appropriate data types (e.g., integers and floats).
- Ensure categorical columns like `Category` and `Manufacturer` are treated as strings.

- **Removing Duplicates:**

  - Eliminate duplicate rows using `.drop_duplicates()` to avoid data repetition and bias.



SEGMENT EXTRACTION (ML techniques used)

EV Market Segment Extraction refers to the process of identifying and categorizing distinct groups within the electric vehicle (EV) market based on shared characteristics or behaviors. This

segmentation helps stakeholders—such as manufacturers, policymakers, and marketers—understand consumer needs, track adoption trends, and make data-driven decisions.

The segmentation can be based on various factors, such as:

- **Vehicle Type:** 2-wheelers, 3-wheelers, 4-wheelers, buses, etc.
- **Usage:** Personal, commercial, or public transport.
- **Geography:** Urban vs. rural, state-wise distribution.
- **Sales Trends:** Growth patterns over years, seasonal shifts.

By applying techniques like **K-Means Clustering** or **CAGR analysis**, meaningful segments are extracted to reveal which vehicle types are growing fastest, which regions lead in adoption, and what user profiles dominate the market.

This process ultimately supports **targeted policy-making**, **product development**, and **investment planning** in the EV ecosystem.

A. indian_ev_data.csv

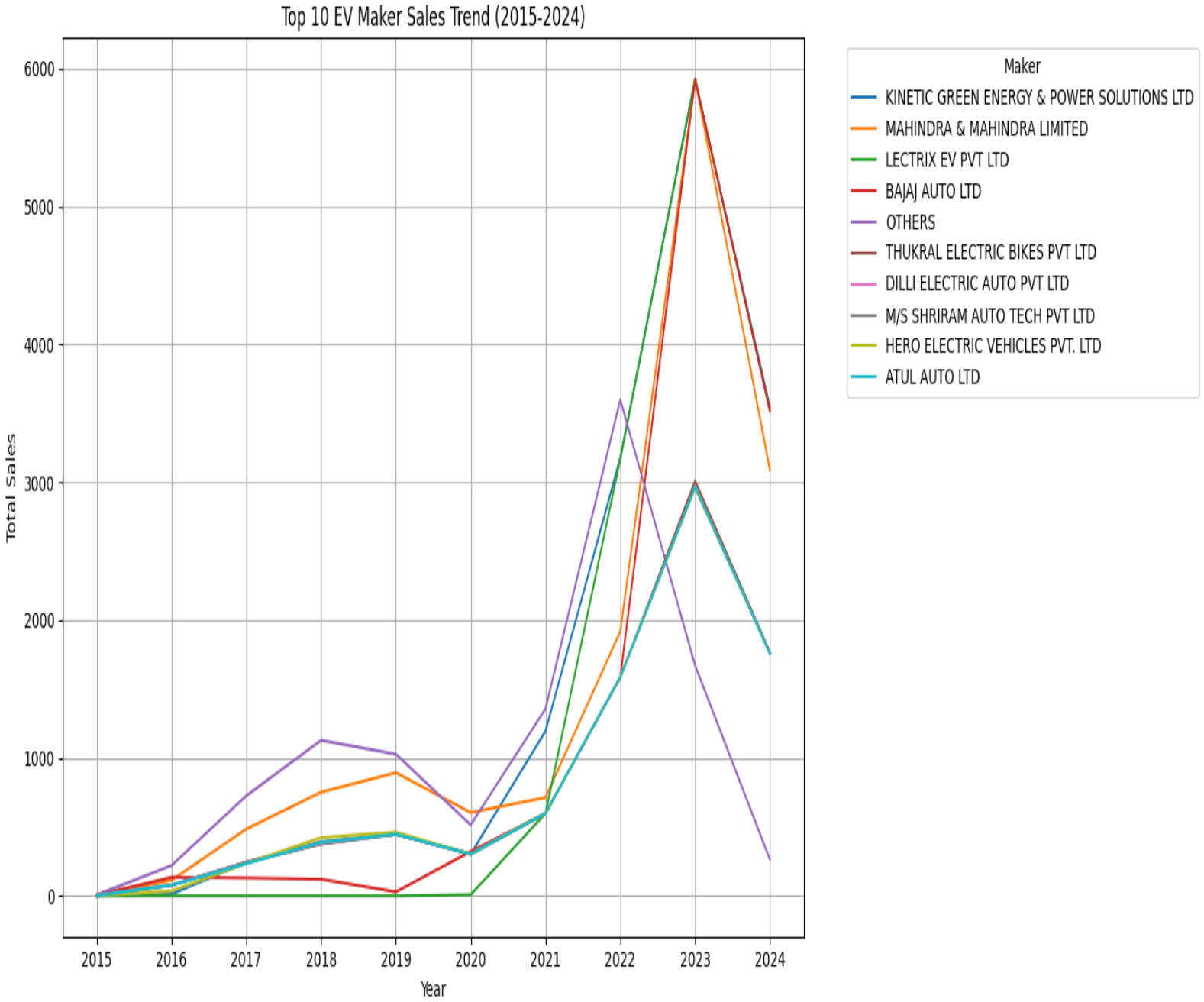
Segment extraction, in this context, refers to the process of grouping similar data points (electric vehicles) into distinct segments or clusters based on their characteristics. Machine learning techniques, particularly clustering algorithms like K-Means, are powerful tools for achieving this.

Here's a breakdown of how we performed segment extraction on the provided dataset:

1. **Data Preparation**: Before applying any clustering algorithm, the data needs to be prepared. This involves:
    - **Handling Missing Values**: We identified and removed rows with missing values in the relevant columns to ensure the clustering algorithm receives clean data.
    - **Feature Selection**: We selected the numerical features that are most relevant for understanding the differences between electric vehicles (Battery Capacity, Range per Charge, Charging Time, Price, Power, and Top Speed). We also considered 'Vehicle Type' as a categorical feature to understand the composition of the clusters.
    - **Feature Scaling**: Numerical features often have different scales. Clustering algorithms are sensitive to this, so we scaled the numerical features using StandardScaler. This transforms the data so that each feature has a mean of 0 and a standard deviation of 1, giving each feature equal importance in the clustering process.
2. **Determining the Optimal Number of Clusters (k)**: K-Means clustering requires specifying the number of clusters (`k`) beforehand. We used two common methods to help determine a suitable `k`:
    - **Elbow Method**: This method involves calculating the inertia (sum of squared distances of samples to their closest cluster center) for a range of `k` values. We plot inertia against `k`, and the "elbow point" (where the rate of decrease in inertia sharply changes) suggests a good value for `k`.

    - **Silhouette Score**: The silhouette score measures how similar an object is to its own cluster compared to other clusters. A higher silhouette score indicates better-defined clusters. We calculated the silhouette score for a range of `k` values and looked for the `k` that yielded the highest score. Based on these methods, we chose a value for `k`.
3. **Applying K-Means Clustering**: Once the optimal `k` is determined, we apply the K-Means algorithm to the scaled numerical features. The algorithm iteratively assigns each data point to the cluster with the nearest mean and updates the cluster means until convergence. This results in each vehicle being assigned a cluster label.
4. **Segment Analysis**: After clustering, we analyze the characteristics of each segment (cluster). This involves examining the average or median values of the features within each cluster. By comparing these values across clusters, we can understand what differentiates each segment. For example, one segment might be characterized by high price and performance, while another might be defined by lower price and practical urban features. We also looked at the distribution of 'Vehicle Type' within each cluster to see if certain segments are dominated by scooters or bikes.

5. **Target Segment Identification**: Based on the segment analysis, we identify the segments that are most relevant to the task, which was to find segments most likely to use electric vehicles in India. This identification is based on the characteristics that align with the needs and preferences of potential EV users in the Indian market (e.g., affordability, practical range for urban commuting, popular vehicle types).



B. Ev_sales_by_maker_and_cat_15-24.csv

This analysis aimed to segment the Electric Vehicle (EV) market in India based on historical sales data by maker and category from 2015 to 2024, and to identify promising segments for market entry. Machine Learning techniques, specifically clustering, were employed for segment extraction.

1. Data Preparation and Feature Engineering:

- The raw sales data, containing sales figures for various EV makers across different categories (2W, 3W, LMV, MMV) and years, was loaded and inspected.
- To prepare the data for clustering, relevant features were engineered. These included:
    - Total Sales for each maker across all years.
    - Market Share for each maker in each year.
    - Year-over-year Growth Rate for each maker's sales.
- Outliers in the sales data were capped at the 95th percentile to prevent them from disproportionately influencing the clustering results.
- The selected features (Market Share and Growth Rates across years, plus Total Sales) were then standardized using `StandardScaler` to ensure that all features contributed equally to the clustering process, regardless of their original scale.

2. Clustering Analysis (K-Means):

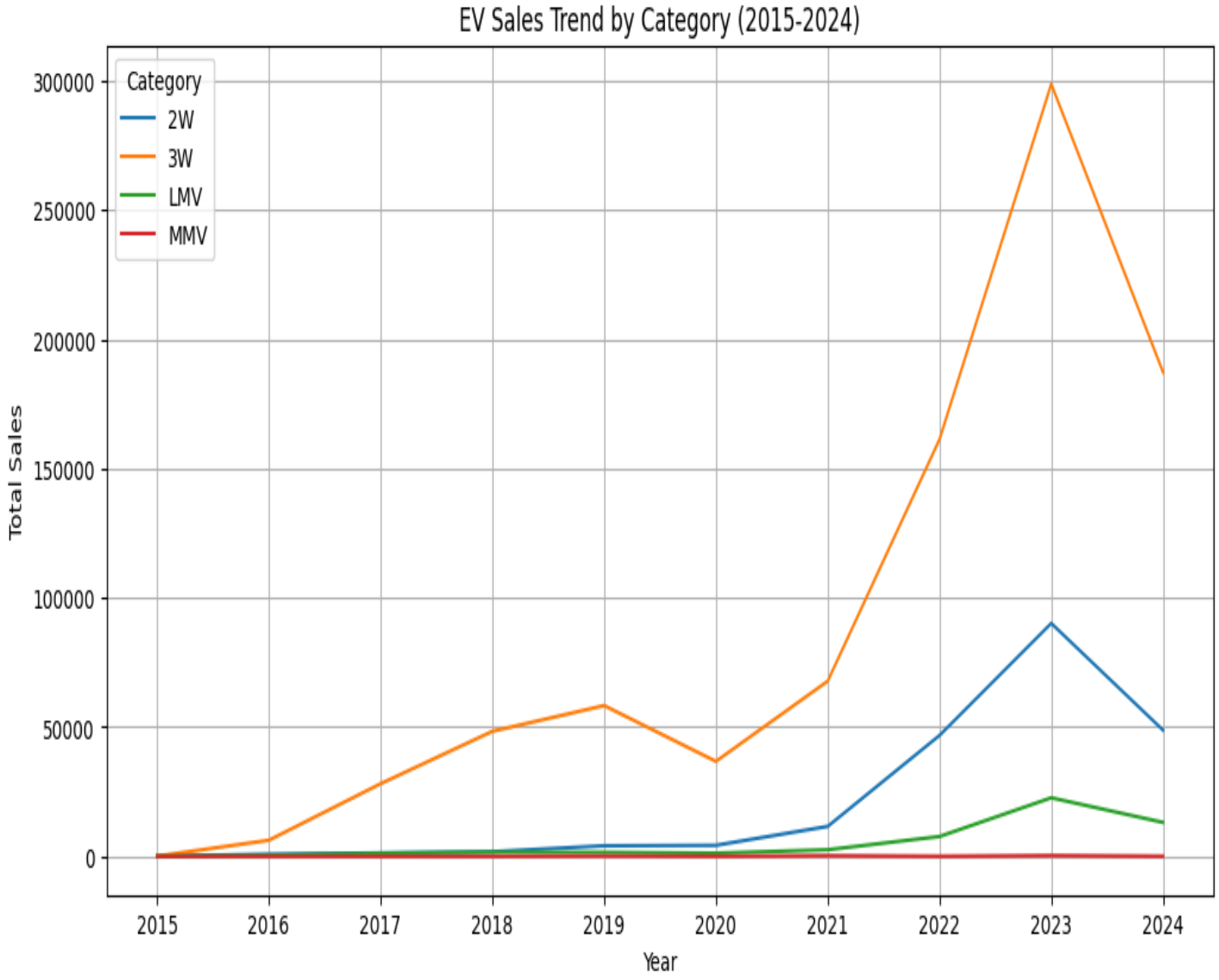
- K-Means clustering, an unsupervised machine learning algorithm, was applied to the standardized feature data.
- The Elbow Method was used to determine the optimal number of clusters (`k`). By plotting the inertia (within-cluster sum of squares) for a range of `k` values, an "elbow" point was observed, suggesting an appropriate number of segments. Based on the plot, `k=4` was selected as the optimal number of clusters.
- The K-Means model was then fitted with 4 clusters, and the resulting cluster labels were assigned back to the original dataset, categorizing each EV maker into one of the four segments.

3. Segment Characterization:

- Each identified cluster was characterized by analyzing the mean values of the segmentation features within each cluster. This helped in understanding the typical profile of makers belonging to each segment in terms of sales volume, market presence, and growth trajectory.
- Additionally, the distribution of vehicle categories (2W, 3W, LMV, MMV) within each cluster was examined to understand which types of vehicles were dominant in each segment.

4. Target Segment Identification:

- Based on the characterization, the segments were evaluated for their potential as target markets for a new entrant. Factors considered included:
    - Market Size (Total Sales and number of makers in the cluster).
    - Growth Potential (Historical and recent growth rates).
    - Alignment with dominant vehicle categories in the overall market.
    - Competitive landscape within the segment.
- **Cluster 3, labeled "Early Growth Leaders,"** was identified as the most promising target segment. This segment exhibited a history of high growth, a significant market size, and a strong presence in the dominant 3W and 2W categories, indicating a market that has been receptive to growth and new offerings.



SEGMENT EXTRACTION (for second question, i.e., market segmentation analysis for ev_sales_by_maker_and_cat_15-24.csv and indian_ev_data.csv)

1. Elbow Method for Optimal K

Definition:

The Elbow Method helps to identify the **ideal number of clusters (K)** to use in K-Means by observing the trade-off between the number of clusters and the WCSS.

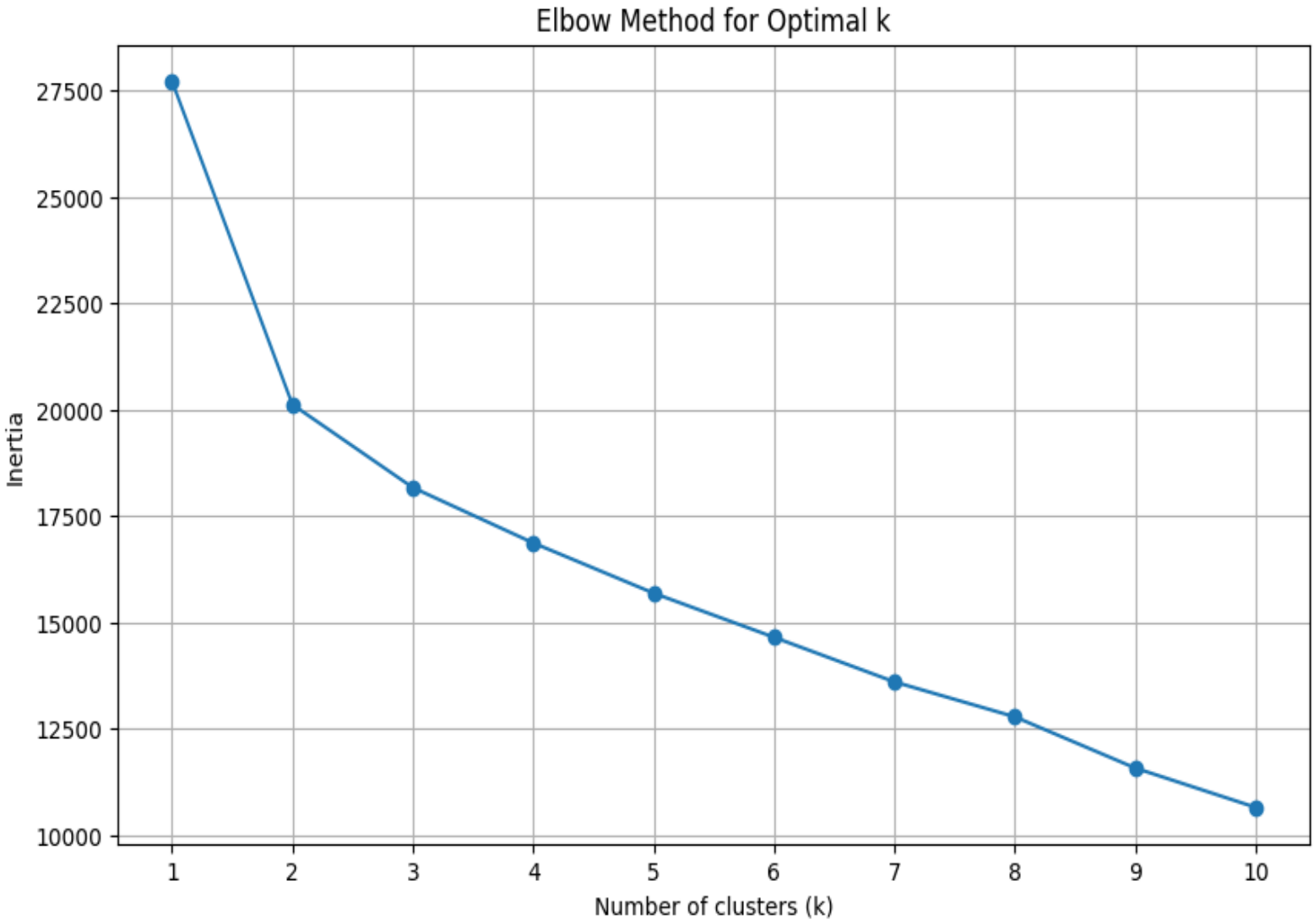
Procedure:

1. Run K-Means for a range of `K` values (e.g., 1 to 10).
2. Record WCSS for each K.
3. Plot `K` vs. `WCSS`.
4. Identify the "elbow" point in the curve — the point after which WCSS decreases more slowly. That point is considered the **optimal K**.

📌 **Why It Matters:**

- Prevents overfitting (too many small clusters) or underfitting (too few, overly general clusters).

- In your EV dataset, this could determine how many **market segments** (e.g., regions or usage types) are naturally present in the data.



2. Principal Component Analysis (PCA)

Definition:

PCA is a **dimensionality reduction technique** used to convert a dataset with many correlated features into a smaller set of **uncorrelated variables** called **Principal Components**.

Key Concepts:

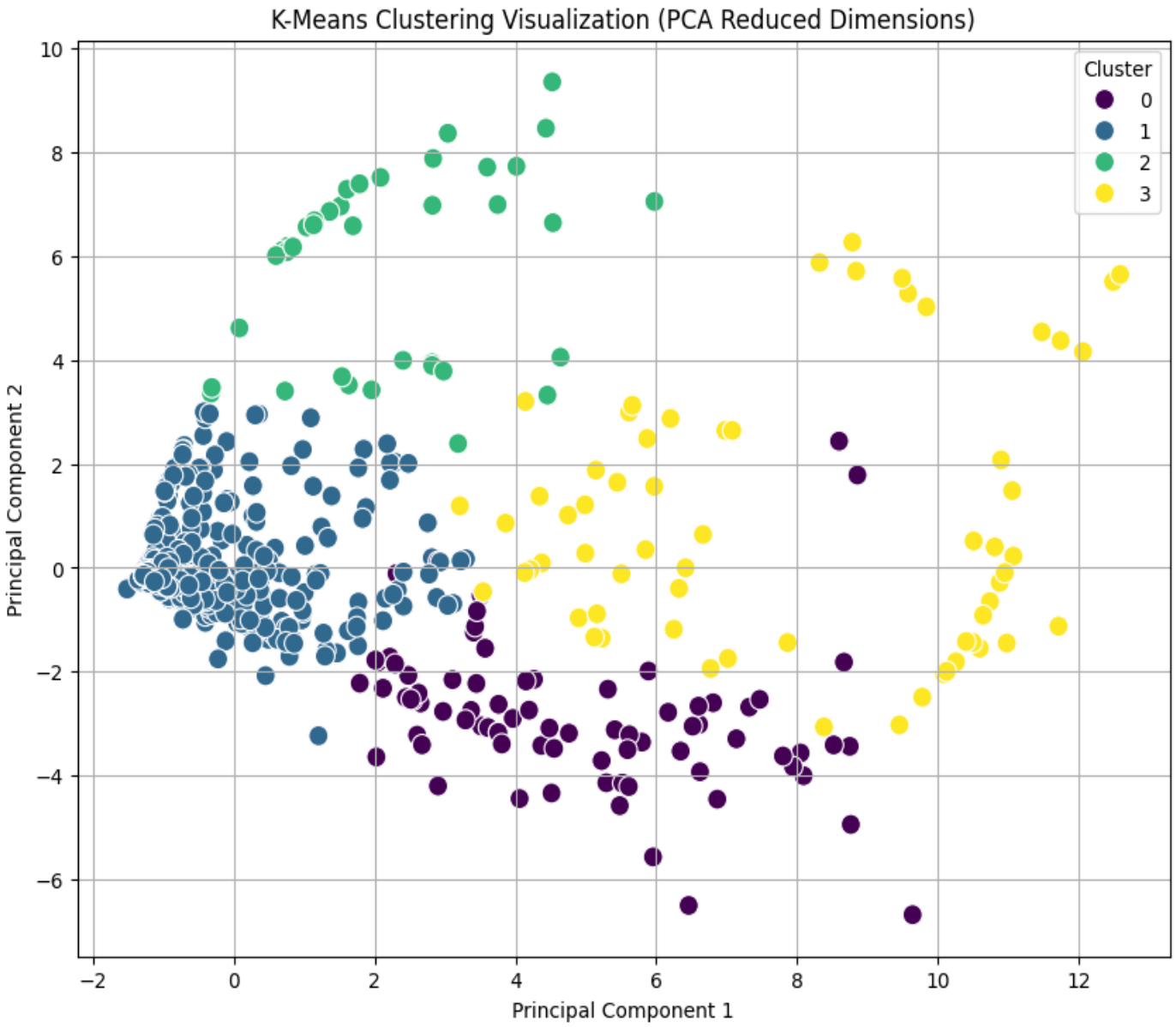
- The **first principal component (PC1)** captures the maximum variance.
- Subsequent components capture the remaining variance orthogonally.
- PCA transforms the dataset without losing significant information.

Steps:

1. Standardize the data (mean = 0, variance = 1).
2. Calculate the **covariance matrix** of the features.
3. Compute **eigenvalues and eigenvectors** of the matrix.
4. Select top `n` principal components based on explained variance.
5. Project data into the new PC space.

Purpose in EV Dataset:

- Simplify high-dimensional data (e.g., year, type, category, state, fuel type).
- Improve performance and visualization of clustering models like K-Means.
- Helps in identifying the **key drivers** of EV adoption patterns.



3. K-Means Clustering

K-Means Clustering is an unsupervised machine learning technique used to group similar data points into distinct clusters. When applied to the **Electric Vehicle (EV) market**, it can be used to segment the market based on **vehicle types**, helping stakeholders understand patterns in consumer preferences and optimize their strategies.

Objective:

To identify distinct market segments in the EV space based on **vehicle type**, such as:

- Electric 2-wheelers (E-bikes, scooters)
- Electric 3-wheelers (rickshaws, loaders)
- Electric 4-wheelers (cars, SUVs)
- Electric commercial vehicles (buses, trucks)

Key Features Used for Clustering:

- Sales volume by vehicle type
- Price range
- Battery capacity and range
- Usage type (personal, commercial, delivery, public transport)
- Region of purchase (urban, rural, metro)
- Charging time or infrastructure availability

Process:

1. **Data Collection:** Gather data on EV sales and features across different vehicle types.
2. **Preprocessing:** Normalize the data for numerical consistency.
3. **Choosing K:** Use the Elbow Method or Silhouette Score to select the optimal number of clusters.
4. **Apply K-Means:** Cluster the data into K groups based on vehicle characteristics.
5. **Interpretation:** Analyze each cluster to determine dominant vehicle types and their market traits.

Outcome:

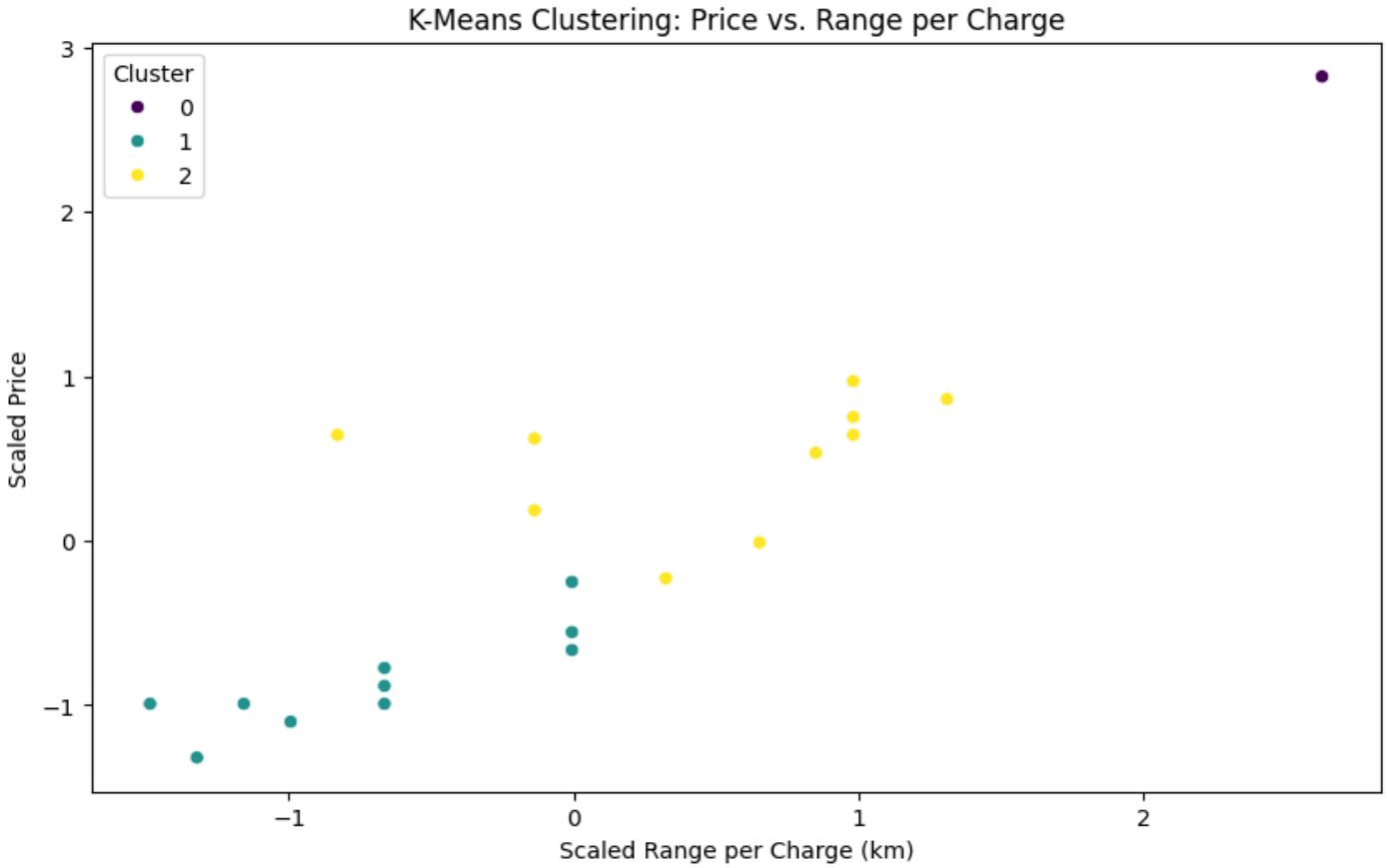
The clustering process may reveal:

- **Cluster 1:** Dominated by low-cost electric 2-wheelers with high adoption in Tier 2 and Tier 3 cities.
- **Cluster 2:** Mid-range electric 4-wheelers used in urban areas for personal use.
- **Cluster 3:** High-capacity electric buses and commercial trucks purchased by fleet operators or public transport agencies.

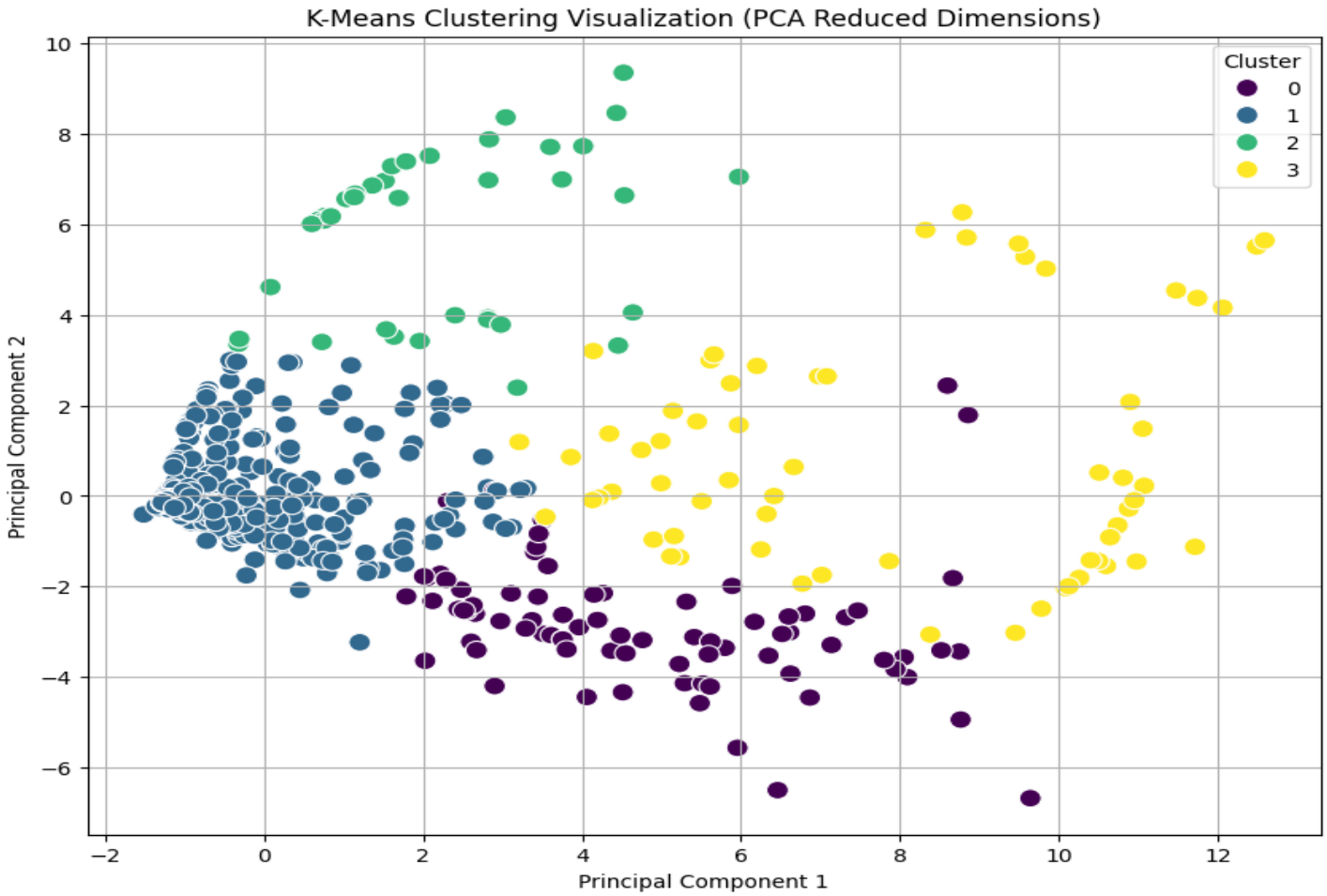
Benefits:

- Helps **manufacturers** target the right vehicle type to the right market.
- Aids **policy makers** in directing subsidies or incentives for specific segments.
- Supports **dealers and marketers** in customizing offerings by region and usage.

1. **indian_ev_data.csv**



2.Ev_ sales _by _maker_and_cat-15-24.csv



4.Profiling and describing potential segments

Based on the K-Means clustering analysis of the Indian EV dataset, we have identified three distinct segments:

A. indian_ev_data.csv :

Cluster 0: The Performance Enthusiasts / Early Adopters:

Characteristics: This segment is characterized by vehicles with the highest battery capacity, range, power, and top speed, and consequently, the highest price. The segment is predominantly composed of electric Bikes.

Profile: These consumers likely prioritize performance, range, and potentially are early adopters of new technology. They are less price-sensitive compared to other segments and are interested in electric vehicles that can offer a thrilling riding experience or longer travel distances.

Likelihood of EV Use: High, but represents a smaller, niche market. These buyers are already inclined towards higher-end EVs.

Cluster 1: The Price-Sensitive Urban Commuters:

Characteristics: This is the most affordable segment, with the lowest battery capacity, range, power, top speed, and charging time. It is heavily dominated by electric Scooters.

Profile: This segment consists of price-sensitive individuals, likely urban dwellers, who need a practical and economical mode of transport for daily commuting. They prioritize affordability, low running costs, and convenience for short to medium distances.

Likelihood of EV Use: Very High. The characteristics of vehicles in this cluster directly align with the needs and financial capabilities of a large portion of the urban Indian population, making them highly likely to adopt electric scooters for their daily commutes.

Cluster 2: The Value-Conscious Mainstream Buyers:

Characteristics: This segment falls between Cluster 0 and Cluster 1 in terms of price, performance, and range. It includes both electric Bikes and a significant number of electric Scooters.

Profile: These consumers are looking for a balance of features, performance, and price. They are willing to pay a moderate amount for an EV that offers better range and performance than the most basic models but are not necessarily seeking the high-end features of the performance segment. This segment is likely broader than Cluster 0 and represents a significant part of the potential mainstream EV market.

Likelihood of EV Use: High. This segment represents consumers who are open to adopting EVs if they offer a good balance of value and features that meet their daily needs beyond just basic commuting.

Summary of Target Segments for Market Entry:

Based on the analysis, Cluster 1 (Price-Sensitive Urban Commuters) and Cluster 2 (Value-Conscious Mainstream Buyers) represent the most promising target segments for market entry in the Indian EV market.

Cluster 1 offers a large volume opportunity due to its focus on affordability and practical urban mobility, which is a major driver for vehicle purchases in India.

Cluster 2 represents the growing mainstream market that is willing to invest in EVs offering a better balance of features and performance, indicating a readiness for more capable electric vehicles at a reasonable price point.

Targeting these two segments allows for addressing both the mass market (Cluster 1) and the expanding middle segment (Cluster 2) of EV buyers in India, providing a strong foundation for market penetration and growth.

5. SELECTION OF TARGET SEGMENT

Selecting a **target segment** in the electric vehicle (EV) market involves identifying the most promising group of consumers to focus marketing, product design, and sales efforts. This decision is based on analyzing factors like **vehicle usage patterns**, **consumer behavior**, **geographic demand**, **income levels**, and **adoption readiness**.

For example:

- **Urban commuters** may prefer affordable 2-wheelers for short-distance travel.
- **Fleet operators** may prioritize 3-wheelers or commercial 4-wheelers with low operating costs.
- **Environmentally conscious families** may opt for electric SUVs with longer range and fast charging.

Key criteria for selecting a target segment include:

- **Market size and growth potential**
- **Profitability and price sensitivity**
- **Competition and differentiation opportunities**
- **Alignment with government incentives and infrastructure**

Careful segment selection enables EV manufacturers and stakeholders to **optimize resources**, **tailor value propositions**, and **accelerate adoption** in a focused, sustainable way

- ## Customizing the Marketing Mix

Based on the analysis of the dataset and the identification of two key target segments (Cluster 1: Price-Sensitive Urban Commuters and Cluster 2: Value-Conscious Mainstream Buyers), here's how the marketing mix (Product, Price, Place, Promotion) can be customized for each segment:

Target Segment 1: Price-Sensitive Urban Commuters (Cluster 1)

Product: Focus on developing and marketing reliable, low-maintenance electric scooters with a practical range suitable for daily urban commutes. Emphasize features like ease of use, durability, and low running costs.

Price: Offer competitive and affordable pricing. Explore options like flexible financing, partnerships with financial institutions, and highlighting potential government subsidies or tax benefits for EV purchases to make them more accessible to this price-sensitive group.

Place: Concentrate distribution channels in urban areas with high commuter traffic. Partner with local dealerships, service centers, and potentially explore direct-to-consumer models or online platforms that cater to urban buyers.

Promotion: Marketing messages should emphasize affordability, cost savings compared to traditional vehicles, environmental benefits, and the convenience of urban mobility. Utilize channels like local advertising, community engagement, social media platforms popular with urban demographics, and testimonials from satisfied customers.

Target Segment 2: Value-Conscious Mainstream Buyers (Cluster 2)

Product: Offer a range of electric vehicles, including both scooters and mid-range bikes, that provide a good balance of features, performance, range, and price. Highlight features like improved battery capacity, moderate speed, and faster charging times compared to basic models.

Price: Set mid-range pricing that reflects the enhanced features and performance. Offer value-added services or warranty packages to justify the price point.

Place: Expand distribution to include both urban and semi-urban areas. Utilize a mix of online and offline channels, including dealerships and experience centers where customers can test ride vehicles and understand the value proposition.

Promotion: Marketing should focus on the balance of features and value, highlighting the enhanced performance and range for various commuting needs. Use a mix of online advertising, content marketing, partnerships with influencers, and participation in auto expos or events to reach a broader audience.

SEGMENT 3: Customizing the Marketing Mix for Early Growth Leaders (Cluster 3) in the Indian EV Market

To effectively enter and compete within the "Early Growth Leaders" segment (Cluster 3) of the Indian EV market, the marketing mix (Product, Price, Place, Promotion) should be tailored to the specific characteristics and needs of this segment, which is primarily focused on 3W and 2W categories with a history of high growth.

1. **Product:**
    - **Focus:** Prioritize the development and offering of high-quality, reliable, and cost-efficient **Three Wheeler (3W)** and **Two Wheeler (2W)** electric vehicles.

- **Features:** Emphasize features crucial for commercial use (for 3Ws) such as battery range, charging time, payload capacity, durability, and low maintenance. For 2Ws, focus on range, performance, charging convenience, and potentially swappable battery options.
- **Innovation:** Explore innovative battery technology (e.g., swappable batteries) and connectivity features that can add value for commercial operators and individual riders

2. **Price:**
    - **Strategy:** Implement a **Value-Based Pricing** strategy that highlights the Total Cost of Ownership (TCO) advantage of EVs over traditional vehicles.
    - **Competitiveness:** Price competitively against existing successful players in Cluster 3 and consider the price sensitivity of commercial buyers.
    - **Financial Solutions:** Offer attractive financing options, leasing models, or battery-as-a-service programs to lower the initial purchase price barrier.

3. **Place (Distribution):**
    - **Network:** Build a robust and accessible dealer and service network, focusing initially on urban and semi-urban areas with high commercial activity and existing EV adoption.
    - **Partnerships:** Establish strong partnerships with fleet operators, logistics companies, and e-commerce platforms that are major users of 3W EVs.
    - **Service & Support:** Ensure readily available after-sales service, spare parts, and trained technicians to address concerns about maintenance and downtime.

4. **Promotion:**
    - **Messaging:** The core marketing message should emphasize the **economic benefits** (significant savings on fuel and maintenance), **reliability**, **environmental advantages**, and potential for **increased earnings** (for commercial users) offered by the EVs.
    - **Channels:** Utilize targeted marketing channels such as commercial vehicle expos, local dealerships, online platforms frequented by commercial vehicle buyers and drivers, and potentially partnerships with financing institutions.
    - **Trust Building:** Offer test drives, provide transparent information on battery performance and warranty, and engage with local communities and driver associations to build trust and awareness.
    - **Incentive Communication:** Clearly communicate and assist customers in accessing available government subsidies and incentives.

8. A. Most Optimal Market Segments (Based on Research)

1. Urban Two-Wheeler (2W) Commuters:

- **Why Optimal:** Consistently the highest-selling EV category due to affordability, ease of charging, and suitability for congested cities.
- **Target Audience:** Young professionals, delivery riders, students.
- **Cities:** Delhi, Bengaluru, Pune, Hyderabad.
- **Key Drivers:** Low cost, daily commute needs, strong subsidies (FAME II), battery-swapping stations.

2. Commercial Electric Three-Wheelers (3W) for Last-Mile Transport:

- **Why Optimal:** Rapid growth in shared mobility, logistics, and e-commerce deliveries.
- **Target Audience:** Fleet operators, gig workers, auto drivers.
- **Regions:** Tier-1 and Tier-2 cities with delivery networks.
- **Key Drivers:** Low operational cost, government support, high daily mileage economics.

3. Compact Family Electric Four-Wheelers (4W):

- **Why Optimal:** Increasing adoption in middle-income households due to rising fuel prices and EV incentives.
- **Target Audience:** Environmentally conscious families, city commuters.
- **States:** Maharashtra, Karnataka, Tamil Nadu, Gujarat.
- **Key Drivers:** State-level subsidies, growing charging infrastructure, improved range in compact EV cars.

4. Electric Public Transport Buses (Govt. & Urban Fleets):

- **Why Optimal:** Driven by public procurement under schemes like FAME-II and state EV policies.
- **Target Audience:** Municipal corporations, state transport undertakings.
- **Cities:** Delhi, Mumbai, Kochi, Lucknow.
- **Key Drivers:** Bulk orders, centralized charging depots, environmental mandates, low emission zones.

📌 Summary Table

Segment	Use Case	Key Drivers	Region Focus
Urban 2W Commuters	Personal, Delivery	Low cost, ease of use, subsidies	Tier-1 Cities
Commercial 3W (Last-mile)	Shared/Logistics	Daily savings, demand in logistics	Tier-1 & Tier-2 Cities
Compact 4W for Families	Personal Mobility	Mid-range pricing, city commuting	Urban States
Public EV Buses	Govt. Transport	Bulk demand, pollution control goals	Metro Cities

---

B. Top 4 Optimal Market Segments (Based on Segmentation)

1. Electric Two-Wheelers for Urban Personal Use

- **Segment Traits:**
  - High volume of sales from 2018 onwards.
  - Popular among middle-income commuters.
  - Dominant in cities with traffic congestion and short commutes.
- **Segmentation Markers:** `Category = 2W`, `Usage = Personal`, `Region = Urban`
- **Why Optimal:** Affordable, low maintenance, easily chargeable at home.

2. Electric Three-Wheelers for Commercial Logistics & Shared Mobility

- **Segment Traits:**
  - Sharp CAGR growth, especially in states like Uttar Pradesh, Bihar, and Delhi.
  - Utilized heavily in last-mile delivery and passenger services.
- **Segmentation Markers:** `Category = 3W`, `Usage = Commercial`, `State = High adoption (e.g., Delhi, UP)`
- **Why Optimal:** High daily usage, fuel savings, strong ROI for operators.

3. Compact Electric Four-Wheelers for Middle-Class Families

- **Segment Traits:**
  - Steady growth in cities with better infrastructure (Bangalore, Pune, Hyderabad).
  - Increasing preference due to fuel price hikes and environmental awareness.
- **Segmentation Markers:** `Category = 4W`, `Usage = Family/Personal`, `Region = Metro Cities`
- **Why Optimal:** Best fit for intra-city family travel; subsidy-eligible.

4. Electric Buses in State and Municipal Transport Fleets

- **Segment Traits:**
  - High value, low volume segment.
  - Mostly driven by government procurement and policy mandates.
- **Segmentation Markers:** `Category = Bus`, `Usage = Public Transport`, `Initiative = Govt-backed`
- **Why Optimal:** Pollution reduction targets, central/state funding (e.g., FAME-II), depot-based operations.

Summary Table

Segment Name	Category	Usage	Region Focus	Key Reasons for Optimality
Urban 2W for Personal Use	2W	Personal	Metro/Tier-1 Cities	Volume leader, low-cost entry, wide adoption
3W for Commercial & Logistics	3W	Commercial	North & East States	High ROI, last-mile demand, low maintenance
4W for Middle-Income Family Use	4W	Personal	South & West Cities	Clean energy image, rising fuel costs
Electric Buses for Govt. Transport	Bus	Public Use	Metro Cities	Policy-driven, bulk demand, emission goals

---