EV Market Segmentation

- by Ahaan Giriya

Explanation of the ML Model Used

In the second project, K-Means clustering was used to group Electric Vehicles based on various features. K-Means is an unsupervised machine learning algorithm that partitions data into K distinct non-overlapping subsets (clusters) based on feature similarity.

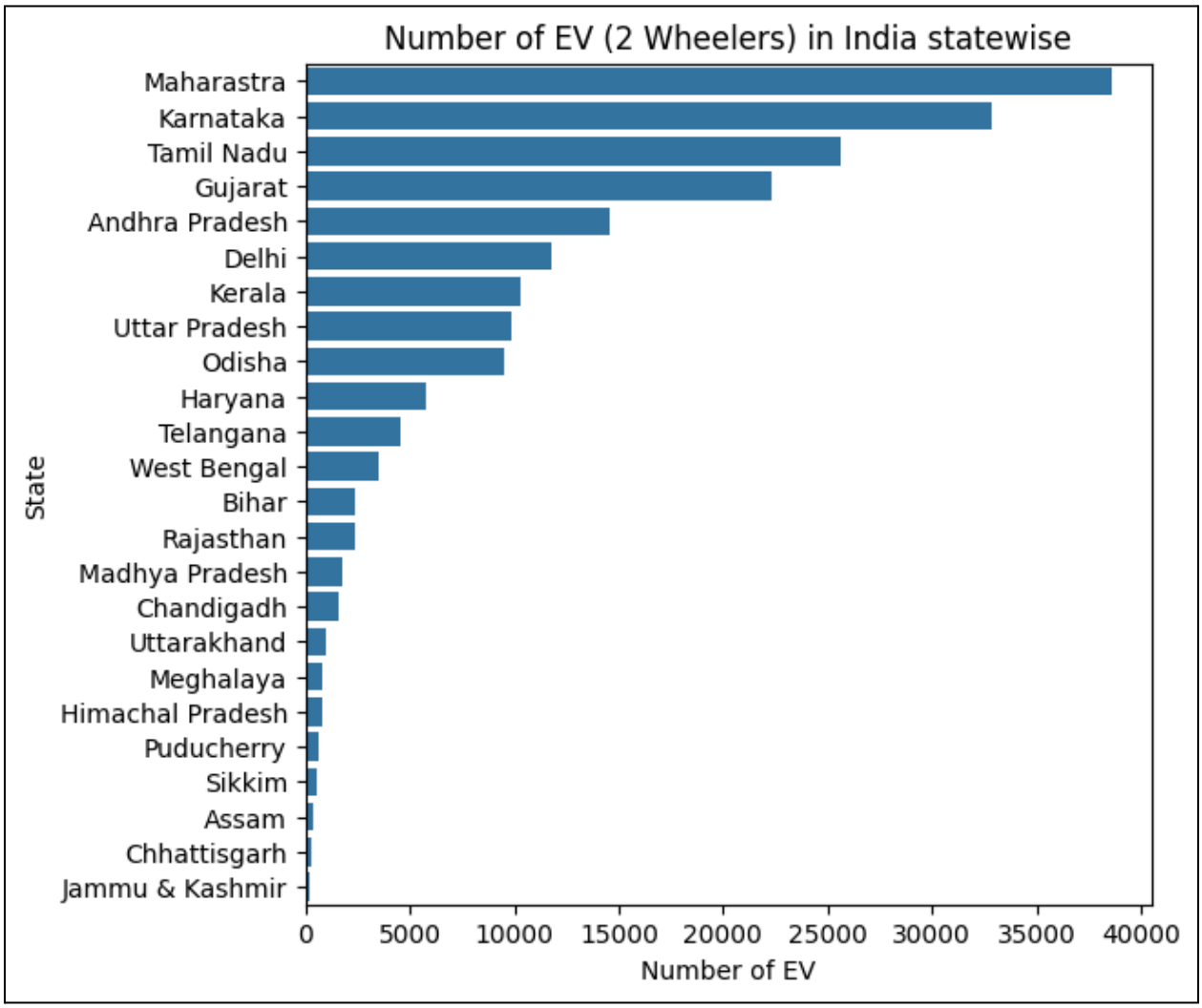
The goal was to group EVs into clusters based on their characteristics to identify patterns and similarities among different types of EVs available in the market.

Steps and Explanation:

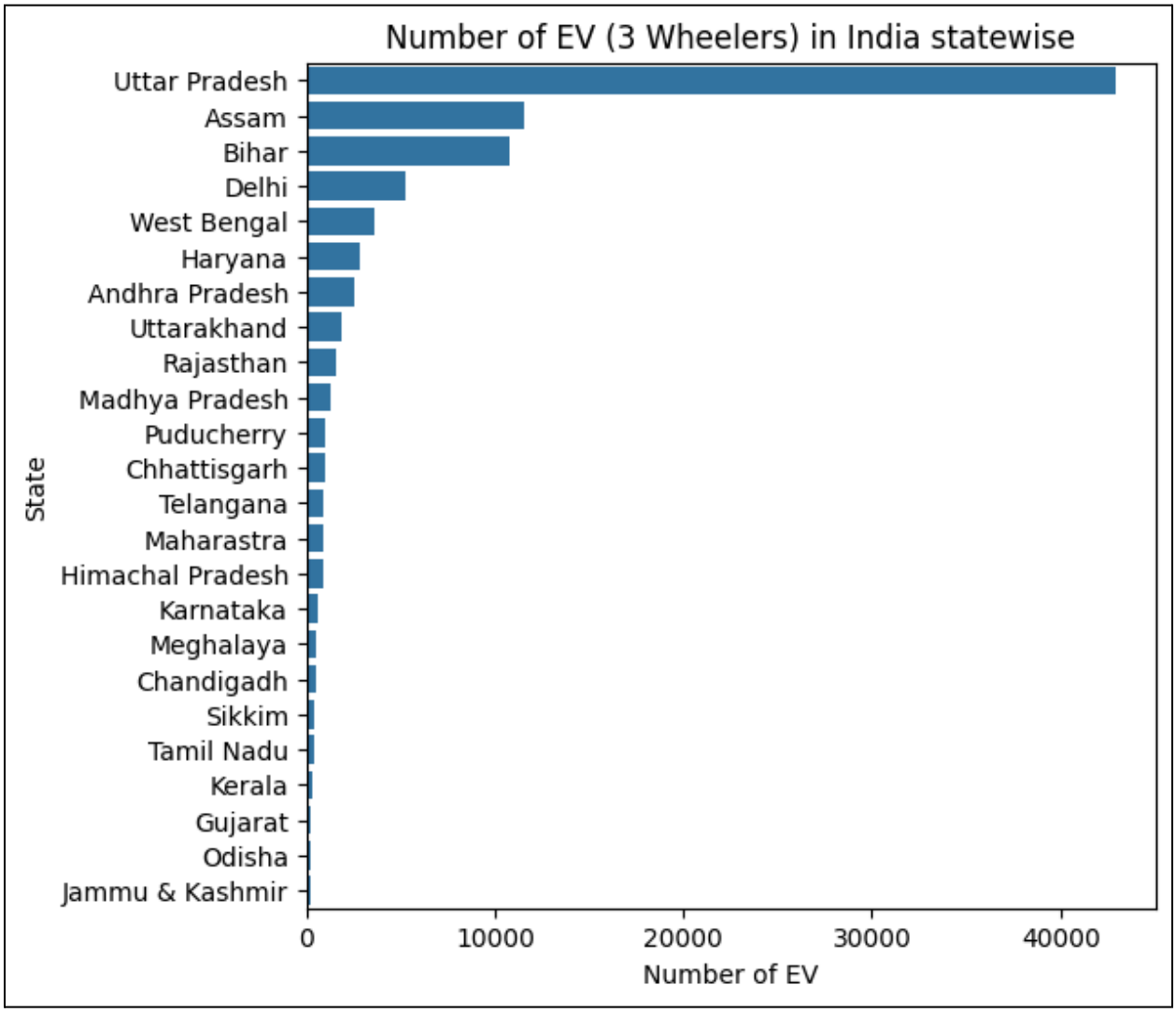
1. **Data Preprocessing:**
    - **Feature Selection:** Selected relevant features for clustering: `AccelSec`, `TopSpeed_KmH`, `Efficiency_WhKm`, `FastCharge_KmH`, `Range_Km`, `RapidCharge`, `Seats`, `PriceEuro`, and `PowerTrain`.
    - **Data Transformation:** Encoded categorical features (`PowerTrain` and `RapidCharge`) to numerical values.
    - **Standardization:** Standardized the features using `StandardScaler` to ensure each feature contributes equally to the distance calculations.
2. **Dimensionality Reduction:**
    - Applied Principal Component Analysis (PCA) to reduce the dimensionality of the dataset while retaining most of the variance in the data. This helps in visualizing the clusters more effectively.
3. **Determining Optimal Number of Clusters:**
    - Used the Elbow Method to determine the optimal number of clusters by plotting the Within-Cluster Sum of Squares (WCSS) against the number of clusters. The optimal number of clusters is where the elbow point occurs.
4. **Applying K-Means Clustering:**
    - Chose 4 clusters (based on the Elbow Method) and applied K-Means to the PCA-transformed data.
5. **Visualization:**
    - Visualized the clusters using a scatter plot of the first and ninth principal components (PC1 and PC9), highlighting the cluster centroids.

Conclusion and Insights Gained from the Research/Analysis Work

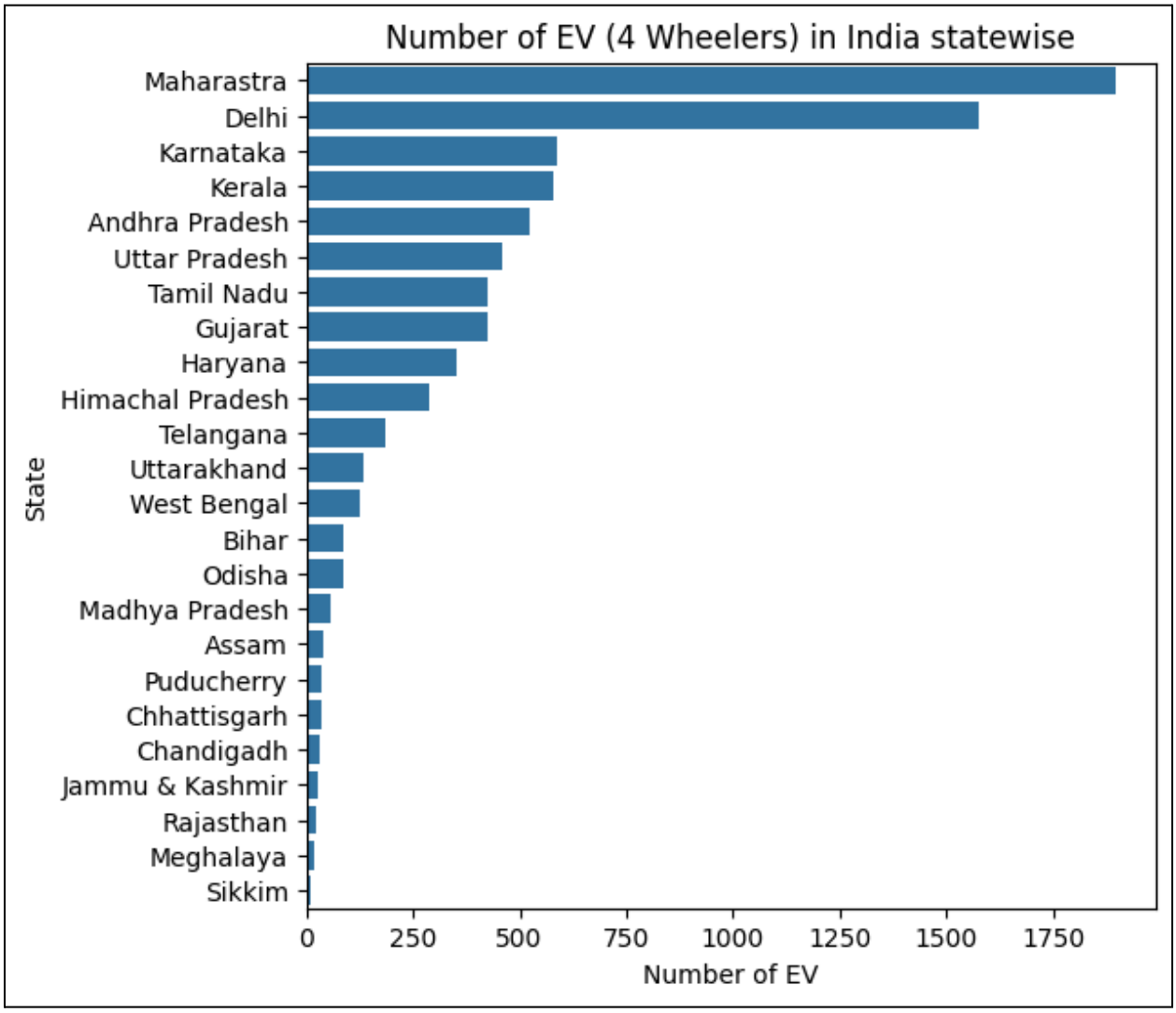
1. **State-wise Distribution of EVs:**
    - **2-Wheeler EVs:** States with the highest number of 2-Wheeler EVs were Maharashtra, Karnataka and Tamil Nadu while with the least were Jammu Kashmir, Chhattisgarh and Assam.



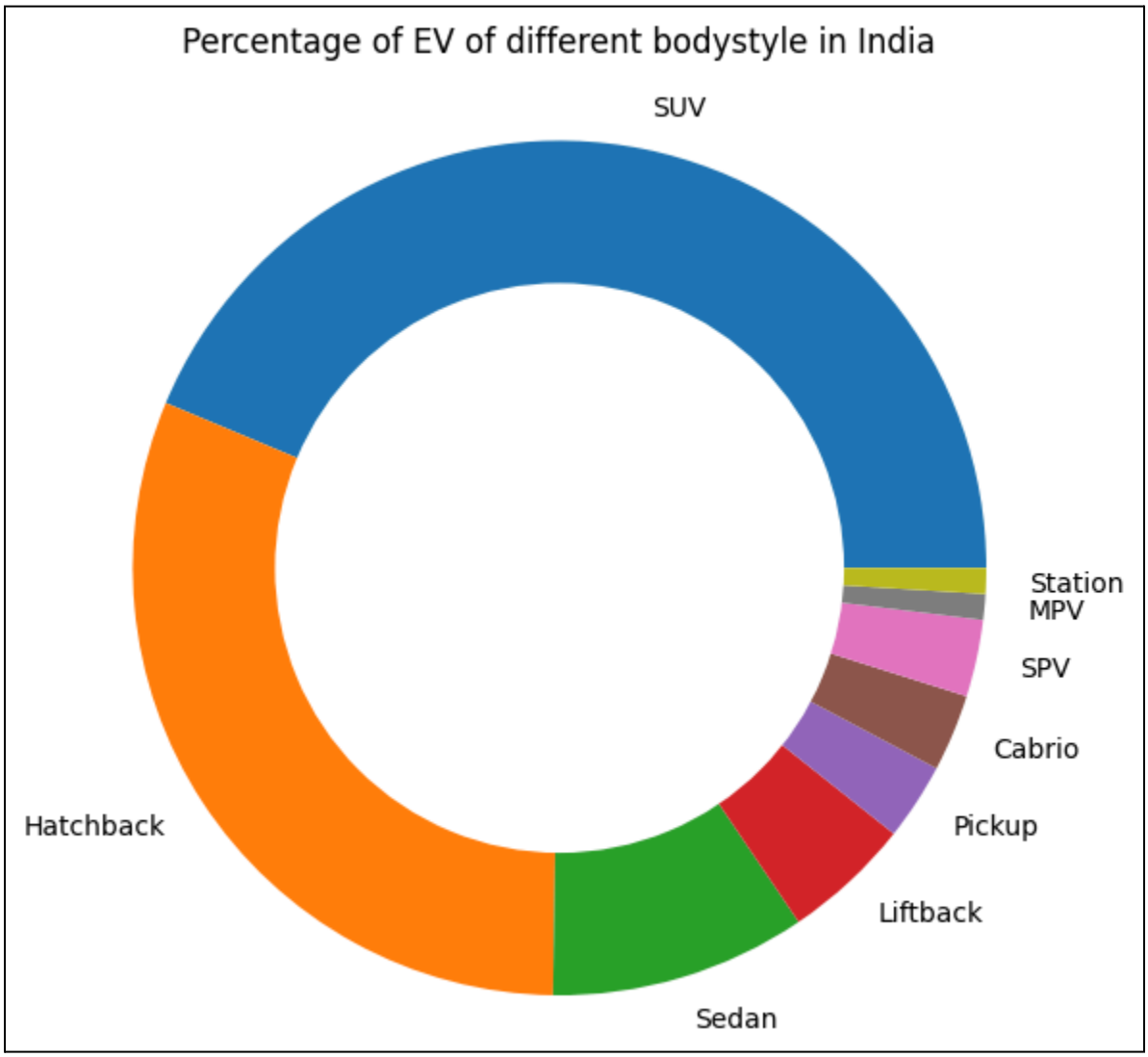
- **3-Wheeler EVs:** States with the highest number of 3-Wheeler EVs were Uttar Pradesh, Assam, Bihar while with the least were Jammu Kashmir, Odisha and Gujarat.



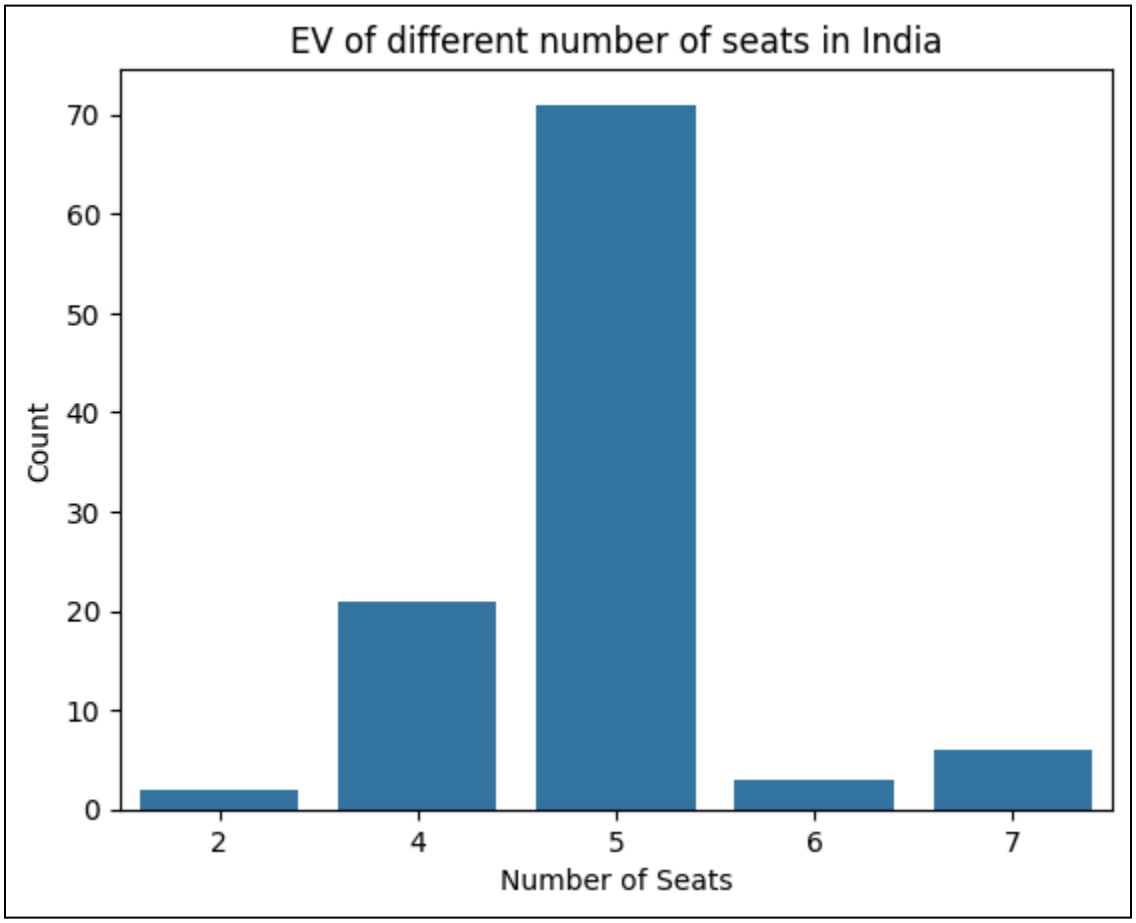
- **4-Wheeler EVs:** States with the highest number of 4-Wheeler EVs were Maharashtra, Delhi and Karnataka while with the least were Sikkim, Meghalaya and Rajasthan.



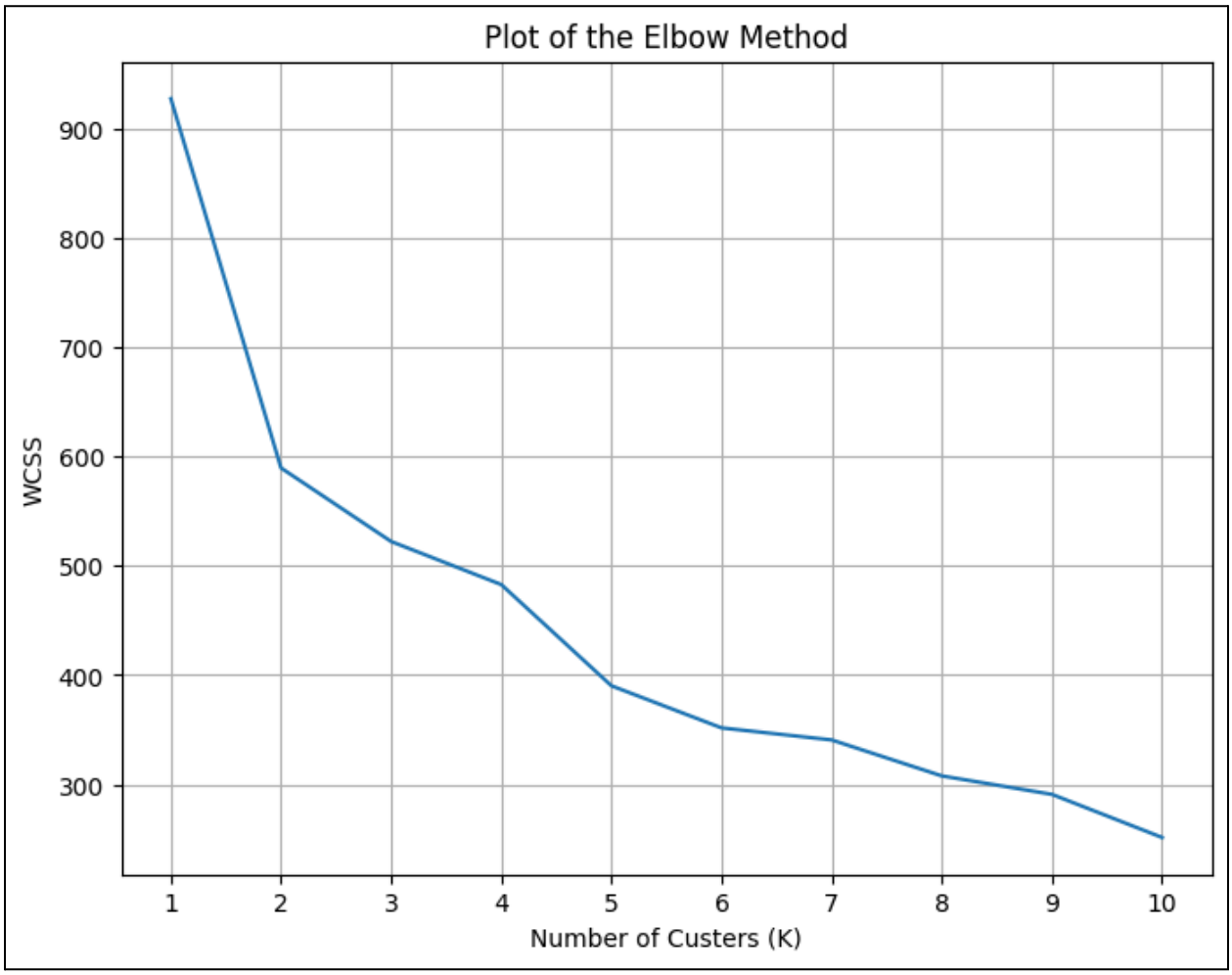
2. **Body Style and Seat Distribution:**
    - **Body Style Distribution:** The pie chart analysis highlighted the market share of different EV body styles. SUV, Hatchback and Sedan were the majority.



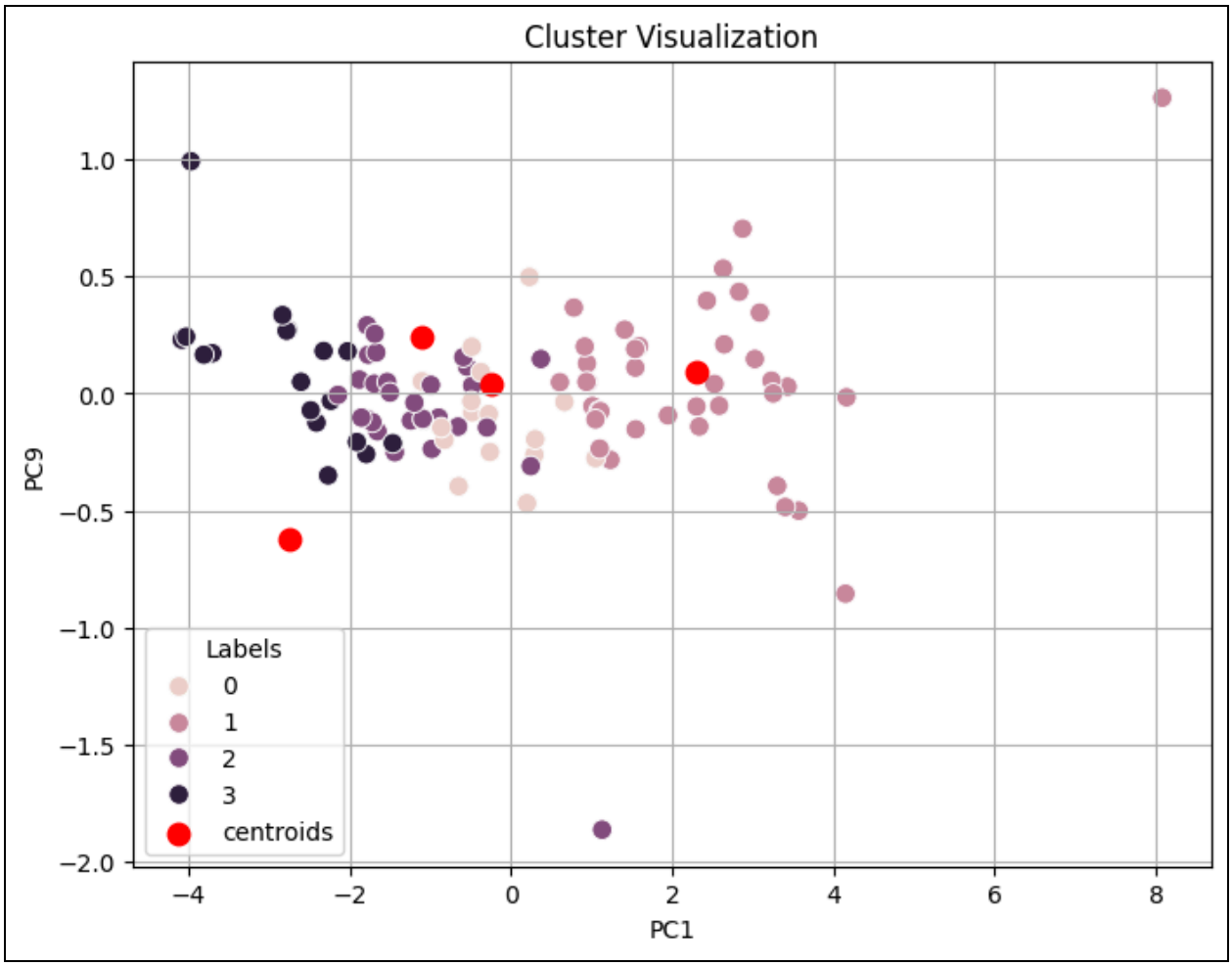
- **Seat Count Distribution:** The count plot for the number of seats in EVs revealed popular configurations. 5 seaters followed by 4 seaters were the most popular.



3. **Clustering of EV Features:**
    - **Optimal Clusters Determination:** Using the Elbow Method, the optimal number of clusters was identified. 4 clusters provided a balance between granularity and interpretability, capturing the variance in the data effectively.



    - **Cluster Analysis:** The scatter plot visualization of the clusters using the first and ninth principal components (PC1 and PC9) illustrated distinct groupings of EVs. Each cluster represented a unique combination of features, such as acceleration, top speed, efficiency, range, and price.
    - **Centroid Analysis:** The centroids of the clusters indicated the average characteristics of each cluster. This information can be used to identify the defining features of each group, helping manufacturers and marketers to tailor their strategies.



Improvements for the Market Segmentation Project

Expanded Datasets

1. **Demographic Data:** Age, Gender, Income Level, Education Level, Occupation
2. **Geographic Data:** Urban/Rural Indicator, Region, Climate Zone
3. **Behavioral Data:** Daily Mileage, Vehicle Usage, Environmental Concerns, Brand Loyalty
4. **Infrastructure Data:** Charging Station Availability, Charging Station Types, Government Incentives
5. **Vehicle-Specific Data:** Battery Life, Warranty Period, Maintenance Costs, Customer Reviews

Additional ML Models

1. **Advanced Clustering Techniques:**
    - Gaussian Mixture Models (GMM): For flexible cluster shapes.
    - DBSCAN: For varying shapes and handling noise.
2. **Dimensionality Reduction Techniques:**
    - t-SNE: For high-dimensional data visualization.
3. **Ensemble Methods:**
    - Combine results from different clustering models for robust clustering.

Estimated Market Size for the EV Market in India

The estimated market size for the non-segmented EV market in India highlights a robust growth trajectory. With a market value projected to reach USD 15.1 billion by 2027 and annual sales expected to exceed 3 million units, the EV market in India presents significant opportunities for manufacturers, investors, and policymakers. This growth is underpinned by strong government support, increasing consumer awareness, and advancements in EV technology and infrastructure.

Top 4 Variables for Optimal Market Segmentation in the EV Market

1. **Vehicle Range (Km):**
    - Importance**:** Addresses range anxiety and commuting needs.
    - Usage**:** Segments by daily commuting, long-distance travel.

2. **Price:**
    - Importance**:** Differentiates income groups and budget preferences.
    - Usage**:** Segments luxury vs. budget-conscious buyers.

3. **Charging Infrastructure Availability:**
    - Importance**:** Influences feasibility and attractiveness of EV ownership.
    - Usage**:** Segments by regions with varying charging station access.

4. **Demographic Factors (Age, Income Level, Occupation):**
    - Importance**:** Affects purchasing decisions and usage patterns.
    - Usage**:** Identifies target customer profiles for personalized marketing.