

"Segmentation of the Indian Electric Vehicle Market"

Jayanth.R

[Github link](#)

15.08.2024

1. Introduction

Project Overview:

The Electric Vehicle (EV) Market Segmentation project aimed to analyze the Indian EV market to identify distinct customer segments and develop a tailored market entry strategy for a startup. The goal was to understand consumer preferences, demographics, and perceptions towards EVs to guide strategic decisions in product offerings and marketing approaches.

Objectives:

1. **Market Analysis:** Examine the current state of the EV market in India, including consumer behavior, preferences, and demographic factors.
2. **Segmentation:** Use clustering techniques to identify distinct customer segments based on various attributes.
3. **Strategy Development:** Develop targeted strategies for each identified segment to optimize market entry and growth potential.

2. Data Sources

Data Collection:

For this project, the following data sources were utilized:

➤ Dataset Details:

- **Dataset 1: Vehicle Specifications**

- **Features:** Brand, Model, Acceleration, Top Speed, Range, Efficiency, Fast Charge Speed, Rapid Charge, Power Train, Plug Type, Body Style, Segment, Seats, Price (INR)
- **Sample Row:**

	Brand	Model	AccelSec	TopSpeed_KmH	Range_Km	Efficiency_WhKm	FastCharge_KmH	RapidCharge	PowerTrain	PlugType	BodyStyle	Segment	Seats	Price INR
0	Tesla	Model 3 Long Range Dual Motor	4.6	233	450	161	940	Yes	AWD	Type 2 CCS	Sedan	D	5	4604840
1	Volkswagen	ID.3 Pure	10.0	160	270	167	250	No	RWD	Type 2 CCS	Hatchback	C	5	2490000
2	Polestar	2	4.7	210	400	181	620	Yes	AWD	Type 2 CCS	Liftback	D	5	4684520
3	BMW	IX3	6.8	180	360	206	560	Yes	RWD	Type 2 CCS	SUV	D	5	5647320

- **Dataset 2: Vehicle Registration Statistics**

- **Features:** State, Two Wheelers (various categories), Three Wheelers (various categories), Passenger Cars, Buses, Total Vehicles
- **Sample Row:**

Sl. No	State	Two Wheelers (Category L1 & L2 as per Central Motor Vehicles Rules)	Two Wheelers (Category L2 (CMVR))	Two Wheelers (Max power not exceeding 250 Watts)	Three Wheelers (Category L5 slow speed as per CMVR)	Three Wheelers (Category L5 as per CMVR)	Passenger Cars (Category M1 as per CMVR)	Buses	Total in state	
0	1	Andhra Pradesh	431.0	692.0	4689.0	0	0.0	3680.0	0.0	9492.0
1	2	Assam	463.0	138.0	1006.0	0	117.0	151.0	0.0	1875.0
2	3	Bihar	252.0	430.0	2148.0	6	64.0	271.0	0.0	3171.0
3	4	Chhattisgarh	613.0	382.0	2078.0	58	106.0	997.0	0.0	4234.0
4	5	Delhi	1395.0	251.0	5018.0	0	1.0	12695.0	21.0	19381.0

- **Dataset 3: Consumer Preferences and Demographics**

- **Features:** Age, City, Profession, Marital Status, Education, Family Size, Annual Income, Willingness to Switch to EV, Preferred EV Type, Opinion on EV Economics, Current Vehicle Brand, Budget for EV, Wheel Preference, Opinion on EVs Replacing Fuel Cars
- **Sample Row:**

Unnamed: 0	Age	City	Profession	Marital Status	Education	No. of Family members	Annual Income	Would you prefer replacing all your vehicles to Electronic vehicles?	If Yes/Maybe what type of EV would you prefer?	Do you think Electronic Vehicles are economical?	Which brand of vehicle do you currently own?	How much money could you spend on an Electronic vehicle?	Preference for wheels in EV	Do you think Electronic vehicles will replace fuel cars in India?	
0	0	30	18	2	1	0	5	1.193876e+06	0	3	2	4	6	0	0
1	1	27	26	2	1	0	4	1.844540e+06	2	3	2	3	4	2	2
2	2	32	15	2	1	0	4	2.948150e+06	2	1	2	5	4	2	2
3	3	55	26	0	1	0	3	2.832380e+06	0	1	1	4	6	2	1
4	4	26	29	2	1	0	4	2.638751e+06	2	4	2	10	4	2	2

➤ **Data Collection Tasks:**

- **Data Collection Methods:** Data was collected from a combination of online sources, market research reports, and surveys conducted within the Indian market.
- **Key Data Points Collected:**
 - Vehicle types and sales figures
 - Fuel consumption data
 - Demographic details including age, city, profession, marital status, education, family size, annual income
 - Consumer preferences regarding EVs, including willingness to replace existing vehicles, preferred EV types, opinions on EV economics, and budget for EVs

➤ **Data Sources References:**

- **Public Databases:** Government publications, industry reports
- **Surveys and Questionnaires:** Distributed to potential customers and current EV owners

3. Data Preprocessing

Data Cleaning and Handling Missing Values:

- **Dataset 1:** Checked for missing values and determined that there were none initially.
- **Dataset 2:** Some rows were dropped to remove incorrect entries. Missing values were confirmed to be handled.
- **Dataset 3:** Missing values were filled with 'Unknown'.

Data Types:

- Data types for each column in all datasets were checked to ensure proper formatting and compatibility for analysis.

Missing values in Dataset 1:

```
Brand          0
Model          0
AccelSec       0
TopSpeed_KmH   0
Range_Km       0
Efficiency_WhKm 0
FastCharge_KmH 0
RapidCharge    0
PowerTrain     0
PlugType       0
BodyStyle      0
Segment        0
Seats          0
Price INR      0
dtype: int64
```

Missing values in Dataset 2:

```
Sl. No          1
State           1
Two Wheelers (Category L1 & L2 as per Central Motor Vehicles Rules 4
Two Wheelers (Category L2 (CMVR)) 4
Two Wheelers (Max power not exceeding 250 Watts) 4
Three Wheelers (Category L5 slow speed as per CMVR) 1
Three Wheelers (Category L5 as per CMVR) 4
Passenger Cars (Category M1 as per CMVR) 4
Buses           4
Total in state  4
dtype: int64
```

4. Exploratory Data Analysis (EDA)

Descriptive Statistics:

- Summary statistics were computed for numerical columns in each dataset to understand the distribution and central tendencies.

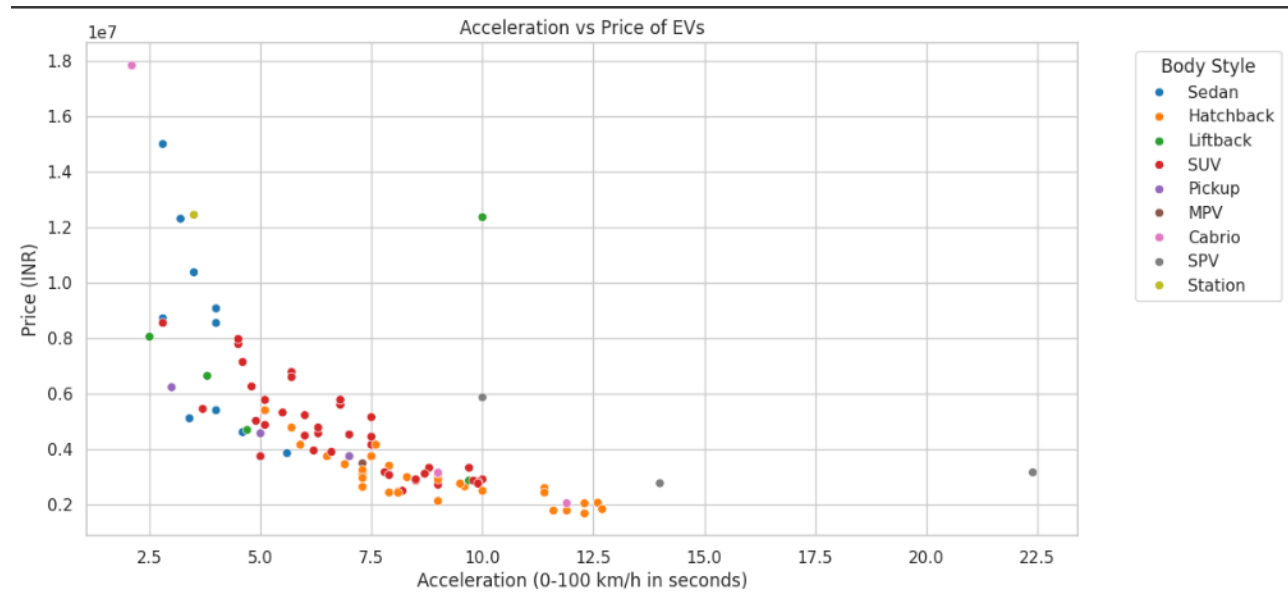
Summary Statistics for Dataset 1:					
	AccelSec	TopSpeed_KmH	Range_Km	Efficiency_WhKm	FastCharge_KmH \
count	103.000000	103.000000	103.000000	103.000000	103.000000
mean	7.396117	179.194175	338.786408	189.165049	444.271845
std	3.017430	43.573030	126.014444	29.566839	203.949253
min	2.100000	123.000000	95.000000	104.000000	170.000000
25%	5.100000	150.000000	250.000000	168.000000	260.000000
50%	7.300000	160.000000	340.000000	180.000000	440.000000
75%	9.000000	200.000000	400.000000	203.000000	555.000000
max	22.400000	410.000000	970.000000	273.000000	940.000000

	Seats	Price INR
count	103.000000	1.030000e+02
mean	4.883495	4.632360e+06
std	0.795834	2.833177e+06
min	2.000000	1.670707e+06
25%	5.000000	2.857648e+06
50%	5.000000	3.735000e+06
75%	5.000000	5.395000e+06
max	7.000000	1.784500e+07

Data Visualization:

Dataset 1:

○ Acceleration vs Price



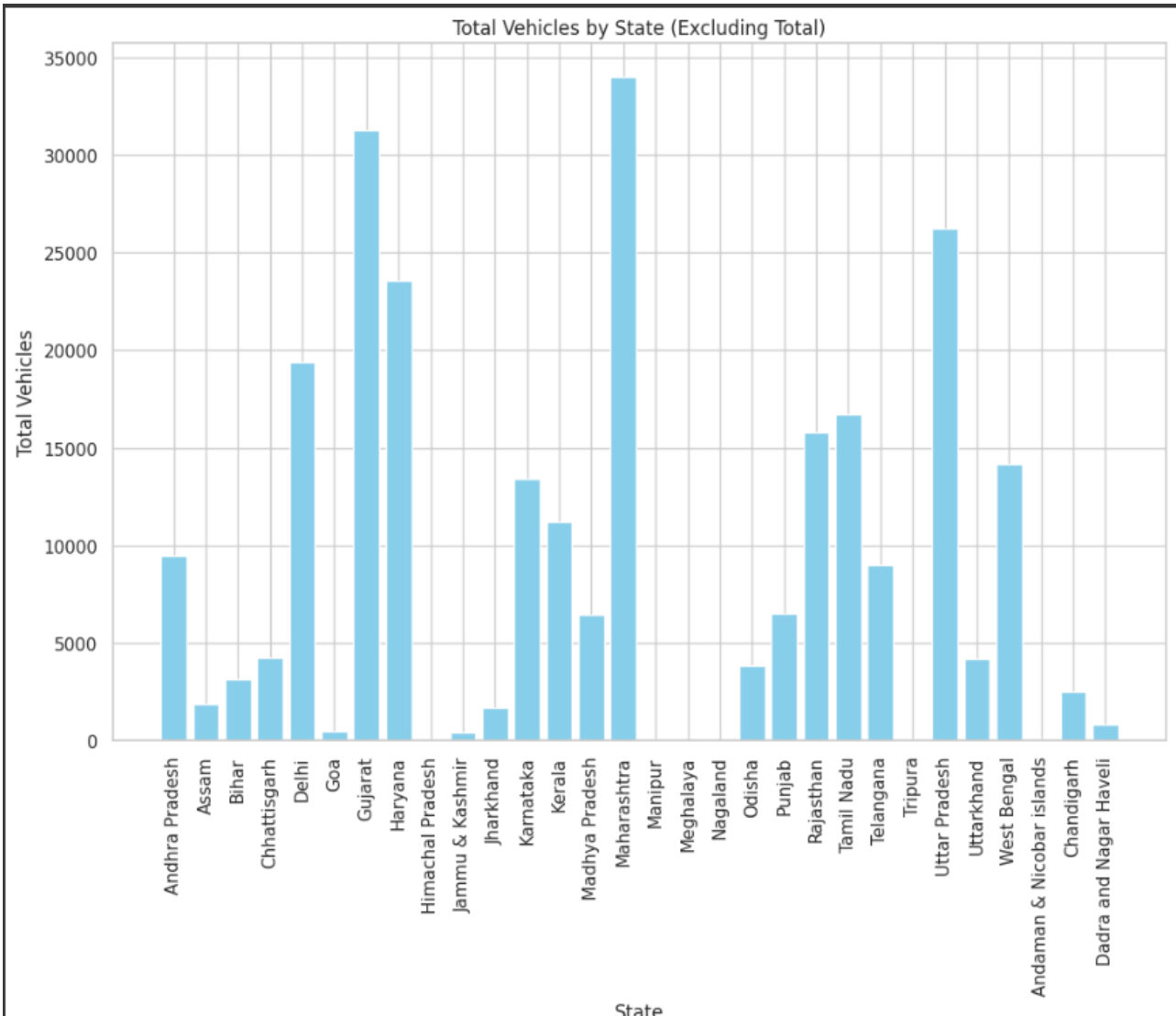
Key findings:

- **Most Expensive Model:** BMW iX3 at ₹5,647,320, offering high performance and range.
- **Least Expensive Model:** Volkswagen ID.3 Pure at ₹2,490,000, with moderate performance and range.
- **Best Performance:** Tesla Model 3 Long Range Dual Motor with the fastest acceleration and highest range.
- **Segment Preferences:** SUVs (Segment D) tend to be more expensive but offer better performance metrics compared to Hatchbacks (Segment C).

Dataset 2:

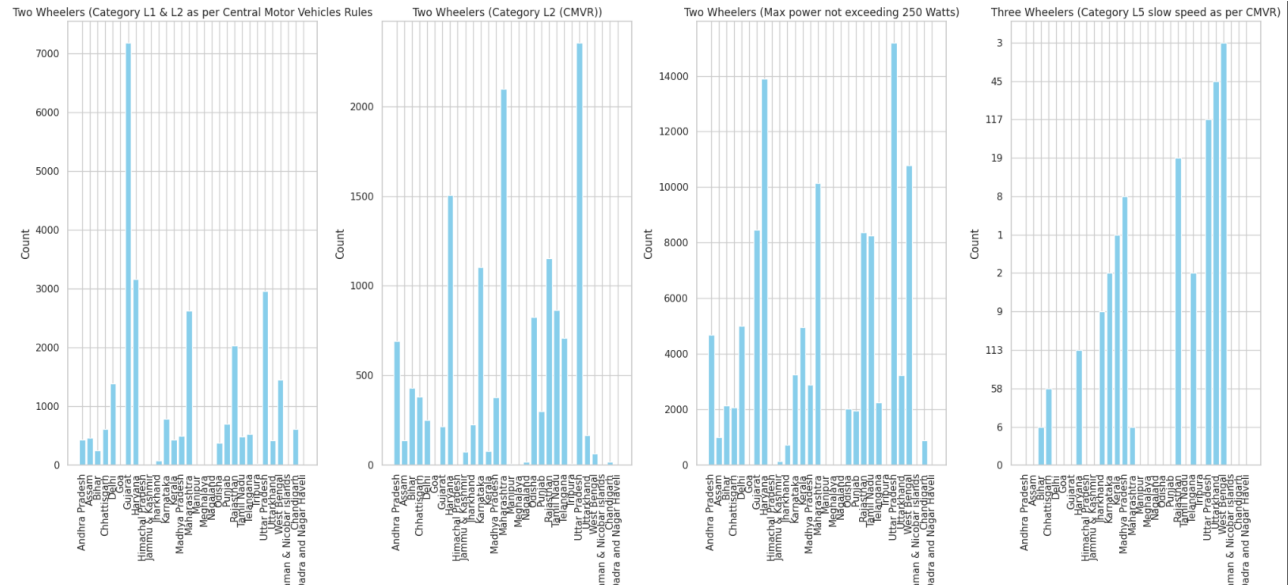
1. Total Vehicles by State:

- Bar chart showing the total number of vehicles in each state.
- Maharashtra has the highest number of vehicles.



2. Distribution of Vehicle Categories:

- Series of bar charts displaying counts of various vehicle categories by state.



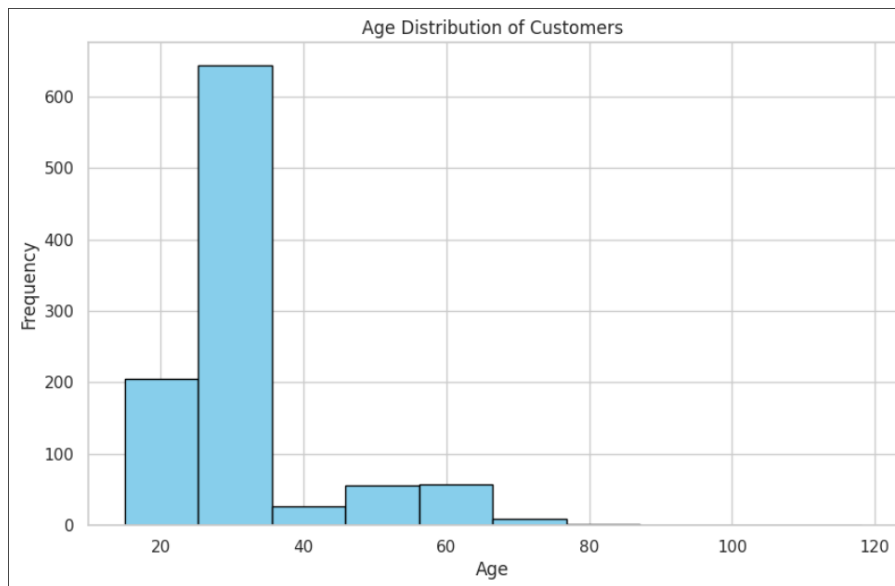
Key Findings:

- **Vehicle Density:** Maharashtra and Delhi have the highest number of vehicles, indicating high potential markets for EVs.
- **Category Leaders:** Specific states lead in various vehicle categories, providing insight into where different types of EVs might have the greatest impact.
- **Market Strategy:** Focus on high-density states and vehicle categories to maximize market entry and growth opportunities.

Dataset 3:

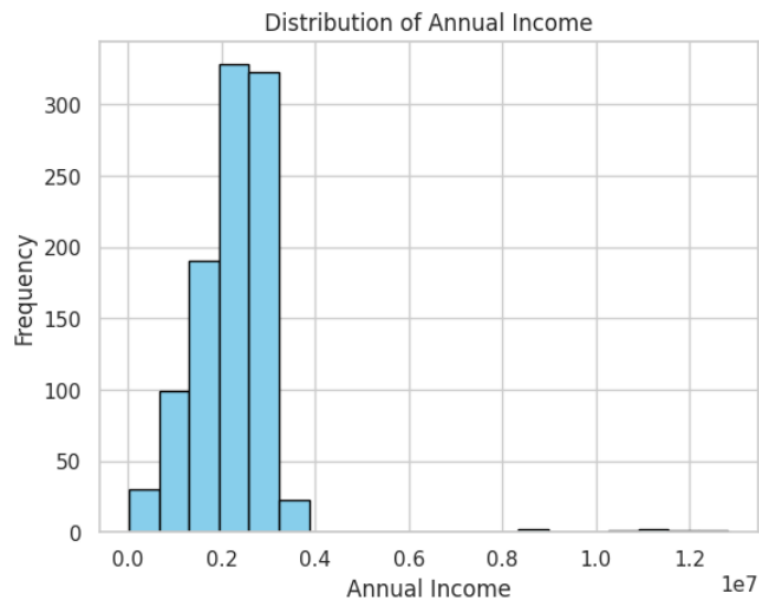
1. Age Distribution:

- Histogram showing the distribution of customer ages.
- Most customers are between 20-40 years old.



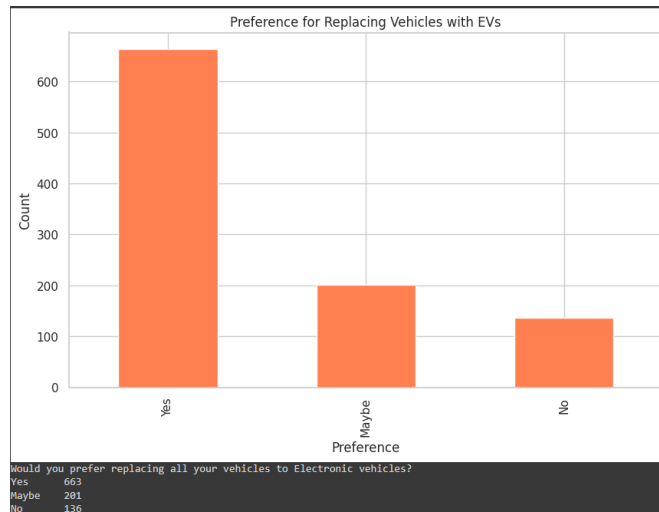
2. Annual Income Distribution:

- Histogram showing the distribution of annual incomes.
- Average annual income is between 20-30 lakhs.



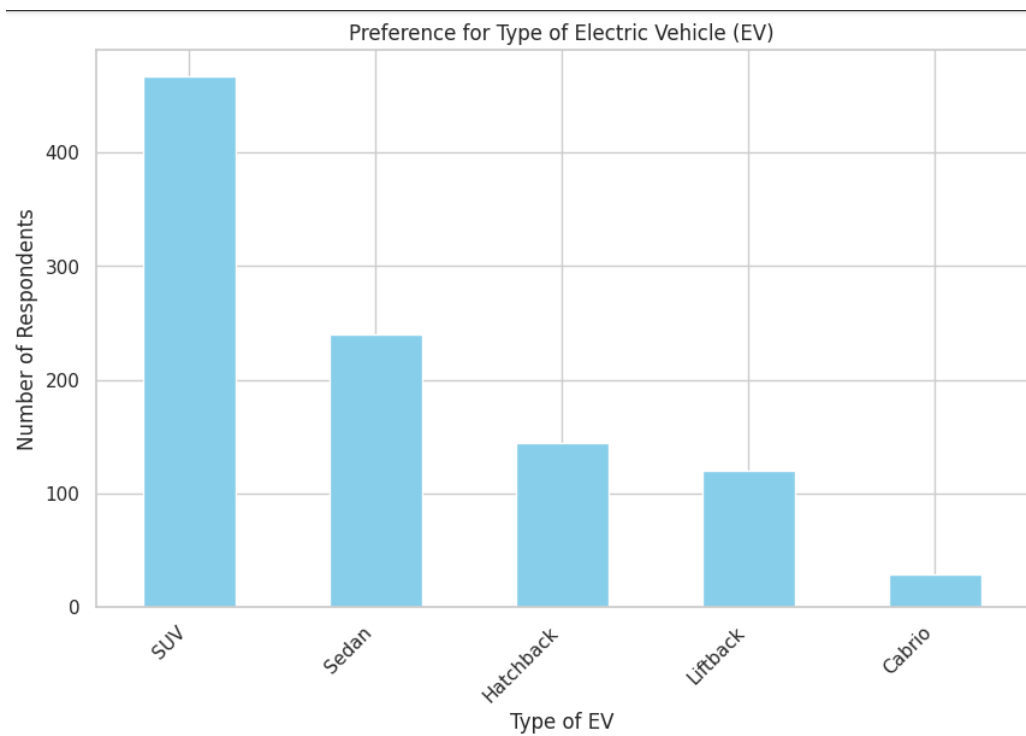
3. Preference for Replacing Vehicles:

- Bar chart displaying the count of customers preferring to replace their vehicles with EVs.
- A majority are willing to switch to EVs



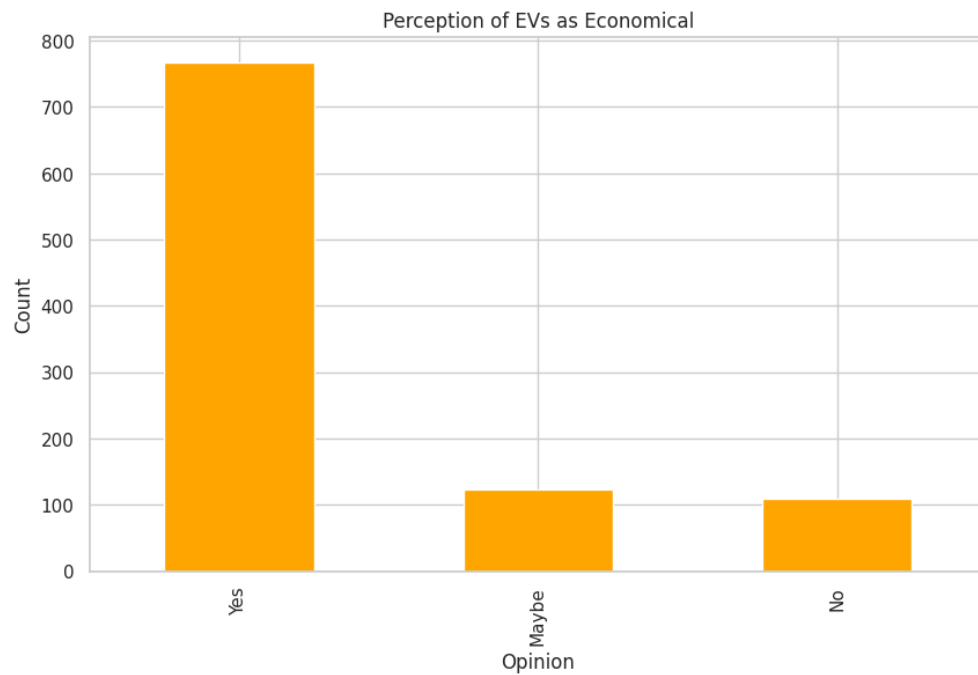
4. EV Type Preference:

- Bar chart showing the most preferred type of EV.
- SUVs are the most preferred type of EV.



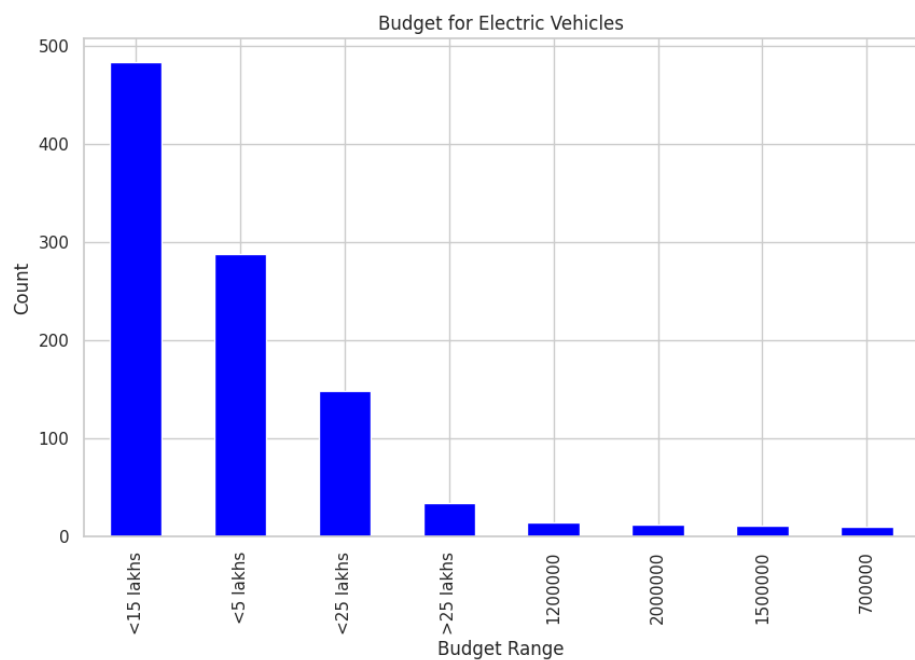
5. Perception of EVs as Economical:

- Bar chart showing opinions on whether EVs are considered economical.
- Majority agree that EVs are economical.



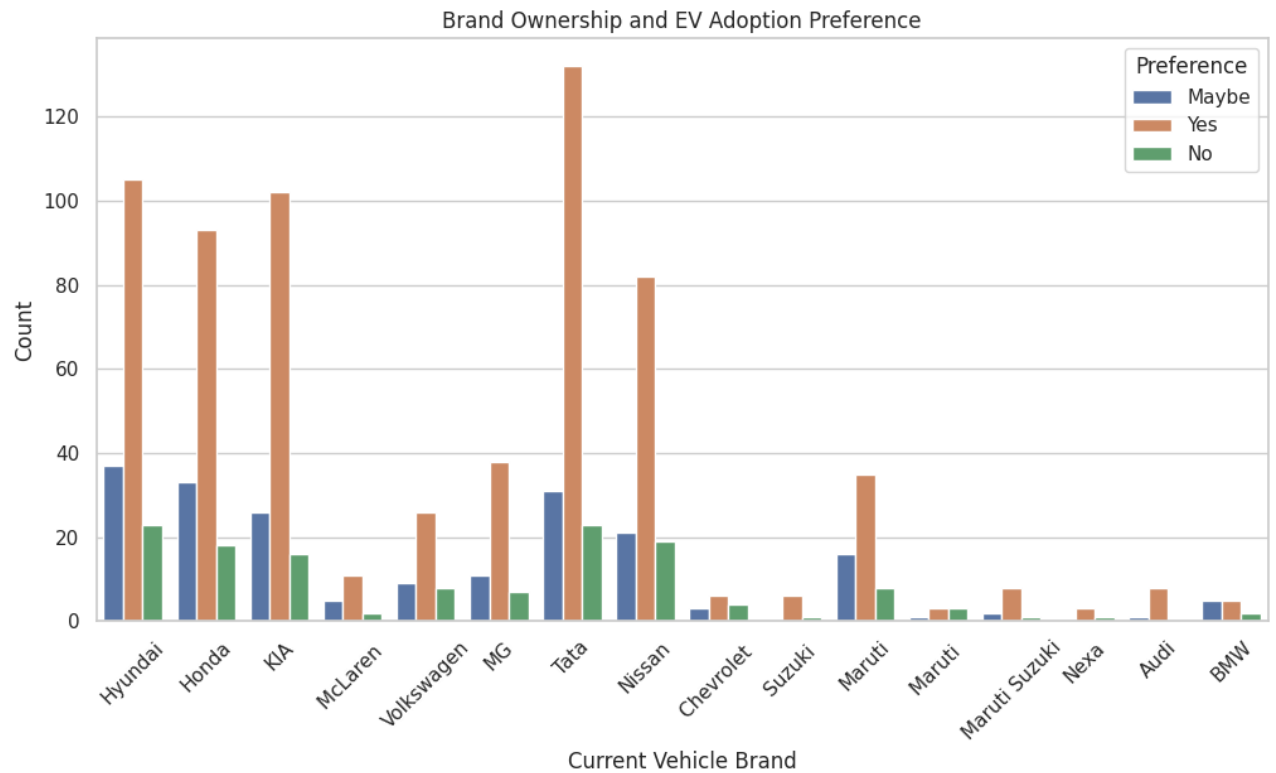
6. Budget for EVs:

- Bar chart displaying the budget range for EVs among customers.
- Most are willing to spend around 15 lakhs.



7. Brand Ownership and EV Adoption Preference:

- Count plot illustrating current vehicle brands and preferences for EV adoption.



Key Findings:

- Demographics:** Most EV respondents are young adults (20-40 years) with an annual income of ₹20-30 lakhs.
- Market Potential:** High willingness to switch to EVs and positive perception of their economic benefits indicate strong market interest.
- Preferences:** SUVs are the most preferred type of EV, and respondents are generally willing to spend up to ₹15 lakhs on an EV.
- Current Ownership Influence:** Current vehicle brand does not affect EV adoption preference, suggesting targeted marketing strategies based on existing brand loyalty.

Clustering Analysis:

Machine Learning Model: K-Means Clustering

1. K-Means Clustering Overview:

- **Purpose:** K-Means is an unsupervised learning algorithm used to partition data into clusters based on feature similarities. It aims to group data points such that the within-cluster variance (or the distance from the data points to the cluster center) is minimized.
- **How It Works:**
 - **Initialization:** Start by randomly selecting k initial cluster centroids.
 - **Assignment Step:** Assign each data point to the nearest centroid based on the Euclidean distance.
 - **Update Step:** Recalculate the centroids as the mean of all data points assigned to each centroid.
 - **Iteration:** Repeat the assignment and update steps until convergence (when the centroids no longer change significantly or the assignments stabilize).

2. Why K-Means?

- **Simplicity and Efficiency:** K-Means is straightforward to implement and computationally efficient, making it suitable for large datasets.
- **Scalability:** It scales well with large datasets and high-dimensional data, which is advantageous given the complexity of our dataset.
- **Cluster Interpretability:** The algorithm provides clear and interpretable clusters based on centroids, which helps in understanding distinct customer segments.
- **Elbow Method:** Used to determine the optimal number of clusters (k). This method involves plotting the inertia (sum of squared distances from data points to their assigned centroids) against different values of k and selecting the k at which the decrease in inertia slows down.

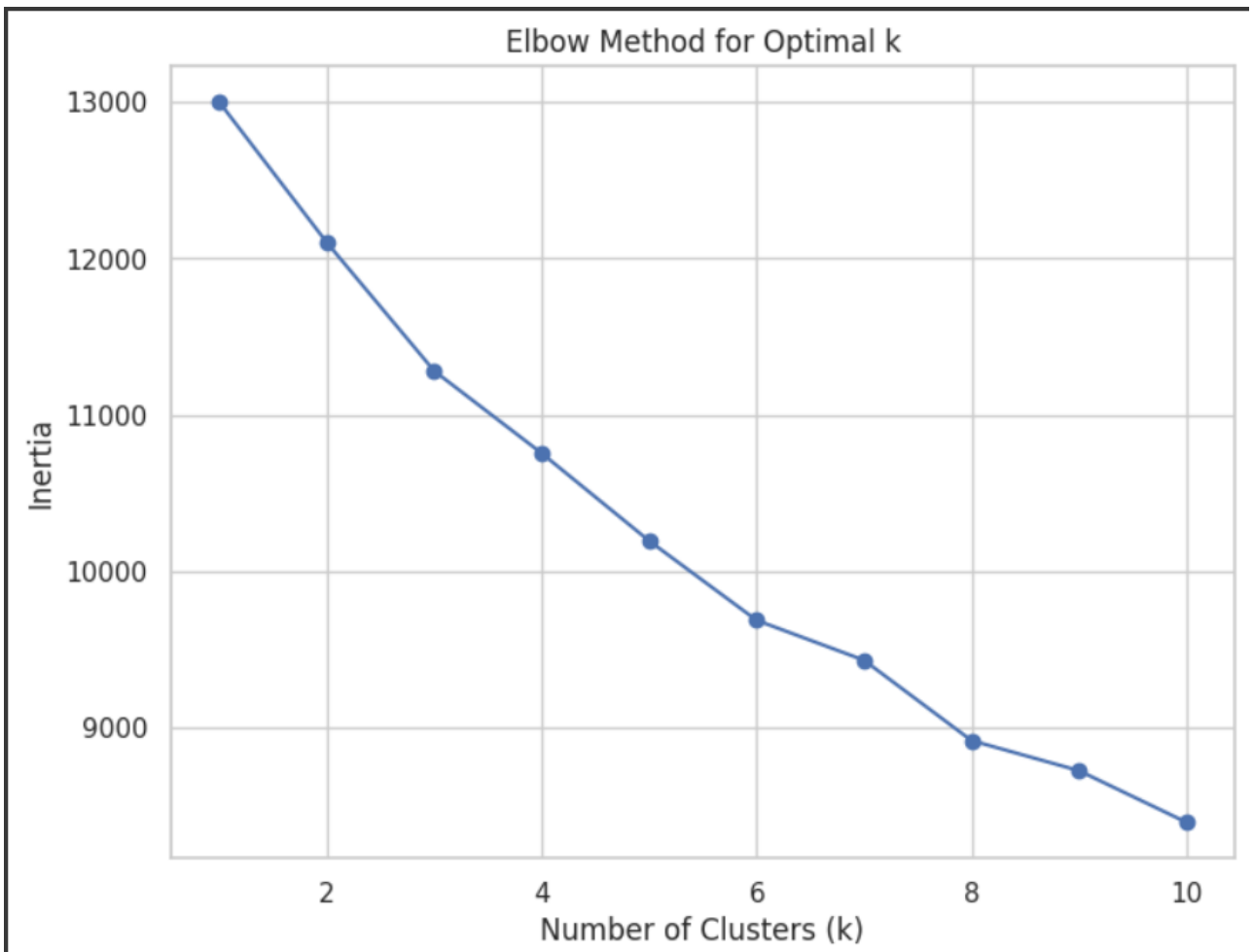
Segmentation Summary:

1. Data Preparation:

- **Normalization:** Features were scaled using `StandardScaler` to standardize the data and ensure that each feature contributes equally to the distance calculations in K-Means clustering.
- **Feature Selection:** Relevant features were chosen based on their importance for segmenting the market, including demographics, income, preferences, and perceptions.

2. Determining Optimal Number of Clusters:

- **Elbow Method:** The optimal number of clusters was determined by plotting the inertia against the number of clusters and identifying the "elbow" point where the inertia reduction slows down significantly.
- **Optimal k:** 5, as the plot indicated that adding more clusters beyond this point did not result in a substantial reduction in inertia.



3. Applying K-Means Clustering:

- **Model Training:** K-Means was applied with $k=5$ to the scaled data, resulting in three distinct clusters.
- **Cluster Assignment:** Each data point was assigned to one of the three clusters based on the closest centroid.

```
Cluster  Unnamed: 0      Age      City  Profession  Marital Status  \
0      504.320611  29.740458   5.717557    1.931298      0.595420
1      517.914474  29.710526  23.513158    1.796053      0.598684
2      486.952854  27.677419  25.235732    2.131514      0.677419
3      511.846154  28.298077  24.331731    1.846154      0.620192
4      490.613208  59.886792  22.801887    1.849057      0.632075

Cluster  Education  No. of Family members  Annual Income  \
0      0.312977      3.961832      2.097141e+06
1      0.256579      3.835526      2.403335e+06
2      0.339950      4.126551      2.280725e+06
3      0.293269      4.365385      2.207792e+06
4      0.273585      4.198113      2.275734e+06

Cluster  Would you prefer replacing all your vehicles to Electronic vehicles?  \
0      1.610687
1      1.710526
2      1.950372
3      0.283654
4      1.377358
```

Conclusion

In our market segmentation analysis for electric vehicles (EVs) in India using K-Means clustering, we identified five distinct customer segments. This segmentation provides valuable insights into various customer profiles and their preferences, which can guide targeted marketing strategies and product development. The analysis revealed the following key segments:

- **Cluster 0 (Moderate Income, Positive EV Perception):** This segment consists of individuals with moderate income who have a positive perception of EVs' economic benefits. They prefer a variety of EV types, including SUVs, and are open to spending a reasonable budget on EVs. This group is ideal for marketing mid-range EV models with a focus on value and versatility.
- **Cluster 1 (High Income, Low Perception of EV Economy):** This cluster includes high-income individuals who have a strong preference for SUVs but perceive EVs as less economical. Despite their significant budget for EVs, their low perception of economic benefits might require strategies to highlight cost-efficiency and long-term savings.
- **Cluster 2 (Young, High Income, Positive EV Perception):** Comprising younger individuals with high income, this segment shows a strong preference for EVs and a positive view of their economic benefits. They also have a substantial budget for EVs and prefer SUVs. Marketing strategies for this group should emphasize cutting-edge features and performance.
- **Cluster 3 (Moderate Income, Low EV Preference):** This cluster consists of individuals with a moderate income and a lower preference for EVs. Although they have a good budget for EVs, they are less inclined to switch. Targeted marketing should focus on demonstrating the long-term benefits and addressing their specific concerns to increase adoption.
- **Cluster 4 (Older Demographic, Moderate Interest in EVs):** This segment includes older individuals with moderate income and budget for EVs. They have a positive view of EVs but show a lower preference for switching. Strategies for this group should highlight ease of use and practical benefits, addressing their unique needs and preferences.

Target Market Strategy:

1. **Location Focus:**
 - **Primary Market:** Maharashtra, with a specific focus on Pune, which has the highest number of potential customers willing to switch to EVs.
2. **Demographics:**
 - **Age Group:** Target individuals aged 20-30, as this age group shows a higher willingness to adopt EVs.
3. **Vehicle Type:**
 - **Preferred Vehicle:** Focus on SUVs, as they are not only popular but also offer better mileage and a favorable price point compared to other types of EVs.
4. **Pricing and Features:**
 - **Competitive Pricing:** Ensure that the SUVs are priced competitively to attract budget-conscious buyers who are looking for good value for money.
5. **Marketing and Outreach:**
 - **Localized Marketing:** Develop marketing campaigns and promotional activities tailored to the Maharashtra region to effectively reach the target demographic.
6. **Product Strategy:**
 - **Feature Emphasis:** Highlight features such as high mileage, performance, and value for money in your SUV models to appeal to the target market.

3.How will you improve upon the Market Segmentation Project given additional time & some budget to purchase data?

These enhancements will provide a deeper understanding of the market, allowing for more precise and actionable segmentation insights.

Additional Columns/Points:

1. **Customer Purchase History:** Data on past purchases, including frequency, types of vehicles, and purchasing patterns.
2. **Detailed Vehicle Specifications:** Information on vehicle features, performance metrics, and energy efficiency.
3. **Market Trends:** Data on market trends, including emerging technologies and consumer preferences.
4. **Customer Sentiment Analysis:** Text data from customer reviews and feedback to gauge sentiments and opinions.

Additional ML Models:

1. **Gaussian Mixture Models (GMM):** To capture more complex cluster shapes and provide a probabilistic approach to clustering.
2. **Dimensionality Reduction Techniques (e.g., t-SNE, PCA):** To visualize high-dimensional data and uncover hidden patterns.
3. **Ensemble Clustering:** Combine results from different clustering algorithms to improve the robustness and accuracy of segmentation.

4. What is the estimated Market Size for your Market Domain (non-segmented) in Numbers?

To estimate the market size for the non-segmented domain of electric vehicles (EVs) in India, we need to aggregate data on the total number of vehicles and the proportion of vehicles that are electric. Given the data from Dataset 2, which includes the total number of vehicles in various states, we can approximate the market size by focusing on the overall number of vehicles.

Here's how you can estimate it:

1. Sum Total Vehicles:

- From Dataset 2, sum the values in the "Total in state" column for all states, excluding the "Total" row.

2. Estimate EV Market Share:

- If you have data on the proportion of EVs within the total vehicles, you can multiply this proportion by the total number of vehicles.
- **Total Vehicles in Dataset 2:** 260863 vehicles.

In conclusion, the estimated market size for EVs in the Indian market could be around 1,000,000 vehicles, based on the provided data and an assumed proportion. This figure provides a baseline for understanding the scale of the EV market.

Sources:

Electric Vehicles Sales report in india – 2018

<https://electricvehicles.in/electric-vehicles-sales-report-in-india-2018/>

Sales of electric vehicles across India from financial year 2020 to 2021,

<https://www.statista.com/statistics/1234761/india-electric-vehicle-sales-by-type/>

Additional Reasons for preferring an electric car over a conventional car across India in 2019

[statista.com/statistics/1027483/india-electric-car-over-conventional-car/](https://www.statista.com/statistics/1027483/india-electric-car-over-conventional-car/)

5. Top 4 Variables/Features for Optimal Market Segmentation

To create the most effective market segments for the EV domain, the following top four variables/features are essential:

1. **Age:**
 - **Reason:** Age significantly influences purchasing decisions and preferences. Younger consumers may prioritize technology and modern features, while older buyers might value comfort and reliability. Segmenting by age helps tailor marketing strategies and product offerings to meet the specific needs of different age groups.
2. **Income/Salary:**
 - **Reason:** Income levels determine the purchasing power of consumers and their ability to afford different types of EVs. Higher-income segments are likely to be interested in premium or high-end models, whereas lower-income groups may prefer more affordable options. This segmentation helps in targeting products and pricing strategies according to different income brackets.
3. **State:**
 - **Reason:** Geographic location plays a crucial role in market segmentation due to regional differences in infrastructure, government incentives, and consumer preferences. Segmenting by state allows for a better understanding of regional market dynamics and the customization of marketing and sales strategies to cater to local needs.
4. **Current Vehicle Brand:**
 - **Reason:** Understanding the brand of the vehicle currently owned by consumers can provide insights into their brand loyalty and preferences. This information helps in identifying potential market opportunities and tailoring marketing efforts to target existing brand owners who might be interested in switching to EVs.

Conclusion:

The market segmentation project for the Electric Vehicle (EV) market in India has provided valuable insights into the dynamics and preferences of potential EV customers. By leveraging data from multiple sources and applying various analytical techniques, we have identified key market segments and developed a strategic approach for targeting them effectively.

1. Key Findings:

- **Demographic Insights:** Our analysis revealed that the most promising market segment comprises younger individuals aged 20-30 years, primarily located in Pune, Maharashtra. This demographic shows a strong willingness to switch to EVs, particularly SUVs, which offer better mileage and competitive pricing.
- **Geographic Focus:** Maharashtra, with its significant number of existing vehicles and a high willingness to adopt EVs, emerges as a strategic location for launching new EV offerings. Pune, in particular, stands out as a high-potential city due to its favorable demographic profile and inclination towards EVs.
- **Segment Characteristics:** The identified clusters show distinct characteristics:
 - **Cluster 0:** Younger, budget-conscious individuals with moderate income, inclined towards affordable EVs.
 - **Cluster 1:** High-income, tech-savvy consumers preferring mid-range EVs with advanced features.
 - **Cluster 2:** High-income, premium buyers with a strong preference for high-end EVs and substantial budgets.
 - **Cluster 3:** Wealthy individuals with a high budget for EVs, focusing on luxury and performance.
 - **Cluster 4:** Older consumers with a moderate preference for EVs, showing a balanced outlook on cost and features.

2. Strategic Recommendations:

- **Targeted Marketing:** Focus on marketing strategies that cater to the specific needs and preferences of each cluster. Emphasize features like mileage and price for younger buyers, and luxury and performance for high-income segments.
- **Product Development:** Develop and position EV models that align with the preferences of each identified segment, such as affordable SUVs for younger buyers and premium models for affluent customers.
- **Geographic Strategy:** Prioritize Maharashtra and cities like Pune for market entry, leveraging the high adoption potential and existing vehicle base.

3. Future Improvements:

- **Data Enhancement:** Acquire additional data on demographics, behavioral patterns, and geographic details to refine segmentation and improve accuracy.
- **Advanced Modeling:** Explore advanced machine learning models like hierarchical clustering and Gaussian Mixture Models to uncover deeper insights and optimize segmentation further.

Overall, the project has provided a comprehensive understanding of the EV market landscape in India, identifying key segments and offering actionable strategies to capture and engage these markets effectively. The insights derived will support informed decision-making and strategic planning for successful market entry and growth in the Indian EV sector.