Electric Vehicle Market Analysis Report

---

1. Team Members Name:

- Archisman Chakraborti (Team lead)
- Raj Arya
- Nihal Barhaiyya
- Lalkrishna Vinayak Joshi

---

2. Fermi Estimation (Breakdown of Problem Statement):

1. What type of vehicles should the EV business produce?

To tackle this question with a Fermi estimation, we'll break it down into smaller, more manageable questions:

1. **Estimate the total potential market size for EVs nationally**.
    - Consider the total number of vehicles sold nationally.
    - Estimate the percentage of vehicles that are electric.
2. **Break down by vehicle type**.
    - Of the total vehicles sold, estimate the percentage breakdown of types (e.g., SUV vs. Sedan vs. Hatchback).
3. **Consider the growth rate of EVs**.
    - Estimate the annual growth rate of electric vehicles.
    - Adjust the above percentages based on this growth rate.
4. **Factor in customer preferences and market trends**.
    - Consider what the market trends suggest regarding the popularity of vehicle types (e.g., are consumers shifting towards SUVs?).
    - Adjust the percentages based on these trends.

Using these sub-questions, we can make a series of educated guesses to approximate the types of vehicles an EV business should focus on.

2. Which regions are better suited to release the products?

Similarly, for this question, we'll break it down:

1. **Estimate the total number of EVs sold by region**.
    - Consider major markets like the different states.
    - Estimate the percentage of total global EV sales each region represents.

2. **Consider infrastructure and government policies**.
    - Regions with better EV infrastructure (like charging stations) might see higher sales.
    - Similarly, regions with government incentives for EVs might also have higher sales.
3. **Factor in regional growth rates**.
    - Some regions might be experiencing faster growth in EV adoption than others.
4. **Consider cultural and economic factors**.
    - Some regions might prefer certain types of vehicles over others due to cultural or economic reasons. For instance, larger vehicles like SUVs might be more popular in less populated states, while compact cars might be preferred in densely populated cities.

---

3. Data Sources (Data Collection):

All team members actively participated in data collection. Our primary dataset for the analysis was sourced from:

- **Dataset Link**: ElectricCarData_Norm.csv
- **Dataset Link**: Domestic Automobile sales
- **Dataset Link**: Provided data by FeynLabs
- **Dataset Link**: Statewise sales data

---

4. Data Pre-processing:

Data Preprocessing

---

1. Handling Missing Values:

The dataset was first inspected for any missing or null values. Missing data can adversely affect the performance of machine learning algorithms and lead to misleading results.

Action Taken: All rows containing null values were removed to ensure data integrity.

---

2. Data Scaling:

Given the diversity of features with different scales in the dataset, scaling was essential, especially for clustering algorithms like K-means that rely on distances.

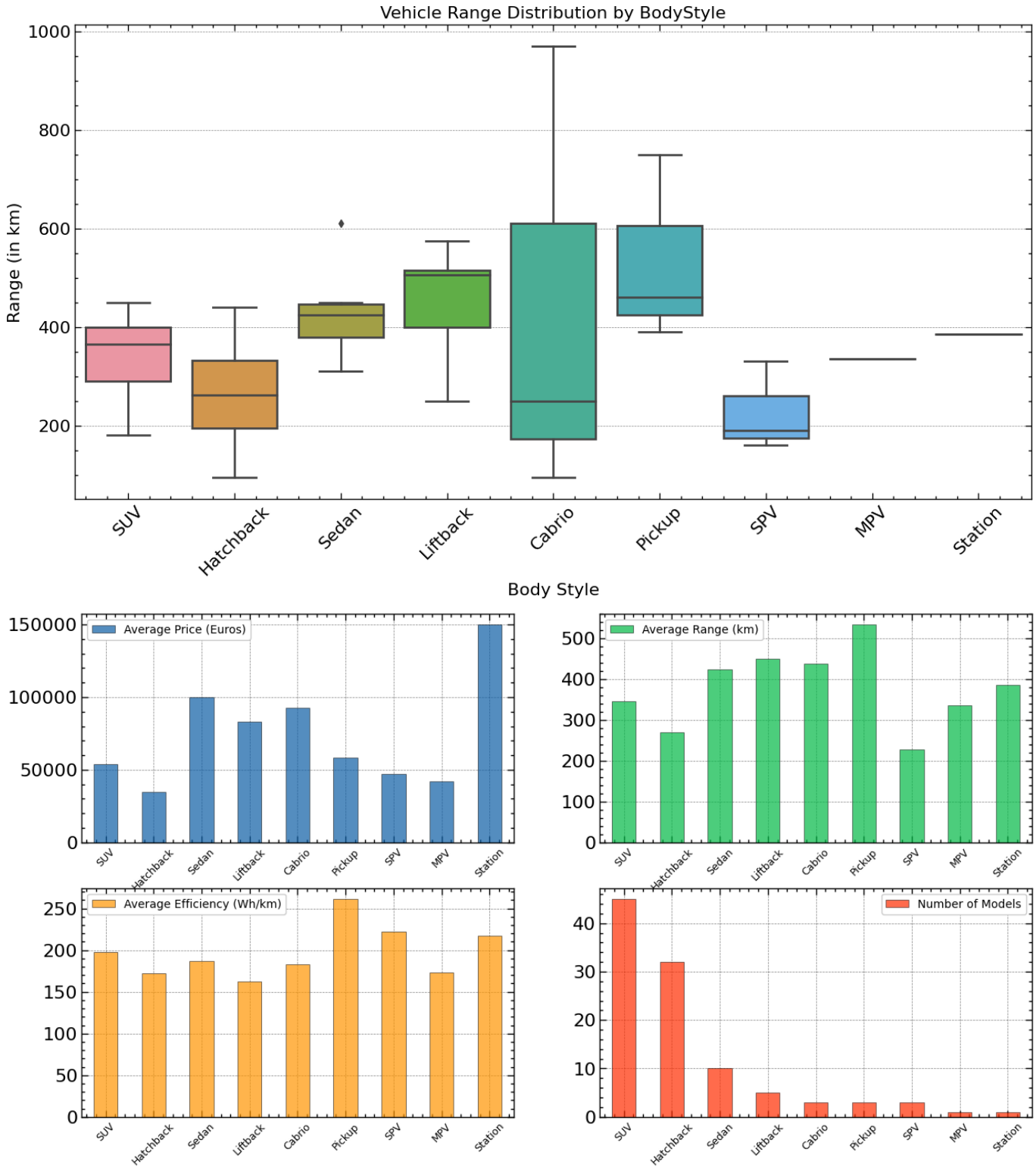
Action Taken: The `StandardScaler` from the `sklearn.preprocessing` library was employed to standardize the dataset. Standard scaling adjusts the features so they have a mean of 0 and a standard deviation of 1.

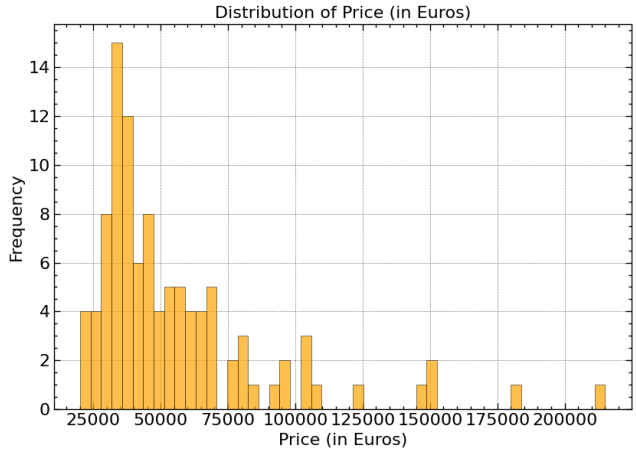
3. Feature Selection:

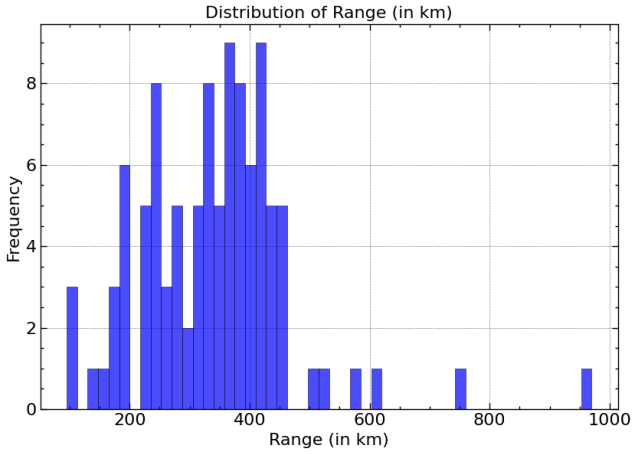
Feature selection is crucial to ensure that only relevant features are used in the analysis, which can improve the performance of machine learning algorithms and provide clearer insights.

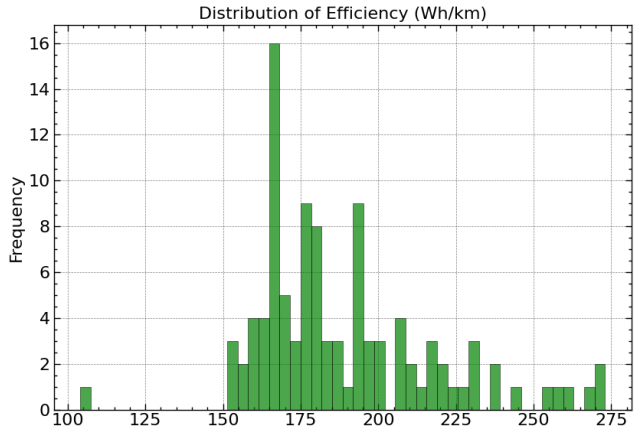
Action Taken: Features were selected based on domain knowledge. Irrelevant or redundant features that wouldn't provide significant value to the clustering analysis were excluded.

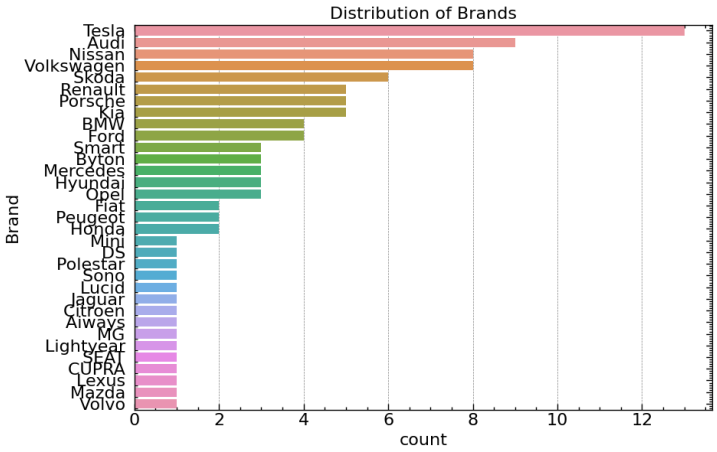
This data preprocessing ensured that the dataset was clean, appropriately scaled, and ready for further analysis with only the most pertinent features included. These steps are foundational to any data analysis workflow, ensuring the reliability and robustness of subsequent analyses.

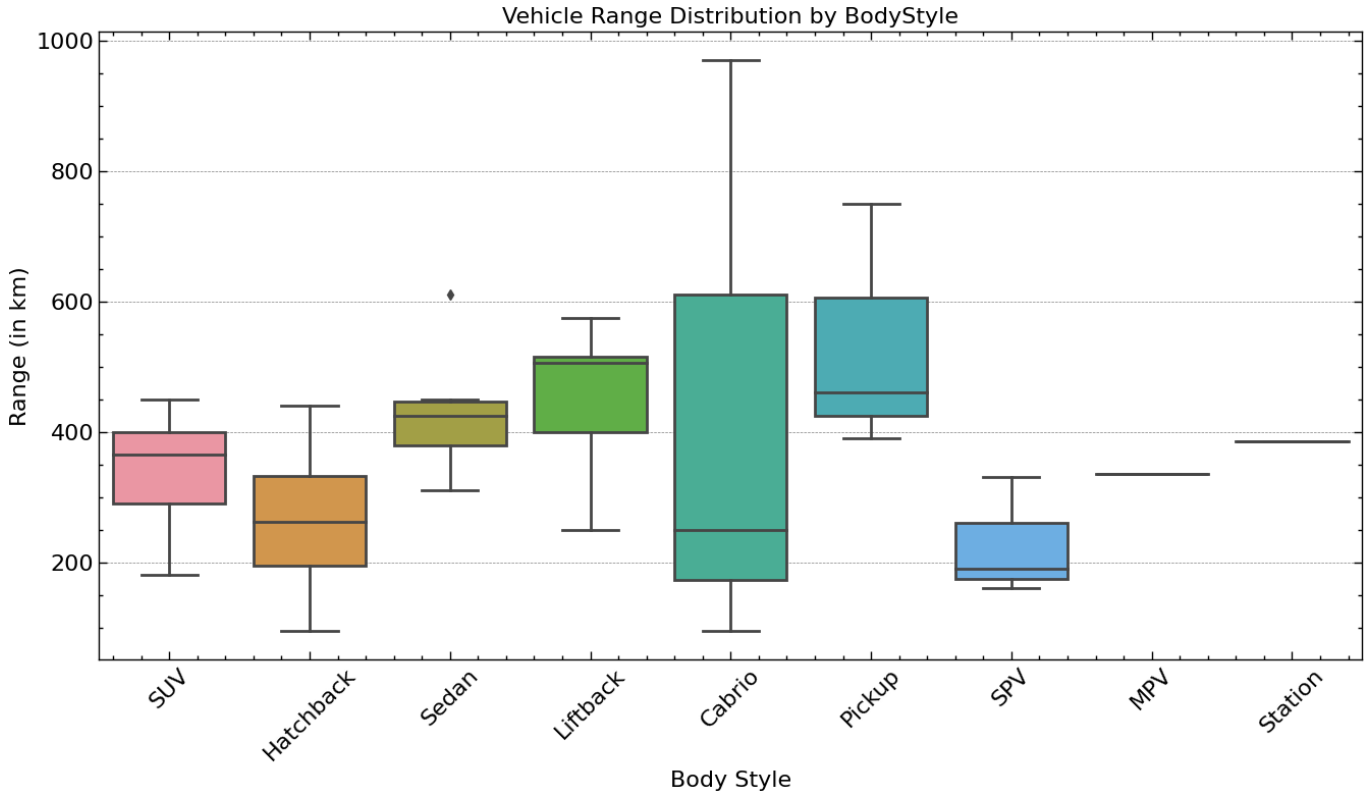










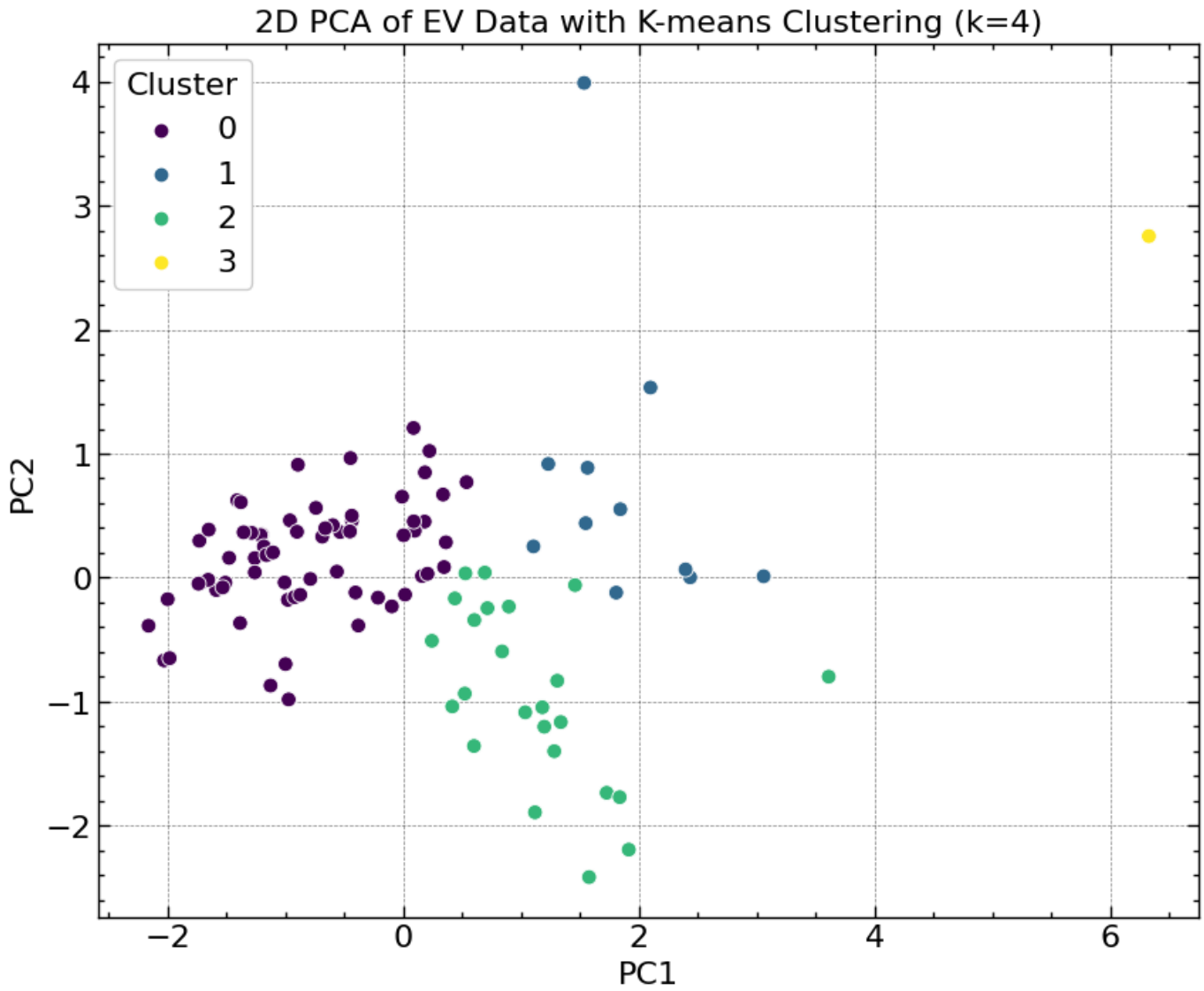


5. Segment Extraction (ML techniques used):

Segment Extraction Using K-means Clustering

K-means Clustering:

K-means is a partitioning clustering algorithm that divides a dataset into (K) distinct, non-overlapping subsets (or clusters). The goal is to minimize the distance between data points within the same cluster and maximize the distance between different clusters.



Determining Optimal Number of Clusters:

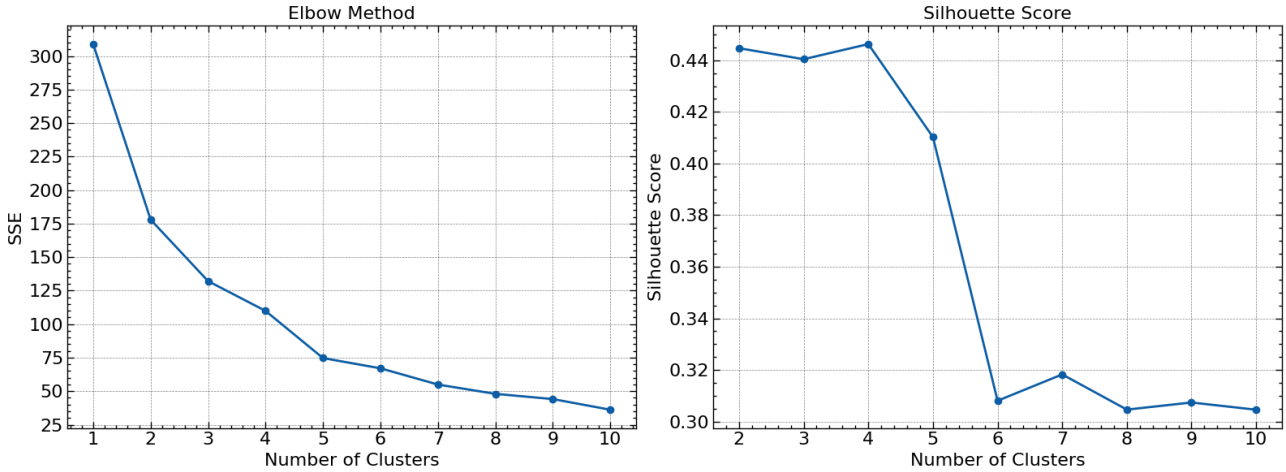
To decide the best number of clusters K, two methods were employed:

1. **Elbow Method**:

   By plotting the cost function J against different values of K, we can look for an "elbow" in the plot. The elbow point represents an optimal value for K (a balance between precision and computational cost).

2. **Silhouette Score**:

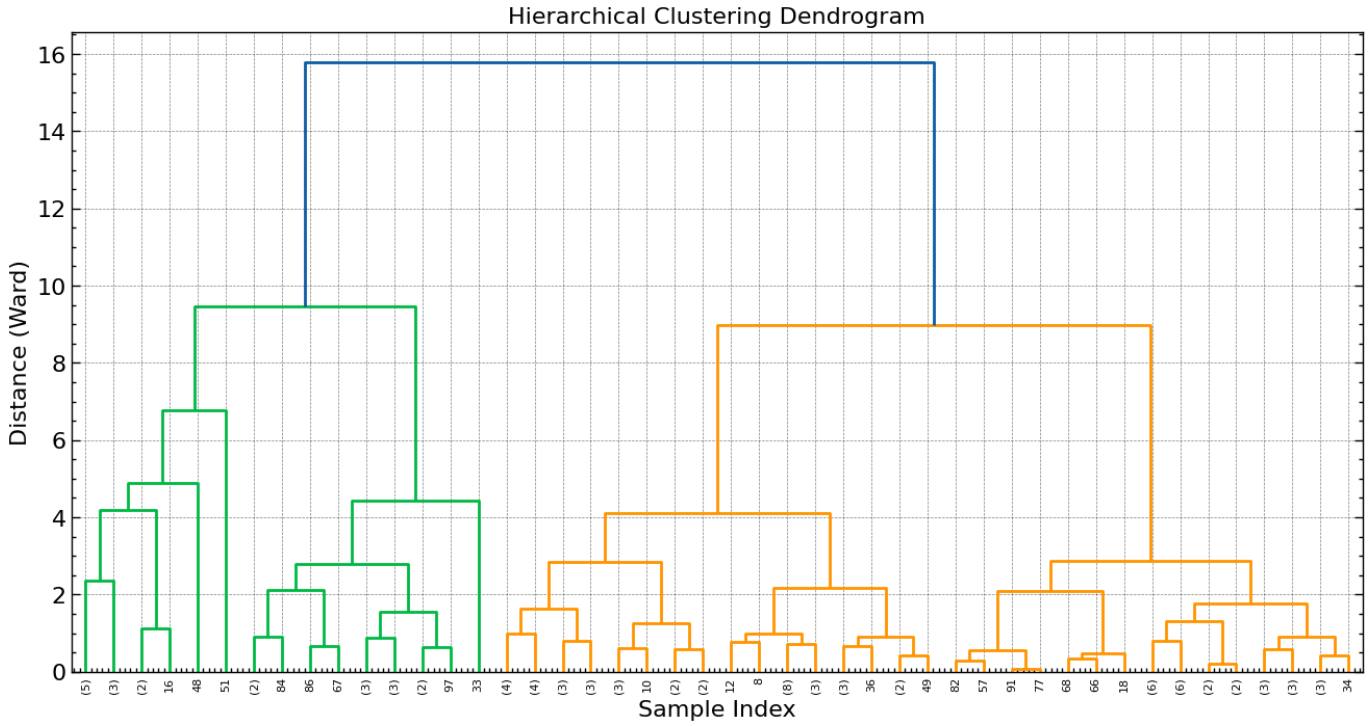
   The silhouette score measures the similarity of an object to its own cluster compared to other clusters. The value ranges from -1 to 1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters.



Here, we took 4 clusters.

Hierarchical Analysis for Segment Profiling:

Hierarchical clustering was used to understand the hierarchical structure of the segments. This technique builds a dendrogram or tree-like diagram that breaks down the dataset into smaller clusters, providing a visual representation of data groupings.



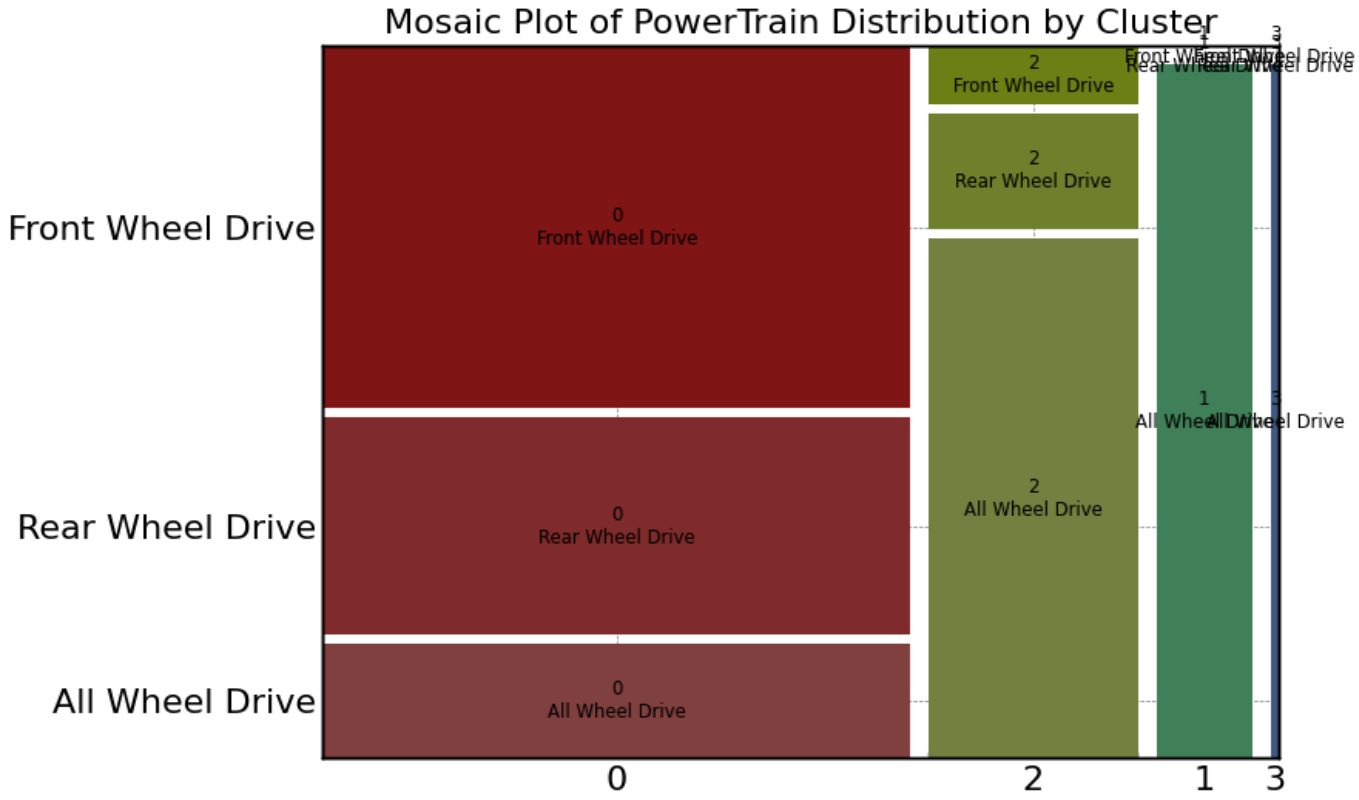
From the dendrogram, we can understand the hierarchy and relationships between clusters, which aids in segment profiling. This detailed structure allows for a granular view of the segments and their characteristics.

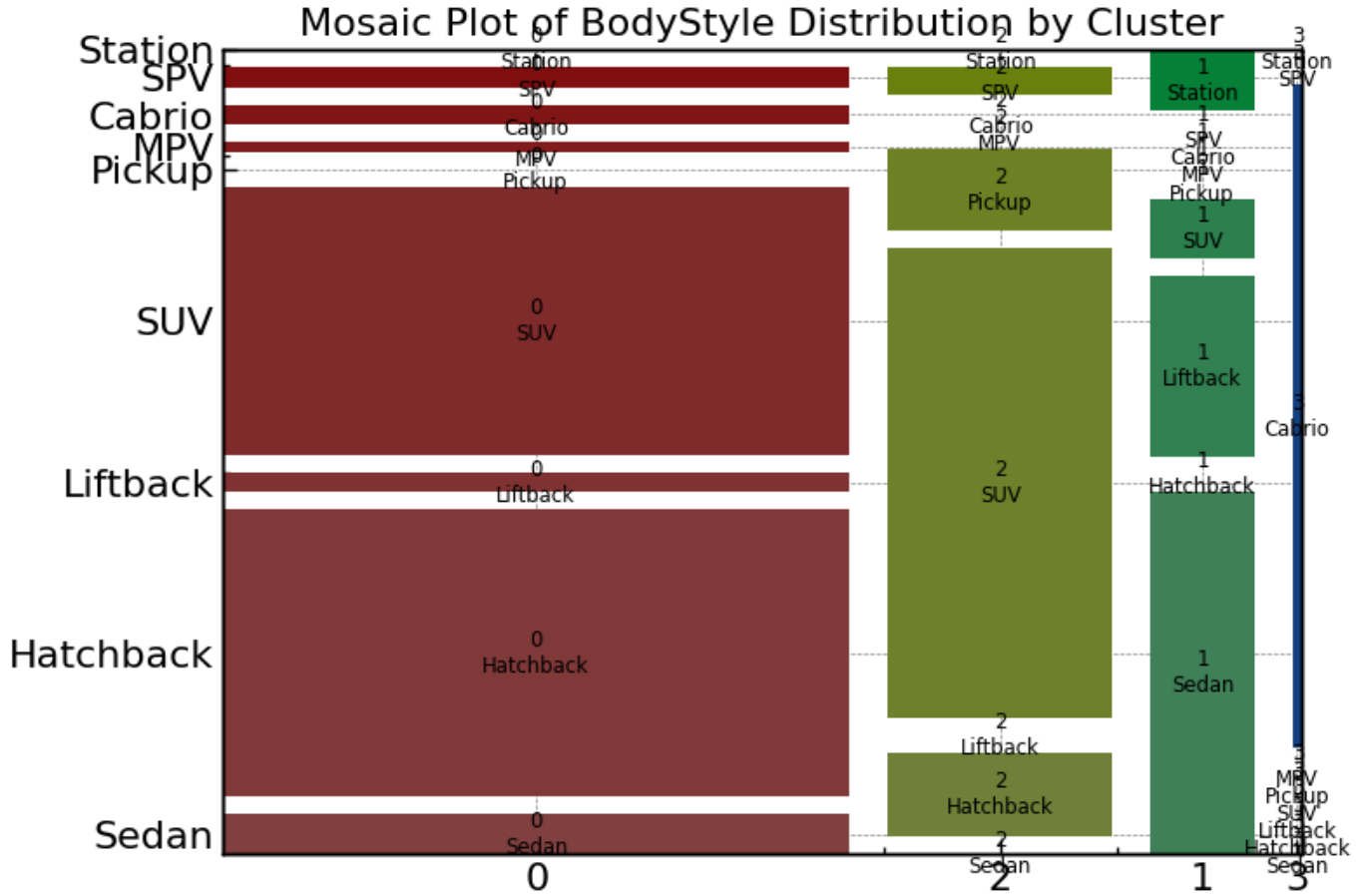
---

6. Profiling and Describing Potential Segments:

Profiling and Segment Analysis

By analyzing the clusters formed during the segment extraction, we can profile and describe the key characteristics associated with each cluster. This provides a clearer understanding of the different market segments and their preferences.





Cluster Descriptions:

- **Cluster 0**: This cluster is predominantly characterized by SUVs, but there's also a notable presence of Sedans. A smaller segment of this cluster comprises Hatchbacks. It appears that this segment has a preference for larger vehicles, with the SUVs taking the lead.
- **Cluster 1**: Representing a more diverse group, Cluster 1 features a mix of SUVs, Sedans, and Hatchbacks. Among these, SUVs are slightly more dominant, suggesting a versatile market segment with varied preferences.
- **Cluster 2**: Standing out for its clear preference, Cluster 2 primarily consists of Hatchbacks. This suggests a segment that favors compact, city-friendly vehicles. There's a minor presence of SUVs and Sedans, indicating some variability within this cluster.
- **Cluster 3**: Overwhelmingly dominated by SUVs, Cluster 3 showcases a strong inclination towards this body style, with only a minor presence of other body styles. This suggests a segment that is heavily oriented towards larger, perhaps more premium vehicles.

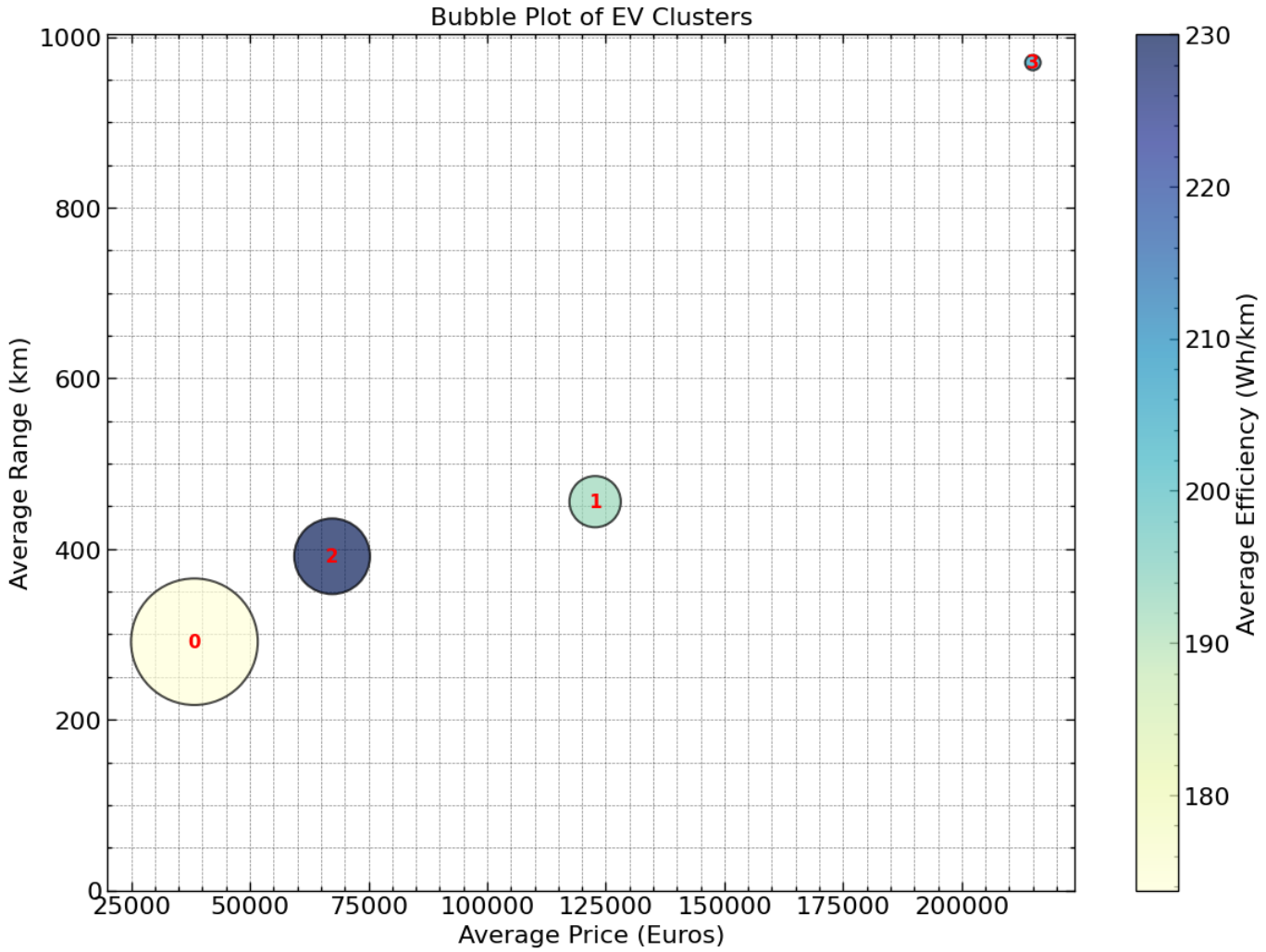
Inferences:

- **Cluster 0** and **Cluster 3**: Both these clusters exhibit a strong preference for SUVs. This could imply that they represent segments of the market that are inclined towards the higher end. Their choices suggest a preference for spaciousness, luxury, and perhaps off-road capabilities which are often associated with SUVs.
- **Cluster 1**: This cluster's balanced mix of popular body styles points to a more general market segment. It suggests a diverse group of consumers with varied needs, from the spaciousness of SUVs to the efficiency of Sedans and the compactness of Hatchbacks.
- **Cluster 2**: With its pronounced preference for Hatchbacks, this cluster likely represents urban dwellers who prioritize compact vehicles that are easy to maneuver and park in city settings. It hints at a segment that values efficiency, affordability, and practicality for city commuting.

The profiling and segment analysis provides a comprehensive understanding of the different market segments, helping stakeholders make informed decisions tailored to each segment's unique preferences and needs.

---

7. Selection of Target Segment:



The bubble plot provides a comprehensive visualization of the clusters based on several attributes:

- **X-axis (Average Price)**: Indicates the average price of vehicles in each cluster.
- **Y-axis (Average Range)**: Represents the average range of vehicles in each cluster.
- **Size of the Bubble**: Denotes the number of models in each cluster, with larger bubbles indicating a higher number of models and thus potentially a more popular or dense segment.
- **Color**: Signifies the average efficiency of vehicles in each cluster, with darker colors indicating better efficiency.

Target Cluster Selection

The identification of a target segment is a critical decision for businesses, enabling them to tailor their offerings, marketing strategies, and operations to best serve and appeal to that particular segment. Based on our cluster analysis and profiling, we can make the following recommendations for target cluster selection:

- **Broad Market Focus - Cluster 1**: For businesses that wish to cater to a wide audience and tap into popular market segments, Cluster 1 emerges as the ideal choice. This cluster is characterized by a balanced mix of vehicle attributes, encompassing a range of body styles. Additionally, the sheer number of models in this cluster underscores its broad market appeal. Companies aiming for widespread market penetration would find Cluster 1 aligning well with their objectives.

- **Premium Segment Focus - Cluster 3**: If a business's strategy is rooted in targeting the premium end of the market, Cluster 3 is the go-to segment. Dominated by SUVs and characterized by a preference for larger, perhaps more luxurious vehicles, this cluster caters to consumers who are likely to prioritize luxury, advanced features, and brand prestige. Companies with premium offerings would resonate well with this segment.

- **Compact City Vehicle Focus - Cluster 2**: Urban environments, characterized by bustling streets and limited parking, create a demand for compact, efficient vehicles. Cluster 2, with its pronounced inclination towards Hatchbacks, represents this very segment of the market. Businesses that specialize in compact, city-friendly vehicles or are looking to tap into the urban commuter segment would find Cluster 2 to be the most apt.

- **Mid-to-High Market Focus - Cluster 0**: Occupying the space between the general and premium segments is Cluster 0. This cluster offers an opportunity for businesses aiming at the mid-to-high market segment. With a clear preference for range and a mix of SUVs and Sedans, it caters to a segment that seeks a blend of luxury, performance, and practicality without necessarily venturing into the premium bracket.

In conclusion, the selection of a target segment should align with a company's strategic objectives, product offerings, and brand positioning. The detailed cluster analysis provides a robust foundation for businesses to make informed decisions and tailor their strategies for optimal market engagement.

8. Customizing the Marketing Mix:

To cater to the identified target segment, companies can:

- **Product**: Offer a mix of SUVs, Sedans, and Hatchbacks.
- **Price**: Position products in the mid-price range, based on the segment's average price.
- **Place**: Focus on regions with a high concentration of the target segment.
- **Promotion**: Emphasize the vehicle's efficiency, range, and environmental benefits.

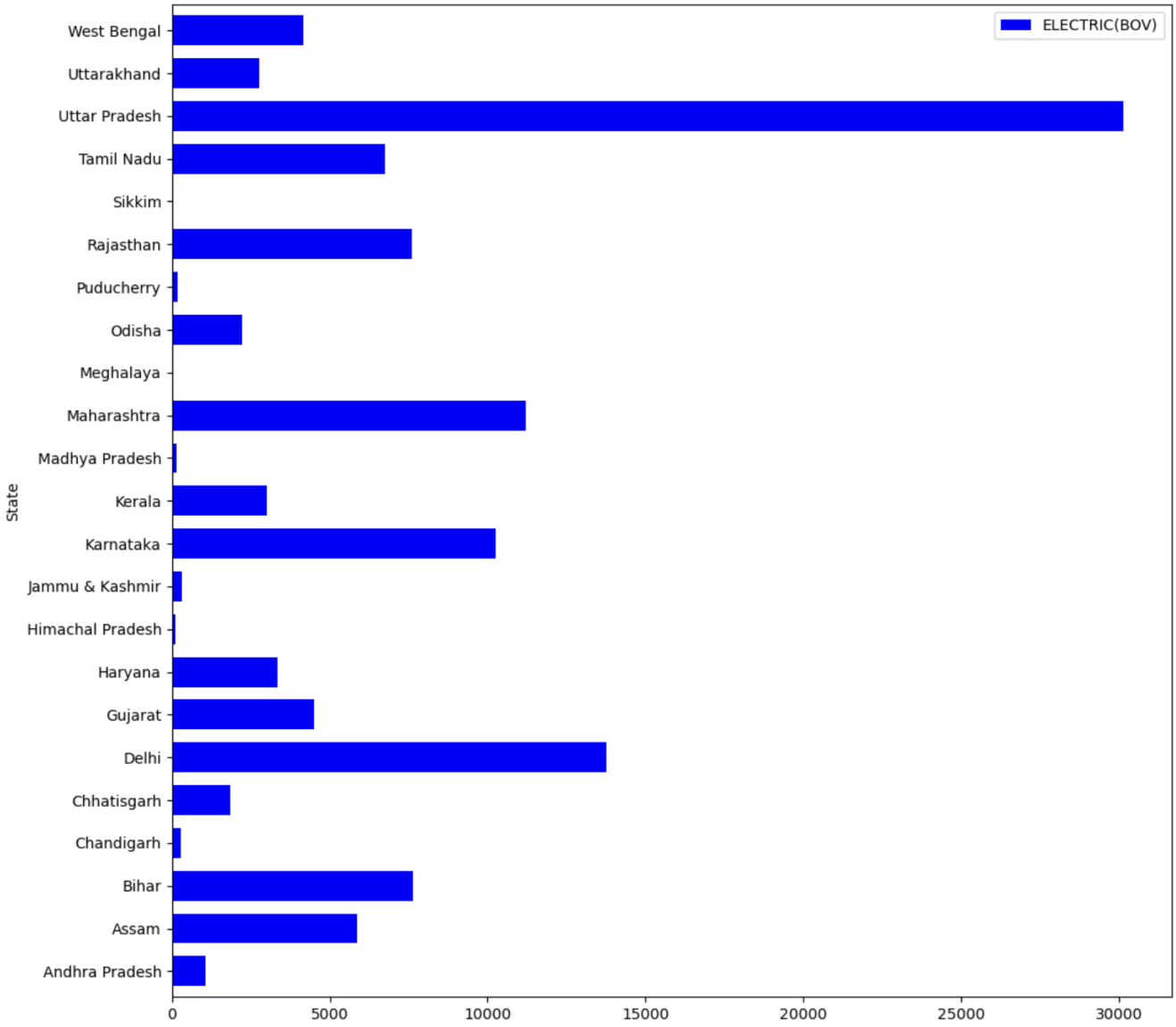
---

10. The MOST OPTIMAL MARKET SEGMENTS:

From our research:

- **Cluster 1** represents the most optimal market segment. It captures a diverse audience with its mix of body styles and is primarily focused on BEVs.

Geographical Targeting Strategy



From our visual analysis, we derived a clear understanding of the regional preferences and market sizes. To optimize market penetration and ensure effective resource allocation, businesses should prioritize their targeting efforts in the following order:

1. **Maharashtra**: As one of the most economically advanced states with a large urban population, Maharashtra represents a significant market for electric vehicles. Cities like Mumbai and Pune can be primary hubs for marketing and distribution efforts.

2. **Uttar Pradesh**: With its vast population and emerging urban centers, Uttar Pradesh offers a substantial market opportunity. The state's diverse demographics also allow for a varied marketing approach.

3. **Karnataka**: Home to tech hubs like Bengaluru, Karnataka is a burgeoning market for electric vehicles, especially among the tech-savvy urban population.

4. **Gujarat**: Known for its industrial prowess and rapidly growing urban centers, Gujarat is a promising market for EV adoption.

5. **Delhi**: The national capital, with its focus on reducing pollution and supportive government policies, presents a significant opportunity for electric vehicle businesses.

6. **Rajasthan**: With its mix of urban and semi-urban regions, Rajasthan can be a strategic market, especially for models tailored for longer distances between cities.
7. **Haryana**: Adjacent to Delhi and home to rapidly growing cities like Gurugram, Haryana is an emerging market for electric vehicles, especially in the corporate sector.

By targeting these regions in the specified order, businesses can maximize their reach, cater to the most promising markets first, and sequentially expand their footprint. This geographical strategy complements the segment targeting, providing a comprehensive approach for businesses to make impactful inroads into the electric vehicle market.

11. Link to GitHub Profile:

All the code, datasets, and related documentation are available in our GitHub repository:

- GitHub Repository Link-Archisman Chakraborti
- GitHub Repository Link-Nihal Barhaiyya
- GitHub Repository Link-Lalkrishna Vinayak Joshi
- GitHub Repository Link-Raj Arya