

# ELECTRIC VEHICLE MARKET SEGMENTATION IN INDIA

Name - Prathamesh Patil

GitHub Link :

<https://github.com/PRATHAMESHPATIL123654/Electric-Vehicle-Market-Segmentation.git>

## Introduction

The electric vehicle (EV) market in India is growing rapidly. In 2022, the market size was valued at USD 2.2 billion and is expected to reach USD 152.21 billion by 2030, growing at a CAGR of 94.4%. The growth of the EV market in India is being driven by a number of factors, including:

- **Government incentives:** The Indian government has introduced a number of incentives to promote the adoption of EVs, including tax breaks, subsidies, and exemptions from registration fees.
- **Rising fuel prices:** The rising cost of fuel is making EVs more attractive to consumers.
- **Increasing environmental awareness:** There is growing awareness among consumers about the environmental benefits of EVs.
- **Improved technology:** The technology of EVs has improved significantly in recent years, making them more affordable and efficient.

The EV market in India is still in its early stages, but it is growing rapidly. The two-wheeler segment is the largest segment of the EV market in India, followed by the three-wheeler segment. The passenger car segment is also growing rapidly and is expected to become the largest segment of the market by 2030.

The growth of the EV market in India is creating new opportunities for businesses. There is a growing demand for EV components, such as batteries, motors, and chargers. There is also a growing demand for EV services, such as charging infrastructure and fleet management. The government has set a target to achieve 30 percent electrification of the country's vehicle fleet by 2030.

## Key Trends Shaping the EV Market in India:

- **Growth of the two-wheeler and three-wheeler segments:** These segments are expected to continue growing rapidly in the coming years, as they are the most affordable and accessible forms of EVs.
- **Increasing popularity of passenger cars:** The passenger car segment is growing rapidly, as consumers are becoming more aware of the environmental benefits of EVs.
- **Growth of the commercial vehicle segment:** The commercial vehicle segment is also growing rapidly, as businesses are looking to reduce their operating costs by switching to EVs.
- **Development of charging infrastructure:** The development of charging infrastructure is essential for the growth of the EV market in India. The government is taking steps to develop charging infrastructure, and private companies are also investing in this area.

- **Growth of the EV component manufacturing industry:** The growth of the EV market in India is creating new opportunities for businesses in the EV component manufacturing industry.

The EV market in India is a rapidly growing market with significant potential. The government is taking steps to promote the adoption of EVs, and the private sector is also investing in this area. The growth of the EV market in India is creating new opportunities for businesses in a variety of sectors.

### Problem Statement

An EV startup is developing electric vehicles (EVs) and is trying to decide which vehicle/customer space to target. The Indian EV market is growing rapidly but is still in its early stages. There are a number of different segments in the market, each with its own unique characteristics.

We need to analyze the market and identify the segments that are most likely to adopt EVs. We also need to develop a feasible strategy to enter the market and compete in these segments.

The Indian EV market can be segmented based on the following criteria:

- **Vehicle type:** The EV market can be segmented into two-wheelers, three-wheelers, passenger cars, and commercial vehicles. Two-wheelers are the most popular type of EV in India, followed by three-wheelers. Passenger cars and commercial vehicles are still in the early stages of development.
- **Customer:** The EV market can also be segmented by customer type. Customer behavioral data can be used to perform market segmentation.

### Analysis of Indian Automobile Buying Behaviour Data

#### Dataset Columns:

- **Age:** The age of the buyer.
- **Profession:** The occupation of the buyer.
- **Marital Status:** The marital status of the buyer.
- **Education:** The highest level of education of the buyer.
- **No of Dependents:** The number of dependents of the buyer.
- **Personal Loan:** Whether the buyer has a personal loan.
- **House Loan:** Whether the buyer has a house loan.
- **Wife Working:** Whether the wife of the buyer is working.
- **Salary:** The salary of the buyer.
- **Wife Salary:** The salary of the wife of the buyer.
- **Total Salary:** The combined salary of the buyer and wife.
- **Make:** The make of the automobile purchased.
- **Price:** The price of the automobile purchased.

## Data Preprocessing

The following data pre-processing steps were carried out using the pandas library:

1. The *Wife Working* variable had an extra category with only one data point. This category was merged with the "Yes" category.

## Inferences from the Exploratory Data Analysis (EDA):

- The salary of customers is grouped around ₹1,000,000 to ₹2,500,000, suggesting that most customers in the dataset belong to the upper middle class or upper class.
- For most customers, the wife's salary is zero. For those whose wives are working, the salary ranges between ₹500,000 and ₹1,500,000.
- The total household salary for most customers is between ₹1,000,000 and ₹3,000,000, indicating a comfortable financial situation.
- The price of cars purchased ranges between ₹500,000 and ₹2,000,000, suggesting that most customers bought mid-range cars.
- 65% of customers are salaried professionals, while 35% have their own business, indicating a majority of employed professionals.
- Most customers in the dataset are married, suggesting stable family situations.
- 55% of customers hold a post-graduate degree, while 35% have a graduate degree, indicating a well-educated customer base.
- Most customers have 2-3 dependents, suggesting that they have young families.
- 68% of customers do not have a personal loan, indicating they could afford cars without borrowing.
- 61% of customers do not have a house loan, indicating they could buy homes without loans.
- 52% of customers' wives are working, showing a balanced employment trend among partners.
- The most popular vehicles purchased are SUVs, followed by Baleno and Creta. Luxury cars were purchased by very few customers, indicating interest in practical, fuel-efficient cars.

## Market Segmentation

The Indian Automobile Buying Behaviour dataset was used to perform EV market segmentation.

## Data Preprocessing

1. Categorical variables were one-hot encoded using the `pd.get_dummies` method.
2. Continuous variables were scaled using the `MinMaxScaler`, ensuring values lie between 0 and 1.

## KMeans Clustering

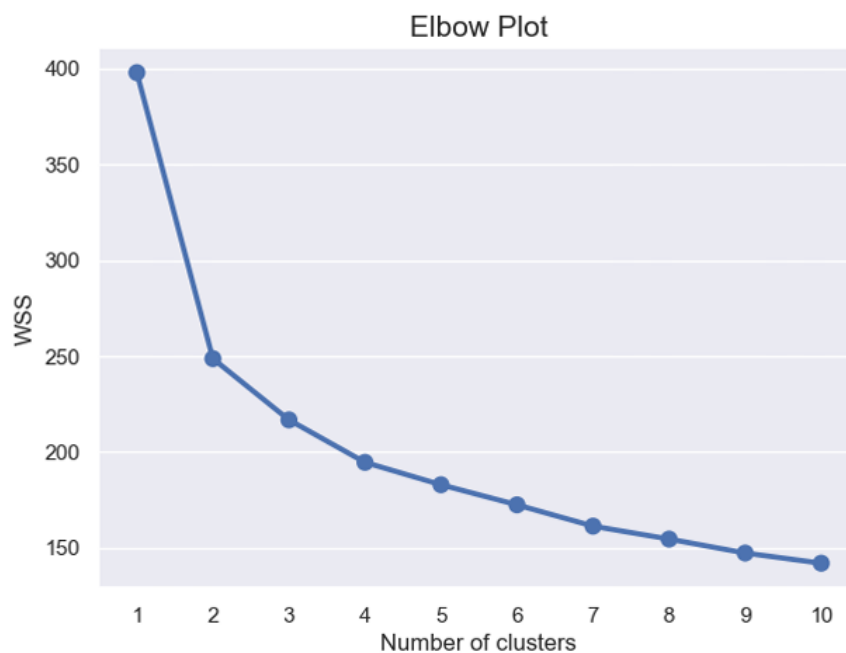
The **K-means clustering** algorithm was applied to segment car buyers. The algorithm iteratively assigns data points to clusters with the closest mean, then recomputes the cluster means. The K-

means algorithm identifies hidden patterns, making it easier to understand different types of car buyers.

### Elbow Plot

To determine the optimum number of clusters, an **Elbow Plot** was used, where clusters are plotted on the X-axis, and within-cluster sum of squares (WSS) is plotted on the Y-axis.

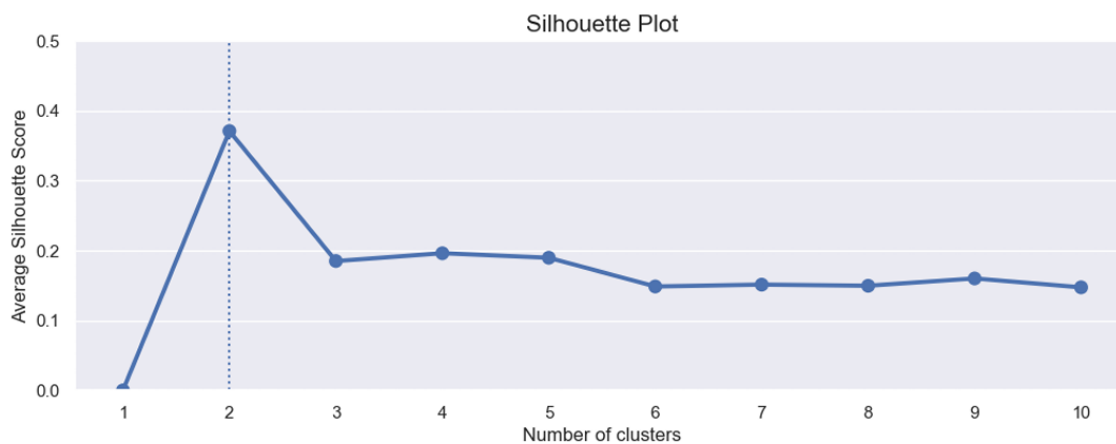
- The plot suggested the optimal number of clusters as 2, as the WSS drop from clusters 1 to 2 was significant, while the drop after cluster 2 was not.



*Figure 1: Elbow Plot*

### Silhouette Plot

The **Silhouette Method** was used to measure clustering tightness and inter-cluster distance. The plot showed that the highest silhouette score was for 2 clusters, confirming the optimal number of clusters as 2.



*Figure 2: Silhouette Plot*

Visualizing Clusters Using PCA

**Principal Component Analysis (PCA)** was used to reduce dimensionality for visualizing the clusters. A scatter plot with color coding for the 2 clusters showed that the clusters are well-separated, with one (yellow) cluster being smaller than the other (purple).

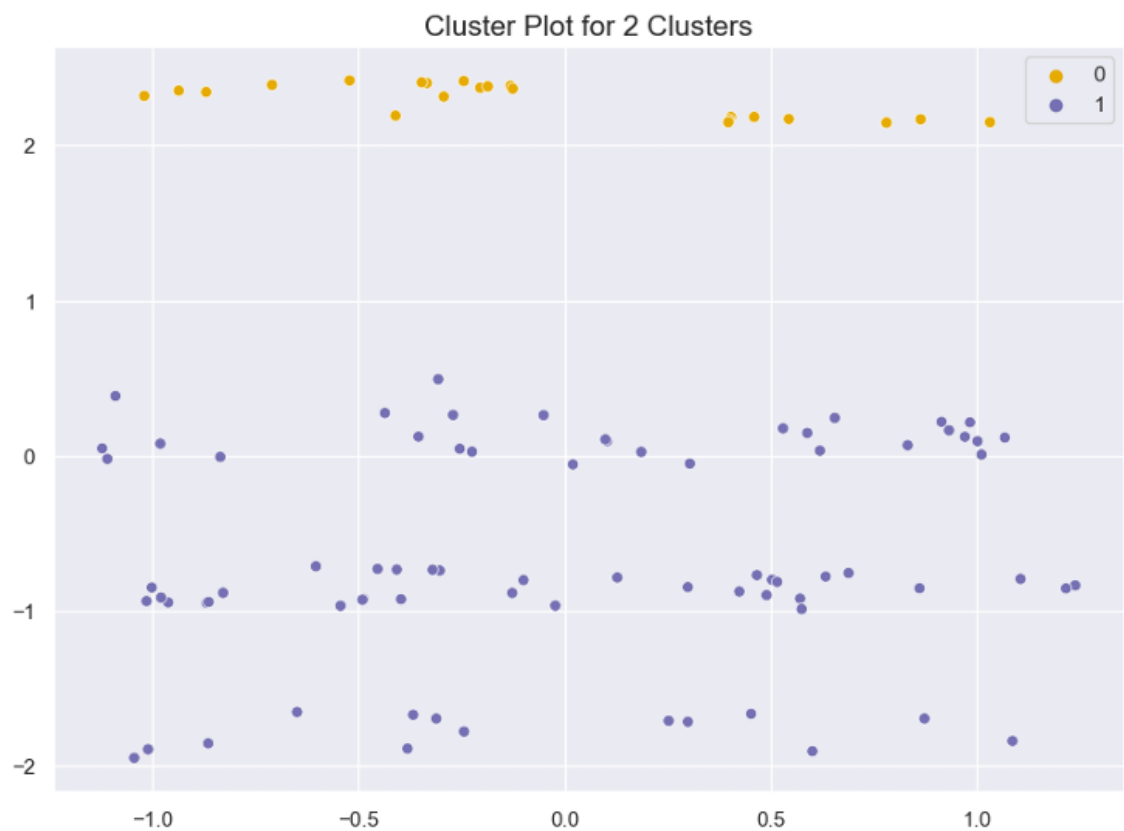


Figure 3: Cluster Plot for 2 Clusters

Profiling the Segments

A summary of the two clusters is presented in the table below:

Feature	Cluster 1	Cluster 2
Average Age	30	40
Most Common Profession	Salaried	Salaried

Feature	Cluster 1	Cluster 2
Marital Status	Single	Married
Level of Education	Post Graduate	Post Graduate
Number of Dependents	0	3
Loan Type	No Loan	No Loan
Wife Working Status	No	Yes
Average Salary	₹1,300,000	₹1,900,000
Average Wife Salary	₹300,000	₹600,000
Average Total Salary	₹1,600,000	₹2,500,000
Most Popular Make	Baleno, Ciaz, Creta	SUV
Average Car Price	₹1,100,000	₹1,200,000

#### Key Variables for Market Segmentation

The top 4 variables to create the most optimal market segments for the Indian car market are:

- **Age:** Younger buyers prefer affordable cars; older buyers look for more expensive models.
- **Income:** Higher-income buyers can afford more expensive cars.
- **Education Level:** Higher education levels correlate with a preference for fuel-efficient, environmentally friendly cars.
- **Family Size:** Larger families tend to choose spacious, comfortable cars.

#### Conclusion

The two clusters represent distinct buyer groups:

- **Cluster 1:** Younger, less well-off professionals looking for affordable cars.
- **Cluster 2:** Middle-aged, well-off professionals looking for more luxurious, powerful cars.

#### Marketing Mix Strategy

The ideal target segment for the EV startup is younger, less well-off professionals looking for affordable cars. The marketing mix should focus on:

- **Product:** Affordable, efficient, stylish EV with a range suitable for Indian driving conditions.

- **Price:** Competitive pricing to make the EV accessible to a wide range of buyers.
- **Place:** Availability in urban and semi-urban dealerships with charging infrastructure.
- **Promotion:** Focus on affordability, environmental benefits, using online/offline channels.
- **People:** Experienced, passionate professionals familiar with the Indian market.
- **Process:** Smooth, efficient sales process with excellent customer service.
- **Physical Evidence:** Well-designed dealerships reflecting the brand's quality and commitment.

#### **Further Possible Improvements**

- Collect additional data points (e.g., buyer satisfaction, EV feature ratings).
- Try additional machine learning models (e.g., hierarchical clustering, Gaussian mixture model).
- Conduct more detailed analyses (e.g., chi-squared test, discriminant analysis).

# **Market Segmentation Analysing the Electric Vehicle Market in India using Segmentation Analysis**

***By Greeshma Haridas***

**Github Link - :**

**[https://github.com/GreeshmaHarids/FeynnLab\\_Project-2.git](https://github.com/GreeshmaHarids/FeynnLab_Project-2.git)**



## Electric Vehicle Market Segmentation Analysis for India

### 1. Summary

This report provides a comprehensive analysis of the electric vehicle (EV) market in India, focusing on state-wise sales and vehicle categories (2-wheelers and 4-wheelers). The objective of this analysis is to:

- Identify state-level EV sales trends across India.
- Understand the penetration rate of electric vehicles in each state.
- Segment states into clusters based on EV sales performance and penetration.
- Analyze vehicle category distribution within each cluster.

The findings will support strategic decision-making, such as where to focus marketing efforts and expand EV infrastructure. Through Exploratory Data Analysis (EDA), Principal Component Analysis (PCA), and K-means clustering, this report identifies key regions and customer segments for early market entry.

Our findings highlight that 2-wheelers are driving the bulk of EV sales, with significant growth in states like Maharashtra, Karnataka, Tamil Nadu, Gujarat, and Rajasthan.

#### Key Segments:

- **Vehicle Categories:** 2-Wheelers and 4-Wheelers.
- **Geographical Focus:** Maharashtra, Karnataka, Tamil Nadu, Gujarat, and Rajasthan are leading in EV adoption and sales.

#### Market Dynamics:

- **Growth Drivers:** Incentives for EVs, rising fuel costs, technological advancements in battery life, and a shift towards sustainable transportation.
- **Consumer Trends:** Increasing preference for eco-friendly vehicles and expanding EV infrastructure.

#### Insights:

- **2-Wheelers** dominate the market in terms of sales and growth compared to 4-wheelers.
- **Regional Insights:** High-potential states like Maharashtra, Karnataka, Tamil Nadu, Gujarat, and Rajasthan are prime for expansion due to their significant EV sales and growth rates.

---

### 2. Market Overview

The Indian electric vehicle market has seen significant growth due to rising environmental awareness, supportive government policies, and increasing fuel costs. The report focuses on segmentation by region and vehicle category (2-wheelers vs. 4-wheelers).

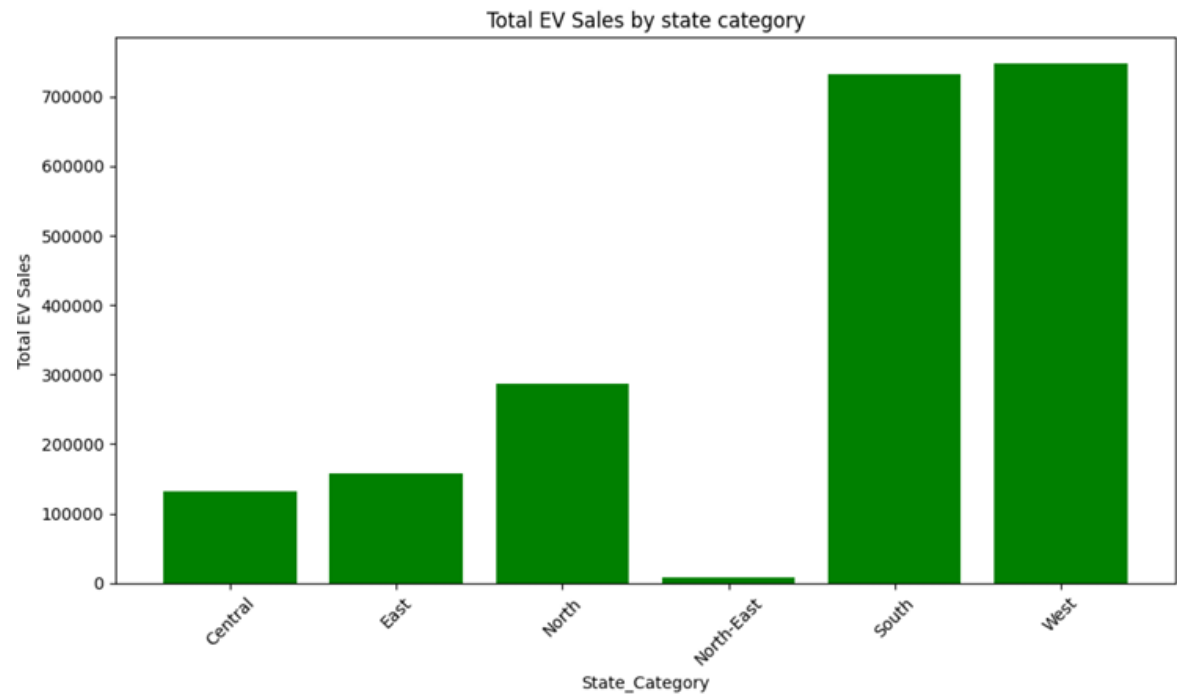
**Dominant Categories:**

- 1. **2-Wheelers:** Account for the majority of EV sales, driven by affordability and practicality for urban commuting.
- 2. **4-Wheelers:** While growing, their penetration remains limited compared to 2-wheelers.

**Regional Contributions:**

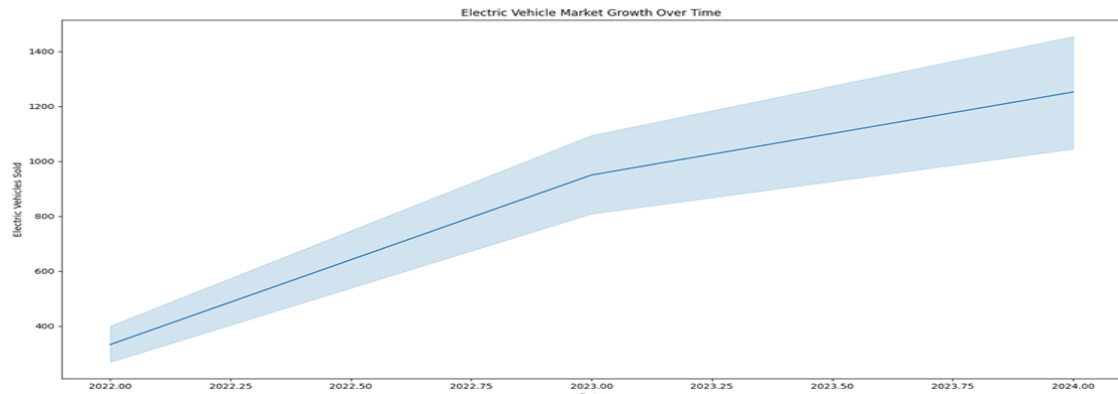
- **West and South Regions:** Largest contributors to the EV market, showing higher adoption rates and sales volumes.
- **North Region:** Moderate contribution, growing but not as fast as West and South.
- **East, Central, and North-East Regions:** Slower adoption rates and market penetration, reflecting growth potential.

**Regional Contributions:**

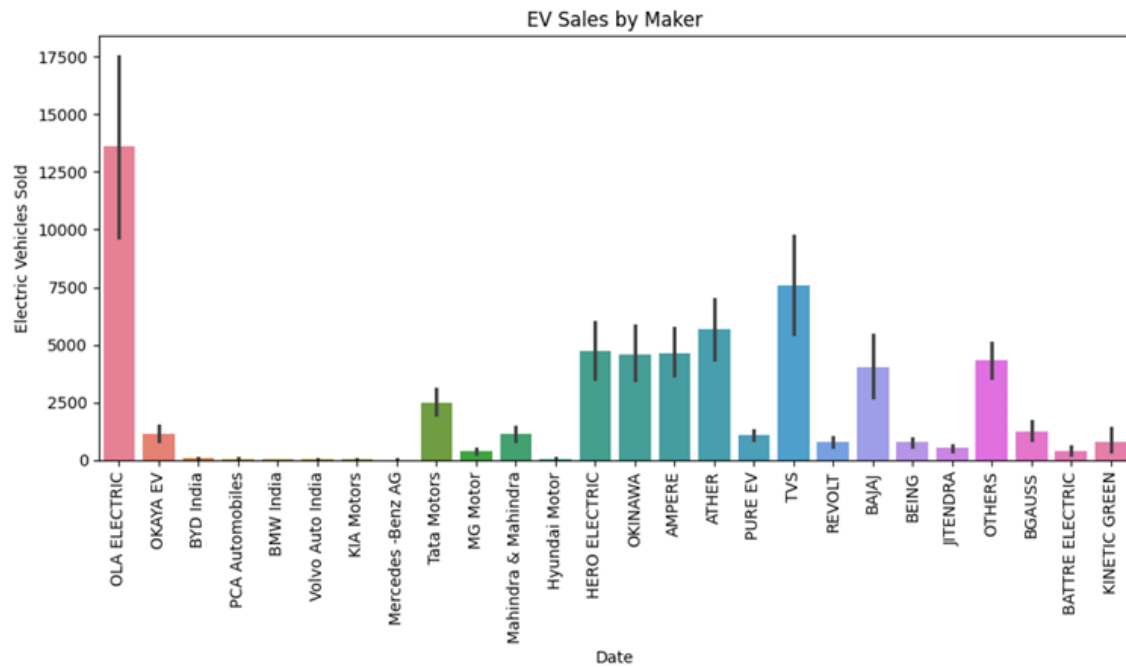


**Major EV Makers in India:**

Leading players include **OLA Electric**, **TVS**, and **Ather**, with others like **Ampere**, **Okinawa**, and **Hero Electric** playing significant roles. 2-wheeler



s dominate the market, driven by affordability and utility.



### 3. Data Overview and Preprocessing

Three datasets were used for this analysis:

1. **Electric Vehicle Sales by State:** State-wise EV sales, vehicle categories, and total vehicles sold.
2. **Electric Vehicle Sales by Makers:** Sales data from different EV manufacturers.
3. **Date Data:** Temporal data for mapping sales across specific time periods.
4. **State Categories:** States categorized into regions: West, North, East, South, Central, and North-East.

**Key Features:**

- **State-wise Sales Data:** EV sales per state for different categories (2-wheelers and 4-wheelers).
- **Penetration Rate:** Percentage of electric vehicles sold out of total vehicles sold in each state.
- **Categorization:** States categorized by region for market analysis.

**Data Preprocessing:**

- Merged datasets (state sales data, date information, state category, and vehicle maker data).
- Encoded categorical variables (e.g., vehicle category, quarter, region).
- Handled missing values using region-based imputation.
- Key features include **Penetration Rate** and **Vehicle Categories** (2-wheelers vs. 4-wheelers).

**4. Exploratory Data Analysis**

**State-wise and Category-wise Trends:**

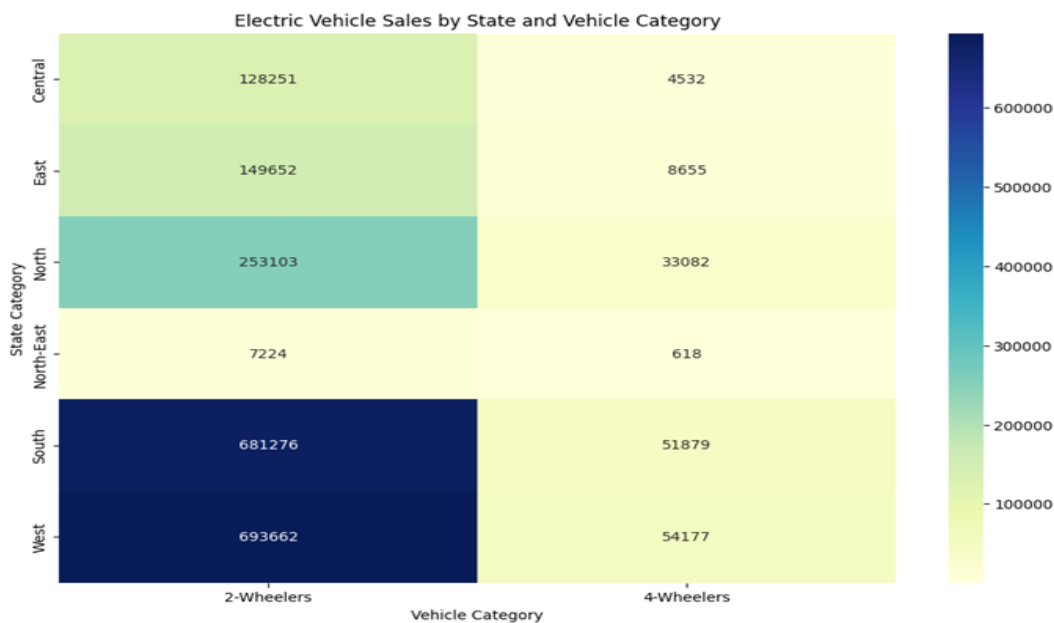
The majority of EV sales are concentrated in specific regions, with **Maharashtra, Karnataka, Tamil Nadu, Gujarat, and Rajasthan** emerging as the top states in 2-wheeler sales, constituting over 70% of total EV sales.

**Penetration Rates:**

The average penetration rate across states is 2.22%, with:

- **High Penetration States:** Maharashtra, Karnataka, Tamil Nadu, leading with rates above 8%.
- **Low Penetration States:** North-Eastern states show low adoption rates, signalling growth potential.

**Regional Sales Distribution:**



---

## 5. Segmentation Analysis

### PCA Results:

- **First Principal Component (PC1):** Explains 99.99% of the variance, capturing state-wise and category-wise sales.
- **Second Principal Component (PC2):** Minimal contribution, highlighting the dominance of 2-wheelers in EV sales.

### K-Means Clustering:

States were segmented into three clusters based on total electric vehicles sold, penetration rate, and vehicle category distribution.

### Cluster Profiles:

- **Cluster 1:** High EV sales and penetration rates, predominantly developed states with strong demand for 2-wheelers.
  - States: Maharashtra, Karnataka, Tamil Nadu, Gujarat, Rajasthan, Kerala, etc.
  - **Strategy:** Focus on scaling operations, expanding product portfolios, and enhancing charging infrastructure.
- **Cluster 0:** Low EV sales and penetration rates, indicating early-stage market development.
  - States: North-Eastern states.
  - **Strategy:** Implement awareness campaigns and invest in infrastructure development to boost adoption.

---

## 6. Strategic Recommendations

### Target States for Market Entry:

- **High-Potential States:** Maharashtra, Karnataka, Tamil Nadu, Gujarat, Kerala, and Rajasthan.
- **Emerging Markets:** States like Assam and Bihar show growth potential for future market development.

### Vehicle Category Focus:

- **2-Wheelers:** Should be the primary focus due to higher sales and faster growth rates.
- **4-Wheelers:** Demand is expected to grow in developed states (Cluster 1).

### Regional Pricing and Marketing Strategy:

- **Premium Pricing:** In Cluster 1 states, premium pricing for advanced models is feasible.
  - **Incentive-Based Marketing:** For Cluster 0 states, government subsidies, low-interest financing, and early adopter incentives can accelerate market penetration.
-

## **7. Marketing Mix**

### **Product:**

Focus on a range of electric vehicles, emphasizing 2-wheelers. Enhance features like fast charging and extended warranties.

### **Price:**

Tiered pricing strategy catering to different market segments—competitive pricing for entry-level models and premium pricing for advanced features.

### **Place:**

Prioritize high-sales regions (Maharashtra, Karnataka, Tamil Nadu) and expand into emerging markets (North-East, Central regions). Utilize both physical dealerships and online sales.

## **8. Conclusion**

The Indian electric vehicle market is poised for rapid growth, particularly in the 2-wheeler segment. States like Maharashtra, Karnataka, and Tamil Nadu present significant opportunities for market entry, while regions like the North-East offer long-term growth potential. A targeted approach, focusing on regional differences and leveraging state-wise sales trends, will enable effective market penetration and sustained growth.

# Electric Vehicle Market Segmentation

**Name - Hemanth Kumar**

**GitHub Link:** <https://github.com/HemanthKumarBodduboina/electric-vehicle-segmentation.git>

---

## Introduction

This report provides an in-depth analysis of India's electric vehicle (EV) market, focusing on the segmentation and targeting of two primary consumer groups—Segment 1 and Segment 2. The analysis identifies distinct preferences within these groups and outlines a strategic approach to product customization, pricing, promotion, and distribution.

---

## Market Segments Overview

### 1. Segment 1 (39% of Consumers)

- Target Price: ₹70,688 to ₹1,29,063
- Riding Range: 89 to 180 km
- Top Speed: 58 to 116 kmph
- Weight: 76 to 120 kg
- Battery Charging Time: 3 to 5 hours
- Power Rating: 1200 to 5500 W

Segment 1 represents budget-conscious consumers seeking affordable EVs for daily commuting. They prefer faster charging options and heavier vehicles with moderate riding ranges. Given the significant market share, this segment offers a strong profit opportunity.

### 2. Segment 2 (33% of Consumers)

- Target Price: ₹51,094 to ₹1,37,890
- Preferences: Visual appeal, reliability, service experience, and comfort

Segment 2 prioritizes affordability but values enhanced features such as visual design and comfort. This segment presents an opportunity for customization in product offerings, focusing on value for money and quality features.

---

## Product Customization

Customization is vital to aligning electric two-wheelers with the distinct needs of each segment. For Segment 1, addressing specific dissatisfaction points through feature enhancements, such as faster charging and improved reliability, will foster customer loyalty. Segment 2 calls for a greater focus on visual appeal, comfort, and overall value.

---

### Pricing Strategy

- **Segment 1:** Competitive pricing between ₹70,688 and ₹1,29,063 to attract price-sensitive buyers.
- **Segment 2:** A slightly higher price point can be justified by additional features and enhanced value for money, within the range of ₹51,094 to ₹1,37,890.

---

### Promotion & Distribution

- **Promotion:** Targeted advertising tailored to the preferences of each segment. For Segment 1, emphasizing affordability and fast charging. For Segment 2, highlighting visual appeal and reliability.
- **Distribution:**
  - Segment 1: Focus on urban areas with a robust online presence and strong customer support.
  - Segment 2: Distribution efforts should target suburban and semi-urban regions, with an emphasis on online accessibility.

---

### Customer Service and Process Customization

Customer service representatives should be trained to address the specific concerns of each segment. Ensuring efficient processes for customization requests and service appointments will enhance customer satisfaction and brand loyalty.

---

### Potential Market Opportunity

- **Segment 1:** With a target price of ₹1,20,000, the potential profit from this segment is ₹39.60 crores.
- **Segment 2:** At a target price of ₹1,10,000, the potential profit is ₹30.47 crores.

---

### Conclusion

The electric vehicle market in India presents a significant opportunity, particularly within Segment 1, which holds 39% of the consumer base. By focusing on tailored product development and marketing strategies that align with the preferences of this segment, there is a clear path to substantial market growth. Segment 2 offers additional expansion opportunities, particularly by focusing on enhanced features and customer experience. This strategic approach, grounded in detailed market segmentation, ensures that our electric vehicle offerings are positioned for success in India's rapidly evolving EV landscape.



# Electric Vehicle - Market Segmentation

**Name - *Sevantkumar S Huggi***

**GitHub Link:** <https://github.com/Sevantkumar/Electric-vehicle-data-segmentation>

---

## 1. Introduction

The global automotive industry is transitioning towards sustainable mobility, with electric vehicles (EVs) at the forefront. In India, the growing demand for EVs is driven by rising fuel prices, government incentives like the Faster Adoption and Manufacturing of Hybrid and Electric Vehicles (FAME) scheme, and increasing environmental awareness among consumers.

However, the market faces challenges such as high upfront costs, limited charging infrastructure, and low consumer awareness. Identifying the right target market is crucial to overcoming these obstacles and ensuring a successful entry into the Indian EV market.

---

## 2. Problem Statement

The Indian EV market is rapidly expanding, driven by technological advancements and government initiatives. For an emerging EV startup, determining a strategic market entry point is critical. This requires a detailed market segmentation analysis to identify the most viable customer and vehicle segments.

The goal is to understand which groups are most likely to adopt EVs and what drives their interest. This will allow the startup to tailor products and services to meet consumer preferences, maximizing the chances of success.

---

## 3. Analysis of Indian Automobile Buying Behaviour Data

### 3.1 Dataset

The dataset contains features related to various electric car models, including:

- **Brand:** Brand of the car.
- **Model:** Model name.
- **AccelSec:** Time to accelerate from 0 to 100 km/h (in seconds).
- **TopSpeed KmH:** Top speed (in km/h).
- **Range Km:** Driving range on a single charge (in km).
- **Efficiency WhKm:** Energy efficiency (in Wh/km).
- **FastCharge KmH:** Charging speed (km/h) for fast charging.
- **RapidCharge:** Whether rapid charging is supported (Yes/No).

- **PowerTrain:** Drivetrain type (AWD, RWD, FWD).
- **PlugType:** Charging plug type (Type 2, CCS, CHAdeMO).
- **BodyStyle:** Body style (SUV, Sedan, Hatchback).
- **Segment:** Market segment (A, B, C, D, E, F).
- **Seats:** Number of seats.
- **Price:** Price of the car (in Euros).

### 3.2 Data Preprocessing

Key data preprocessing steps included:

- **Handling Missing Data:** Missing values were handled through row removal or imputation with mean/median values.
  - **Encoding Categorical Variables:** Categorical columns like PowerTrain, PlugType, BodyStyle, and Segment were one-hot encoded for KMeans clustering.
  - **Feature Scaling:** StandardScaler was applied to normalize features with different units, such as Price (in Euros) and Efficiency (in Wh/km).
  - **Feature Selection:** Key features selected for clustering were Model, Price, FastCharge, Range, PowerTrain, Efficiency, TopSpeed, and Acceleration.
- 

### 3.3 KMeans Clustering Algorithm

KMeans clustering was applied to segment the cars into distinct groups based on pricing, performance, efficiency, and charging capabilities.

#### Choosing the Optimal Number of Clusters

The Elbow Method was used to determine the optimal number of clusters. The "elbow" in the plot represents the point where adding more clusters does not significantly reduce the sum of squared distances from each point to its centroid.

---

### 3.4 Segmentation Analysis

Clusters were identified based on the key features:

- **Cluster 1:** High-performance cars with rapid acceleration, high top speed, and premium pricing. Targeting luxury-seeking consumers.
- **Cluster 2:** Affordable electric cars with moderate charging speeds and energy efficiency. Targeting cost-conscious customers.
- **Cluster 3:** Long-range EVs with advanced fast charging capabilities. Targeting long-distance travelers.
- **Cluster 4:** Budget-friendly EVs with lower performance metrics. Targeting customers prioritizing affordability.

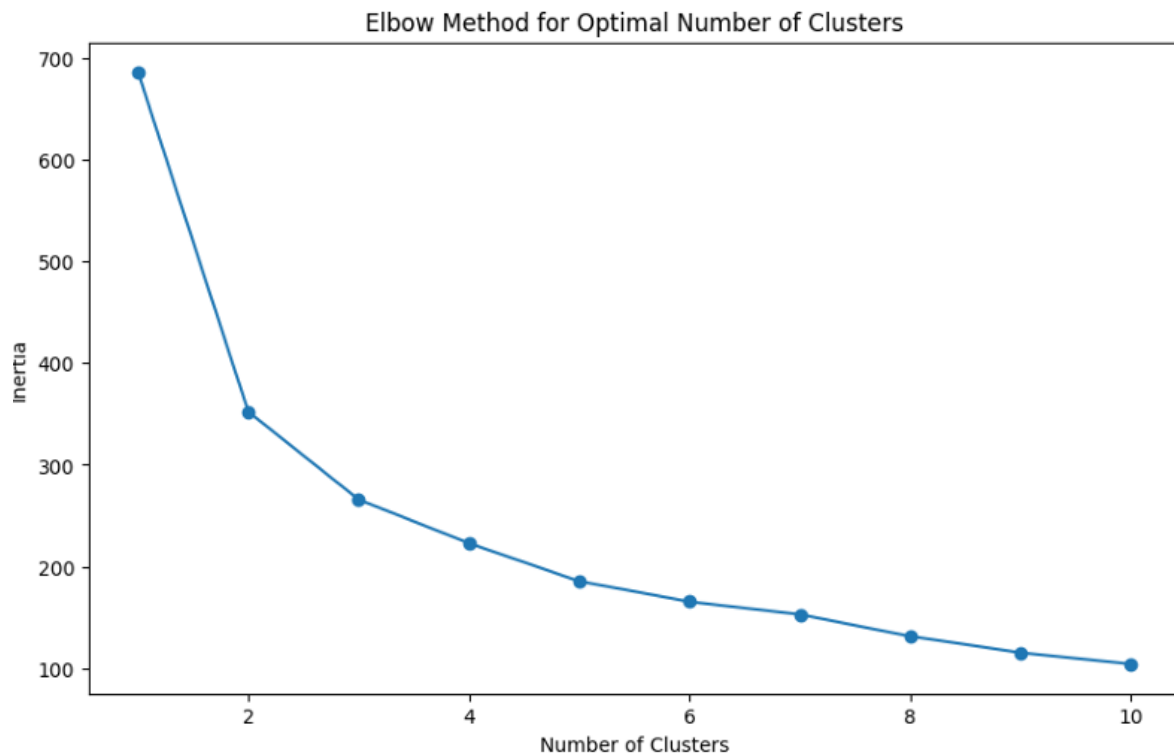


Figure 1: Elbow Plot

### 3.5 Results and Visualization

To visualize the clustering results, PCA (Principal Component Analysis) was used to reduce data dimensions, allowing for a 2D representation of the clusters.

Visualization tools used:

- **Scatter Plot:** To show the segmentation of car models, with colors distinguishing different clusters.
- **Cluster Centroids:** Representing the average characteristics of each cluster.

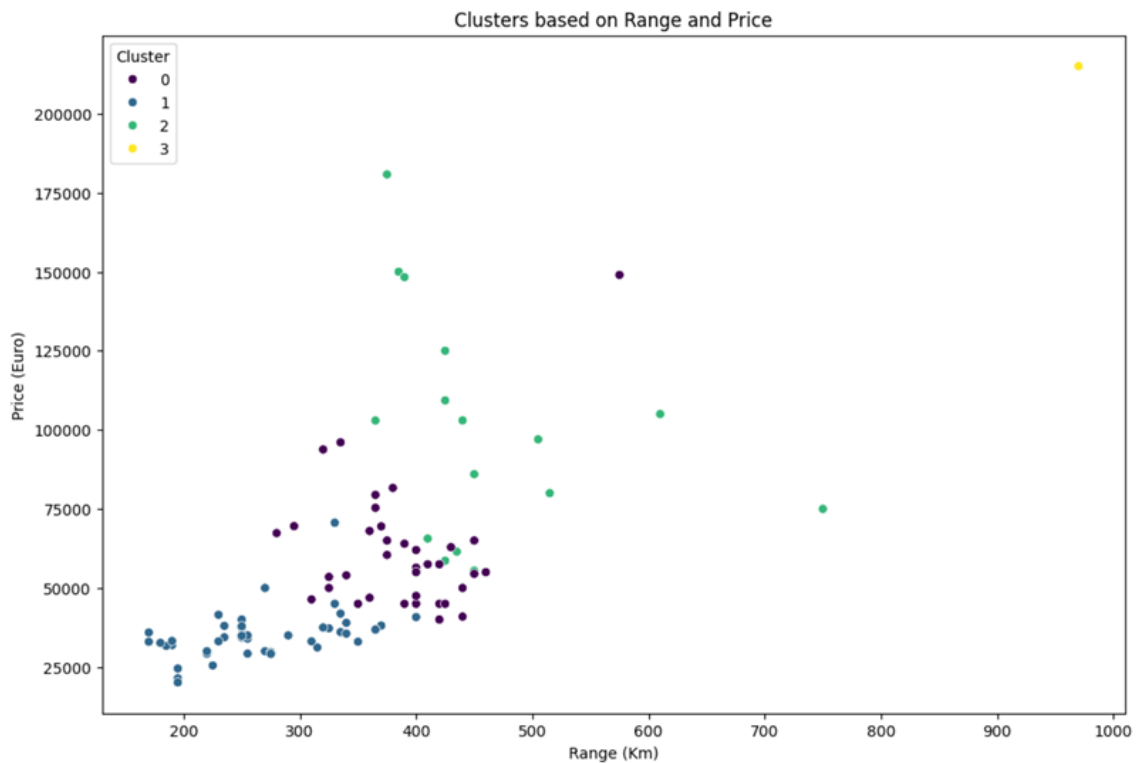


Figure 2: Cluster Plot

---

#### 4. Conclusion

The key variables—Price, Range, Acceleration, and FastCharge—were instrumental in creating the optimal market segments. These features capture essential aspects influencing consumer decisions:

- **Price:** Differentiates between luxury and budget models.
- **Range:** Differentiates between long-distance and short-distance vehicles.
- **Acceleration:** Highlights performance-focused models.
- **Charging Speed:** Addresses convenience concerns.

By focusing on these features, the EV market was segmented into distinct consumer groups, enabling precise targeting for marketing strategies, product development, and overall customer satisfaction.

# Electric Vehicle market Segmentation

Name: Saloni Rai

GitHub Link: <https://github.com/Saloni-Rai19/market-segmentation-analysis>

---

## Introduction:

My problem statement for this project was to segment features for the start-up of Electric Vehicles (EV). I chose a dataset of cars that consists of the following columns:

- **name:** Describes the car model name.
- **year:** Describes the year the customer purchased the car.
- **selling\_price:** Describes the selling price of the car.
- **km\_driven:** Describes the distance driven by the car.
- **fuel:** Describes the type of fuel used in the car.
- **seller\_type:** Describes whether the seller is an Individual, Dealer, or Trust Dealer.
- **transmission:** Describes whether the car transmission is manual or automatic.
- **owner:** Describes whether the owner is the first, second, third, fourth, or a test-driven car.
- **mileage:** Describes the mileage of the cars in kmpl or km/kg.
- **engine:** Describes the engine capacity in CC.
- **max\_power:** Describes the power of the car in bhp.
- **torque:** Describes the torque of the car.
- **seats:** Describes the number of seats in the car.

Before applying any algorithm, we gathered all information about the columns. For this, we used `describe()` and `info()`. Additionally, `isnull().sum()` gave us the number of NaN values in the dataset. These data points had to be removed or filled in. In our dataset, the null values were in **mileage**, **engine**, **max\_power**, **torque**, and **seats**. Specific techniques were adopted to remove them.

---

## Exploratory Data Analysis (EDA):

- The selling price of the cars increased until 2015–2018 and then decreased by 2020.
- Most people use diesel cars, and LPG cars are the least used.
- The majority of seller types are individuals, while Trustmark dealers are the least common.
- The most popular cars are:
  - Maruti Swift Dzire VDI
  - Maruti Alto 800 LXI

- Maruti Alto LX
- 

### Feature Extraction:

Two methods were used for feature extraction: **kModes** and **FAMD**.

My dataset consists mostly of columns with categorical data types and a few with numerical data types. Hence, these methods are suitable.

---

### kModes:

The **K-Modes algorithm** is a clustering technique specifically designed for categorical data. Unlike traditional clustering algorithms like K-Means, which use distance metrics and means, K-Modes uses modes (the most frequent values) to define cluster centroids.

#### How It Works:

1. **Initialization:** Randomly select K initial modes (centroids) from the dataset.
2. **Assignment:** Assign each data point to the cluster with the closest mode, based on the dissimilarity measure.
3. **Update:** Update the modes of the clusters by finding the most frequent values for each attribute within the cluster.
4. **Iteration:** Repeat the assignment and update steps until the modes no longer change or the maximum number of iterations is reached.

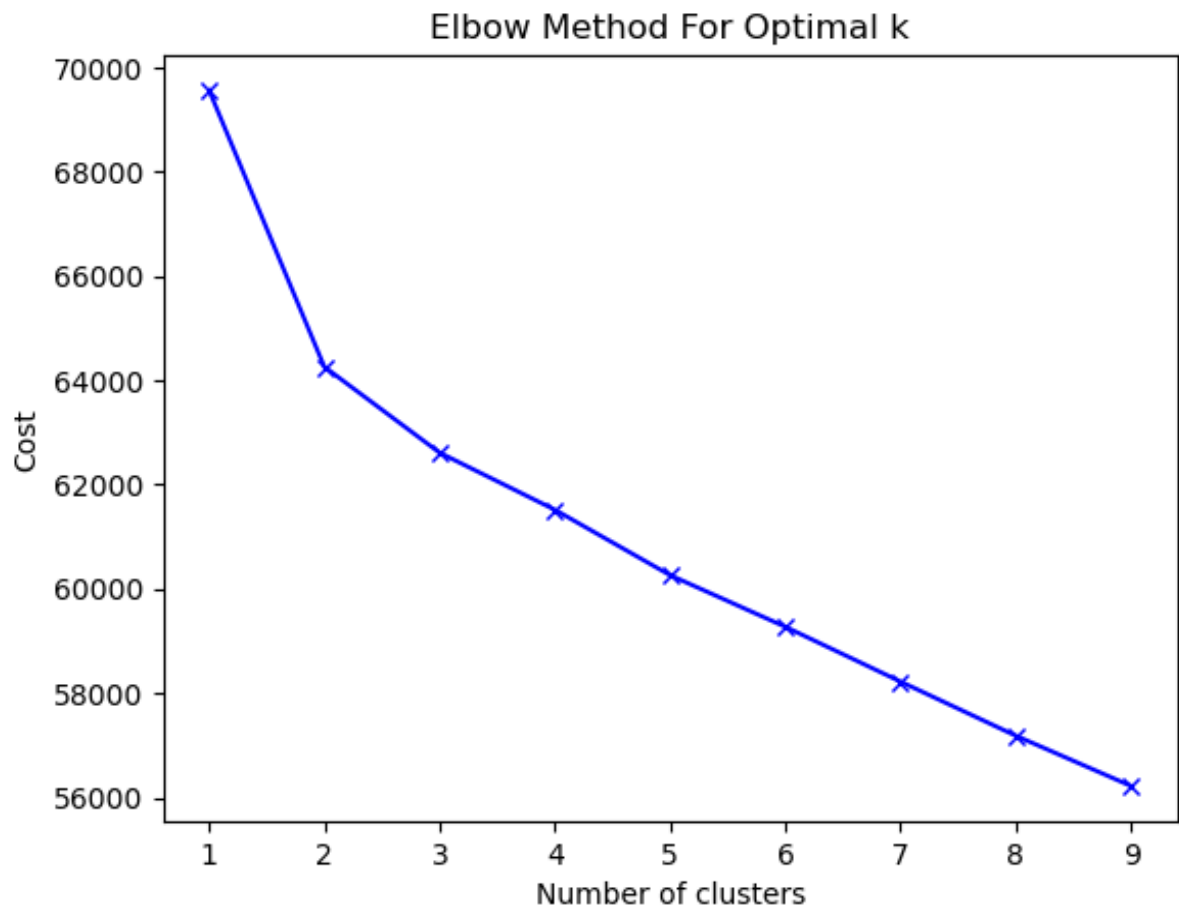
Before applying KModes, we first calculated the number of clusters using the **Elbow Method**.

---

### Elbow Method:

Key features of this method are:

- **Graphical Approach:** Involves plotting the within-cluster sum of squares (WCSS) against the number of clusters. WCSS measures the variance within each cluster.
- **Elbow Point:** As the number of clusters increases, WCSS decreases. The optimal number of clusters is identified at the “elbow point” of the plot, where the rate of decrease sharply slows down.



According to the graph of the Elbow Method, the optimal number of clusters is **3**.

---

**Cluster Information:**

Columns	Cluster 0	Cluster 1	Cluster 2
name	Maruti Alto 800 LXI	Maruti Swift Dzire VDI	BMW X4 M xDrive20d Sport X
year	2016	2013	2017
selling_price	300000	250000	600000
km_driven	70000	120000	120000
fuel	Petrol	Diesel	Diesel
seller_type	Individual	Individual	Individual
transmission	Manual	Manual	Manual
owner	First owner	Second owner	First owner
mileage	18.9	12.99	15.96

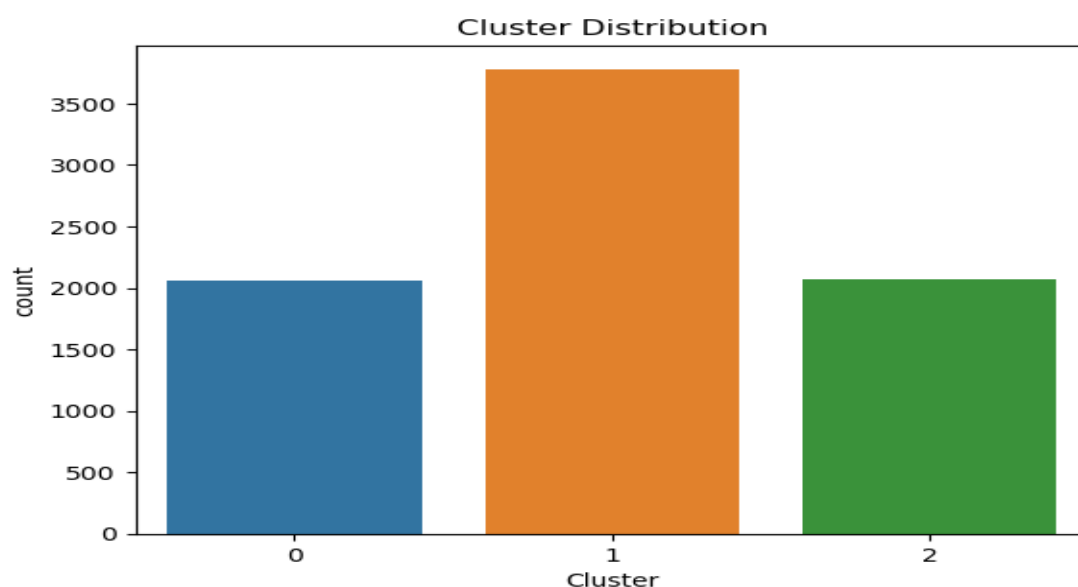
Columns	Cluster 0	Cluster 1	Cluster 2
engine	1197.0	1248.0	1248.0
max_power	81.8	74.0	88.5
torque	90Nm @ 3500rpm	190Nm @ 2000rpm	200Nm @ 1750rpm
seats	5.0	5.0	5.0

#### Cluster Explanation (Simplified):

- **Cluster 0:** This cluster mainly consists of **smaller, fuel-efficient cars** like the Maruti Alto 800 LXI. These cars are typically affordable, driven by **first owners**, and have lower mileage, which makes them a popular choice for city commuting. The fuel type here is mostly **Petrol**.
- **Cluster 1:** This group includes **slightly older, higher-mileage cars**, like the Maruti Swift Dzire VDI. These vehicles are **used more intensively** (as seen in the high km driven), often by **second owners**. Diesel cars dominate this cluster, reflecting their **longer life spans** and higher mileage compared to petrol cars.
- **Cluster 2:** This cluster has **more premium cars** like the BMW X4 M xDrive20d. These cars are generally newer, more expensive, and driven less compared to those in other clusters. Despite being premium, the **fuel type is still diesel**, possibly due to its efficiency in high-performance cars.

#### Cluster Distribution:

The distribution of these clusters is shown in the following plot, indicating that **Cluster 1** contains the largest portion of the dataset. This could be due to the fact that second-hand car buyers often go for higher-mileage, relatively older cars that still provide value.





---

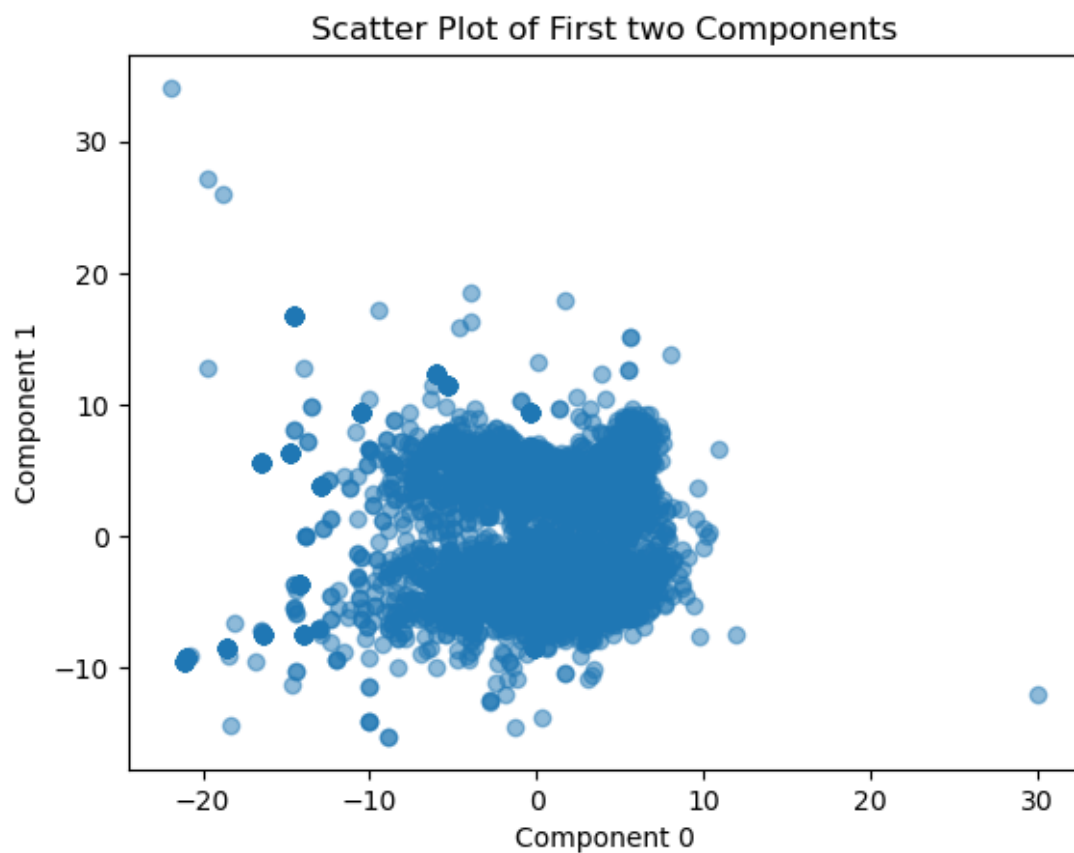
### FAMD Analysis:

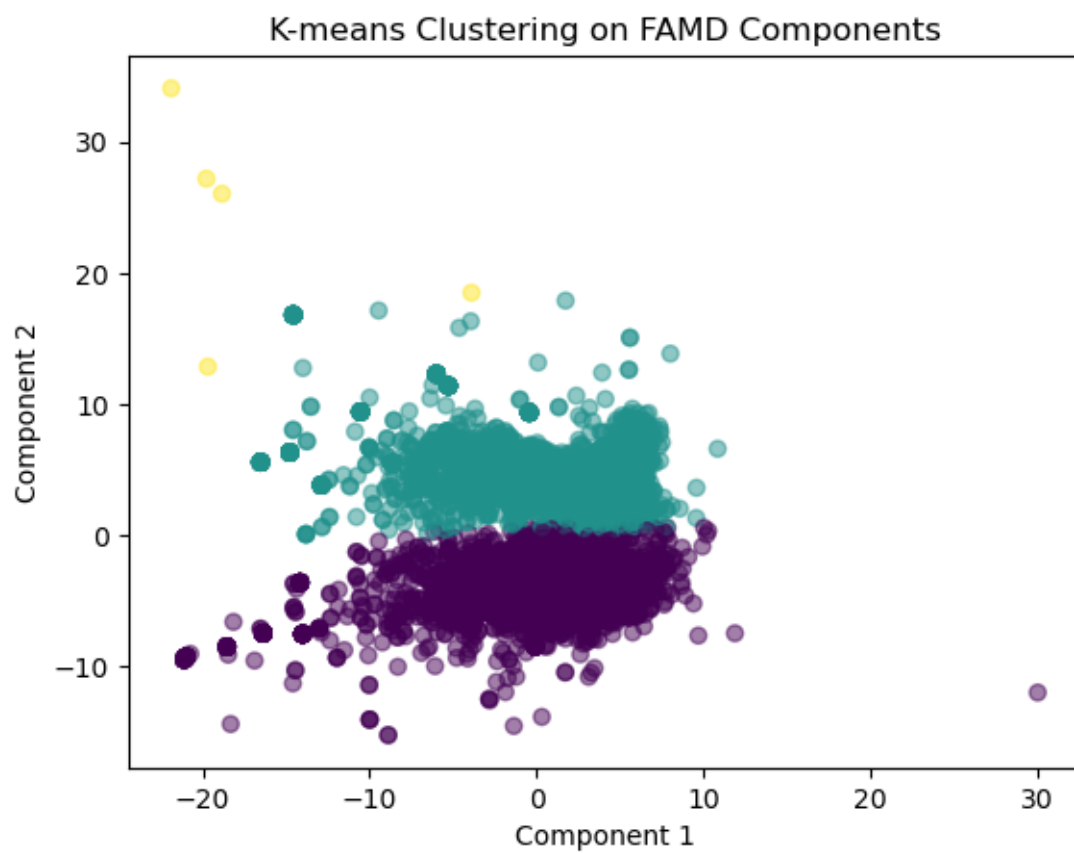
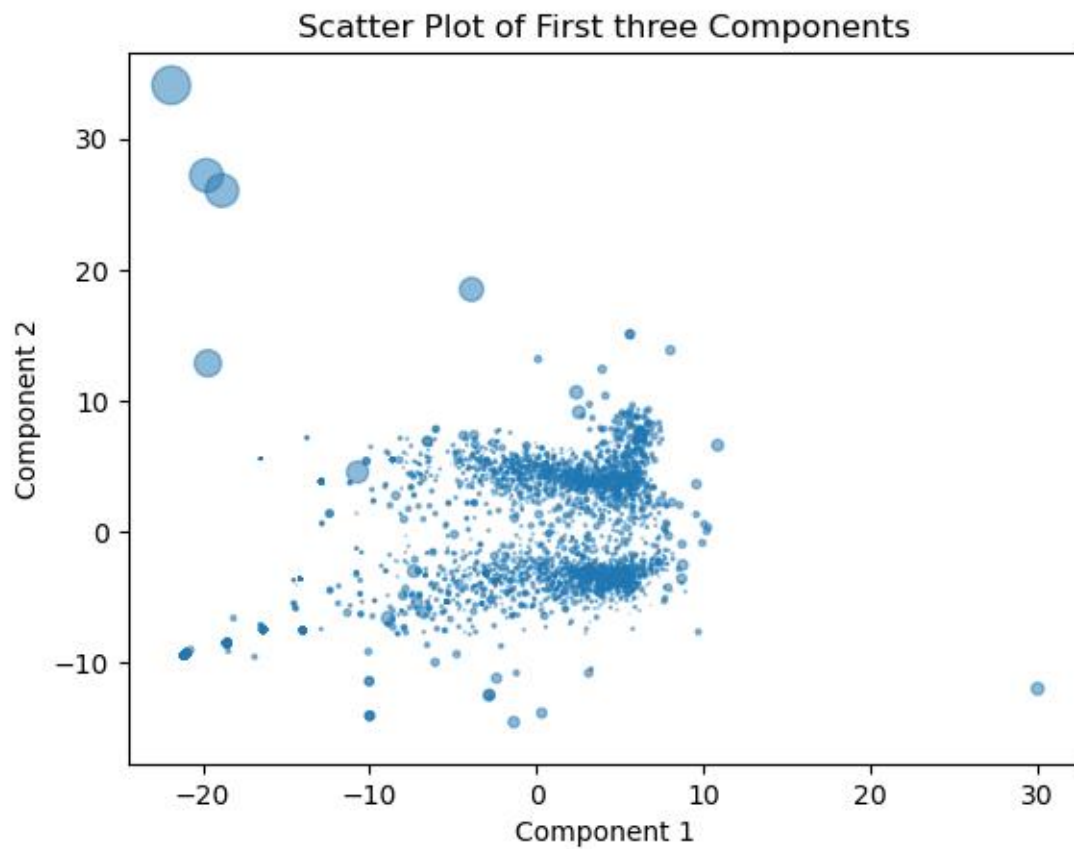
**Factor Analysis of Mixed Data (FAMD)** is a statistical technique designed to analyze datasets containing both quantitative and qualitative variables. It combines elements of **Principal Component Analysis (PCA)** for quantitative data and **Multiple Correspondence Analysis (MCA)** for qualitative data, making it a versatile tool for mixed data types.

---

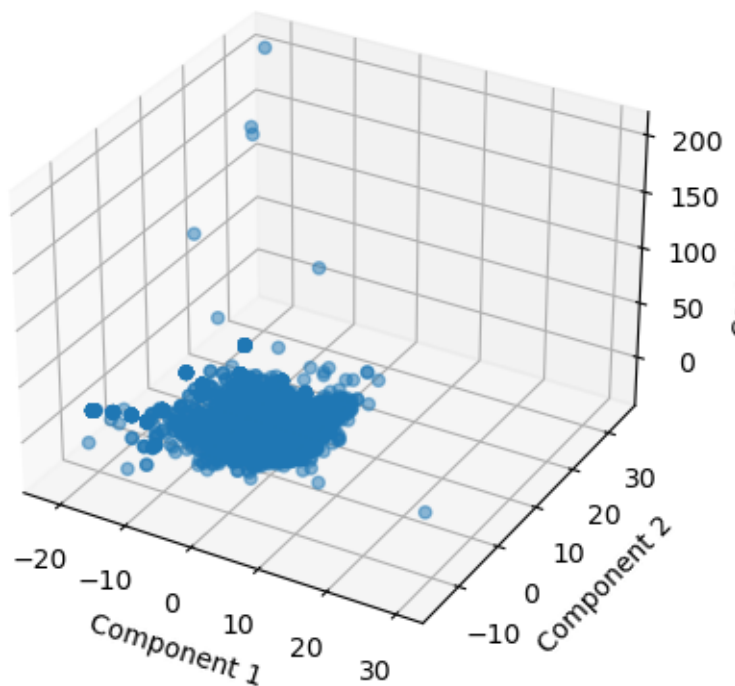
### Scatter Plots:

- **Scatter Plot** for the first two PCA components (indexing starts at 0):
- **Clustering** of these components:
- **3D Scatter Plot** for all three components:





### 3D Scatter Plot of First Three Components



#### Conclusion:

From the discussion above, we can conclude the following key insights:

1. **Key Factors for Car Selection:** The most significant factors influencing the choice of cars in this dataset are **selling\_price**, **km\_driven**, and **mileage**. These factors represent the financial and practical considerations buyers make when purchasing a car. Buyers often seek a balance between affordability, how much the car has been used, and its fuel efficiency.
2. **Cluster-Specific Insights:** The segmentation of the car models into three clusters reveals patterns in buyer preferences:
  - **Cluster 0** represents buyers looking for **small, efficient cars** that are affordable and good for short city commutes.
  - **Cluster 1** consists of buyers seeking **older, more used vehicles**, which are still reliable but come at a lower price. Diesel engines dominate this group, likely due to their longer lifespan.
  - **Cluster 2** targets **premium buyers** who want higher-end cars, like BMWs, but are still looking for fuel efficiency through diesel options.
3. **Market Implications for EV Startups:** For electric vehicle (EV) manufacturers, focusing on these identified segments can help capture a broader market. Buyers in **Cluster 0** may be looking for affordable, compact EVs, while **Cluster 1** customers may focus on the **durability and lower running costs** of EVs compared to traditional cars. **Premium EVs** could appeal to Cluster 2 buyers, who want a blend of luxury and performance but are environmentally conscious.