Charutosh Usakoyal

EV MARKET SEGMENTATION REPORT

Problem Statement

We have conducted market segmentation on electric vehicles to analyze the distribution of various cars based on features such as price range, seating capacity, and brands. EV market segmentation seeks to identify distinct customer groups within the electric vehicle market by understanding their preferences and behaviors. This enables businesses to customize their marketing strategies, product offerings, and services to cater to the specific needs of each segment. Ultimately, this approach enhances customer satisfaction and drives sales growth.

Dataset

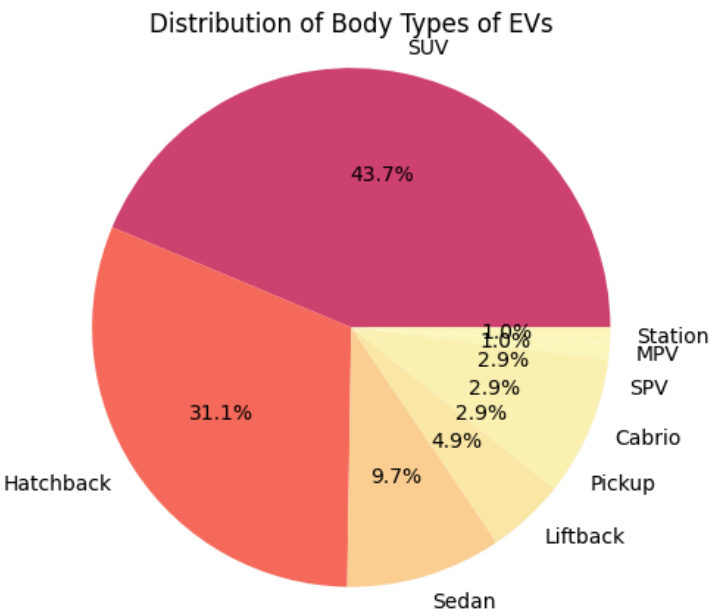
The dataset titled "Electric_Vehicle_Data.csv" contains 14 columns: ‘Brand’, ‘Model’, ‘Accel’, ‘TopSpeed’, ‘Range’, ‘Efficiency’, ‘FastCharge’, ‘RapidCharge’, ‘PowerTrain’, ‘PlugType’, ‘BodyStyle’, ‘Segment’, ‘Seats’, and ‘PriceEuro’. The columns ‘Brand’, ‘Model’, ‘RapidCharge’, ‘PowerTrain’, and ‘PlugType’ are categorical, while the remaining columns are continuous.

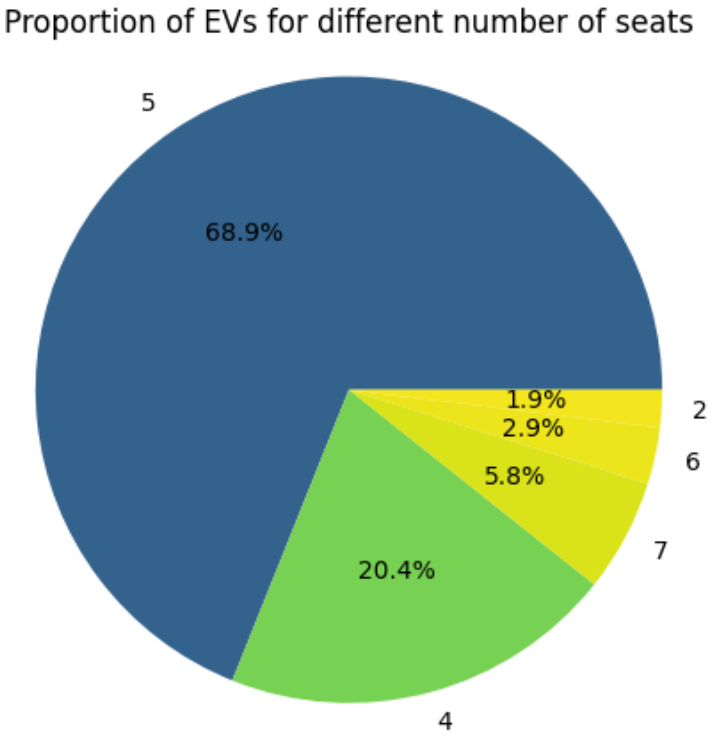
Data Preprocessing

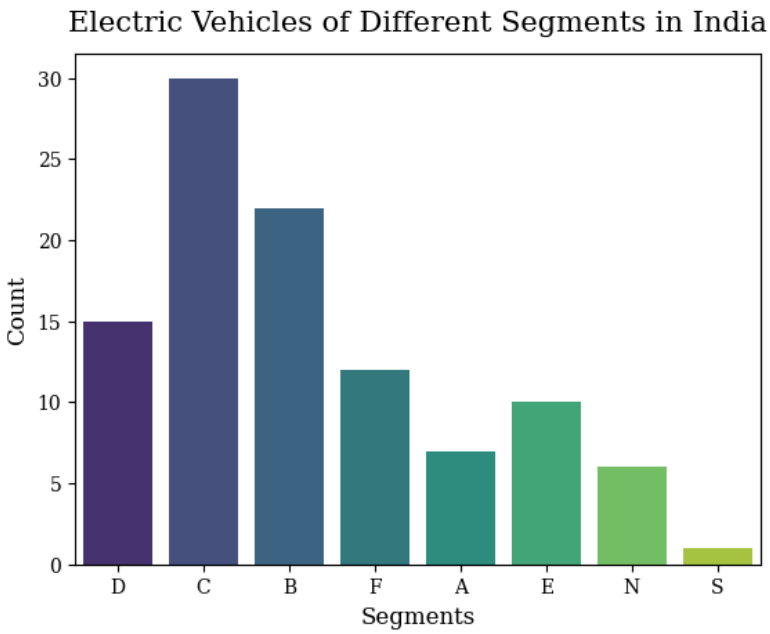
The dataset comprises 103 rows and 14 columns. We converted columns such as 'TopSpeed,' 'Efficiency,' 'Range,' 'Accel,' and 'FastCharge' to the float data type. Additionally, we performed encoding (i.e., converting categorical columns to continuous columns) on the 'PowerTrain' and 'RapidCharge' columns.

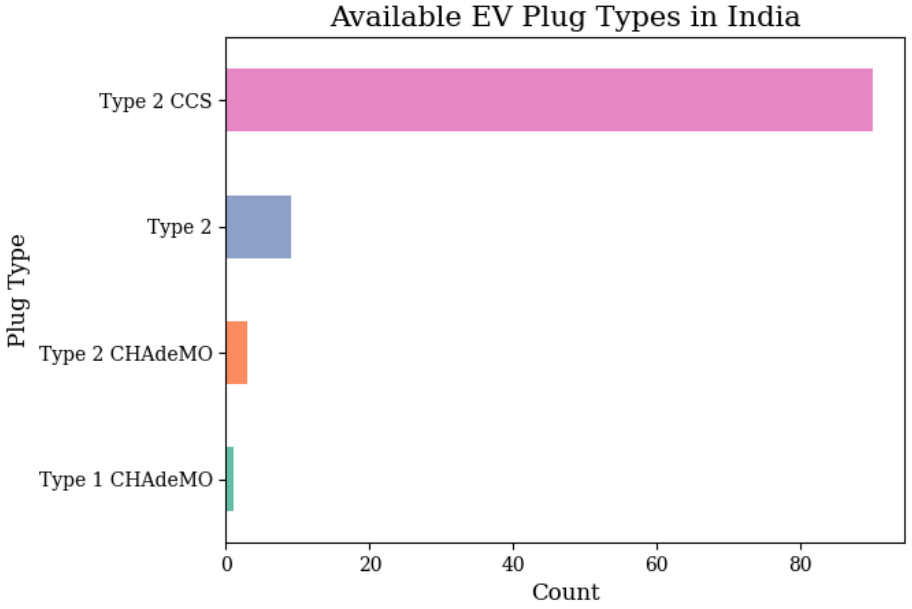
Exploratory Data Analysis

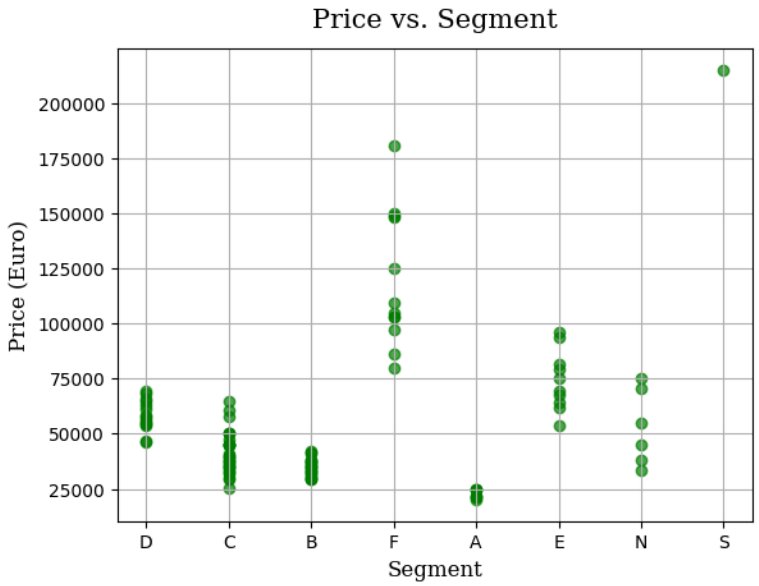
In this analysis, we examined various aspects of the electric vehicle (EV) market, such as the 'Number of EV Models Manufactured by a Brand,' which revealed that Tesla produces the most EV models. When analyzing the distribution of EV body styles, we found that SUVs account for the highest percentage (43.1%), followed by hatchbacks at 31.1%. In India, the majority of electric vehicles are 5-seaters, representing 68.9%, while 2-seater electric vehicles are the least common at 1.9%. The EV market is divided into eight segments: ‘A’, ‘B’, ‘C’, ‘D’, ‘E’, ‘F’, ‘N’, and ‘S’. The C segment has the highest number of EVs with 29, whereas the S segment has the fewest with only 2. The 'Type 2 CCS' plug type is the most widely available for EV charging, while 'Type 1 CHAdeMO' has the fewest plugs available Additionally, we created a box plot for the price distribution of EVs, finding the average price to be around 20k-50k Euros. Segment S has the highest price range, while Segment A has the lowest price range in the EV market.


Distribution of Body Types of EVs
SUV
43.7%
Station
MPV
SPV
Cabrio
Pickup
Liftback
Sedan
Hatchback
31.1%
9.7%
4.9%
2.9%
2.9%
2.9%
1.0%
1.0%


Proportion of EVs for different number of seats
5
68.9%
2
6
7
4
1.9%
2.9%
5.8%
20.4%


Electric Vehicles of Different Segments in India
Count
Segments
0
5
10
15
20
25
30
D
C
B
F
A
E
N
S


Available EV Plug Types in India
Plug Type
Count
Type 2 CCS
Type 2
Type 2 CHAdeMO
Type 1 CHAdeMO
0
20
40
60
80



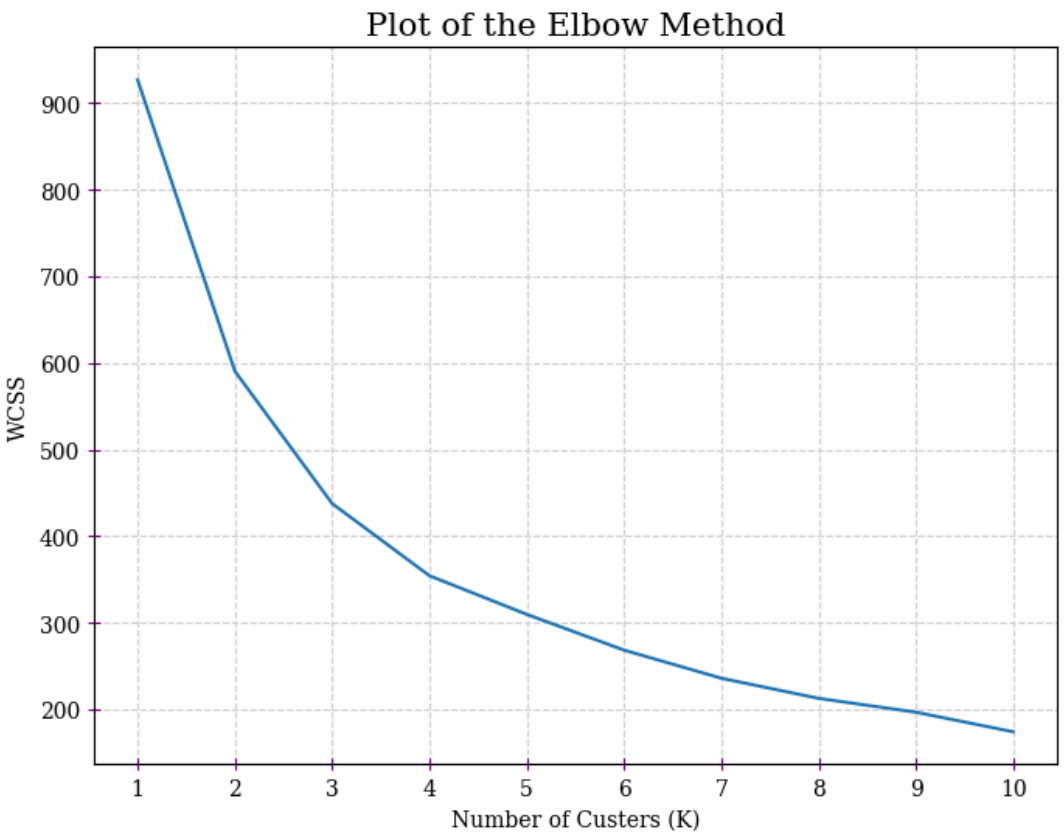
Market Segmentation

The K-Means machine learning model proved instrumental in conducting market segmentation of EV vehicles using a comprehensive dataset featuring 14 key columns: ‘Brand’, ‘Model’, ‘Accel’, ‘TopSpeed’, ‘Range’, ‘Efficiency’, ‘FastCharge’, ‘RapidCharge’, ‘PowerTrain’, ‘PlugType’, ‘BodyStyle’, ‘Segment’, ‘Seats’, and ‘PriceEuro’. Initially, K-Means enabled us to cluster EV models based on their quantitative attributes such as acceleration, top speed, range, efficiency, and charging capabilities. By grouping similar vehicles into clusters, K-Means facilitated the identification of distinct market segments within the EV industry, each characterized by unique performance and functional attributes. Moreover, K-Means clustering allowed for segmentation based on categorical features such

as brand, powertrain type, plug type, body style, segment classification, seating capacity, and pricing. This segmentation strategy provided a nuanced understanding of consumer preferences and market trends across different segments of the electric vehicle market. For instance, it helped highlight which brands or segments dominate specific performance metrics or price ranges, aiding in strategic decision-making for manufacturers and marketers alike. Furthermore, K-Means clustering provided insights into market dynamics by revealing correlations and patterns among various attributes, thereby supporting targeted marketing strategies and product positioning. By segmenting EVs into meaningful clusters, businesses can tailor their product offerings and marketing campaigns to better meet the diverse needs and preferences of distinct customer groups. Ultimately, the application of K-Means in this analysis enhanced our ability to uncover actionable insights and optimize business strategies in the rapidly evolving electric vehicle industry.

We created 9 principal components using PCA (Principal Component Analysis) before performing K-means clustering which is beneficial because PCA reduces the dimensionality of the dataset by transforming correlated variables into a set of linearly uncorrelated components. This transformation helps in capturing the maximum variance in the data while minimizing information loss. By using PCA, we can focus on the most significant features that explain the variability within the dataset, making the clustering process more robust and accurate. Additionally, PCA preprocesses the data by removing noise and irrelevant features, enhancing the clustering algorithm's effectiveness in identifying meaningful clusters based on the reduced set of principal components.

The elbow method serves as a crucial tool in determining the optimal number of clusters for K-means clustering through a systematic evaluation of within-cluster sum of squares (WSS). By plotting the WSS against varying numbers of clusters, typically from a predefined range, the method visually identifies a point where the rate of WSS reduction begins to diminish significantly, resembling an elbow in the graph. This point signifies the ideal number of clusters, where adding more clusters yields diminishing returns in terms of explaining variance within the data. This approach helps in striking a balance between model complexity and explanatory power, ensuring that the selected number of clusters effectively partitions the data into meaningful and distinct groups. Consequently, the elbow method provides a quantitative criterion for determining the most suitable clustering solution, guiding data analysts and researchers in optimizing the performance and interpretability of their K-means models.



The final insights gained from the segmentation are that we created 4 clusters named Cluster 0, Cluster 1, Cluster 2, and Cluster 3. Each and every cluster contains properties like Number of EVs, Brands, Models, PowerTrain Types, Plug Types, Body Styles, Seats Range, Price Range (Euro). The maximum number of EVs are in the Cluster 0 i.e., 55 and it has a price range of around 20129-68040 Euros. The lowest number of EVs are in the Cluster 2 and it has a price range of around 21387-38000 Euros. Below following is the cluster summary for different types of EVs and their properties :-

clusters	Accel	TopSpeed	Range	Efficiency	Seats
0	8.603636	156.163636	292.909091	174.563636	4.818182
1	5.573333	196.266667	392.666667	219.066667	5.433333
2	14.240000	131.000000	116.000000	175.400000	3.400000
3	3.861538	255.769231	494.230769	187.230769	4.461538

clusters	PriceEuro
0	37540.800000
1	65793.533333
2	26154.400000
3	121482.230769

Given additional time and budget for the EV Market Segmentation Project, I would strategically enhance the dataset collection by focusing on acquiring comprehensive demographic data including income levels, urbanization metrics, and environmental attitudes. These variables are crucial for refining segmentation accuracy and identifying target customer segments effectively. Moreover, integrating behavioral data such as charging patterns, travel preferences, and purchase intent would provide deeper insights into consumer behavior and decision-making processes. In terms of model selection, leveraging advanced machine learning algorithms such as gradient boosting machines (GBM) and deep neural networks (DNN) could significantly improve predictive performance and segmentation precision. GBM algorithms, like XGBoost or LightGBM, excel in handling complex interactions between variables and can capture nonlinear relationships inherent in consumer preferences for electric vehicles (EVs). Additionally, exploring unsupervised learning techniques such as clustering algorithms (e.g., k-means or Gaussian mixture models) would enable the discovery of natural groupings within the data, facilitating more nuanced segmentation strategies.

Furthermore, investing in data enrichment through external sources like third-party behavioral data providers or partnerships with transportation agencies could enrich the dataset with real-time mobility patterns and infrastructure utilization, enhancing the project's predictive power and strategic insights. Implementing a robust data governance framework and ensuring compliance with data privacy regulations will be pivotal in maintaining data integrity and fostering stakeholder trust. Overall, these enhancements aim to elevate the EV Market Segmentation Project by fostering a data-driven approach to understanding consumer behavior and optimizing marketing strategies in the burgeoning electric vehicle market landscape.

Estimating the market size for the Electric Vehicles Market Domain dataset, which comprises 30 rows and 10 columns, involves several key considerations. Market size estimation typically begins by assessing the scope and representativeness of the dataset relative to the entire market landscape. With 30 rows and 10 columns, the dataset's dimensions reflect a limited sample of the broader electric vehicle market, suggesting it captures a small fraction of the total diversity and variability within the industry. To extrapolate market size, analysts may employ statistical techniques such as sampling methodologies, regression analysis, or data modeling to infer trends and characteristics observed in the dataset onto the larger market context. Additionally, external data sources and industry reports provide supplementary insights into market dynamics, consumer behavior, and competitive landscape, aiding in refining estimations. However, the accuracy of market size estimates hinges on the dataset's representativeness and the robustness of analytical methods used to generalize findings to the entire market domain. Ultimately, while the dataset provides foundational insights, comprehensive market size determination requires integrating diverse data sources and rigorous analytical approaches to ensure reliability and relevance in decision-making processes within the electric vehicles sector.

In the domain of Electric Vehicles (EVs), identifying optimal variables for market segmentation involves selecting features that profoundly influence consumer preferences and market dynamics. Four pivotal variables include:

1. **Powertrain Technology**: The type of powertrain, whether battery electric (BEV), plug-in hybrid electric (PHEV), or hybrid electric (HEV), significantly shapes vehicle performance, efficiency, and environmental impact. Consumer segments prioritize different powertrain technologies based on factors such as driving habits, environmental concerns, and government incentives.
2. **Range Capability**: The range an EV can cover on a single charge is crucial, impacting consumer confidence in vehicle usability for daily commuting, long-distance travel, and regional variations in charging infrastructure availability. Market segments often differentiate based on range preferences, with urban commuters favoring shorter ranges and long-distance travelers prioritizing extended ranges.
3. **Price Point**: EV pricing remains a pivotal factor influencing market segmentation, with varying affordability thresholds across consumer demographics. Factors influencing price include battery capacity, technology integration, government incentives, and production scale economies, shaping consumer accessibility and market penetration strategies.
4. **Charging Infrastructure**: The availability, speed, and compatibility of charging infrastructure heavily influence EV adoption patterns and market segmentation. Consumer segments differentiate based on access to public charging networks, home charging capabilities, and the convenience of fast-charging options, influencing purchasing decisions and lifestyle compatibility.

These variables collectively inform market segmentation strategies tailored to distinct consumer needs, preferences, and behavioral patterns within the EV market. Analyzing these features enables stakeholders to refine product offerings, pricing strategies, and marketing campaigns effectively, fostering greater consumer satisfaction and market competitiveness. By leveraging data-driven insights from these critical variables, businesses can navigate the complexities of the EV market landscape, optimize resource allocation, and capitalize on emerging opportunities in sustainable mobility solutions.