



EV MARKET SEGMENTATION

Report Work Submitted By:

Rohan Kharwar

(Machine Learning Intern, Feynn Labs)

29th July

PROBLEM STATEMENT

Question.

Based on market analysis, the segmentation challenge can be summarized into two key questions:

1. What kind of electric vehicle (EV) will the company manufacture?
2. Who is the target customer?

In this context, the focus is specifically on the 4-wheeler EV market, which is considerably larger and more dynamic compared to others. Therefore, the objective is to determine which types of 4-wheelers EVs a company should allocate its resources towards.

Approach.

The objective is to analyze the electric vehicles (EV) market in India using segmentation analysis and develop a viable strategy for entering the market, focusing on the segments most likely to adopt their product. The analysis will consider geographic, demographic, psychographic, and behavioral factors.

In this report, I have examined the four-wheeler electric vehicles market in India using segmentation analysis to address some of the critical questions. Additionally, I have performed segmentation on customers and their income data. The segmentation was conducted using Principal Component Analysis (PCA) and K-Means clustering, with hierarchical clustering also demonstrated. Finally, the potential segments for an EV company to target are evaluated based on selected features.

GitHub Link to the Analysis:

<https://github.com/Rohan1305-hub/rohankharwarfynnlab.git>

DATA COLLECTION

Demographic dataset and EV bikes dataset are collected from <https://www.kaggle.com/>.

The bikes dataset needed to be modified to add some missing values like products from popular brands.

PSYCHOLOGICAL AND BEHAVIOURAL SEGMENTATION

In this section, I outline the methods used to perform demographic segmentation on the dataset. Initially, an Exploratory Data Analysis (EDA) is conducted. EDA is a comprehensive examination that reveals the underlying structure of a dataset, making it crucial for a company. It helps uncover trends, patterns, and relationships that are not immediately obvious.

Exploratory Data Analysis (EDA)

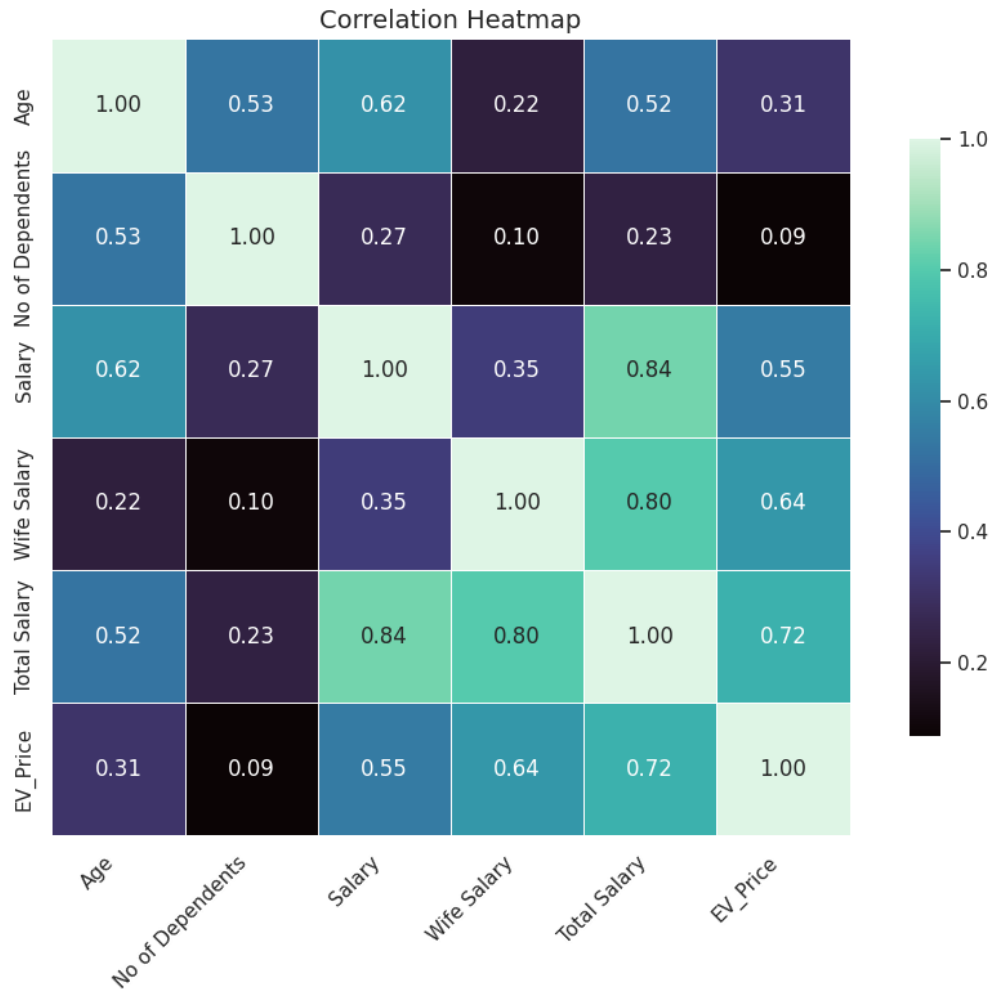
The datasets: <https://www.kaggle.com/>

	Name	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	House Loan	Wife Working	Salary	Wife Salary	Total Salary	Make	Price
0	Aarav	30	Salaried	Single	Post Graduate	0	Yes	No	No	800000	0	800000	i20	800000
1	Vivaan	40	Salaried	Married	Post Graduate	2	Yes	Yes	Yes	1400000	600000	2000000	Ciaz	1000000
2	Vihaan	48	Business	Married	Graduate	4	Yes	Yes	No	1800000	0	1800000	Duster	1200000
3	Kiara	41	Business	Married	Post Graduate	3	No	No	Yes	1600000	600000	2200000	City	1200000
4	Kiara	32	Salaried	Married	Post Graduate	2	Yes	No	Yes	1800000	800000	2600000	SUV	1600000

The analysis of the dataset, with columns such as Age, No. of dependence, salary, Wife salary, total salary and price indicates that consumers are generally satisfied with all attributes. However, the reviews suggest that consumers are particularly satisfied with the appearance of the EVs compared to the value for money they provide. This highlights the importance of aesthetic appeal in influencing consumer satisfaction within the electric vehicle market.

	Age	No of Dependents	Salary	Wife Salary	Total Salary	Price
count	99.000000	99.000000	9.900000e+01	9.900000e+01	9.900000e+01	9.900000e+01
mean	36.636364	2.181818	1.736364e+06	5.343434e+05	2.270707e+06	1.194040e+06
std	7.354652	1.335265	6.736217e+05	6.054450e+05	1.050777e+06	4.376955e+05
min	22.000000	0.000000	2.000000e+05	0.000000e+00	2.000000e+05	1.100000e+05
25%	31.000000	2.000000	1.300000e+06	0.000000e+00	1.550000e+06	8.000000e+05
50%	37.000000	2.000000	1.600000e+06	5.000000e+05	2.100000e+06	1.200000e+06
75%	40.000000	3.000000	2.200000e+06	9.000000e+05	2.700000e+06	1.500000e+06
max	56.000000	4.000000	3.800000e+06	2.100000e+06	5.200000e+06	3.000000e+06

A correlation matrix is generated to identify how numerical features relate to one another. A heatmap is then plotted using Seaborn's heatmap function. Strong correlations (positive or negative) are visually evident through color intensity. For instance, a high correlation between total salary and EV price would suggest that income strongly influences purchase behavior. This step aids in selecting impactful features for clustering and eliminating redundant variables..

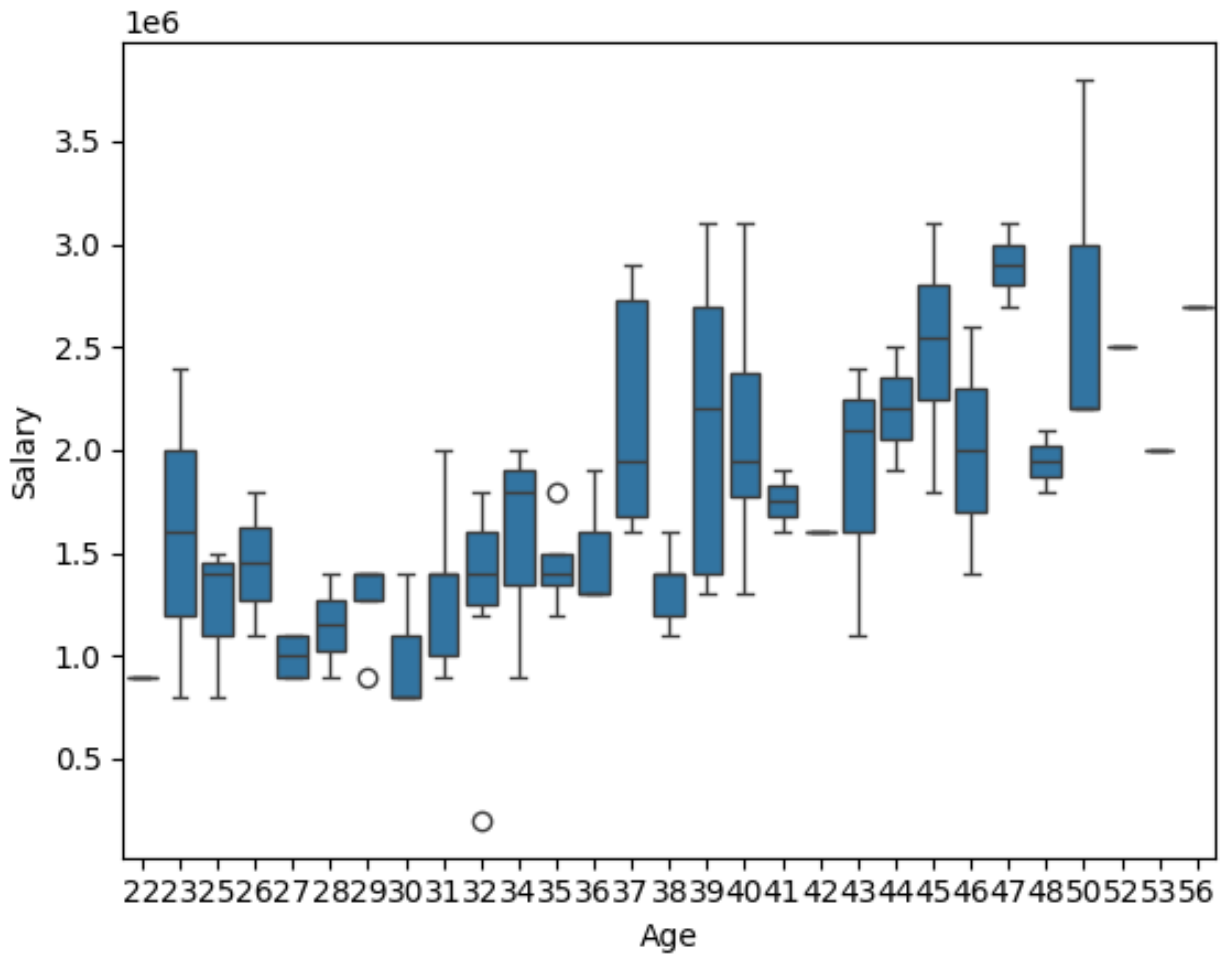


The Kernel Density Estimate (KDE) plot illustrates that there is a high correlation between the appearance and comfort of electric vehicles. It also shows that consumer ratings are not correlated with the vehicle's fuel economy. The plot highlights that vehicles perceived as offering better value for money tend to receive higher ratings. Moreover, it underscores the significant roles that performance and value for money play in consumer satisfaction.

The boxplots uses to examine how user demographics like age, dependents, profession, education, and salary relate to their preferred EV make. By analyzing these visual comparisons, the goal is to find patterns such as which age group prefers which EV model, or whether higher-earning individuals are more inclined toward specific brands. It deepens the psychographic and behavioral understanding necessary for customer segmentation.

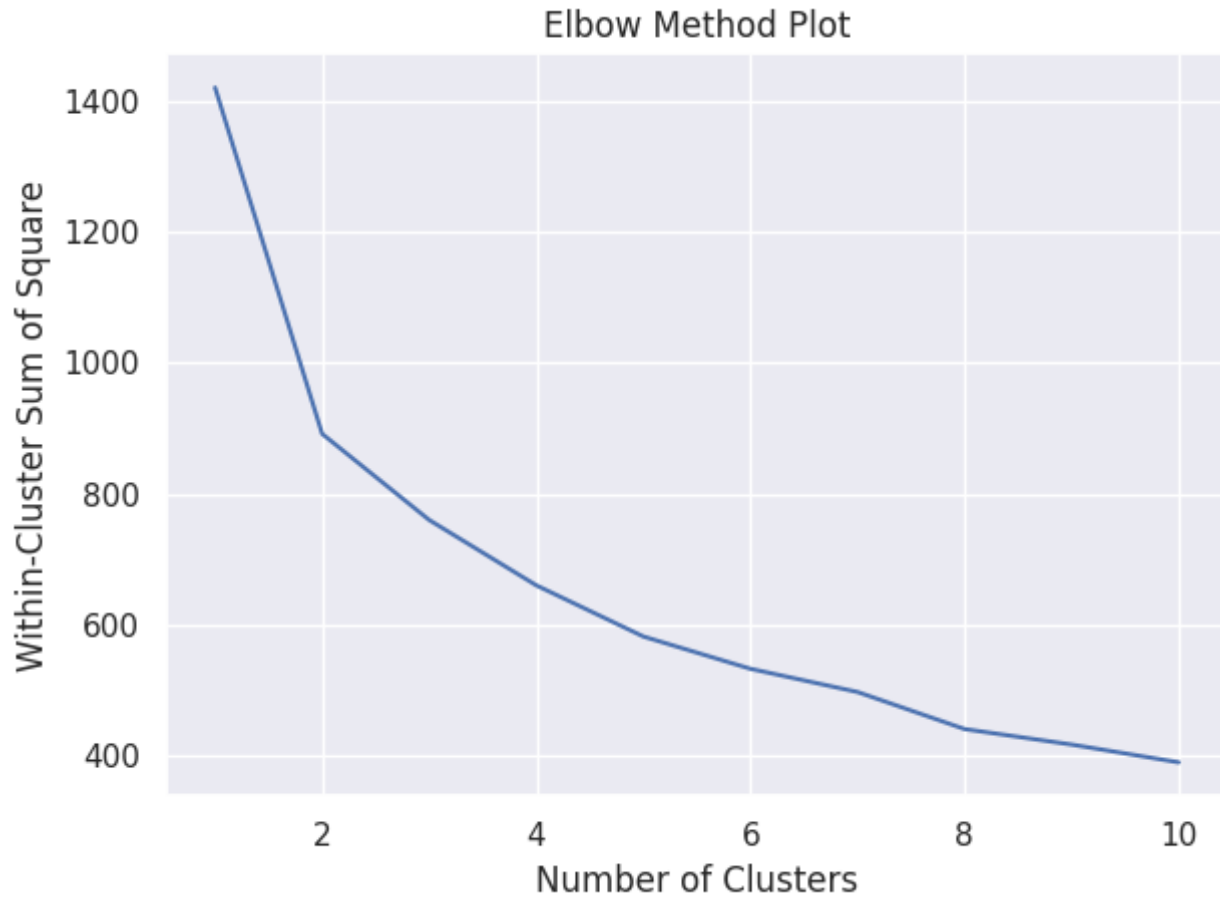
A data to show percentage of candidates having different electric vehicle options given by the company.

The findings reveal that the EV has the largest market presence and is the best-performing vehicle, achieving the highest average ratings overall. In contrast, the Tata Tigar EV has the lowest market presence and the poorest consumer satisfaction. The Hyundai Kona, while competitive, has a lower market presence.



K-Means Clustering

I performed K-means clustering for various cluster counts and plotted the Elbow curve to determine the optimal number of clusters, as the algorithm requires this input. It's worth noting that I used the "k-means++" initialization method, which smartly initializes clusters rather than doing so randomly as in the standard K-means.



Segment Extraction

To identify distinct customer segments in the Indian electric vehicle (EV) market, this report employed unsupervised machine learning techniques—primarily **K-Means Clustering**, supplemented by **Hierarchical Clustering** and **Principal Component Analysis (PCA)**. These techniques enabled a structured, data-driven approach to uncover hidden patterns in consumer behavior and preferences.

1. K-Means Clustering

Purpose: To partition the dataset into groups where each data point belongs to the cluster with the nearest mean.

Preprocessing: The dataset was standardized using `StandardScaler` to ensure all numerical features were on the same scale.

Optimal K Selection: The *Elbow Method* was used to determine the ideal number of clusters (**K=3**), balancing intra-cluster similarity and inter-cluster separation.

Result: Three meaningful customer segments were extracted, each with unique behavioral and demographic traits influencing their EV preferences.

2. Hierarchical Clustering

Purpose: To visually validate the structure and depth of customer groupings without pre-defining the number of clusters.

Method Used: Agglomerative clustering with dendrogram visualization.

Insight: Confirmed the presence of two to three distinct clusters, supporting the K-Means results and enhancing segmentation confidence.

3. Principal Component Analysis (PCA)

Purpose: Dimensionality reduction to simplify complex multivariate relationships and enable better cluster visualization.

Application: PCA helped visualize the spread and separation of clusters in a 2D space, validating the integrity of K-Means clusters and identifying outliers or overlaps.

Outcome: Enhanced interpretability of segments by focusing on the most influential features, such as *total salary*, *car price*, and *satisfaction parameters*.

By combining K-Means, Hierarchical Clustering, and PCA, the segmentation process yielded robust and interpretable clusters that align with real-world customer behavior. These machine learning techniques ensured that the identified segments were not only statistically valid but also strategically actionable—providing a solid foundation for market targeting and product development in India's EV landscape.

DEMOGRAPHIC SEGMENTATION

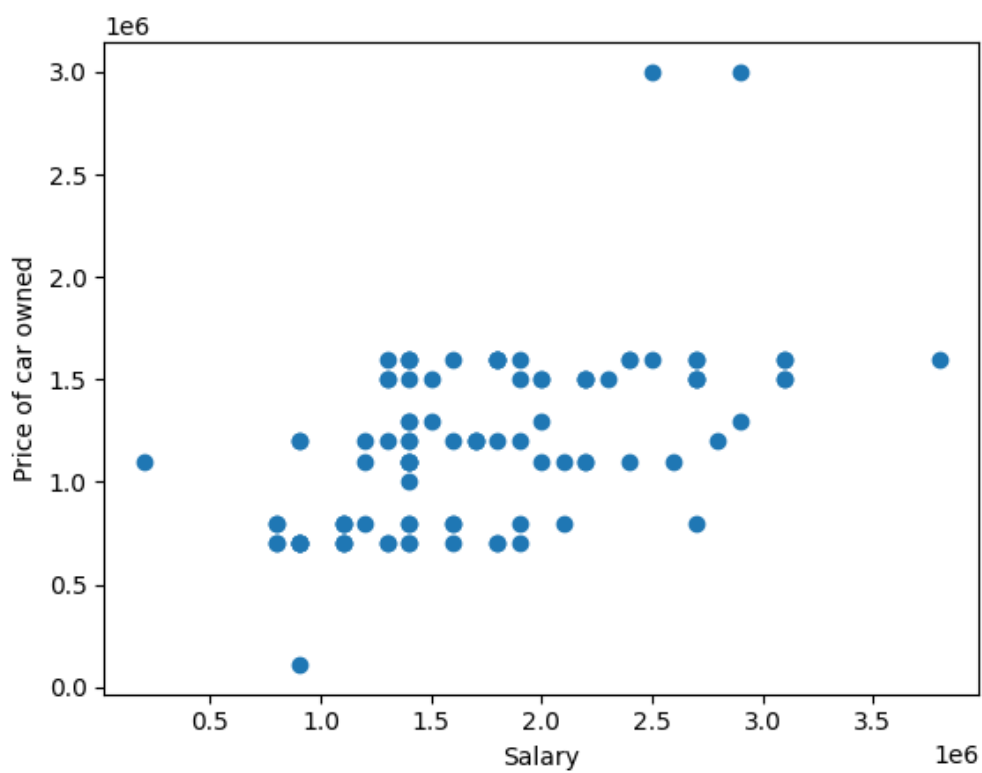
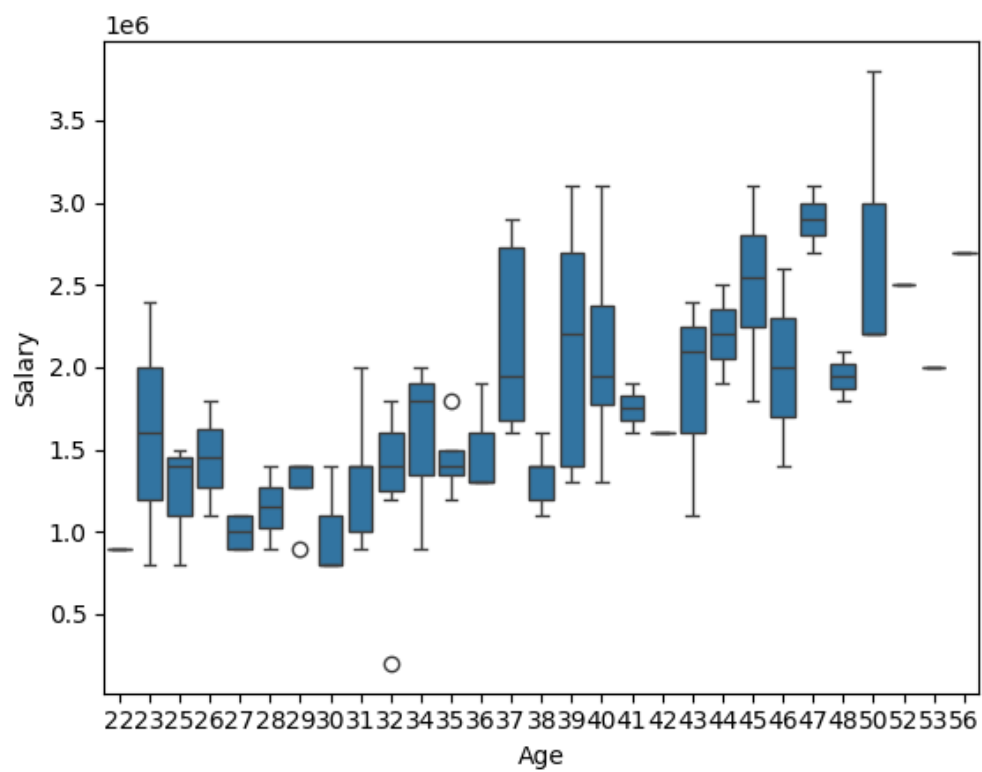
Loan:

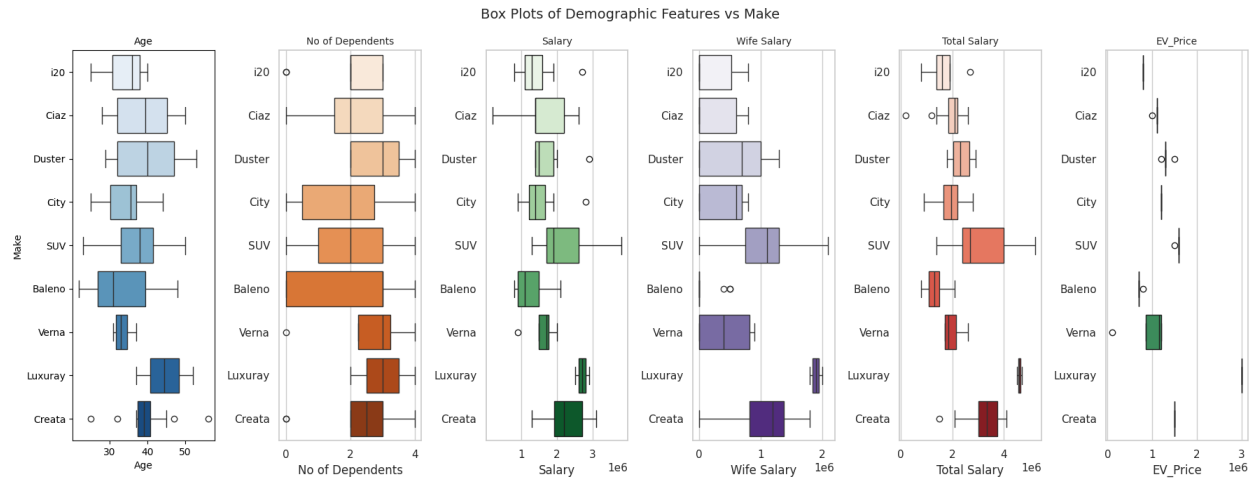
Although Personal and home loans do not seem to have any significant impact on EV purchasing patterns.

Salary:

Regardless of even high salaries, average to low price-ranged cars are preferred.

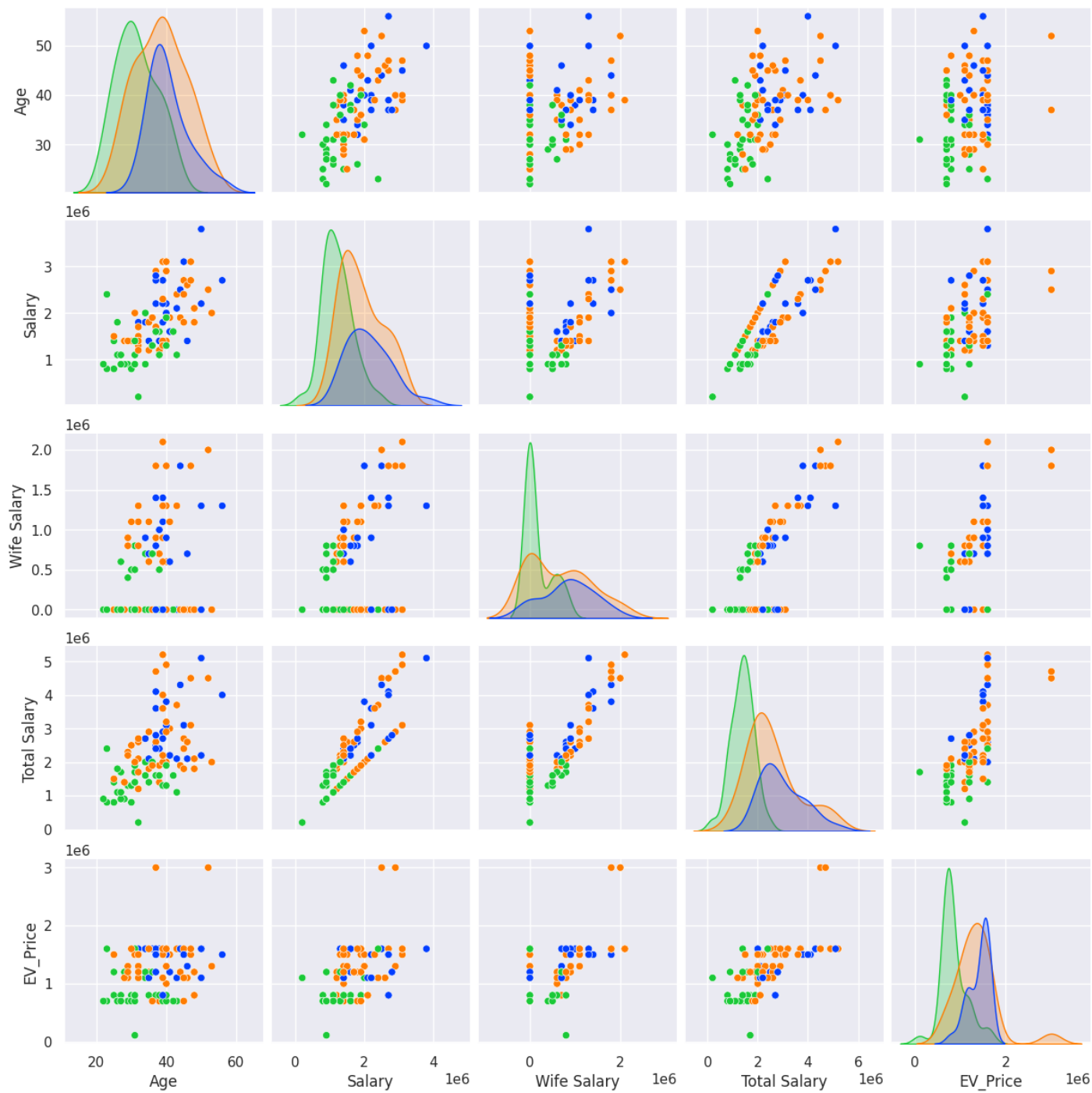
We see people buying over a range of prices when it comes to Verna, whereas in the case of other brands, probably some specific models are preferred.





Notable Findings:

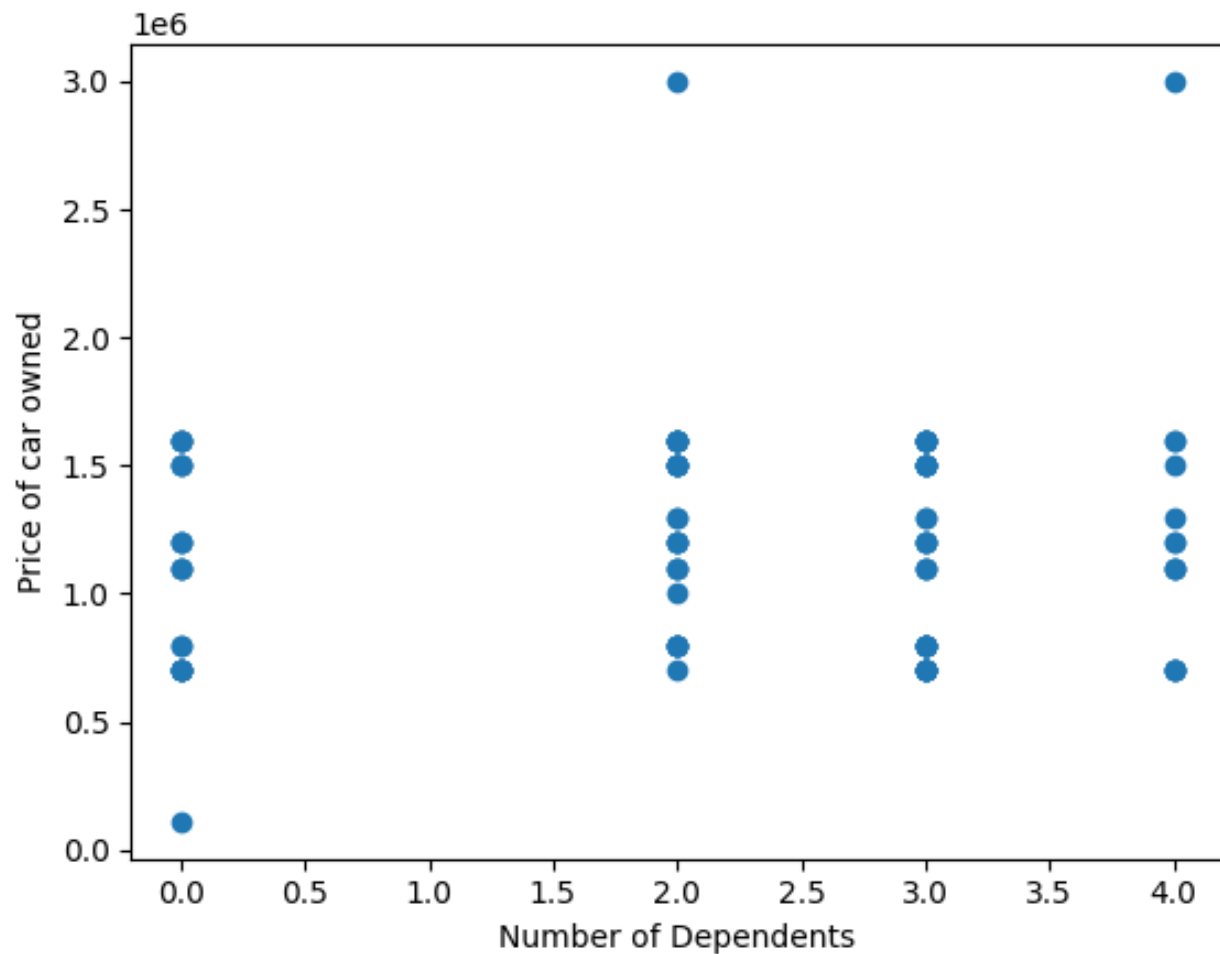




Clusters:

- **Cluster 1 (Orange):** Prefer low to moderately priced EVs, Age 20–30, Moderate salary, Wife's salary Low
- **Cluster 2 (Green):** Prefer low to moderately priced EVs, Age 30–45, Average to high salary, Wife's salary Low
- **Cluster 3 (Blue):** Prefer moderately to high-priced EVs, Age 30–45, Average to very high salary, Wife's salary High

Relationship Between Number of Dependents and Price of Car Owned



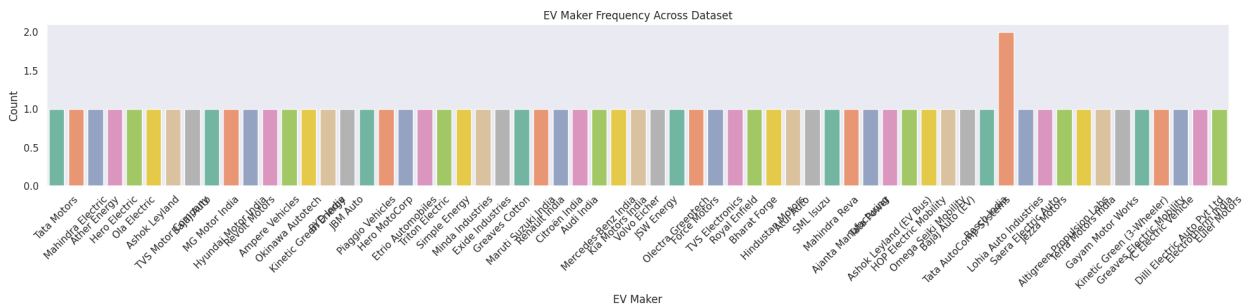
Key Observations:

- No clear trend or correlation is observed between the number of dependents and the price of the car.
- For each dependent count (0 to 4), car prices are widely distributed, including both low-priced and high-priced vehicles.
- High-end vehicles (up to 30 lakhs) are owned by individuals across multiple dependent categories, including those with 2 and 4 dependents.
- This suggests that having more dependents does not significantly influence the car price a customer opts for.

The absence of a visible correlation implies that factors other than the number of dependents—such as income, lifestyle preference, or professional needs—may play a more substantial role in influencing the choice of vehicle.

Key Conclusions

Analysis based on efficiency achieved by a brand:



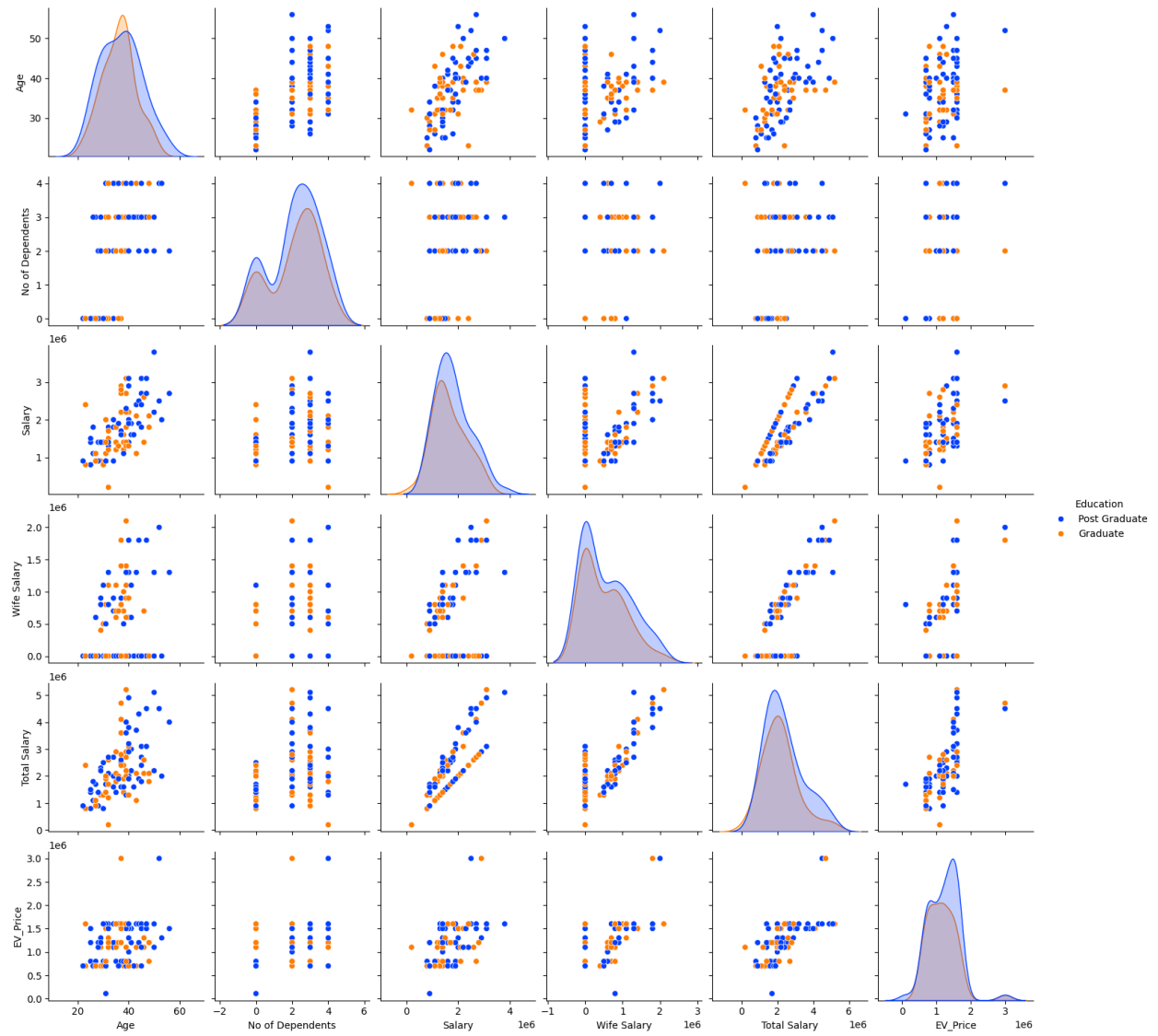
- Moderately priced EVs are the most preferred across all the clusters.
- Most purchasing activity is seen in the age range of 26–38.
- Personal and home loans do not significantly impact EV purchases.
- High correlation between total salary and price.

Optimal Target Profile

Impact of Demographic and Financial Features on EV Purchase Behavior (by Education Level)

Key Insights:

- **Age Distribution:** Both education groups are primarily aged between 30 and 45. Post Graduates show a slightly wider and more uniform age spread, indicating a more diverse buyer profile. This age range forms the core EV buyer demographic.



- **Dependents:** The number of dependents shows no clear relationship with other variables. Buyers across all dependent categories exhibit similar distributions of income and EV price.
- **Salary, Wife Salary, and Total Salary:**
 - A strong linear relationship is seen between Salary and Total Salary, especially where Wife Salary is significant.
 - Post Graduates tend to fall into higher income brackets and are more likely to have spouses contributing financially.
- **EV Price Correlation:** EV Price demonstrates a positive correlation with both Salary and Total Salary. Individuals with higher incomes prefer higher-priced EVs, and this trend is especially evident among Post Graduates.
- **Education-Level Influence:**
 - Post Graduates (blue) predominantly occupy the upper-right portions of the plots, suggesting higher household income and preference for premium EVs.
 - Graduates (orange) cluster around middle-income levels and moderate EV price ranges, with fewer in the luxury segment.

Overall, education level appears to be a significant factor in EV purchase behavior, with Post Graduates showing stronger purchasing power and a greater tendency to invest in high-end electric vehicles.

Demographic Profile:

- Cluster 0: Young Age, No Business, Graduate, Single
- Cluster 1: Young age, Working Professional, Graduate, Married
- Cluster 2: Middle aged (30+), Businessmen, Post Graduate
- Cluster 3: Middle aged, Salaried professional, Married

Geographic Profile:

- **High EV Adoption, High Infrastructure:** Karnataka, Delhi
- **High EV Adoption, Moderate Infrastructure:** Tamil Nadu, Chhattisgarh
- **Moderate EV Adoption, Moderate Infrastructure:** Maharashtra, Rajasthan, West Bengal, Gujarat, Kerala
- **High EV Adoption, Low Infrastructure:** Uttar Pradesh
- **Low EV Adoption, Low Infrastructure:** Odisha, Punjab, Bihar, Assam, Haryana, Ladakh, Sikkim, Jharkhand, Puducherry, Goa, Jammu Kashmir

Possible Improvements

While the current analysis using **K-Means** and **Hierarchical Clustering** has provided valuable insights into customer and market segmentation, there remains scope for refinement and deeper exploration.

1. Exploring Alternative Clustering Methods

To enhance the precision and granularity of segmentation, alternative clustering algorithms can be explored:

- **DBSCAN (Density-Based Spatial Clustering)**: Useful for identifying clusters of varying density and for detecting noise or outliers in consumer behavior data.
- **Gaussian Mixture Models (GMM)**: Allows for soft clustering where a data point can belong to multiple clusters with certain probabilities, capturing overlapping consumer profiles.

These methods may better capture the *non-linear patterns* and *varied density structures* present in India's diverse and evolving EV market.

2. Enhancing Data Quality and Variety

The quality of segmentation results is heavily dependent on the quality and richness of input data. In the current analysis, certain key behavioral and psychographic attributes were either limited or unavailable. These include:

- Brand loyalty and environmental consciousness
- Vehicle usage patterns (e.g., daily commute vs. leisure)
- Financing preferences and budget flexibility

3. Data Enrichment Strategies

To address these limitations, future iterations of the project can incorporate data from:

- **Online surveys or structured interviews** with EV users and prospects.
- **Web scraping** of automotive reviews, customer feedback platforms, and EV community forums.
- **Government databases or industry reports** containing socio-economic and vehicle registration data.

These additions would enable more **nuanced segmentation** and empower EV manufacturers to develop highly targeted and adaptive strategies for specific consumer clusters. This ultimately facilitates better product-market fit and improved competitive positioning within the Indian electric vehicle ecosystem.

Market Mix Strategy for EV Adoption in India

To effectively penetrate and grow in the electric vehicle (EV) market, it is essential to focus on four key strategic pillars: **Innovation**, **Infrastructure**, **Customer Experience**, and **Sustainability**. These elements form the foundation of a successful market entry and long-term customer engagement strategy.

1. Innovation

Constant innovation is a driving force in the EV sector. To remain competitive and meet the dynamic expectations of Indian consumers, EV manufacturers must emphasize:

- Advanced battery technologies for extended range and quicker charging.
- Autonomous driving features to improve safety and convenience.
- Smart system integration enabling real-time connectivity, diagnostics, and user personalization.

2. Infrastructure

Widespread EV adoption in India heavily depends on the development of accessible and reliable infrastructure. Key focus areas include:

- Widespread charging station availability, especially in urban and semi-urban areas.
- Efficient service and maintenance networks to ensure hassle-free ownership.
- Smart grid integration for optimizing energy use and enabling renewable energy support.

3. Customer Experience

A seamless and supportive customer journey is critical to retaining and expanding the EV customer base. This includes:

- Robust after-sales service to handle technical support and part replacements.
- Intuitive mobile apps for real-time vehicle monitoring, bookings, and servicing.
- Consumer education and support to bridge the gap between traditional vehicles and EV technology.

4. Sustainability

Sustainability is not only a value proposition but also a differentiator in the Indian EV landscape. Focusing on environmentally conscious practices will appeal to the eco-aware segment. Key initiatives include:

- Use of eco-friendly materials in manufacturing and interiors.
- Battery recycling programs to minimize environmental impact.
- Reducing carbon footprint through cleaner production and logistics.

This market mix, tailored to insights derived from customer segmentation and demographic analysis in this report, supports the strategic positioning of EV products for maximum adoption across India's emerging consumer segments.

Conclusion

This report provides a comprehensive analysis of India's four-wheeler electric vehicle (EV) market through demographic, behavioral, and psychographic segmentation. By leveraging machine learning techniques such as **K-Means Clustering**, **Hierarchical Clustering**, and **Exploratory Data Analysis (EDA)**, we identified key customer segments based on factors such as income, education, profession, vehicle preference, and buying behavior.

The findings reveal that **Postgraduate, married salaried professionals** with higher household incomes and working spouses are the most prominent adopters of electric vehicles—particularly favoring SUVs and premium models. In contrast, **younger and mid-income consumers** show a preference for hatchbacks and more moderately priced EVs. Interestingly, factors such as *personal and house loans* or *number of dependents* have minimal influence on EV buying decisions.

Additionally, segmentation insights suggest that **total salary** has the strongest correlation with EV price, making it a crucial targeting parameter. The geographic segmentation also highlights **states like Karnataka, Delhi, and Tamil Nadu** as high-potential EV markets due to better infrastructure and adoption rates.

While the analysis delivers meaningful direction, further improvements can be made by introducing advanced clustering techniques such as **DBSCAN** or **Gaussian Mixture Models**, and by incorporating more granular behavioral data through **surveys, web scraping, or industry databases**.

Ultimately, the insights drawn here can assist EV manufacturers in making **data-driven strategic decisions**, aligning product development and marketing efforts with the most promising consumer segments in India's rapidly evolving electric vehicle ecosystem.