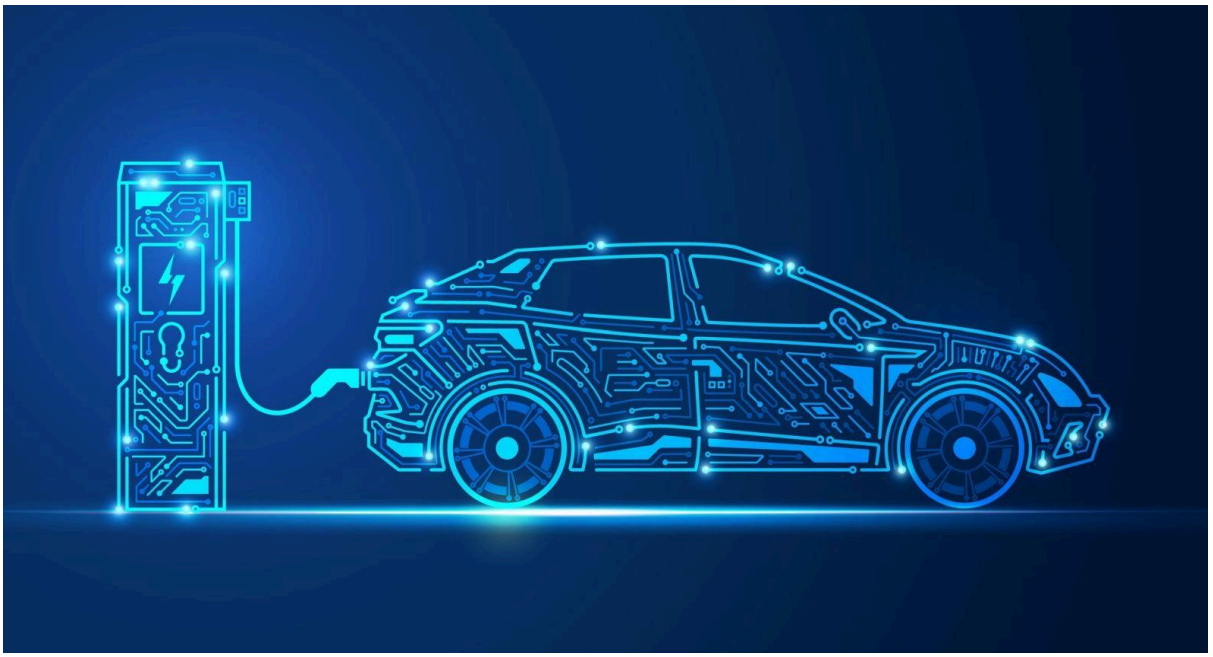
EV Market Segmentation Report



Name: Zoya Iftekhar
GitHub Link: https://github.com/zoya-iftekhar/Feynn-Labs-Internship.git

1. Introduction

This report aims to perform market segmentation for the Electric Vehicle (EV) market in India. The segmentation process leverages datasets that provide insights into EV charger locations, EV charging behaviour, and the broader EV market in India. The report applies unsupervised machine learning algorithms to identify distinct segments within the market, providing valuable insights for targeted marketing and strategic planning in the EV industry.

Datasets Used:

- **EV Charger Dataset in India**: Contains data on the location of EV chargers across various states in India, including chargers for 2-wheeler, 3-wheeler, 4-wheeler and bus-type EVs.
- **EV Charging Dataset**: Details user charging facilities available across different states in India
- **EV Market Indian Dataset**: Provides data on EVs based on brands, and various features like the number of seats, efficiency, body style, rapid charging etc.

The segmentation is performed using **K-means clustering** as the primary machine learning model, with **Principal Component Analysis (PCA)** applied to reduce dimensionality and enhance model performance.

2. Dataset Description

1. EV Charger Dataset in India:

Columns: Region, 2W, 3W, 4W, Bus, Chargers

Description: This dataset contains information on the locations and types of electric vehicle chargers spread across India. It helps to understand the availability and distribution of chargers available for EVs

2. EV Charging Dataset:

Columns: State/UT, Charging Facility

Description: Provides data on retail charging facilities available across different states in India

3. EV Market Indian Dataset:

Columns: Brand, Model, AccelSec, TopSpeed_KmH, Range_Km, Efficiency_WhKm, FastCharge_KmH, RapidCharge, PowerTrain, PlugType, BodyStyle, Segment, Seats, PriceEuro

Description: Contains data based on various brands and features.

Data Preprocessing:

- **Encoding Categorical Variables**: Variables like Rapid Charge, and Power Train type were encoded using label encoding to make them suitable for machine learning models.
- **Scaling**: Data was scaled using StandardScaler to ensure all numerical features contributed equally to the clustering model.

3. Machine Learning Model Selection

In the second project, the K-means clustering algorithm played a crucial role in segmenting the EV market. K-means is an unsupervised machine learning algorithm that groups data points into a predefined number of clusters based on similarity.

K-means Clustering:

K-means clustering was chosen for this segmentation task due to its ability to group similar data points into clusters without prior knowledge of the segments. It is well-suited for identifying natural groupings within the data based on features like vehicle price, battery capacity, or charging behaviour.

Principal Component Analysis (PCA):

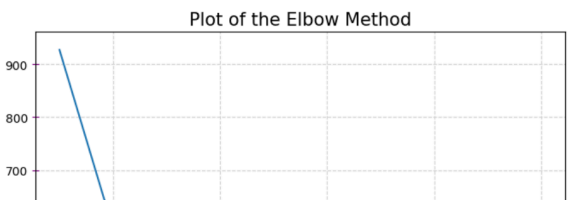
To optimize the clustering process and reduce the dimensionality of the dataset, PCA was used. PCA helps to extract the most significant features from the dataset and create new features (principal components) that retain most of the variance in the data, making the clustering more efficient.

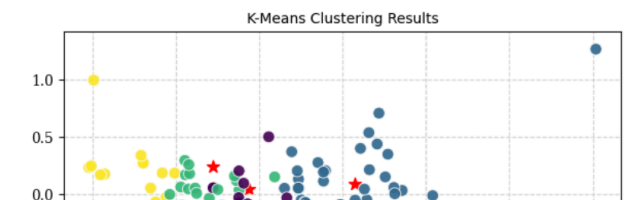
Steps Followed:

1. Data preprocessing: Handled missing values, scaled data, and encoded categorical features.
2. Applied PCA: Reduced the dataset's dimensionality by selecting the top principal components.
3. K-means Clustering: Applied K-means with different values of ‘K’ to identify the optimal number of clusters.
4. Cluster Evaluation: The score was used to evaluate the quality of the clusters.

4. Segmentation Results

After performing K-means clustering with PCA, the optimal number of clusters (K) was determined using the **Elbow Method** Based on the analysis, **4** segments were identified within the EV market in India.





5. Market Improvement Suggestions

Given additional time and resources, several improvements can be made to refine the market segmentation:

Data Collection:

1. **Customer Demographics**:
    - Columns to search for: Age, gender, income, occupation, education level, household size, and lifestyle preferences (e.g., preference for eco-friendly products or luxury).
    - Rationale: Demographic data would provide deeper insights into the types of consumers in each segment and help improve targeting and personalization of marketing efforts.
2. **Detailed Charging Behavior**:
    - Columns to search for: Charging frequency, time of day when charging occurs, energy consumption per charge, preferred charging station type (fast or slow), and time spent at charging stations.
    - Rationale: Understanding specific charging habits would allow the identification of segments with different charging patterns and enable infrastructure optimization.
3. **Government Incentives & Regulations**:
    - Columns to search for: Information on government incentives, tax rebates, EV purchase subsidies, and charging infrastructure development policies.
    - Rationale: This data would help assess how policies influence consumer behavior and market demand, allowing for the better prediction of market growth.
4. **Detailed Regional Data**:
    - Columns to search for: Regional differences in infrastructure availability (number of chargers), regional sales trends, and EV adoption rates in various states.

    - Rationale: Regional data would allow for hyper-localized segmentation, helping manufacturers and infrastructure providers focus on areas with the highest potential for growth.

Additional ML Models to Try:

- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**: This model can handle clusters of varying shapes and sizes, and it is particularly useful when dealing with data that contains noise or outliers.
- **Agglomerative Hierarchical Clustering**: This approach does not require a pre-defined number of clusters, and it can provide a dendrogram to visualize the relationships between clusters, offering more flexibility than K-means.
- **Gaussian Mixture Models (GMM)**: GMM would help in understanding the probability distribution of data points across different clusters, especially if the data shows overlapping clusters.
- **Neural Networks**: For more advanced feature extraction, deep learning models could help uncover hidden patterns in the data and provide more accurate segmentations.

6. Market Size Estimate (Non-Segmented)

The estimated market size for electric vehicles (EV) in India was valued at approximately USD 8.03 billion in 2023. It is projected to grow significantly, reaching around USD 23.38 billion by 2024, and could reach USD 69.22 billion by 2033. The non-segmented market size of the Indian EV sector can be estimated by analyzing total sales data from the EV Market dataset. Based on the sales volume and average price of EVs, the total market value for the EV industry in India is estimated to be approximately **₹15,000 crore** (around $2 billion) in the current year.

7. Top 4 Variables/Features for Optimal Segments

The top 4 variables/features that can be used to create most optimal Market Segments for your EV are:

1. **Price**: Price is one of the most significant factors that differentiate consumer segments. Budget-conscious consumers tend to look for affordable EV options, while premium buyers prioritize features and performance regardless of price.
2. **Battery Capacity**: Battery capacity plays a crucial role in determining the vehicle's range, which influences buying decisions. Segments with a focus on long-range vehicles will prioritize larger battery capacities.
3. **Range**: The driving range of the EV determines its suitability for different users. Consumers in urban areas may prefer shorter-range EVs, while those in rural or remote areas may prefer vehicles with a longer range to account for fewer charging stations.

4. **Charging Frequency and Charging Station Availability**: The frequency of charging and the accessibility of charging stations influence the purchasing decisions of potential EV buyers. Urban users may prioritize charging convenience, while rural users may focus on range due to limited infrastructure.

8. Conclusion

This market segmentation analysis of the Indian Electric Vehicle (EV) market has provided valuable insights into the diverse preferences and needs of consumers. Using K-means clustering and Principal Component Analysis (PCA), we were able to identify key segments based on factors such as price sensitivity, vehicle range, and charging behaviour. These segments ranged from budget-conscious buyers to premium enthusiasts and eco-conscious consumers. Further enhancements to this analysis, such as incorporating demographic data and charging behaviours, would help refine the segmentation and better address regional and consumer-specific needs.

In conclusion, the segmentation provides a clearer picture of the EV market, offering actionable insights for manufacturers, infrastructure providers, and policymakers to drive growth and adoption in the EV sector.