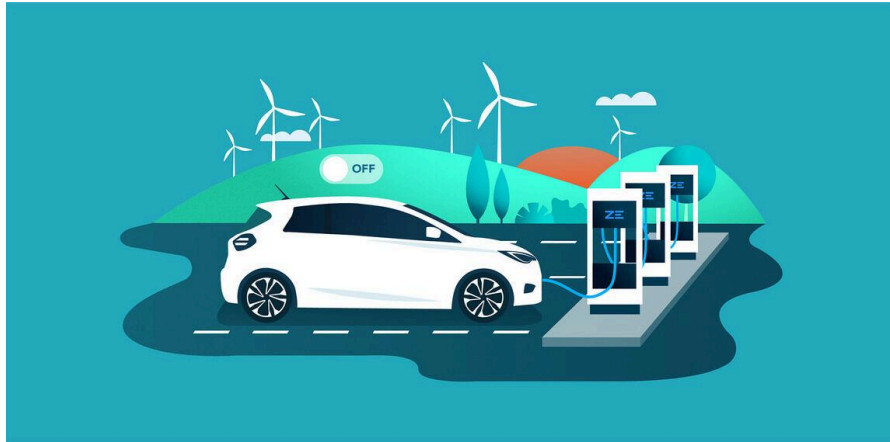


Indian EV Market Segmentation Based on Customer Reviews and Product Attributes

Manan Sonali Kher

4th February 2025



Abstract

The rapid expansion of the Indian Electric Vehicle (EV) market has led to an increasing need for data-driven insights into consumer preferences and product positioning. This report presents a market segmentation study using machine learning techniques, including customer sentiment analysis, text embeddings, and clustering algorithms. The objective is to uncover distinct customer and product segments to help EV manufacturers, marketers, and policymakers make informed decisions.

By leveraging customer reviews, ratings, and product attributes, we analyze purchasing behavior, key decision factors, and sentiment-driven insights. Techniques such as BART-based sentiment analysis, BERTopic topic modeling, and clustering (K-Means, GMM, t-SNE visualization) are employed to derive meaningful market segments. The findings provide actionable recommendations for improving customer experience, refining marketing strategies, and enhancing EV product offerings.

Github Link: <https://github.com/MacaroniMutton/Indian-EV-Market-Segmentation>

1. Introduction

1.1 Background

India's EV industry is witnessing a significant transformation, driven by government initiatives, increasing environmental awareness, and advancements in battery technology. With more consumers shifting towards electric vehicles, understanding consumer behavior, pain points, and key decision factors becomes crucial for business success.

Traditional market segmentation approaches rely on demographics or price range, but this study adopts an AI-driven approach by analyzing real-world customer feedback to uncover hidden patterns in preferences, satisfaction, and concerns.

1.2 Objective of the Study

This report aims to:

- ✓ Segment Indian EV customers based on reviews, ratings, and product perceptions.
- ✓ Identify key factors influencing EV purchase decisions.
- ✓ Profile product-based clusters to differentiate EV models based on attributes and customer experiences.
- ✓ Provide actionable insights for targeted marketing, product improvement, and customer engagement.

1.3 Research Methodology

To achieve the study's objectives, we employ:

- ◆ Exploratory Data Analysis (EDA): Identifying trends in customer feedback.
- ◆ Sentiment Analysis: Extracting emotions and opinions using BART-based aspect-based sentiment analysis.
- ◆ Topic Modeling: Discovering themes using BERTopic to group common concerns and discussions.
- ◆ Clustering Algorithms: Applying K-Means, Gaussian Mixture Models (GMM), and t-SNE visualization to segment data.
- ◆ Actionable Business Insights: Profiling segments and recommending targeted strategies.

2. Scope of Report

2.1 Industry Focus

This report focuses on the Indian two-wheeler EV market, specifically analyzing user reviews from the BikeWale website. The insights derived from this study will help manufacturers, marketers, and policymakers understand customer sentiment, product perception, and market segmentation for electric scooters and bikes in India.

The key EV models covered in this study include:

- ✓ TVS iQube
- ✓ Ola S1 Pro
- ✓ Ather 450X
- ✓ Bajaj Chetak
- ✓ Hero Vida V1, among others.

By analyzing customer feedback, this report will provide a data-driven segmentation of EV users and products, offering actionable recommendations for enhancing product features, optimizing marketing strategies, and improving customer satisfaction.

2.2 Data Sources

The dataset used in this study was extracted from BikeWale, a popular vehicle review platform, and consists of customer reviews and ratings for electric two-wheelers. The data includes:

- ✓ Customer-written reviews about their experiences with EVs.
- ✓ Ratings on various aspects like performance, comfort, service experience, and value for money.
- ✓ Usage details (e.g., how long the vehicle was owned, how far it has been ridden).

In addition to raw data, derived features were created to enhance the analysis, including:

- ✓ Aspect-based Sentiment Analysis: Extracting sentiment scores for different aspects of EVs (e.g., battery, performance, comfort).
- ✓ Text Embeddings: Converting textual reviews into numerical representations using Transformer-based models for clustering.
- ✓ Topic Modeling: Identifying themes within reviews to understand key concerns and preferences.

2.3 Target Audience

The insights from this study will benefit multiple stakeholders, including:

- ♦ EV Manufacturers: Understanding customer pain points, product perception, and feature improvement areas.

-
- ♦ Marketers & Sales Teams: Designing targeted campaigns and messaging based on customer segments.
 - ♦ Investors & Policymakers: Gaining data-driven insights into market trends, adoption challenges, and growth opportunities in the Indian two-wheeler EV sector.

3. Data Overview

3.1 Dataset Description

The dataset consists of **844 user reviews** for electric two-wheelers, with **14 features** related to customer feedback and product experience. It consists of **39 distinct EV bike models**. The dataset structure is as follows:

| Feature Name | Description |
|--------------------|--|
| review | Customer-written review about the EV experience. |
| Used it for | Purpose of the EV usage (e.g., daily commute, long rides). |
| Owned for | Duration for which the user owned the EV. |
| Ridden for | Distance traveled by the EV in kilometers. |
| rating | Overall rating given by the user (1-5). |
| Visual Appeal | Rating for the EV's design & looks. |
| Reliability | Rating for durability & dependability. |
| Performance | Rating for speed, power, and acceleration. |
| Service Experience | Rating for after-sales support & servicing. |
| Extra Features | Rating for additional features like smart connectivity. |
| Comfort | Rating for riding comfort. |
| Maintenance cost | Rating for cost-effectiveness of service & repairs. |
| Value for Money | Rating for overall worth of the EV. |
| Model Name | Name of the EV model reviewed. |

3.2 Derived Features

To enhance the analysis, **additional features** were generated using advanced NLP techniques:

- Aspect-Based Sentiment Analysis - Extract sentiment scores (positive, neutral, negative) for specific product aspects like **battery, performance, comfort, and service experience**. **BART-based one-shot sentiment classification** was used to assign aspect-level sentiment scores to each review.

| Aspect | Description | Sentiment Example |
|------------------|---|-------------------------------------|
| Battery Life | Battery performance, range, degradation | "Battery drains fast!" (-1) |
| Performance | Speed, acceleration, ride quality | "Smooth and fast ride!" (+1) |
| Comfort | Seating, ride experience | "Seat is uncomfortable." (-1) |
| Charging Issues | Charging speed, availability | "Charging takes too long!" (-1) |
| Maintenance Cost | Repair and servicing expenses | "Very low maintenance cost." (+1) |
| Customer Service | Dealership & service center experience | "Service center is unhelpful." (-1) |

- Text Embeddings - Convert text reviews into numerical representations for clustering. Used Transformer-based embeddings to capture semantic meaning.
- Topic Modeling (BERTopic) - Identify key themes and recurring topics in customer reviews. BERTopic clustering was applied, revealing the most discussed topics.

4. Segmentation Criteria

This dataset supports behavioral and psychographic segmentation:

- Behavioral: Based on usage patterns (e.g., "Used it for", "Owned for", "Ridden for").
- Psychographic: Based on user priorities (e.g., "Visual Appeal", "Comfort", "Value for Money").
- Product-Based: Using "Model Name" to identify segments tied to specific bike models.

5. Market Segmentation Analysis Step-By-Step

Step 1: Deciding Whether to Segment the Market

Why segment the EV market?

- The Indian EV two-wheeler market is diverse, with customers having different needs, expectations, and purchase behaviors.
- Market segmentation identifies distinct groups based on customer sentiment, product features, and usage patterns.
- Helps manufacturers and marketers optimize product offerings, pricing, and marketing strategies for different customer segments.

Step 2: Identifying the Ideal Target Customer

Who are the potential EV buyers?

- Daily Commuters → Looking for affordability & efficiency.
- Tech Enthusiasts → Want smart features & cutting-edge design.
- Performance Seekers → Prioritize speed & acceleration.
- Eco-Conscious Consumers → Prefer sustainability & government incentives.
- Cost-Conscious Buyers → Focus on low maintenance & long-term savings.

Why is this important?

- Helps tailor product features & marketing campaigns for each segment.
- Ensures higher customer satisfaction and brand loyalty.

Step 3: Collecting Data

Sources of Data

- Bikewale.com user reviews → Real customer feedback on EV models.
- Product specifications → Range, battery life, charging time, etc.
- Feature ratings → Performance, reliability, comfort, service experience.

Data Focus

- Quantitative Data → Ratings, numerical aspects of EV models.
- Qualitative Data → Customer reviews for sentiment & topic analysis.

Step 4: Exploring the Data

The EDA aims to uncover patterns in customer feedback and segment users based on their experiences. Key techniques include **sentiment analysis, text embeddings, and clustering**.

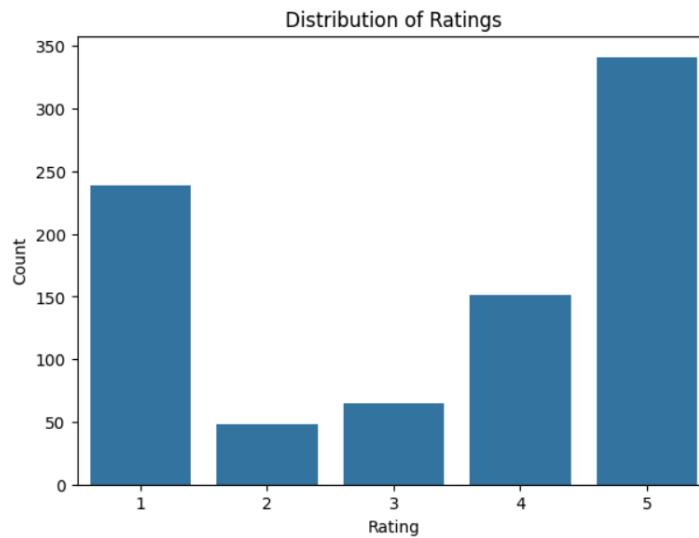
1. Missing Data Heatmap



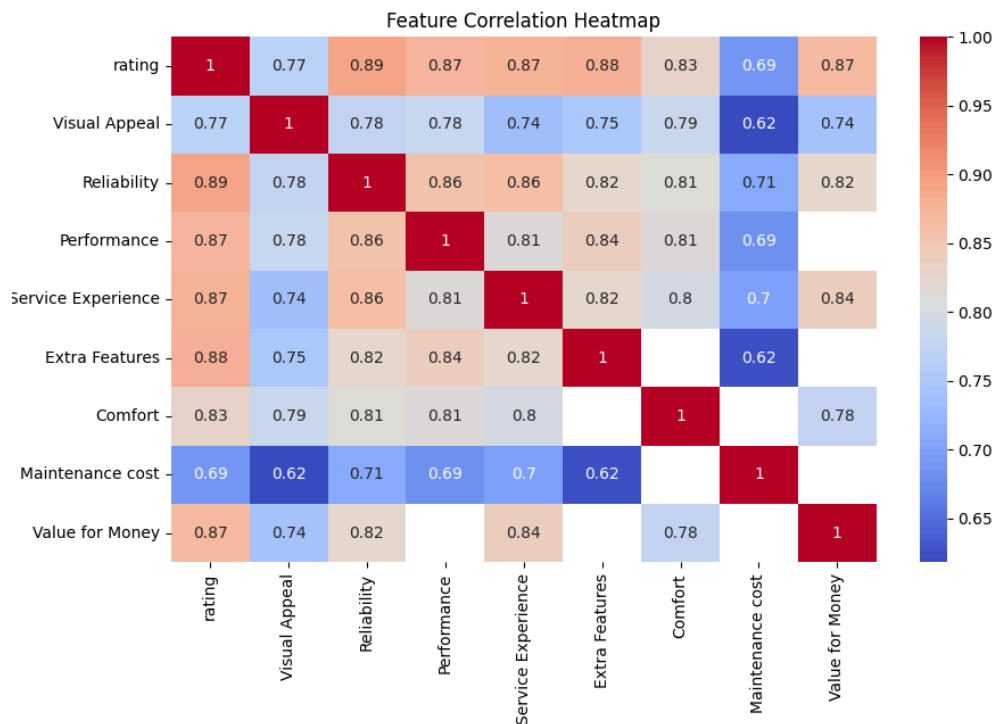
The dataset contains missing values across multiple columns, particularly in user-provided ratings for specific EV aspects. Some key observations:

- Review Text (5.5% missing) → Some users provided ratings but did not write a review.
- Performance Rating (59.1% missing) → Many users skipped performance evaluation, possibly due to limited riding experience.
- Extra Features (78.1% missing) → Indicates that users either did not explore additional features or they were not prominent in the models.
- Maintenance Cost (78.7% missing) → Suggests either a lack of maintenance experience or that maintenance cost is not a major concern.
- Value for Money (53.7% missing) → Could indicate subjective opinions varying among users.

2. The distribution of ratings given in the dataset



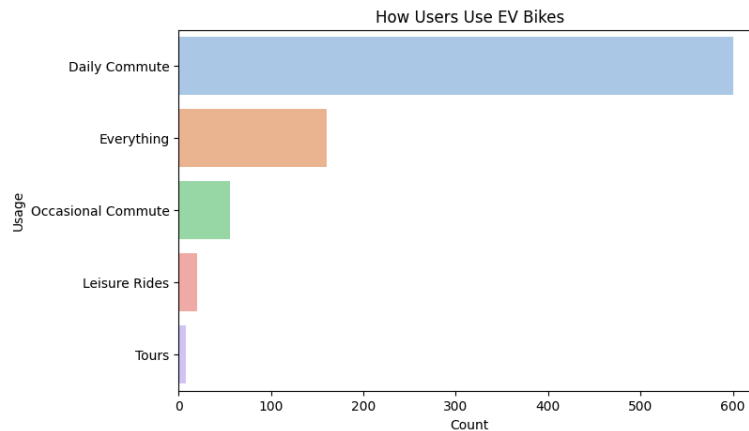
3. Correlation between the Features



The heatmap reveals that reliability, performance, service experience, extra features, and value for money are the strongest factors influencing overall ratings, suggesting that users prioritize functionality over aesthetics. Service experience and value for money are highly correlated, indicating that better after-sales service enhances customer satisfaction.

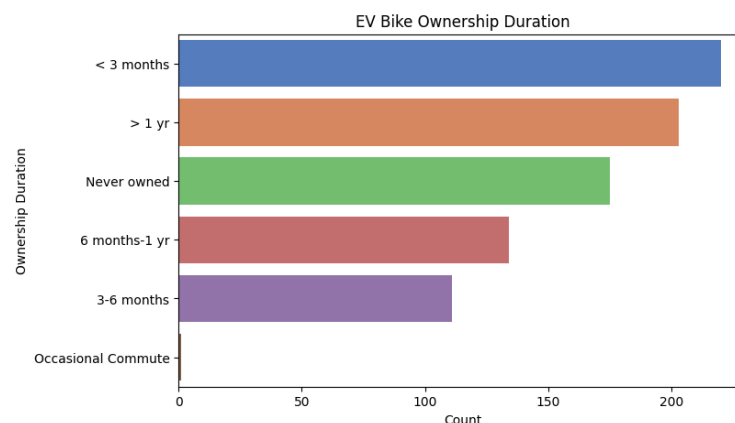
Maintenance cost has the weakest impact on ratings, meaning users may tolerate higher costs if the EV delivers on reliability and features. Companies should focus on improving performance, durability, and customer support to maximize positive reviews and market success.

4. Usage statistics of EV bikes



The count plot reveals that Daily Commute is the dominant use case for EV bikes, highlighting their practicality and cost-effectiveness for regular transportation. The Everything category follows, indicating that many users rely on EV bikes for multiple purposes, such as commuting, errands, and casual rides. Occasional Commute is a smaller segment, suggesting that some users treat EVs as a secondary mode of transport. Leisure rides and tours are the least common, implying that EVs are not yet widely adopted for long-distance travel or recreational use. These insights suggest that manufacturers should prioritize battery life, charging convenience, and comfort for daily commuters while also exploring ways to attract occasional and leisure riders.

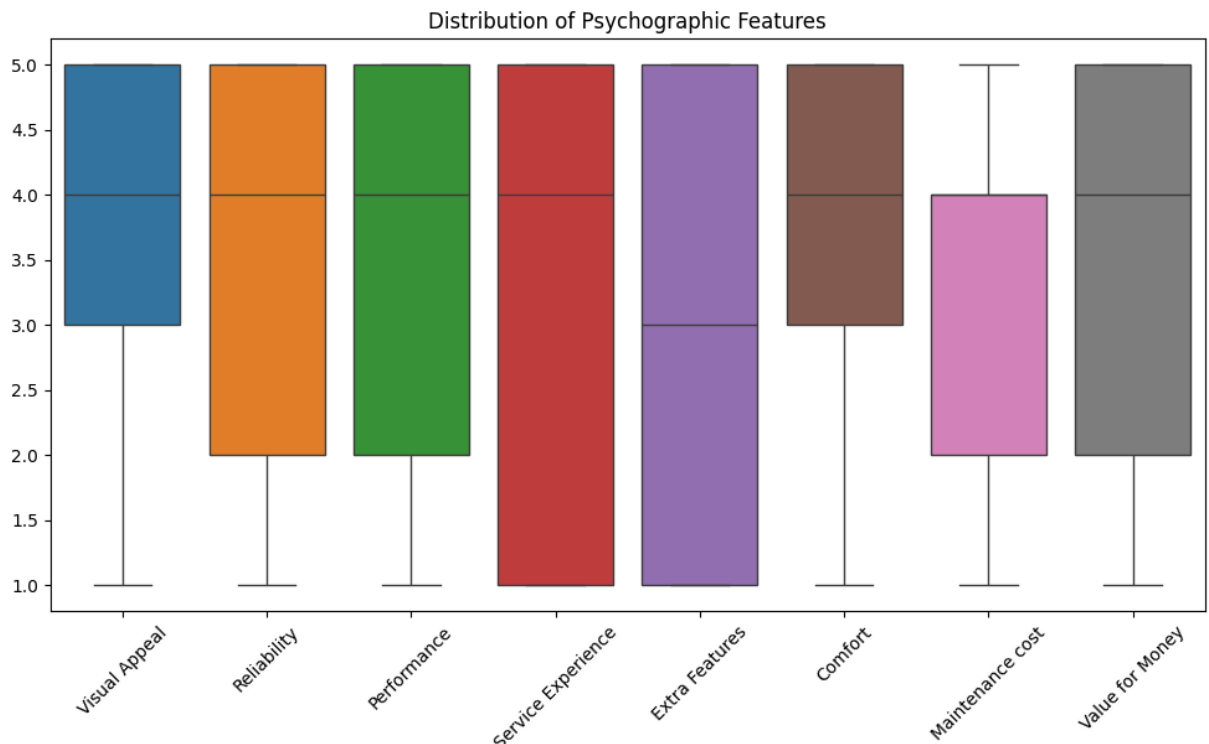
5. Duration for which users have owned an EV bike



Most EV bike owners fall into <3 months and >1 year categories, showing a mix of new adopters and experienced users. A notable share has owned their EVs for 6 months to 1 year, indicating growing retention. The Never Owned group likely represents test riders or potential buyers. The occasional commute entry appears to be a data error. These insights suggest that brand loyalty, long-term performance, and after-sales support are crucial factors for sustained EV adoption.

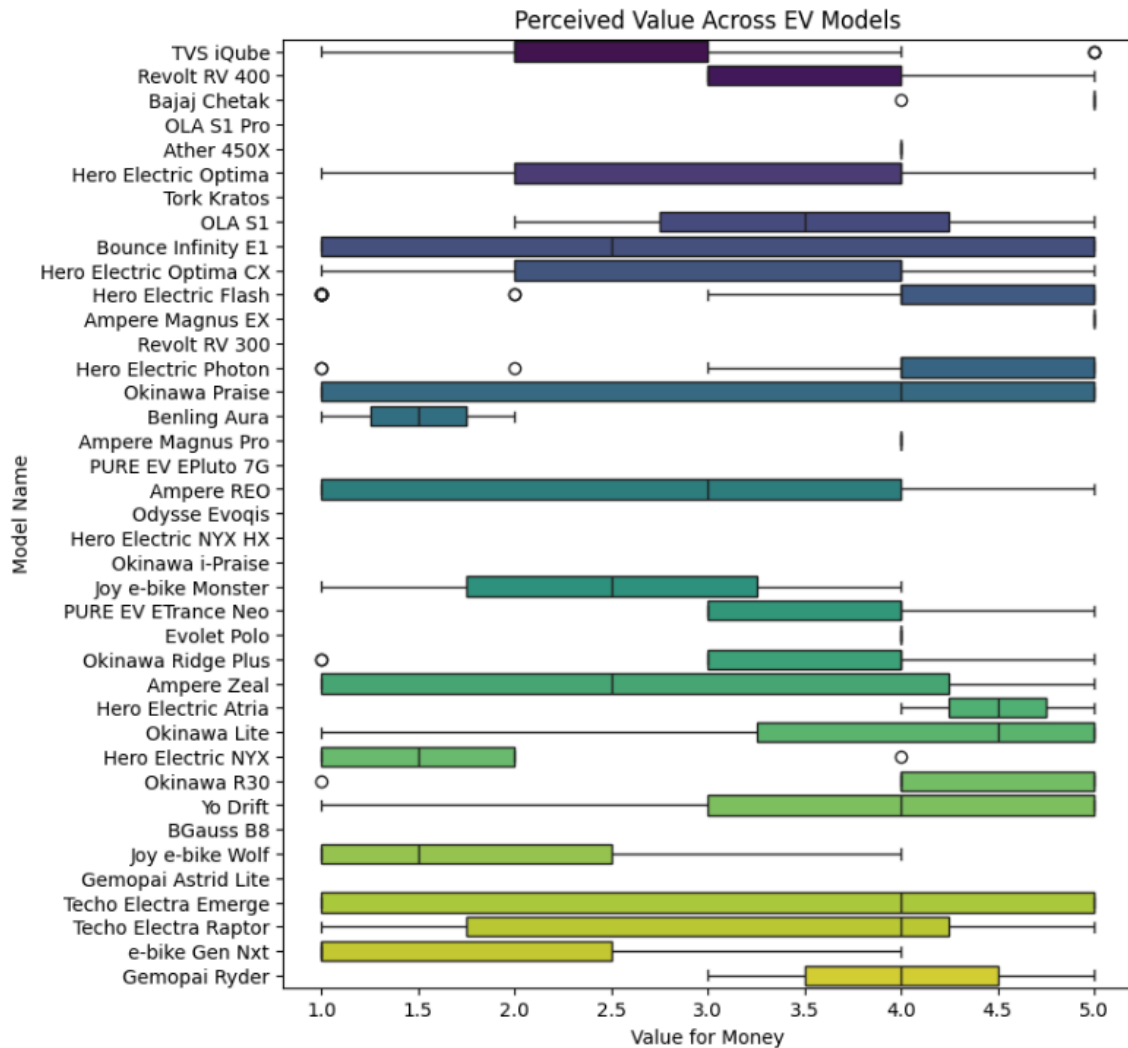
6. Distribution of Psychographic Factors Influencing EV Bike Adoption

This visualization highlights the spread and variability of key psychological and experiential factors—such as Visual Appeal, Reliability, Performance, and Value for Money—that shape user perceptions and satisfaction with EV bikes.



The box plot shows the distribution of ratings for various psychographic features of EVs. Most features have a high median (around 4), indicating generally positive user perceptions. Service Experience and Extra Features have the widest spread, suggesting varied opinions—some users rate them very low while others rate them highly. Maintenance Cost has the lowest median and a large spread, implying that cost is a major concern for users. To enhance customer satisfaction, manufacturers should focus on improving service quality, feature offerings, and cost-effectiveness.

7. User's Perception of Value for Money for various EV bike models



Consumer ratings reveal distinct market segmentation patterns in India's electric two-wheeler sector. Mass market models like Hero Electric Flash and Bounce Infinity E1 demonstrate superior value-for-money ratings (4.0-4.5), indicating strong product-market fit in the price-sensitive Indian market. However, premium segment leaders like Ather 450X and Ola S1 Pro show moderate value ratings despite higher prices, suggesting their target audience prioritizes features and brand value over pure price considerations.

Notably, there's significant rating variance across models. Wide confidence intervals for brands like Revolt RV 400 indicate divided consumer opinions, potentially highlighting inconsistent product experiences or varying expectations across different user segments. In contrast, models like Benling Aura show consistent but lower ratings, suggesting clear misalignment with market expectations.

Negative sentiments (scores below 0) are relatively rare, making up a small portion of the reviews. This suggests that while customers may have varying degrees of satisfaction, outright dissatisfaction is uncommon. The distribution tapers off gradually on both ends, with very few extreme opinions (either highly negative or extremely positive). This balanced but positive-leaning distribution indicates generally favorable market reception of electric two-wheelers in India, though there's still room for improvement in customer satisfaction levels.

Step 5: Creating Market Segments

Market segmentation is the process of dividing consumers into groups based on shared characteristics, enabling businesses to target each segment effectively. In this report, we perform segmentation based on behavioral factors, psychographic factors, and product attributes, leveraging advanced clustering techniques like K-Means, Gaussian Mixture Model (GMM), and PCA-t-SNE for visualization.

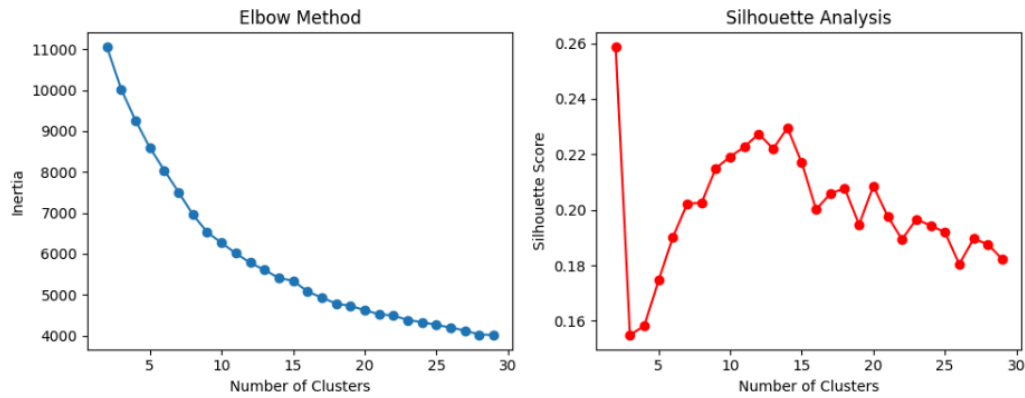
1. Behavioural and Psychographic Segmentation

Behavioral segmentation groups customers based on their usage patterns, purchasing behavior, and sentiment towards EV bikes. This helps in identifying how different consumer groups interact with EV products. Psychographic segmentation categorizes consumers based on attitudinal and emotional aspects influencing their EV purchasing decisions.

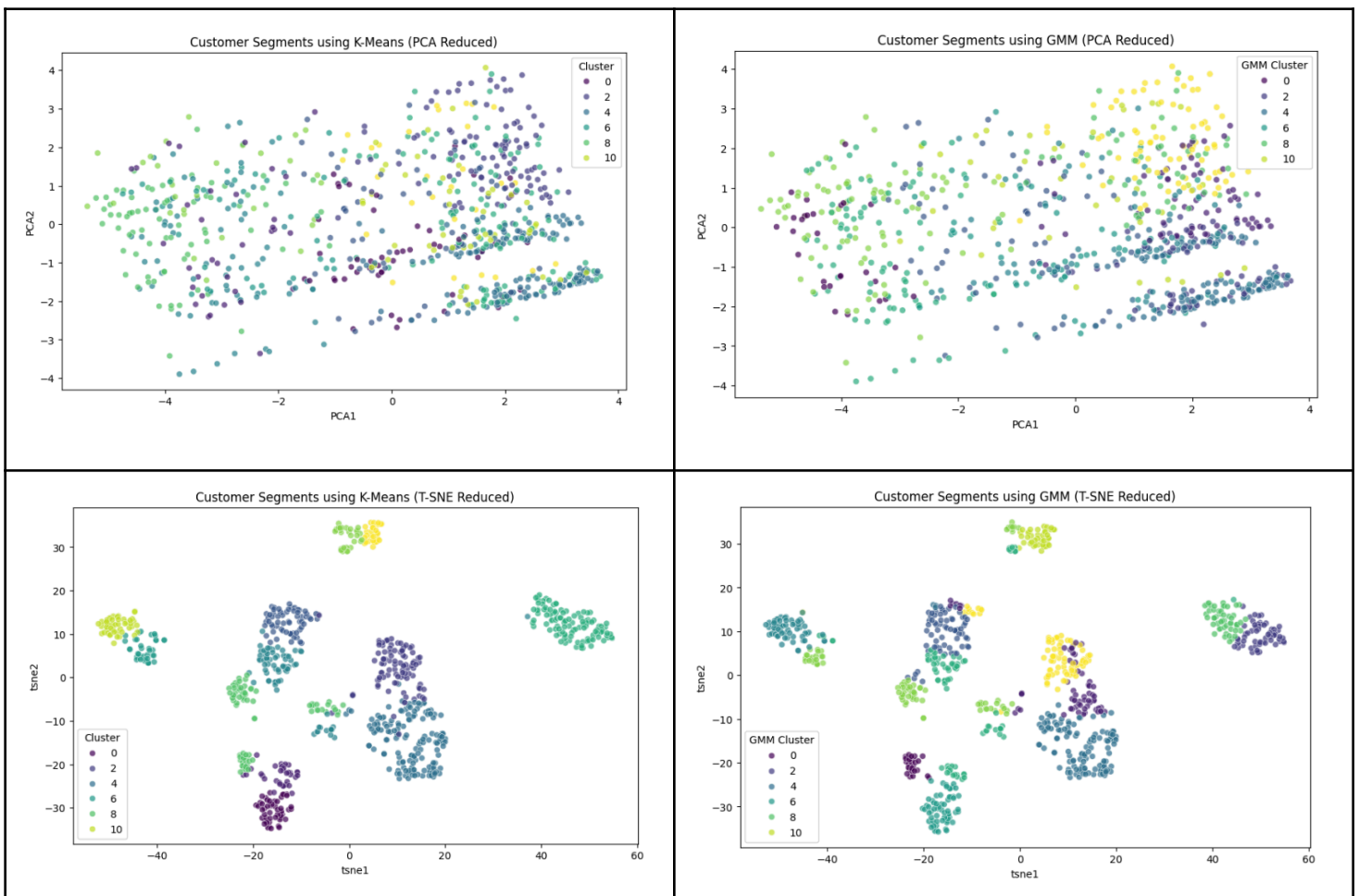
Features Used for Segmentation:

- Usage Behavior: (e.g., "Daily Commute," "Occasional Commute," "Everything")
- Ownership Duration: (e.g., "<3 months", ">1 year", "Never Owned")
- Sentiment-Based Topics: Extracted using BERTopic, capturing key themes in user reviews.
- Aspect-Based Sentiment Analysis: Measuring sentiment for specific product aspects like Battery Life, Performance, Comfort, Charging Issues, Maintenance Cost, and Customer Service.
- Attitudinal Features - Visual Appeal, Reliability, Performance, Service Experience, Extra Features, Comfort, Maintenance Cost, Value for Money

First, we find the optimal K value for performing K-Means clustering and Gaussian Mixture Model. This can be done using the Elbow Method along with Silhouette Scores Analysis, to find the optimal number of clusters.



From the graphs, the optimal K value is around 12. We applied K-Means and GMM, visualizing clusters with PCA and t-SNE. While PCA is useful for dimensionality reduction, its low explained variance limited its ability to capture meaningful structures from embeddings, sentiment scores, and psychographic features. To better reveal non-linear relationships, we used t-SNE, which preserves local similarities and helps distinguish customer segments more effectively.



2. Product-Based Segmentation

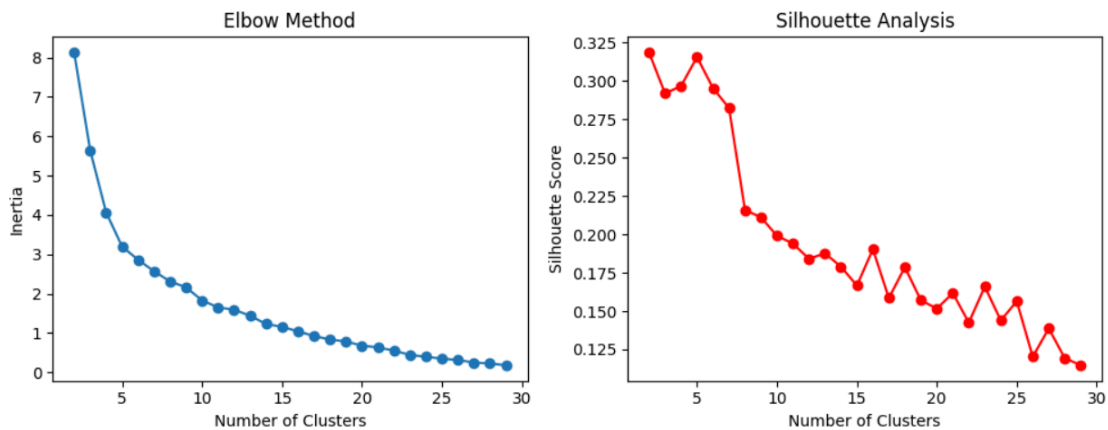
This segmentation groups **EV bike models** based on their technical specifications, performance ratings, and user feedback. It helps in:

- Benchmarking different models
- Identifying gaps in product design
- Understanding how each EV model is perceived by the market

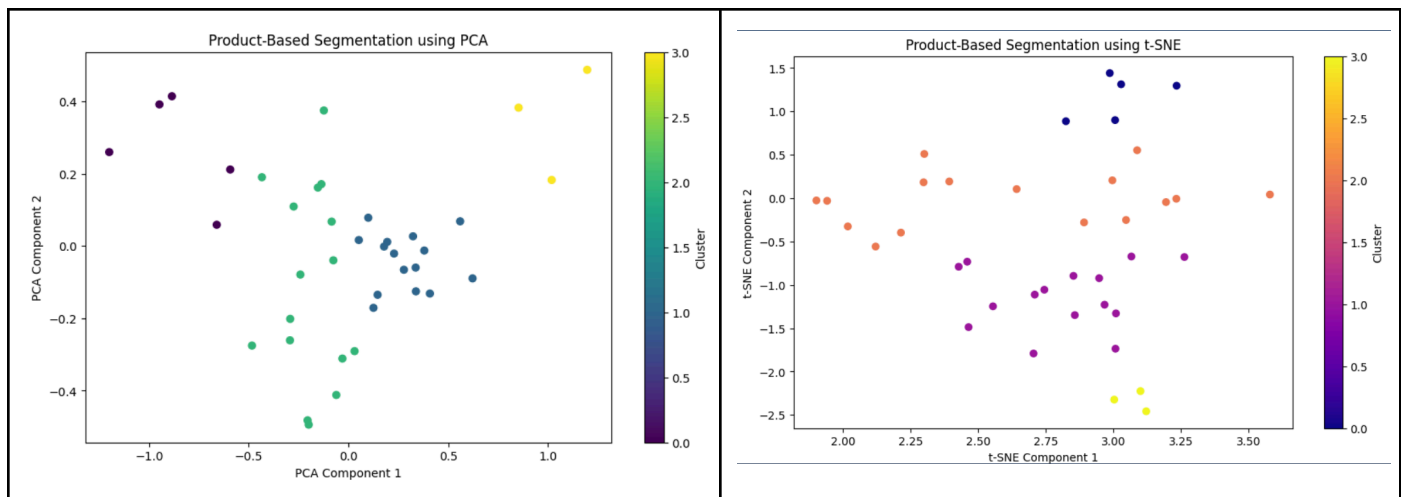
Features Used:

- Performance Ratings (Aggregated from Reviews)
- Maintenance & Service Experience Scores

We group data by EV model and compute mean scores for key features, then follow the same procedure for performing segmentation



The optimal K value is 4 for product based clustering.



Cluster 0:

Ampere REO, Bounce Infinity E1, Hero Electric NYX, Joy e-bike Monster, Joy e-bike Wolf, TVS iQube

Cluster 1:

Ampere Magnus EX, Ather 450X, Bajaj Chetak, Gemopai Ryder, Hero Electric Atria, Hero Electric Flash, Hero Electric NYX HX, Hero Electric Photon, OLA S1 Pro, Okinawa Lite, Okinawa R30, PURE EV ETrance Neo, Revolt RV 300, Revolt RV 400, Tork Kratos

Cluster 2:

Ampere Magnus Pro, Ampere Zeal, BGauss B8, Benling Aura, Hero Electric Optima, Hero Electric Optima CX, OLA S1, Okinawa Praise, Okinawa Ridge Plus, Okinawa i-Praise, PURE EV EPluto 7G, Techo Electra Emerge, Techo Electra Raptor, Yo Drift, e-bike Gen Nxt

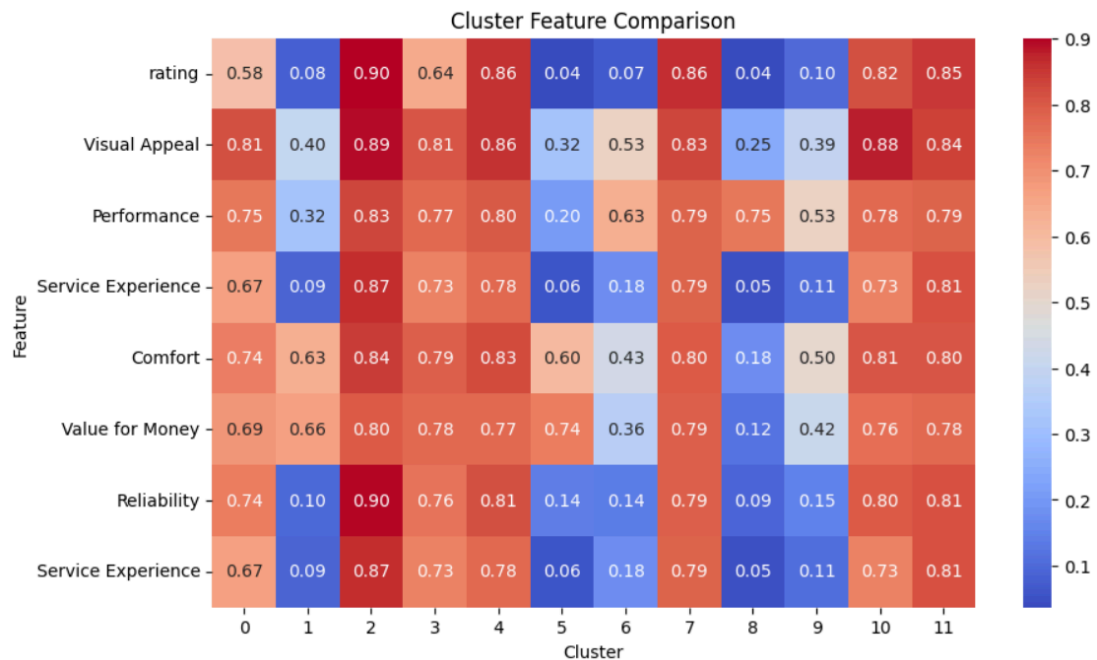
Cluster 3:

Evolet Polo, Gemopai Astrid Lite, Odysse Evoqis

Step 6: Understanding Each Customer Group

Understanding each customer group involves analyzing the **behavioral, psychographic, and product-based segmentation** to identify common patterns in user preferences, needs, and pain points.

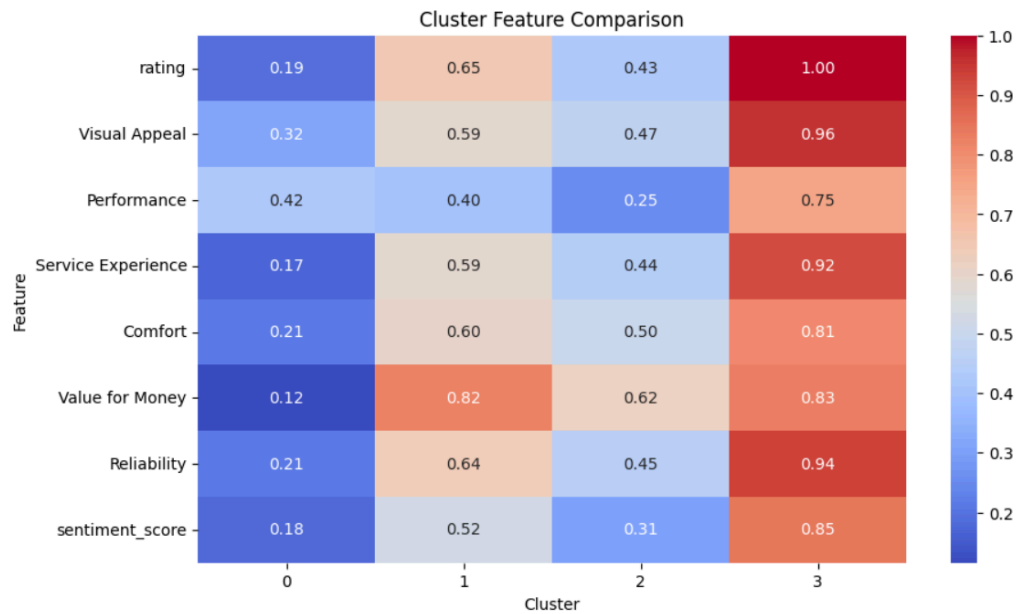
1. Behavioral & Psychographic Segmentation:



Based on the cluster feature heatmap:

- Clusters 2, 4, 7, 10, and 11 exhibit high ratings across performance, reliability, and service experience, indicating a segment of satisfied and brand-loyal users who value premium features and reliability.
- Clusters 1, 5, 6, 8, and 9 have low ratings, suggesting dissatisfied users who may be experiencing reliability or service issues.
- Some clusters (e.g., Cluster 6) show high performance scores but low value-for-money, indicating users who appreciate power but feel the cost is too high.

2. Product-Based Segmentation



- Cluster 3 represents high-end models with top ratings in visual appeal, service experience, and reliability, making it a premium segment.
- Cluster 1 has moderate ratings across all attributes, suggesting a balance between affordability and performance.
- Cluster 0 has low scores across most features, likely representing budget-conscious buyers who compromise on quality.

Step 7: Describing the Segments in Detail

Each segment is characterized by specific behaviors and expectations. The clusters can be broadly categorized into the following:

1. **Premium Enthusiasts** (High Ratings – Clusters 2, 4, 7, 10, 11 in Behavioral/Psychographic & Cluster 3 in Product-Based)
 - Seek top-tier performance, reliability, and service quality
 - Willing to pay a premium for EV bikes with superior technology

-
- Likely to be brand-loyal and prefer advanced features
 - 2. **Balanced Buyers** (Moderate Ratings – Clusters 0, 3 in Behavioral/Psychographic & Cluster 1 in Product-Based)
 - Value comfort, service experience, and affordability
 - Seek a mix of performance and price efficiency
 - May switch brands if better value-for-money options are available
 - 3. **Budget-Conscious Users** (Low Ratings – Clusters 1, 5, 6, 8, 9 in Behavioral/Psychographic & Cluster 0 in Product-Based)
 - Prioritize cost over premium features
 - May experience lower satisfaction due to maintenance issues
 - Require better customer service and value-for-money offerings

Step 8: Choosing the Best Target Segments

Based on the insights, the best segments to target would be:

- **Premium Enthusiasts:**
 - **Why?** High willingness to spend, brand loyalty, potential for upselling new models and premium features.
- **Balanced Buyers:**
 - **Why?** This segment is large and can be persuaded to upgrade if value-for-money is emphasized.

Budget-conscious users have low satisfaction, meaning high churn risk. However, targeted campaigns offering affordable financing or after-sales service improvements may help retain them.

Step 9: Creating a Marketing Plan for Each Group

A tailored marketing approach can be designed for each segment:

1. Premium Enthusiasts:
 - Focus on premium branding and exclusive offers
 - Highlight high performance, extended battery life, and top-tier service
 - Use brand loyalty programs
2. Balanced Buyers:
 - Position models as a balance of affordability and quality
 - Offer exchange programs or trade-in offers
 - Focus on customer service improvements
3. Budget-Conscious Users:
 - Provide budget-friendly financing plans
 - Educate on long-term savings and EV cost efficiency
 - Improve after-sales service reputation

Step 10: Monitoring Performance

To ensure effective segmentation, key metrics should be tracked:

- Customer Satisfaction Scores (NPS, Review Ratings) for each segment
- Retention Rates & Repeat Purchases to measure customer loyalty
- Conversion Rates of Marketing Campaigns targeting each segment
- Customer Support Metrics (e.g., issue resolution times for budget-conscious users)

Regular cluster analysis and heatmap updates should be performed to refine strategies over time.

6. Conclusion and Final Insights

The segmentation analysis provides a data-driven understanding of **customer preferences, behavior, and product choices** in the EV bike market. Through behavioral, psychographic, and product-based segmentation, we identified distinct customer groups with varying expectations and satisfaction levels. The **Premium Enthusiasts** segment emerges as a high-value group with strong brand loyalty and a willingness to pay for superior features, while **Balanced Buyers** represent a key market that can be influenced through value-driven offerings. On the other hand, **Budget-Conscious Users** present challenges in retention but offer opportunities through strategic pricing and improved after-sales service. By leveraging these insights, targeted marketing strategies, product enhancements, and personalized customer engagement plans can be implemented to optimize market positioning and drive long-term customer satisfaction. Continuous monitoring of segment performance and customer feedback will be essential in refining these strategies for sustained growth and competitive advantage.

7. Machine Learning and NLP Models Used in the Market Segmentation Project

For this segmentation analysis, we used a combination of clustering algorithms and NLP techniques:

- Clustering Models:
 - K-Means was applied to group customers based on psychographic and behavioral attributes.
 - Gaussian Mixture Model (GMM) helped capture more complex, probabilistic customer distributions.
 - PCA and t-SNE were used to reduce dimensionality and visualize the clusters.
- NLP Techniques for Sentiment Analysis & Aspect-Based Opinion Mining:

-
- Sentiment Analysis: We extracted sentiment scores from customer reviews using TextBlob. These scores were integrated into the segmentation process to distinguish customer attitudes.
 - Aspect-Based Opinion Mining: Reviews were categorized based on key aspects such as Performance, Reliability, Comfort, Service Experience, and more, using One Shot Classification with Facebook's BART model
 - BERT-based Embeddings: We leveraged SBERT (Sentence-BERT) to generate dense vector embeddings for customer reviews, allowing a more nuanced understanding of customer sentiments.
 - Topic Modeling: BERTopic was used to uncover key themes in customer feedback, helping in product-based segmentation.

By integrating structured data (ratings, purchase patterns) with unstructured text data (reviews, feedback), we created a more refined segmentation model that captured both customer preferences and emotional responses toward EV bikes.

8. Potential Improvements with Additional Time & Budget

With more time and resources, the segmentation project could be improved in several ways:

- Expanded Dataset Collection: Additional data points could improve model accuracy. Key data points to collect:
 - Customer demographics (age, income, location)
 - Riding patterns (average daily distance, preferred terrain)
 - Battery health & longevity data
 - Brand-switching behavior
 - Purchase frequency & spending patterns
- Advanced ML Models:
 - Hierarchical Clustering for better cluster hierarchy visualization
 - DBSCAN for identifying outliers and density-based segments
 - Neural Networks for Representation Learning, particularly Autoencoders, to learn latent factors in customer preferences

9. Estimated Market Size for the EV Bike Market

The global electric two-wheeler market is projected to reach \$100 billion+ by 2030, growing at a CAGR of around 10-12%. The Indian EV two-wheeler market alone is expected to surpass 10 million unit sales by 2030, driven by government incentives, rising fuel costs, and increased environmental awareness. The non-segmented market size provides a broad estimate, but segmentation allows for more granular targeting and market penetration strategies.

10. Top 4 Features for Optimal Market Segmentation

From the analysis, the most influential features for creating optimal market segments are:

1. Performance Rating – Users prioritize speed, acceleration, and motor power.
2. Reliability Score – Brand trust and long-term durability significantly impact retention.
3. Service Experience – Customer satisfaction with maintenance, warranty, and support services.
4. Value for Money – Pricing and perceived worth play a critical role in purchase decisions.

These features strongly differentiate customer groups, making them ideal for refining EV market segmentation strategies.