

# Market Segmentation Analysis of Electric Vehicle market in India

Devam Tanwani

15/08/2024

GitHub Link: <https://github.com/Devam153/Feynn-Labs-Market-Segmentation-Project>



## Defining Objectives

Before diving into the market segmentation analysis, it's crucial to clearly define the objectives of the analysis. This step sets the foundation for the entire process and ensures that the analysis aligns with business goals.

### Primary Objective

Identify distinct customer segments based on their preferences and usage of electric vehicles (EVs).

### Secondary Objectives

1. *Understand Customer Preferences:* Analyze which features (e.g., exterior, comfort, performance, fuel economy) are important to different customer segments.
2. *Identify High-Value Customers:* Find out which segments are more likely to give higher ratings or are willing to pay more for specific features (value for money).
3. *Evaluate Satisfaction Levels:* Identify segments with high or low satisfaction levels (based on overall ratings) to improve customer experience.

By labeling and defining these objectives, it is ensured that the subsequent steps in the market segmentation analysis are focused on and aligned with overall business goals.

## Data Preparation

The data was extracted from Kaggle link mentioned below for EV market segmentation.

<https://www.kaggle.com/deadprstkriish/ev-cars-user-reviews-india?select=4-wheeler-EV-carwale.csv>

The above link contained 3 datasets and the dataset with the name "4-wheeler-EV-carwale dataset" was used.

### Data Preprocessing

Data preprocessing is a crucial phase in data analysis where the precision of results is closely linked to the quality of the dataset. Here is an overview of the steps taken during the preprocessing phase for this project:

```

<class 'pandas.core.frame.DataFrame'>
Index: 120 entries, 0 to 128
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   review                 120 non-null    object
1   Exterior               120 non-null    float64
2   Comfort                120 non-null    float64
3   Performance            120 non-null    float64
4   Fuel Economy           120 non-null    float64
5   Value for Money        120 non-null    float64
6   rating                 120 non-null    float64
dtypes: float64(6), object(1)
memory usage: 7.5+ KB

```

## Initial Data Inspection

The `df.info()` command was used to examine the dataset along with the `df.describe()` to examine the values. This step was essential for identifying the presence of null values and understanding the structure of the dataset, including column types and potential issues.

## Null and Duplicate Values Check

The dataset was found to be well-maintained with no null values or duplicated entries. This indicates that the data was clean and suitable for further analysis.

## Identification and Removal of Problematic Columns

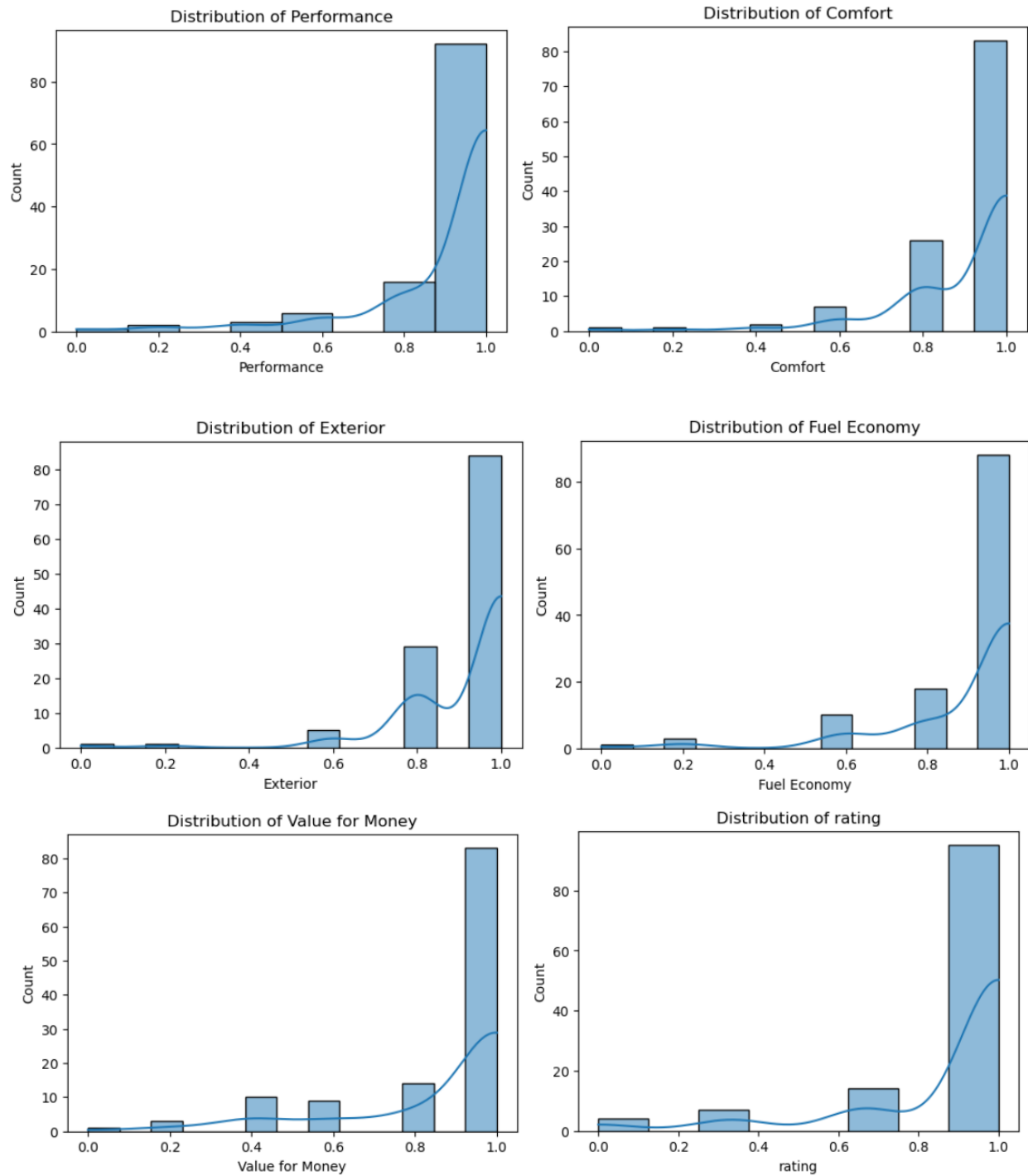
Upon deeper inspection, it was discovered that 8 columns contained a rating of 0, while the remaining columns had valid ratings, and 1 column contained a valid rating and the remaining columns all had a rating of 0. These columns were problematic as they could skew the modelling process. Therefore, these columns were removed from the dataset to improve data quality and ensure more accurate results.

## Numerical Data Extraction and Scaling

A new DataFrame was created to include only the numerical columns from the original dataset. To standardize the range of value, the `MinMaxScaler` from the `sklearn.preprocessing` library was applied.

## Data Visualization

In this step, histograms and boxplots were created that provide initial insights into the distribution of numerical features within the dataset.



### Insights from Histogram Analysis

Histograms allow us to visualize the distribution of each numerical feature and how many times it occurred. By plotting histogram for each of the numerical features in the dataset, we can derive several insights:

### 1. Skewed Distribution:

The histogram of the features shows that most of the data points are clustered towards the higher end of the scale (0.8 - 1.0).

This indicates that most of the vehicles have a good overall rating, which suggests that customers tend to rate these features positively.

*Implication:* This skewness indicates that the EV market generally has a positive perception.

### 2. Clusters of Ratings:

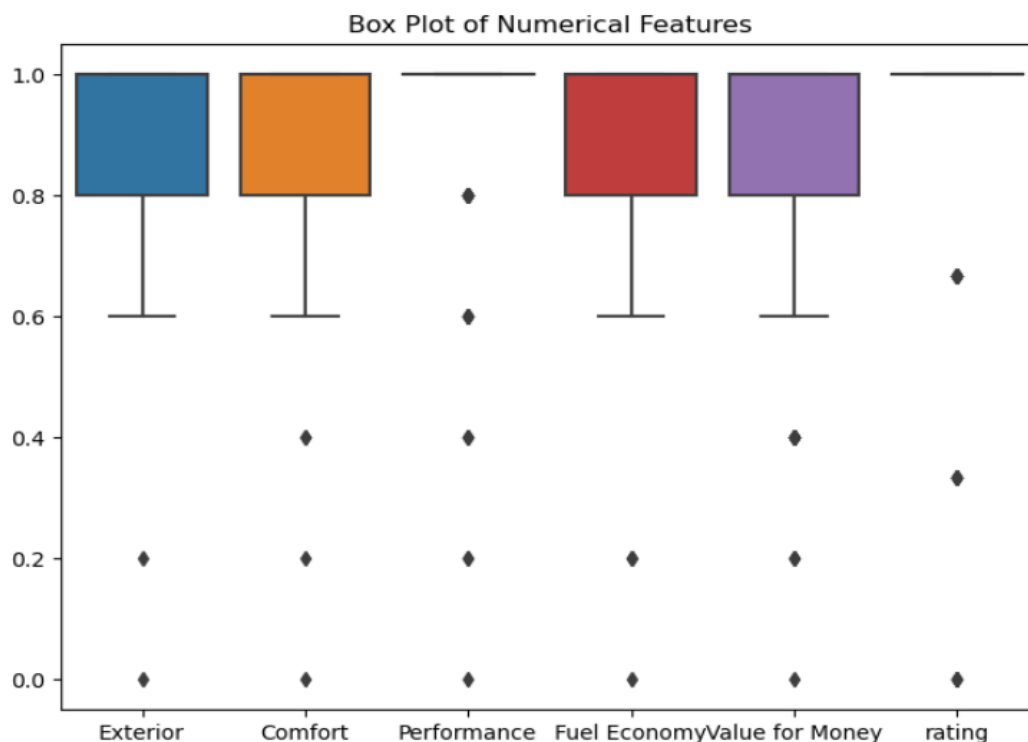
The histogram indicates distinct clusters or modes in the distribution, which can suggest different customer preferences or segments.

*Implication:* These clusters could correspond to different market segments. For instance, some customers may prioritize the exterior, while others might not consider it as critical.

### 3. Tail-End Observations:

The lower end of the histogram has fewer data points, indicating that few features are considered bad.

*Implication:* This can be used in marketing strategies to highlight the general satisfaction with vehicle exteriors.



## Insights from Box Plot Analysis

Box plots provide a summary of the distribution of the data, including the median, quartiles, and potential outliers.

### 1. Central Tendency (Median):

The median for most of the features is quite high. This suggests that most of the ratings given by users for these features are positive.

### 2. Interquartile Range (IQR):

The IQR, represented by the length of the box, is relatively narrow for all the features. This narrow range indicates that the middle 50% of the data points (from the 25th to the 75th percentile) are closely packed, which suggests limited variability in user ratings. This limited variability further supports the observation that user experiences are consistently positive across the various features of the vehicles.

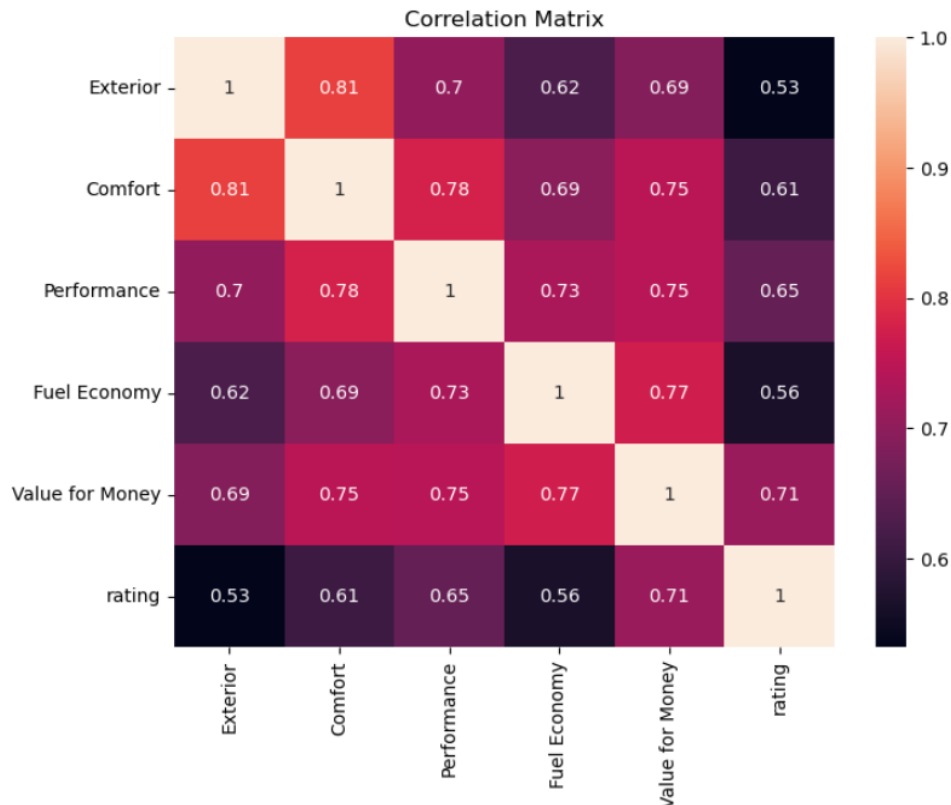
### 3. Outliers:

There are several outliers below the lower whisker (0.25 quantile), especially noticeable in all the features. These outliers represent instances where users rated the vehicles significantly lower than the majority.

The presence of these outliers indicates that while most users are satisfied with the given features, there are a few who have had negative experiences, leading to lower ratings.

These outliers could represent edge cases can tell specific issues with certain vehicle models or user expectations.

## Understanding the Correlation Matrix



A correlation matrix shows the relationship between different variables in a dataset that varies from -1 to 1.

A value of 1 indicates a perfect positive correlation, meaning the two variables move in the same direction.

A value of -1 indicates a perfect negative correlation, meaning the variables move in opposite directions.

A value of 0 indicates no linear relationship between the variables.

Understanding these correlations can help tailor marketing messages to different customer segments.

#### *High Correlations:*

- *Exterior and Comfort (0.81):* The strong correlation between exterior and comfort indicates that customers who like a car's exterior also tend to rate it high on comfort. This suggests that these two features are often seen together as indicators of overall quality in EVs.
- *Value for Money and Fuel Economy (0.77):* This strong correlation suggests that customers who perceive high value for money in an EV also like to rate its fuel economy positively. It indicates that efficiency in energy consumption is closely associated with the perception of value.

#### *Lower Correlations:*

- *Rating and Exterior (0.53):* There is a slightly lower correlation between the overall rating and the exterior of the vehicle, suggesting that while the exterior design might influence ratings, it is not as critical as other factors like performance or value for money.

## **Principal Component Analysis**

Prior to applying K-Means clustering and getting segments, Principal Component Analysis (PCA) was used to reduce the dimensionality of the data.

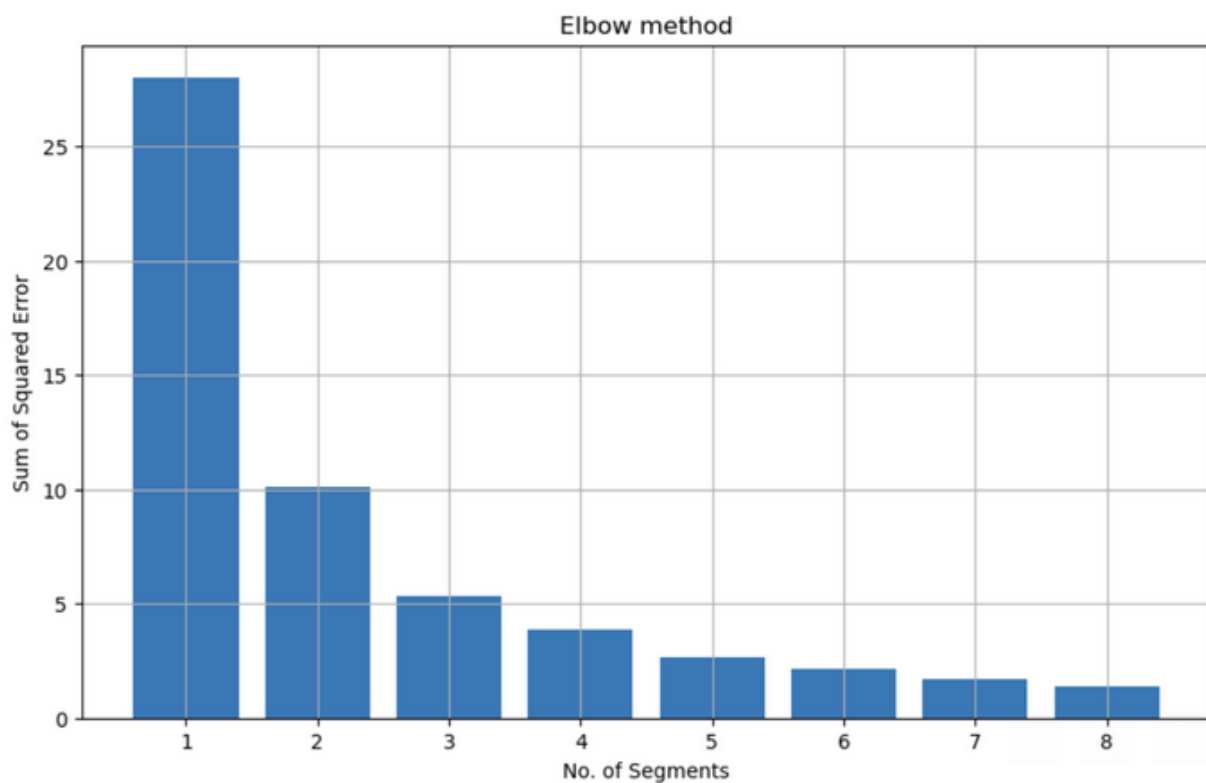
PCA identifies the principal components that capture the most variance in the data. By focusing on these components, we can improve the efficiency and interpretability of the clustering process.

In this specific case, the 2 principal components explain approximately 84% of the total variance. By combining PCA with K-Means, we can identify meaningful customer segments based on the reduced-dimensionality data while preserving the essential information contained in the original dataset.

## Elbow Method

To determine the optimal number of clusters for K-Means clustering, the elbow method is employed. The graph below illustrates the relationship between the number of clusters and the corresponding sum of squared errors (SSE).

- *SSE*: This metric quantifies the sum of squares of the distances of each point from its closest centroid. As the number of clusters increases, the SSE generally decreases.
- *Elbow Point*: The 'elbow' in the graph represents the point where the rate of decrease in SSE starts to lessen significantly. This point is often considered an optimal balance between maximizing explained variance and minimizing the number of clusters.

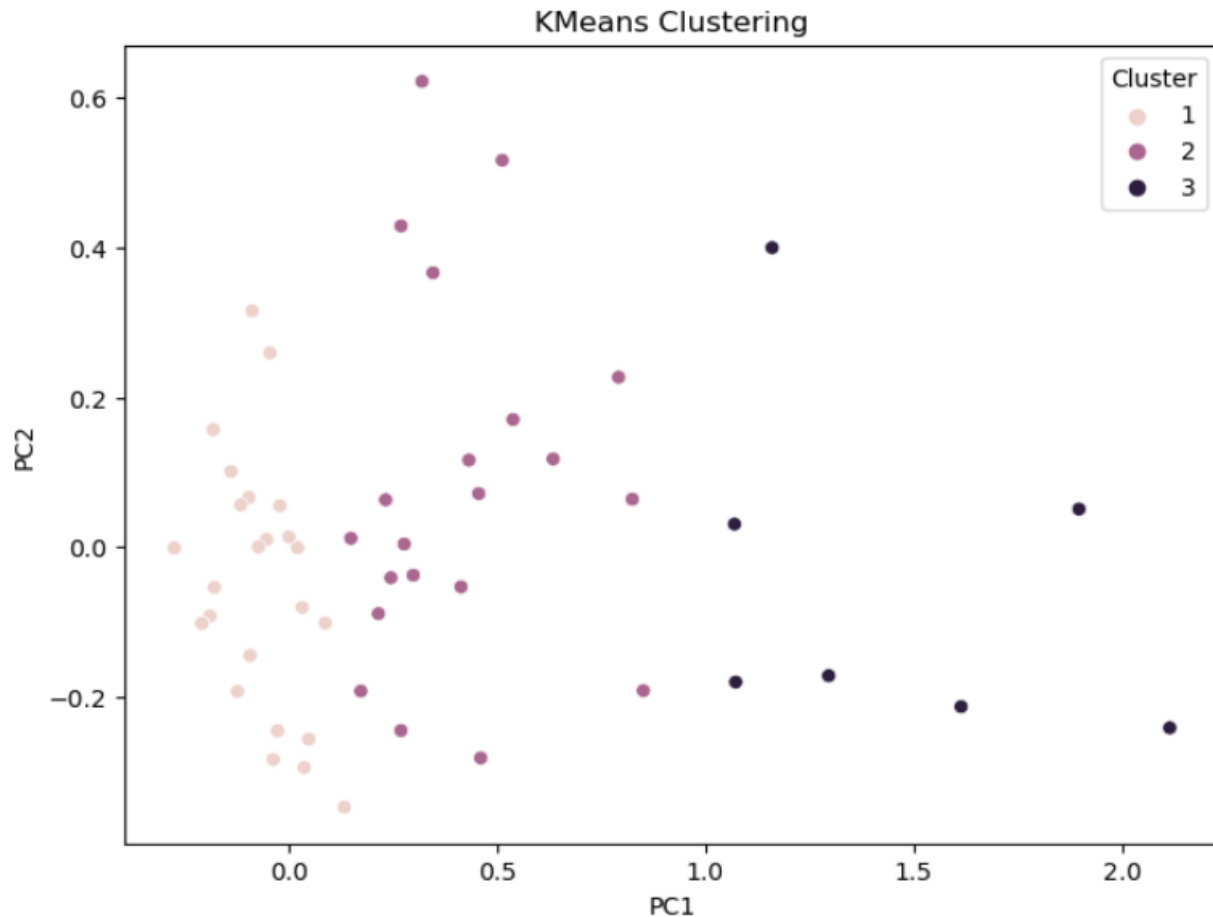


In this case, the elbow appears to be around the 3rd cluster, suggesting that creating 3 segments might be appropriate for this dataset based on the SSE criterion.



## Using Kmeans Clustering

The K-Means algorithm assigned each data point in the reduced-dimensionality space (represented by the first two principal components) to one of the three clusters. These cluster labels were then added back to the original and PCA-transformed dataframes for further analysis.



The scatterplot shown above is generated provides a visual representation of the K-Means clustering results.

The x-axis corresponds to the first principal component (PC1), and the y-axis represents the second principal component (PC2).

By examining the distribution of data points across the clusters, we can start to identify potential customer segments with distinct characteristics based on the original features used in the analysis.

Customers within a particular cluster share similar preferences or usage patterns related to electric vehicles.

## Cluster Analysis

	Exterior	Comfort	Performance	Fuel Economy	Value for Money	rating
Cluster						
1	0.961538	0.958791	0.980769	0.969780	0.978022	0.978022
2	0.809524	0.797619	0.773810	0.761905	0.500000	0.746032
3	0.535714	0.357143	0.321429	0.321429	0.142857	0.142857

### Cluster 1: Enthusiast Customers

*High ratings across all attributes:*

This cluster represents customers who have a high overall satisfaction with their electric vehicles as well as love the concept of their vehicles being electric and not running on fossil fuels.

They value exterior design, comfort, performance, fuel economy, and perceive good value for money.

This cluster includes customers who demand the best in all aspects of an EV.

They are willing to pay a premium for a vehicle that excels in performance, comfort, and other features.

### Cluster 2: Value-Conscious Customers

*Moderately high ratings:*

This cluster represents a group of customers who are generally satisfied with their electric vehicles but have some reservations.

While they appreciate factors like exterior design, comfort, and performance, they are less enthusiastic about fuel economy and especially value for money as electric cars, as of now, tend to be pricier than conventional 4-wheelers.

Thus, this segment might be more price-sensitive and looking for better deals or incentives.

### Cluster 3: Dissatisfied Customers

*Low ratings across all attributes:*

This cluster indicates a group of customers who are generally dissatisfied with their electric vehicles and perhaps its concept.

They have low opinions about various aspects such as exterior design, comfort, performance, and fuel economy.

Understanding the specific reasons for their dissatisfaction could be crucial for improving product offerings and customer satisfaction and targeting this branch of market segments.

## Key Differences and Insights:

- *Cluster 1 vs. Cluster 2:* Both clusters show a preference for exterior design, comfort, and performance, but Cluster 2 is more critical about value for money. This suggests that Cluster 2 might be more interested in affordable options without compromising on key features.
- *Cluster 3:* This cluster clearly represents a group of dissatisfied customers. Further analysis is needed to understand the specific reasons behind their low ratings.

## Further analyzing the clusters

To get further insights from the clusters, cluster labels were added to the original dataset containing the reviews. These reviews can be grouped with the clusters. This tells a more comprehensive analysis of the textual content within each cluster.

Cluster		review
0	1.0	Superb car like as fantastic as petroleum car....
1	2.0	Pretty good car, smooth as a glider fast car, ...
2	3.0	This is very important segment in our Indian c...

By merging the cluster labels with the review data, we can group the reviews based on their respective clusters, summarize them by performing text analysis techniques to extract common themes, sentiments, and other relevant information.

This approach provides a more nuanced understanding of the customer experiences and preferences associated with each cluster.

## Cluster 1

### General Sentiment

The reviews for Cluster 1 reflect a positive sentiment towards electric vehicles, with both commendations and some criticisms highlighted. While the cars are praised for their innovative features and environmental benefits, significant concerns are raised about their practicality, particularly in the context of the Indian market.

### Positive Aspects

- *Performance and Comfort:* Reviewers appreciate the excellent speed, smooth rides, and low maintenance of electric vehicles. Phrases like "excellent speed" and "smooth rides" underscore the high performance and comfort.

- *Design and Technology:* The interior and exterior designs, along with advanced features such as fast charging, are often highlighted. Comments about the "environmentally friendly features" and "advanced technology" reflect a positive reception of these aspects.
- *Environmental Benefits:* The eco-friendly nature of the cars is frequently praised, aligning with the growing awareness of sustainable transportation.

## Concerns

- *Price:* A concern is the relatively high cost of electric vehicles. Many users suggest that reducing the price would make these cars more accessible to people.
- *Charging Infrastructure:* The lack of sufficient charging stations and the long charging times are significant drawbacks. Reviewers express that the current infrastructure is inadequate for long journeys and practical use.
- *Range:* The limited range on a single charge is highlighted as a major issue. Some users feel that cars are not yet suitable as a sole family vehicle due to this limitation.

## Additional Observations

- *Suggestions for Improvement:* Users recommend increasing the number of charging stations, offering more vehicle variants, and potentially integrating features like solar panels to improve the vehicle's utility.
- *Practicality for Long Journeys:* Many reviews emphasize that while electric vehicles are good for daily use, they are not yet practical for extended trips.

## Cluster 2

### General Sentiment

The reviews for Cluster 2 reflect a generally positive sentiment towards the cars. However, there are notable concerns about price and charging infrastructure.

### Positive Aspects

- *Performance and Comfort:* Many reviewers appreciate the smooth driving experience, instant pickup, and sporty feel of the cars. Phrases like "smooth as a glider" and "feels like a sport car" highlight the high performance and comfort.
- *Design:* The cars are described as visually appealing, with comments on their stylish and sporty looks.
- *Driving Experience:* Many reviewers find the driving experience enjoyable, citing smooth and noise-free driving. The ease of handling and decent range for daily use are frequently mentioned.

## Concerns

- *Price:* Several reviewers express concern about the high cost of the cars, suggesting that the pricing could be improved to make them more accessible. Many mention that the car is much more expensive than conventional cars and another highlights the price disparity between different models.
- *Charging and Infrastructure:* Issues with charging time and infrastructure are recurring concerns. Reviewers note that the charging process can be time-consuming and that there is a need for more charging stations. Some mention problems with the car stopping during charging.
- *Battery and Range:* There are mixed feelings about the battery life and range. While some find the range adequate for daily use, others highlight concerns about range anxiety and the car's performance in different weather conditions. Some also suggest that the battery capacity should be larger.

## Additional Observations

- *Service and Maintenance:* A few reviews mention the ease of servicing and maintenance, noting that the cars are low maintenance and easy to handle.
- *Customer Experience:* Positive comments about customer service and quick delivery times enhance overall satisfaction.

## Cluster 3

### General Sentiment

The reviews for Cluster 3 show a predominantly negative sentiment towards the cars, with significant dissatisfaction expressed, particularly regarding price, performance, and value for money.

### Negative Aspects

- *Price:* The most frequently mentioned issue is the high price of the cars. Reviewers feel that the vehicles are too expensive for their size and performance.
- *Performance and Range:* Many reviews highlight dissatisfaction with the range of the electric vehicles. Some find the actual range to be much lower than advertised and the battery consumption in sport mode is noted as problematic.
- *Value for Money:* The low scores in "Value for Money" reflect widespread dissatisfaction with the overall value offered by these cars. Reviewers mention that while the vehicles may be good in some aspects, their cost does not justify the performance and features provided.

## Positive Aspects

- *Comfort and Design:* Despite the overall negative sentiment, there are some positive comments about the comfort and design of the vehicles. For example, Tata Nexon EV is praised for its ride comfort and premium interiors.
- *Brand Loyalty:* Some reviews show strong brand loyalty, with positive remarks about the brand's customer support and the overall driving experience. Hyundai and Tata are appreciated for their contribution to the electric vehicle segment, even if the specific models are criticized.

## Additional Observations

- *Future Prospects:* There is a sense of optimism about future improvements in this segment. Some reviewers express hope that future price reductions and improvements in infrastructure will make these cars more appealing.
- *Infrastructure Needs:* The importance of having adequate charging points is emphasized. Several reviewers note that without proper charging infrastructure, the practicality of owning an electric vehicle is limited.

\*\*\*\*\*