ELECTRIC VEHILCES

Market Segmentation Analysis

Sidharth R

Problem Statement

The objective is to analyze and compare these vehicles based on their technical specifications, performance metrics, and pricing.

Fermi Estimation

Population of India: Approx. **1.4 billion people**.

Car Ownership Rate: In developing countries, car ownership is lower. Assume 1 car for every **10 people**.

- Total cars in India: 1,400,000,00010=140,000,000 cars.\frac{1,400,000,000}{10} = 140,000,000 \, \text{cars}.101,400,000,000=140,000,000cars.

EV Range: From the dataset, the average EV range is **350 km per charge**.

Annual Driving Distance: Assume each car drives about **12,000 km per year** in India (lower than Western countries due to congestion and shorter trips).

Energy Consumption Per EV:
From the dataset, average efficiency is about **180 Wh/km**.

- Annual energy consumption per EV: 2.16MWh.

Total Energy Demand for All EVs:
Multiply energy per EV by the total number of cars

=302.4TWh annually.

Charging Infrastructure: Assume 1 charging station is shared by **5 EVs**.

- Total charging stations needed: 140,000,0005=28,000,000 stations.\frac{140,000,000}{5} = 28,000,000 \, \text{stations}.5140,000,000=28,000,000stations.

EV Lifetime: Assume each EV lasts for **10 years**.

- 14,000,000EVs/year.

Data Collection

The data used here was obtained from Kaggle. The data description are

Unnamed: 0:A serial index for each row in the dataset.

Brand:The name of the company or manufacturer of the EV (e.g., Tesla, Volkswagen, BMW, etc.).

Model:The specific model name of the EV (e.g., Model 3 Long Range Dual Motor, ID.3 Pure, etc.).

Type: Categorical (String).

AccelSec:The time taken by the vehicle to accelerate from 0 to 100 km/h, measured in seconds.

TopSpeed_KmH:The maximum speed the EV can achieve, measured in kilometers per hour (Km/h).

Range_Km:The maximum distance the EV can travel on a single full charge, measured in kilometers (Km).

Efficiency_WhKm:Energy efficiency of the EV, measured as the energy consumed per kilometer traveled in Watt-hours (Wh/km).

FastCharge_KmH:The distance (in Km) an EV can gain per hour of fast charging.

RapidCharge:A binary indicator showing if the EV supports rapid charging.

PowerTrain:The type of drivetrain system used in the EV.

PlugType:The type of charging plug supported by the EV (e.g., Type 2 CCS).

BodyStyle:The body style or design of the vehicle (e.g., Sedan, SUV, Hatchback, etc.).

Segment:A classification of the vehicle into categories such as D, C, or B based on size and type.

Seats:The seating capacity of the EV, i.e., the number of seats available.

PriceEuro:The price of the EV in Euros (€).

Datasetlink:https://github.com/Chaganti-Reddy/EVMarket-India/blob/main/data.csv

Github: https://github.com/sidharthram99/sidharth.git

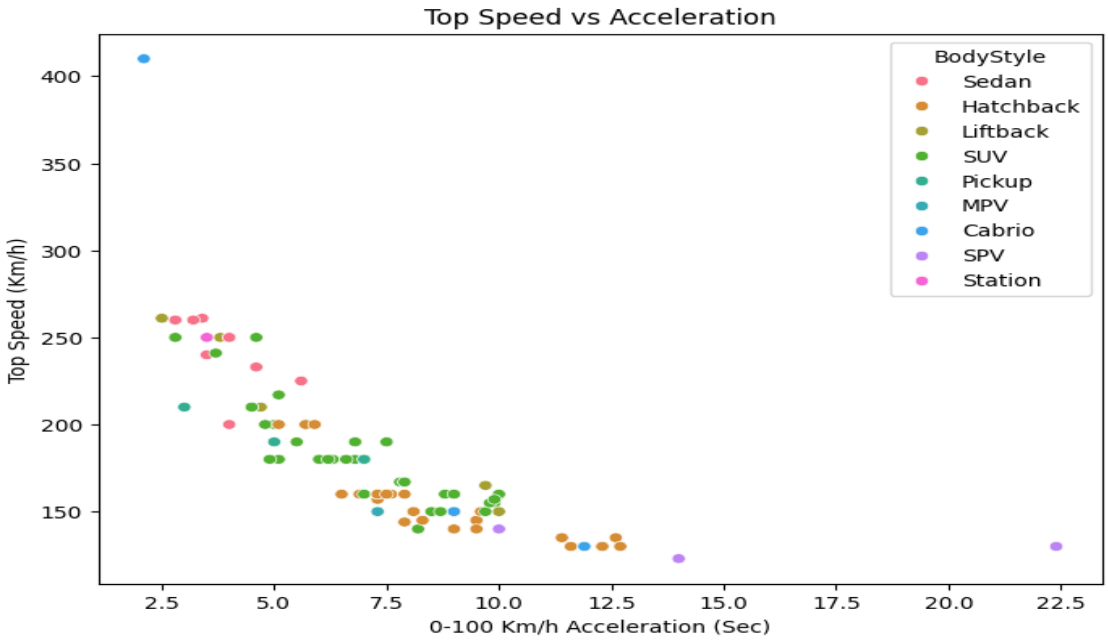
Data Pre-processing

- Yes/ No was re-encoded as 0 and 1
- String values of segment and body style was re-encoded to numeric values
- Price in Euro was converted to INR.
- Columns like unnmaed:0 was dropped.

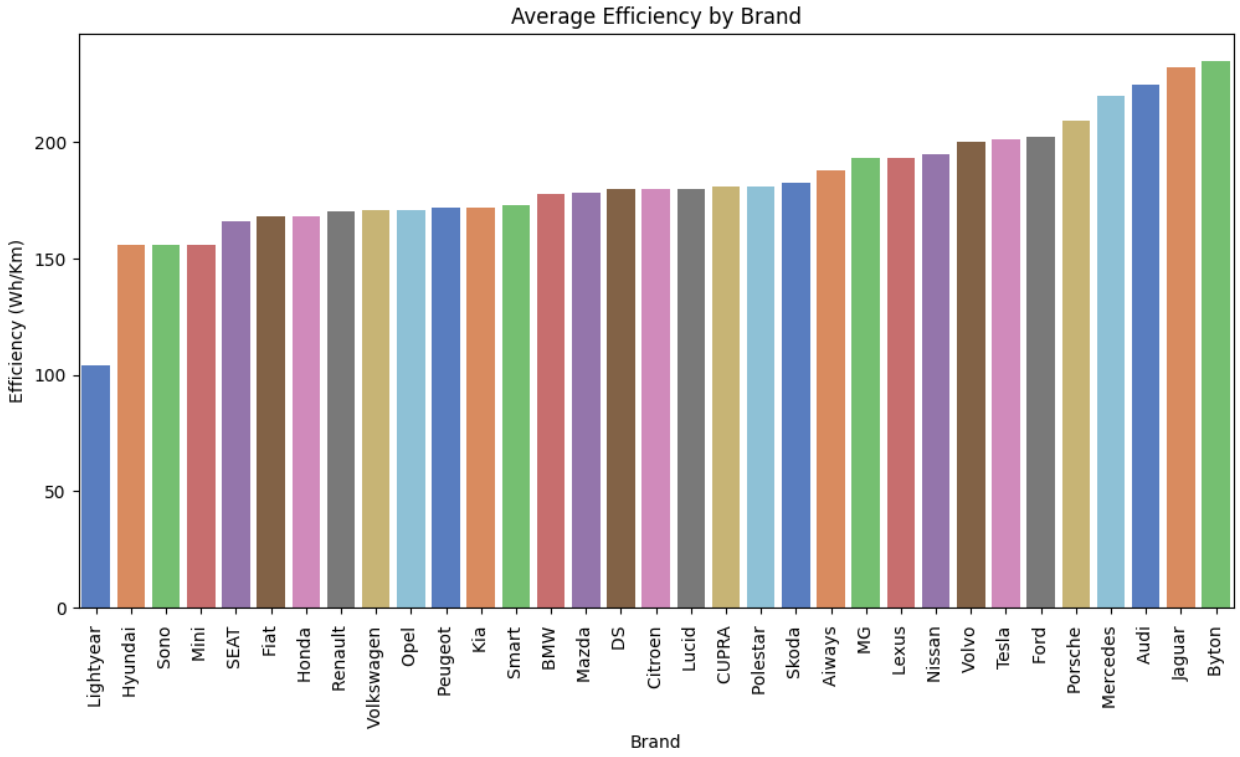
Exploratory Data Analysis

EDA was primarily conducted through profile reports from pandas. This gives an interactive dashboard for analyzing the data. The highlights of EDA are detailed below.

The above graph is a plot between price and range of EV vehicle. From this plot we come to know higher range EV's are more expensive.



Comparing the top speed and acceleration sedans and high performance EV's have better acceleration and top speed.



Brands like Tesla and Polestar show better efficiency (lower Wh/Km). SUVs tend to have a slightly lower range compared to sedans due to weight. Models supporting rapid charging generally have higher prices.

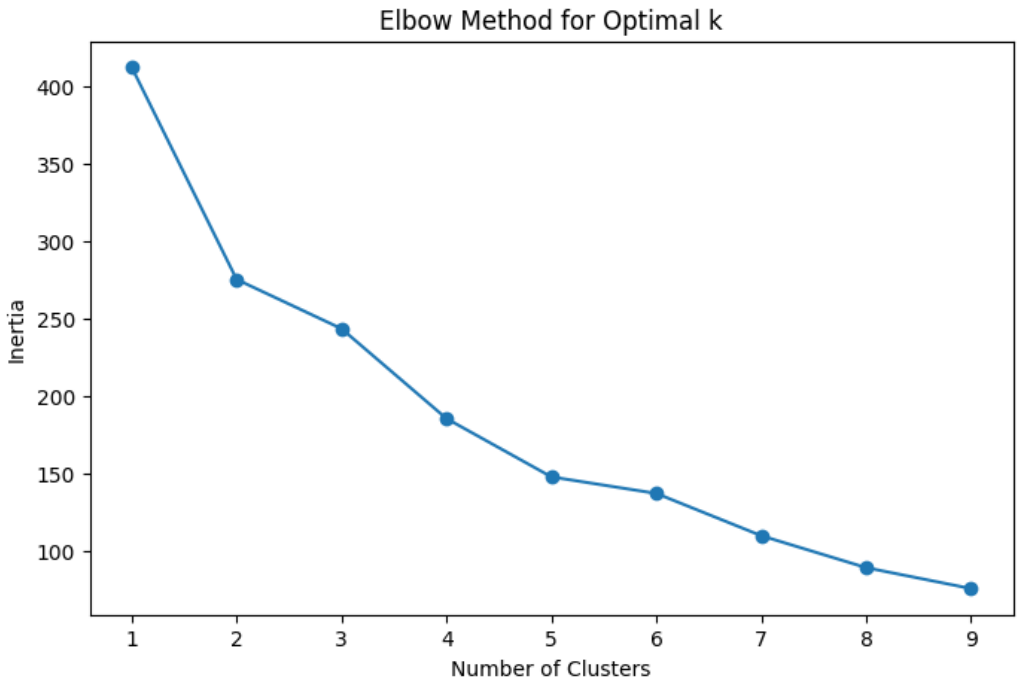
For the segmentation the EDA narrowed down to ideal segmentation variable such as Price, Range, Acceleration, Seats and Power Train.

Segment Extraction

Segmentation was conducted using K-means clustering , an unsupervised machine learning algorithm used for partitioning a dataset into a predefined number of groups, or clusters, based on similarity. The algorithm works by minimizing the distance between data points and the center of their assigned cluster, known as the centroid. It begins by randomly initializing KKK centroids, where KKK represents the desired number of clusters. Each data point is then assigned to the nearest centroid based on a distance metric, typically the Euclidean distance. After assignment, the centroids are recalculated as the mean of the points in their respective clusters. This process of assignment and centroid updating is repeated iteratively until the centroids stabilize or a stopping criterion is met. K-means is widely used for applications such as customer segmentation, image compression, and anomaly detection due to its simplicity and scalability. However, it has limitations, including sensitivity to the choice of KKK, initial centroid positions, and challenges with non-spherical or overlapping clusters. Despite these drawbacks, K-means remains a fundamental and popular clustering technique in data analysis and machine learning.

The **Elbow Method** is a technique used to determine the optimal number of clusters (KKK) in K-means clustering. It works by evaluating the **Within-Cluster Sum of Squares (WCSS)**, which measures the total variance within clusters, for different values of KKK. As KKK increases, WCSS decreases because adding

more clusters reduces the distance between data points and their nearest centroids. However, there is a point where the rate of decrease slows down significantly, forming an "elbow" shape when WCSS is plotted against KKK. This "elbow point" indicates the optimal number of clusters, as adding more clusters beyond this point provides minimal improvement while increasing complexity. The Elbow Method is widely used because it provides a visual and intuitive approach to choosing KKK, though in some cases the elbow point may not be distinct, requiring additional metrics like the **Silhouette Score** for validation.

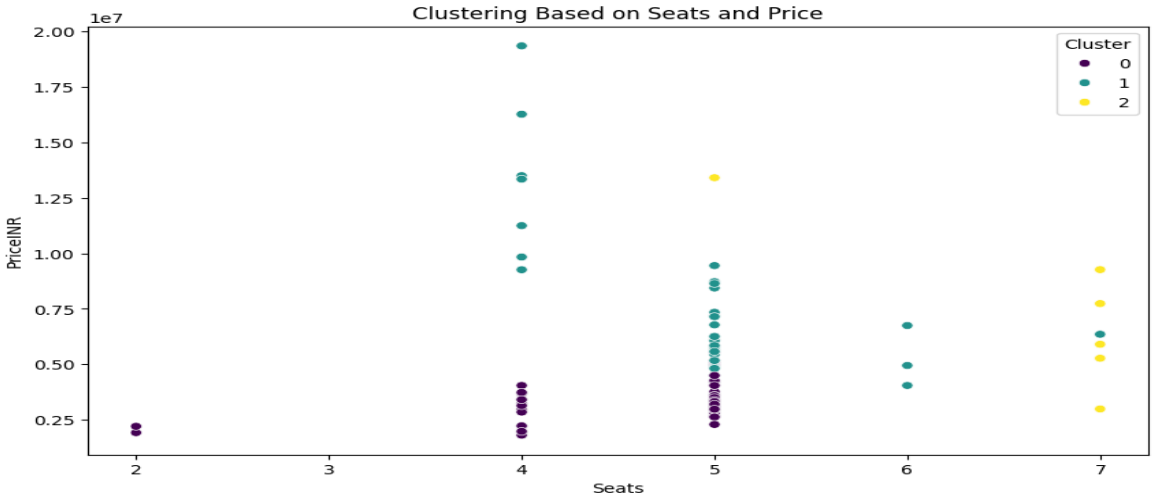


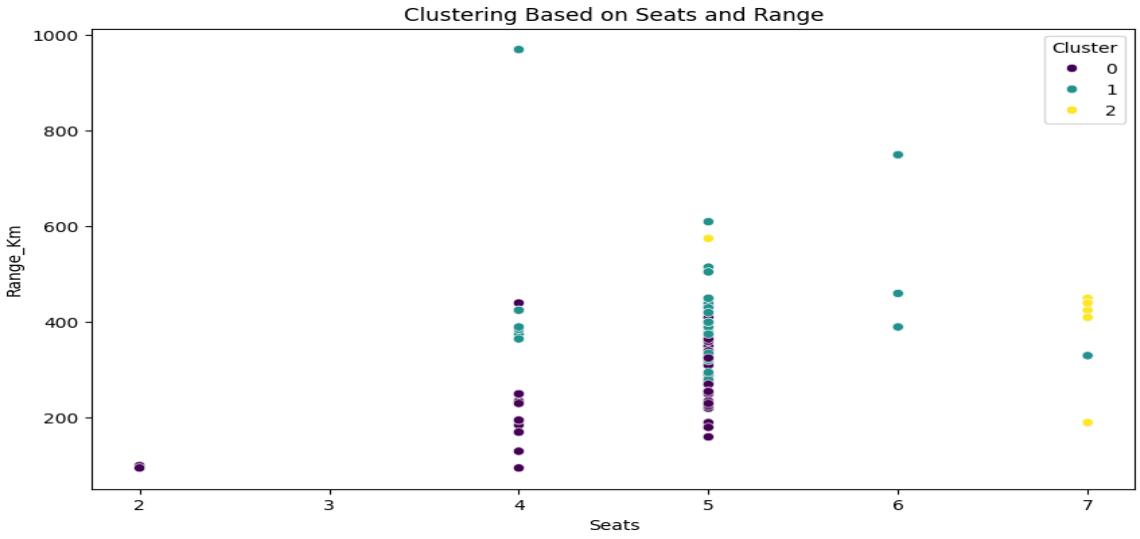
Based on the elbow curve, we assume number of clusters to be optimal around 3.We can confirm this by using the silhouette coefficient ,which is a metric for calculating the goodness of the clustering output. A higher value for the coefficient indicates better cluster.

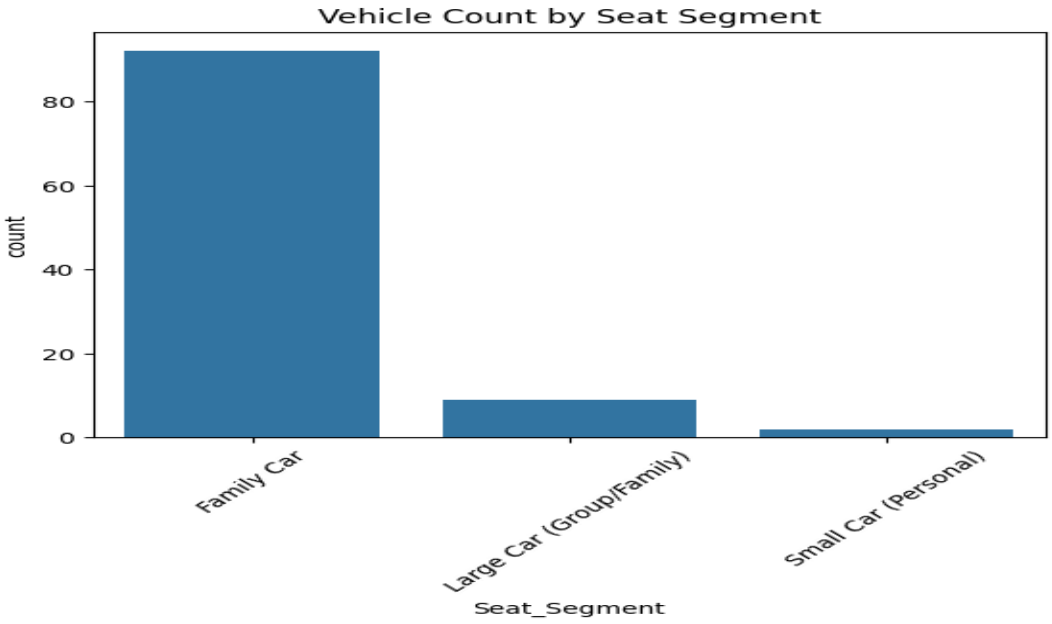
Profiling Potential Segments

Demographic segmentation: Demographic segmentation involves dividing the market based on demographic factors such as family size, household structure, or individual preferences. In our dataset, the "Seats" column represents seating capacity, which can serve as a basis for demographic segmentation.

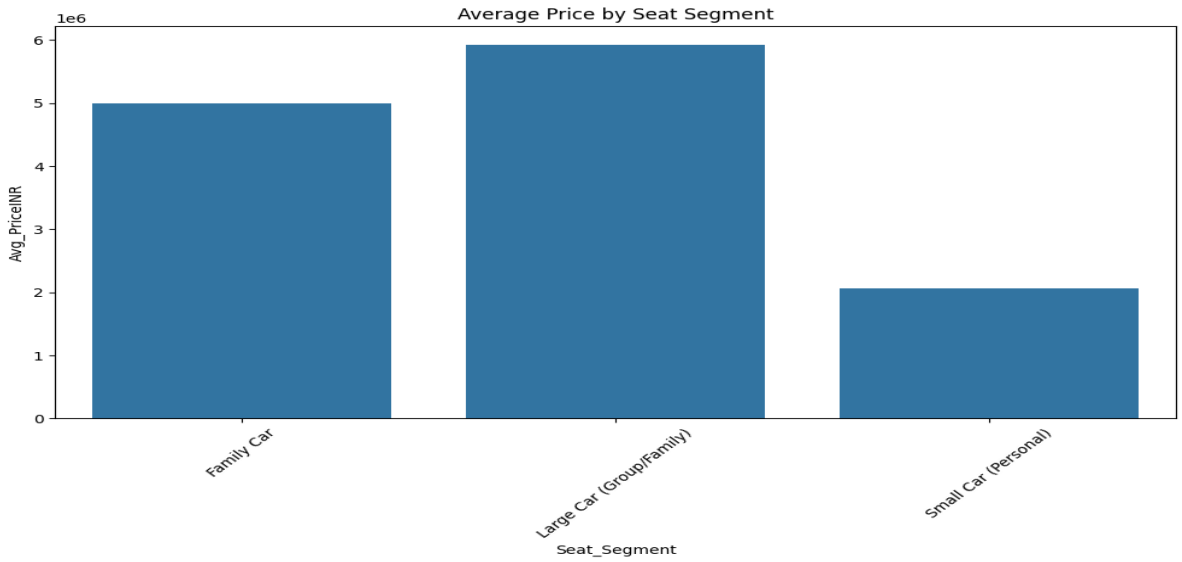
Vehicles are grouped based on their seating capacity. Cluster 0 represents small budget cars(low seats and low prices).Cluster 1 represents Family cars(medium Seats and medium prices) and cluster 2represents premium or large cars(high price and high seats).


Clustering Based on Seats and Price
1e7
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
PriceINR
2
3
4
5
6
7
Seats
Cluster
0
1
2

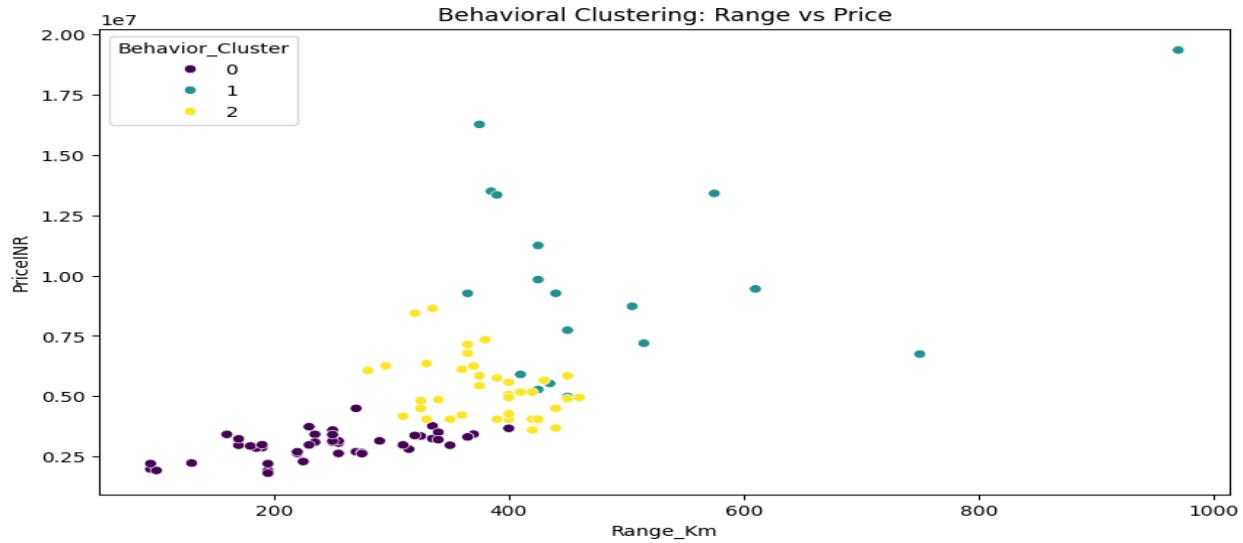

Clustering Based on Seats and Range
1000
800
600
400
200
Range_Km
2
3
4
5
6
7
Seats
Cluster
0
1
2


Vehicle Count by Seat Segment
80
60
40
20
0
count
Family Car
Large Car (Group/Family)
Small Car (Personal)
Seat_Segment

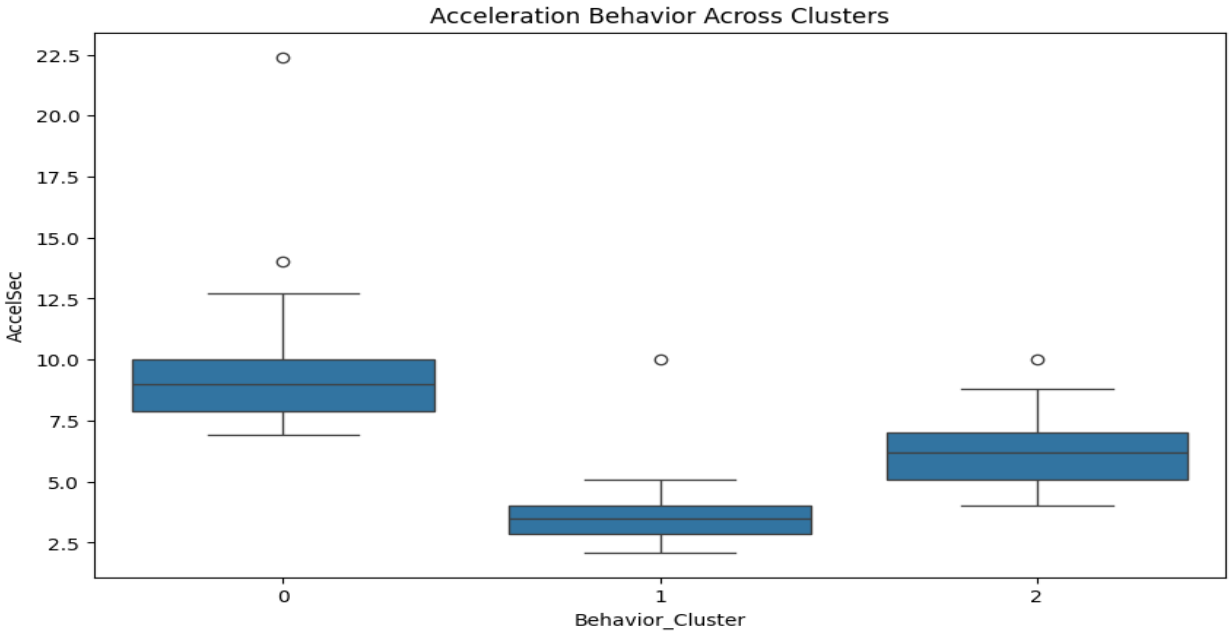
Most of EV vehicles manufactured focus mainly for family and most of the average. Smaller car tends to have lower prices and smaller range. But family cars balance affordability, range and seating.



Behavioral Segmentation: Behavioral segmentation involves grouping customers based on their behaviors, such as usage patterns, preferences, or decision-making factors. For your car dataset, you can segment vehicles based on behavioral attributes such as: Acceleration, Top speed, range, efficiency, fast charge.



There is a **positive correlation** between **range (Km)** and **price (NR)**; higher ranges typically correspond to higher prices. Cluster 0 is tightly packed in the lower range and price segment, whereas Cluster 1 is more dispersed. Cluster 2 seems to act as a transition group between Cluster 0 and Cluster 1.

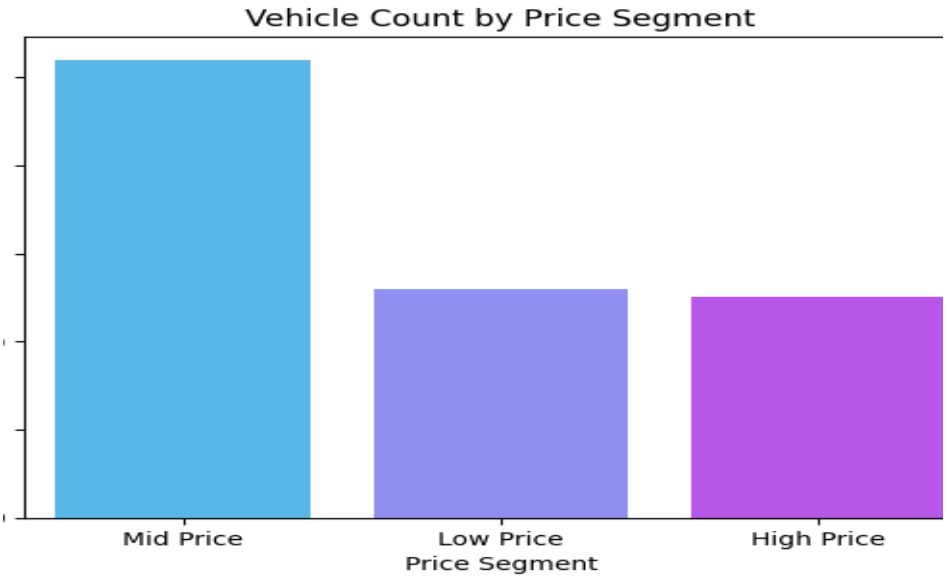


Cluster 1 has the fastest acceleration and the lowest variance, representing high-performance behavior .Cluster 0 has the slowest acceleration with higher variance and outliers, reflecting slower or less consistent performance. Cluster 2 balances between the two, with moderate acceleration performance.

Based on Economy:

The price distributions across the three clusters reveal distinct characteristics. **Cluster 0** exhibits the lowest price range, tightly distributed between approximately 25 and 50, indicating minimal variance in pricing. **Cluster 1** has moderately higher prices with a broader interquartile range (IQR), reflecting

greater variability, with most prices falling between 50 and 100. **Cluster 2** shows the highest price range, predominantly between 100 and 150, with one noticeable outlier reaching nearly 200.



Mid-Price Dominance:

- The majority of vehicles fall into the **Mid-Price** segment, suggesting that this segment is the most competitive and potentially the most appealing to a broad customer base.

Low and High Price Balance:

- The **Low-Price** and **High-Price** segments have a roughly equal number of vehicles, indicating a balanced focus on both budget-conscious buyers and premium customers

Target Segment

Primary Focus:

- Long-Range, Mid-Price EVs (mass appeal).
- Urban Compact EVs (entry-level market with high adoption potential).

Secondary Focus:

- Premium Performance EVs (high-margin luxury buyers).
- Eco-Conscious Family EVs (growing middle-class demand for family-oriented EVs).

Customizing The Marketing Mix

The marketing mix refer to the set of actions or tactics, that a company uses to promote its brand or product in the market. The 4Ps make up a typical marketing mix-Price,Product,Promotion,Place.

1. **Product**: The goods or services offered by a business to meet customer needs or solve problems. This includes the product's design, quality, features, packaging, branding, and any after-sales service.
2. **Price**: The amount customers pay for the product or service. Pricing strategies consider factors like production costs, market demand, competition, and perceived value to determine an optimal price that balances profitability and customer satisfaction.
3. **Place**: The distribution channels and locations where products or services are made available to customers. This includes physical stores, online platforms, or other channels that ensure customers can access the product conveniently.
4. **Promotion**: The communication strategies used to inform, persuade, and remind customers about the product. This includes advertising, sales promotions, public relations, social media campaigns, and other marketing activities designed to generate interest and drive sales.

Together, these four elements form the foundation of a business's marketing strategy, ensuring that the right product reaches the right customer at the right price and through the right channels.

Conclusion

The dataset includes vehicles from a variety of brands, catering to different market segments, from affordable EVs to premium models. Vehicles differ significantly in body styles (e.g., Sedan, Hatchback, SUV) and powertrain options, indicating broad customer preferences. Performance metrics such as acceleration (AccelSec), top speed (TopSpeed_KmH), and range (Range_Km) highlight the variety in capabilities. Price-based segmentation reveals three key customer groups: Low Price, Mid Price, and High Price segments, catering to diverse economic needs. The dataset shows segmentation based on seating capacity, with most vehicles falling into the Medium Car segment (4–5 seats). Smaller cars (2–3 seats) cater to compact city driving, while larger cars (6+ seats) serve family or commercial needs.