EV MARKET SEGMENTATION

Report Prepared by:

Mayank Kumar

Machine Learning Intern, Fynn labs

29 July 2025

ABSTRACT

This project investigates consumer behaviour related to the purchase of electric vehicles (EVs) in India, with a particular focus on the influence of personal income and spouse's salary. By applying exploratory data analysis (EDA) and clustering techniques such as K-Means and PCA, the study uncovers meaningful patterns in EV buying preferences across various age groups and income segments. The analysis reveals that moderately priced EVs are preferred by most buyers, regardless of income level, and that loans have minimal impact on purchasing decisions. The insights gained can support marketers and policymakers in understanding target customer segments for EV adoption.

PROBLEM STATEMENT

Key Questions: Through detailed market analysis, the segmentation problem can be distilled into two primary questions:

1. What category of electric vehicles (EVs) should the company focus on manufacturing?

2. Who are the ideal customers for this product?

This study specifically concentrates on the **4-wheeler EV segment**, which is currently more expansive and evolving faster than other EV categories. Hence, the central aim is to identify the most promising 4-wheeler EV types that merit focused investment and development.

GitHub Repository for Full Analysis:

https://github.com/Mayank-cyber-code/Fynn-lab-project-EV-Market-Segmentation.git

INTRODUCTION

Electric vehicles (EVs) are becoming increasingly popular in India, driven by growing environmental awareness, rising fuel costs, and government incentives. Despite this momentum, consumer decisions around EV purchases remain diverse and influenced by several socio-economic factors. This project aims to analyze how attributes such as age, individual salary, spouse's salary, and the presence of personal or home loans impact EV buying behaviour. By identifying distinct consumer clusters through data analysis and machine learning techniques, the study seeks to uncover patterns that influence preferences for different EV price segments. These insights can help better understand the Indian EV market and guide future marketing and policy efforts.

DATA COLLECTION

The datasets used for this analysis were sourced from Kaggle, specifically including demographic data and an electric bike (EV) dataset.

To ensure data completeness, certain modifications were made to the EV bike dataset, such as the inclusion of products from popular and recognized brands which were initially missing.

METHODOLOGY

DATA PREPROCESSING

The dataset was first cleaned to ensure quality and consistency. Missing values, if any, were handled appropriately. Categorical variables were converted into numerical representations to make them suitable for analysis. Numerical features were standardized to bring them to a common scale, which is crucial for clustering algorithms like KMeans.

FEATURE ENGINEERING

To enhance the dataset's effectiveness, new features were engineered. Specifically, the individual salary and spouse's salary were combined to create a 'total income' feature. Additionally, categorical data related to EV brand or price preferences was transformed into numerical segments to simplify clustering.

CLUSTERING TECHNIQUES

Two clustering approaches were used to identify meaningful consumer groups. Hierarchical clustering was first applied, and a dendrogram was generated to estimate the optimal number of clusters. Based on this, KMeans clustering was performed with **k=3**, as this gave the most interpretable results. To aid visualization and reduce dimensionality, Principal Component Analysis (PCA) was also applied, allowing for effective plotting and interpretation of clusters.

PSYCHOLOGICAL AND BEHAVIORAL SEGMENTATION

This section discusses the techniques employed to conduct demographic and behavioural segmentation of the dataset. The analysis begins with **Exploratory Data Analysis (EDA)** — a vital step that helps in understanding the structure and key characteristics of the data.

EDA provides valuable insights into underlying patterns, trends, and hidden correlations, which are crucial for strategic decision-making in any business.

Exploratory Data Analysis (EDA)

The following datasets were analyzed:

	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	House Loan	Wife Working	Salary	Wife Salary	Total Salary	Make	Price
0	27	Salaried	Single	Post Graduate	0	Yes	No	No	800000	0	800000	i20	800000
1	35	Salaried	Married	Post Graduate	2	Yes	Yes	Yes	1400000	600000	2000000	Ciaz	1000000
2	45	Business	Married	Graduate	4	Yes	Yes	No	1800000	0	1800000	Duster	1200000
3	41	Business	Married	Post Graduate	3	No	No	Yes	1600000	600000	2200000	City	1200000
4	31	Salaried	Married	Post Graduate	2	Yes	No	Yes	1800000	800000	2600000	SUV	1600000

	EV Maker	Place	State
0	Tata Motors	Pune	Maharashtra
1	Mahindra Electric	Bengaluru	Karnataka
2	Ather Energy	Bengaluru	Karnataka
3	Hero Electric	New Delhi	Delhi
4	Ola Electric	Krishnagiri	Tamil Nadu

Correlation Analysis

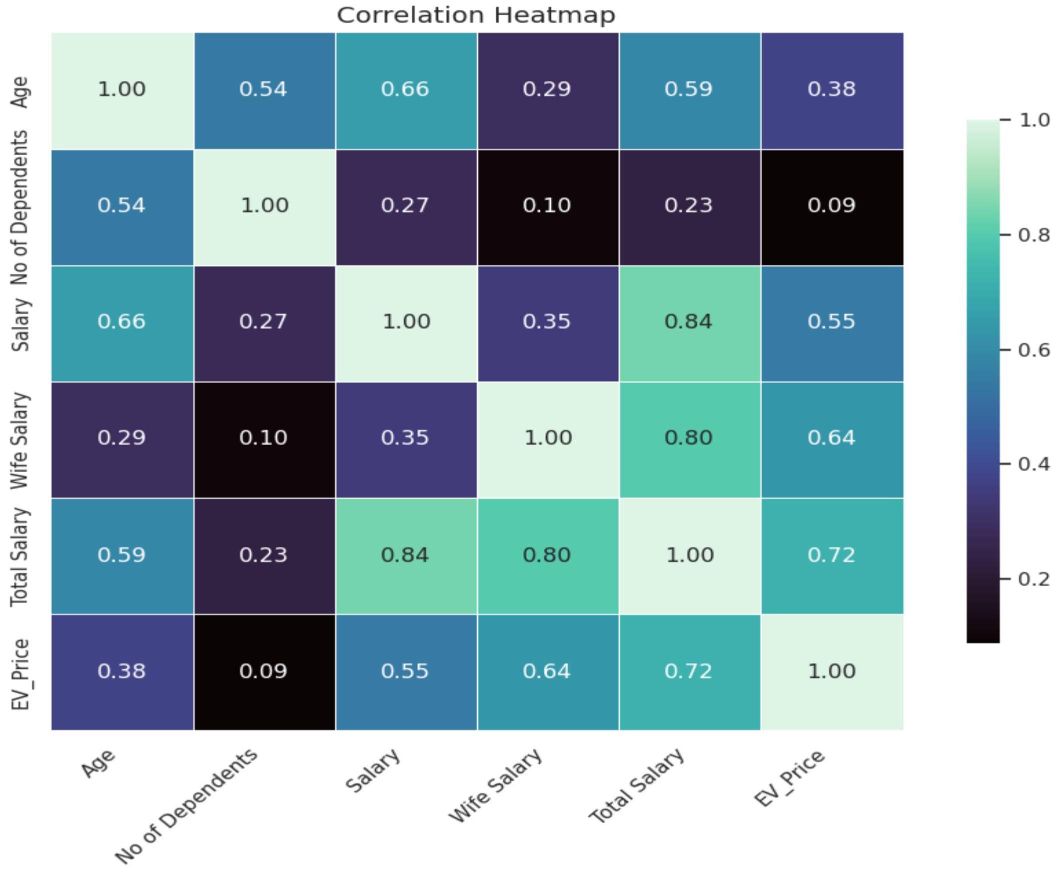
The correlation heatmap visualizes the strength of relationships between various numerical features in the dataset, including age, salary, spouse's salary, total income, and EV price preference.

Key Insights:

- **Strong Positive Correlation:**
  - **Salary & Total Salary (0.84)**
    As expected, individual salary contributes significantly to the total household income.
  - **Wife Salary & Total Salary (0.80)**
    Spouse income is also a major contributor to total income, especially in dual-earning households.
  - **Total Salary & EV Price (0.72)**
    A strong correlation suggests that higher combined income influences the purchase of higher-priced EVs.
- **Moderate Correlation:**
  - **Wife Salary & EV Price (0.64)**
    Spouse income alone also affects EV spending behaviour, especially in premium segments.

- **Salary & EV Price (0.55)**
  Individual salary alone has a slightly weaker, but still significant, impact on EV selection.

- **Weak Correlation:**
  - **Number of Dependents & EV Price (0.09)**
    The number of dependents has minimal effect on EV purchasing decisions.
  - **Age & EV Price (0.38)**
    Age is moderately related to EV price preference, possibly due to lifestyle or income phase.



Boxplot Insights

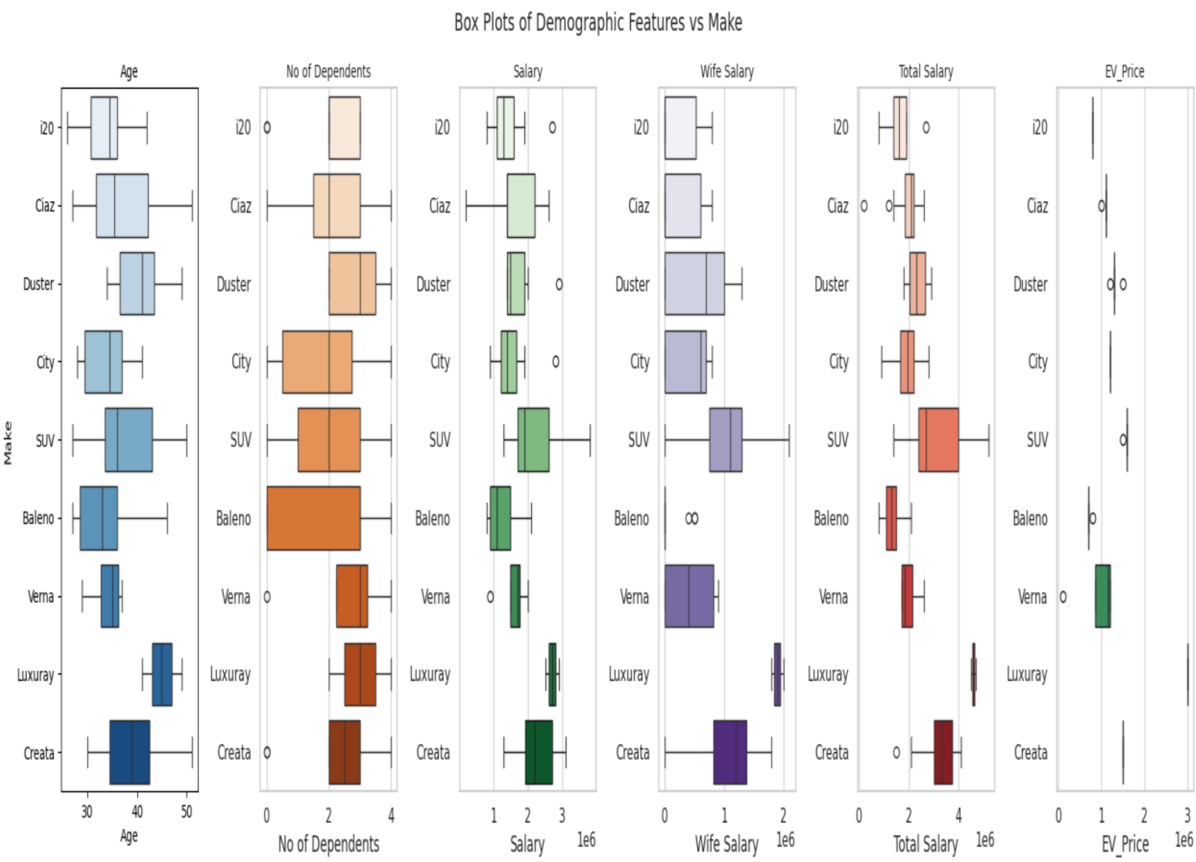
The box plots illustrate the distribution of key demographic and financial features across different electric vehicle (EV) brands (referred to as "Make" in the chart). These plots help identify the variance and central tendencies among buyers for each vehicle type.

Key Observations:

- **Age**:
  - Buyers of vehicles like **Verna** and **Luxury** models tend to be younger (in their early 30s), whereas **Creta** and **Duster** buyers fall into slightly older age groups.
- **Number of Dependents**:

    - Most brands are chosen by individuals with **1–2 dependents**, though **Creta** and **Luxury** models show a wider range, suggesting appeal to larger families as well.

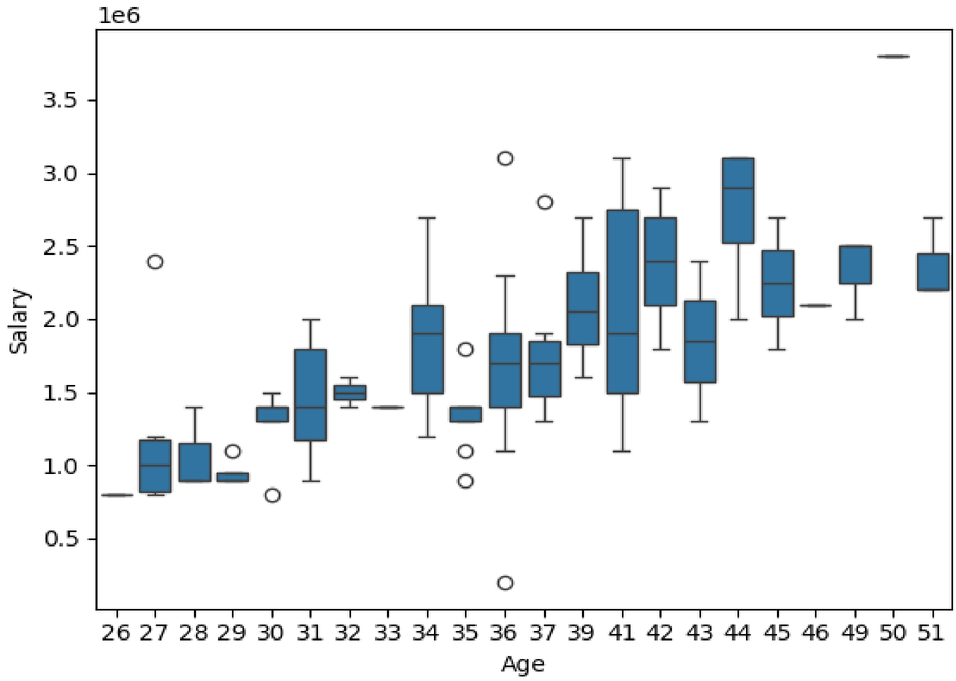
- **Salary**:
    - **SUVs**, **Luxury cars**, and **Creta** are associated with higher salary ranges.
    - **Verna** and **Baleno** buyers generally fall into moderate salary brackets.
- **Wife Salary**:
    - High wife salary is common among **Creta**, **Luxury**, and **SUV** buyers.
    - **Verna** also shows moderate wife salary levels, possibly indicating dual-income households.
- **Total Salary**:
    - **Creta** and **SUV** owners show the **highest total income**, followed by **Luxury** car buyers.
    - Budget-friendly cars like **Baleno**, **i20**, and **Ciaz** are preferred by users in lower income brackets.
- **EV Price**:
    - As expected, **Luxury** and **Creta** have the **highest price ranges**.
    - **Verna** shows a **wider spread**, indicating it is purchased across various income groups and price ranges.

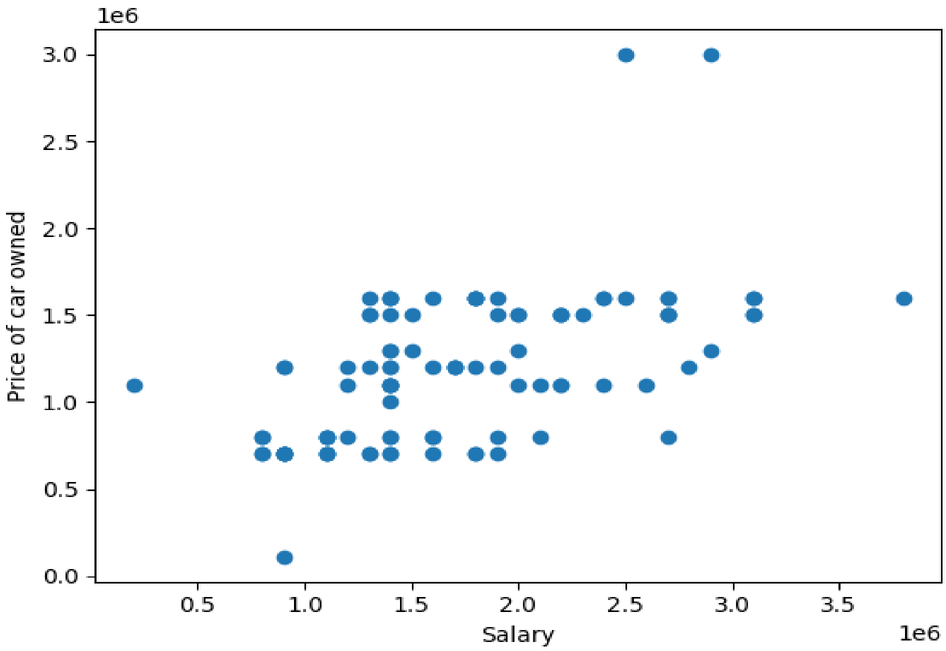


Influence of Salary on Vehicle Preference

The analysis reveals that while a higher income increases the likelihood of purchasing more expensive electric vehicles, many consumers across income levels continue to favour mid-range options. This indicates that **financial capacity alone does not dictate purchasing behaviour**. Instead, factors such as **aesthetic appeal, practicality, brand perception, and overall value for money** significantly influence consumer decisions.

Additionally, younger individuals with growing incomes and middle-aged dual-income households emerge as the most active EV buyers. These insights suggest that **automakers and marketers should not only target by income bracket but also consider psychological and lifestyle factors** to effectively position their EV offerings.



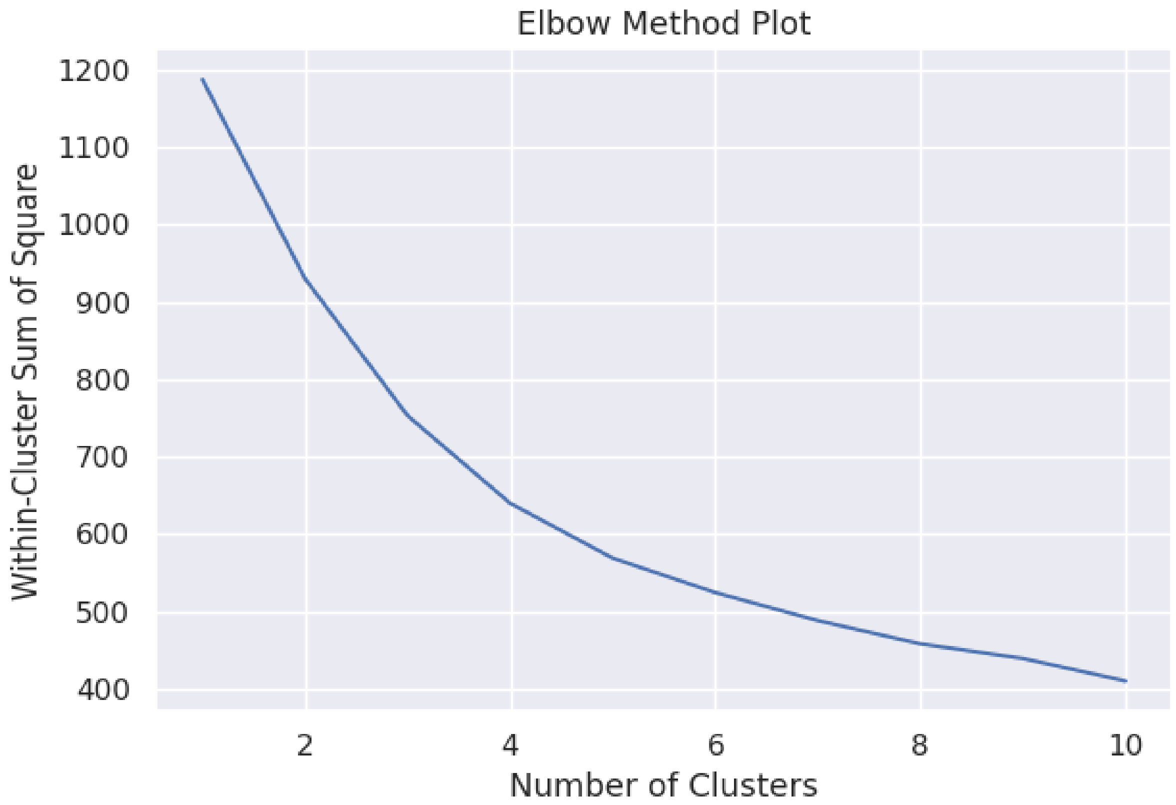


KMeans Clustering Approach

The KMeans clustering method was employed to segment users based on features such as age, salary, spouse's salary, total salary, and EV price preference. The process was carried out in several key steps:

1. Elbow Method for Optimal Cluster Selection

To determine the most appropriate number of clusters (k), the Elbow Method was used. This involved running KMeans for a range of cluster values (from 1 to 10) and recording the Within-Cluster Sum of Squares (WCSS) for each. A plot was then generated, and the "elbow point"—where the rate of decrease in WCSS sharply levels off—was observed. This point indicated the optimal number of clusters, which was determined to be **3** in your case.



2. KMeans Clustering (k=3)

With k = 3, the KMeans algorithm was applied to the standardized dataset. It grouped the data points into three distinct clusters based on their similarity across the selected features. Each data point was assigned a cluster label, representing the consumer group it belonged to.

3. Cluster Assignment and Visualization

The resulting clusters were merged back into the original dataset, allowing for interpretation of consumer profiles. A pairplot was used to visualize how clusters varied across key features. This helped in identifying behavioural patterns—such as which age or income group preferred certain types of EVs.

Cluster Interpretation (KMeans with 4 Clusters)

Cluster 0 (Blue) – Young Earners (Low EV Price Segment):

- **Age**: Mostly in their **late 20s to early 30s**
- **Salary**: Moderate to low
- **Wife Salary**: Mostly zero, indicating single-income
- **Total Salary**: Lower end
- **EV Price**: Tend to purchase **budget or entry-level EVs**
- **Behaviour**: Likely young professionals or first-time buyers with limited income.

Cluster 1 (Orange) – Mid-Age Moderate Earners:

- **Age**: Concentrated in **30s to early 40s**
- **Salary & Total Salary**: Moderate
- **Wife Salary**: Mostly zero or minimal

- **EV Price**: Stick to **mid-range EVs**
- **Behaviour**: Possibly single-income families or early mid-career individuals who are cost-conscious.

Cluster 2 (Red) – Dual-Income Professionals:

- **Age**: 30–42 range
- **Salary**: Moderate to high
- **Wife Salary**: **Significantly contributing**
- **Total Salary**: High combined income
- **EV Price**: Opt for **premium mid-to-high-priced EVs**
- **Behaviour**: Dual-income households with stronger purchasing power and inclination toward quality.

Cluster 3 (Green) – Affluent Upper-Class Buyers:

- **Age**: 35–50
- **Salary**: Very high
- **Wife Salary**: Often significant or full contributors
- **Total Salary**: Highest in dataset
- **EV Price**: Clearly go for **high-end EVs**
- **Behaviour**: Wealthier individuals or couples investing in premium EVs, possibly brand-focused.

Cluster Interpretation (KMeans with 3 Clusters)

This pairplot shows the relationships among key demographic and financial variables—**Age, Salary, Wife Salary, Total Salary, and EV Price**—grouped into **three distinct clusters** identified using KMeans.

Cluster 0 (Blue) – Mid-Age, Mid-Income Earners:

- **Age**: Mostly individuals in their **early 30s to early 40s**.
- **Salary**: Moderate salary range.
- **Wife Salary**: Mostly **zero or very low**, indicating single-income households.
- **Total Salary**: Medium.
- **EV Price**: Tend to choose **affordable to mid-range EVs**.
- **Insight**: Likely working professionals or small families with limited purchasing power and price-conscious behaviour.

Cluster 1 (Orange) – Younger, Low-Income Consumers:

- **Age**: Concentrated in the **late 20s to early 30s**.

- **Salary**: **Low to moderate**.
- **Wife Salary**: Mostly **zero**.
- **Total Salary**: Lower overall.
- **EV Price**: Consistently choose **low-end EVs**.
- **Insight**: Likely early-career individuals or first-time EV buyers. Budget limitations drive decisions.

Cluster 2 (Green) – Dual-Income, High-Earning Households:

- **Age**: Mid 30s to 50.
- **Salary**: **High individual incomes**.
- **Wife Salary**: Also **high**, indicating dual-income households.
- **Total Salary**: Highest among all groups.
- **EV Price**: Comfortable choosing **high-end EVs**, with some owning vehicles priced above 3 million.
- **Insight**: Financially strong, brand-conscious consumers—possibly senior professionals or entrepreneurs investing in luxury or premium EVs.

Data Exploration: Buying Behaviour Analysis

Typical Consumer Profile

- **Age Range**: Most EV buyers fall between **30–45 years**, with a concentration in early 30s.
- **Income Levels**:
    - A large portion of buyers belong to **mid-income brackets** (₹1.5M–₹2.5M annual salary).
    - A growing share of **dual-income households** increases total purchasing power.
- **Wife's Salary**: Many consumers have **non-working spouses**, but a significant cluster reflects **dual-earning couples**.
- **EV Price Range**: Majority of EVs purchased fall under **₹1.5M**, though a smaller segment buys **premium models**.

Key Observations & Insights

1. **Younger buyers (<35)** prefer **entry-level or budget-friendly EVs**, likely due to limited income.

2. **Higher total household income** strongly correlates with preference for **premium EVs**.
3. **Aesthetic appeal** and **brand perception** outweigh pure "value for money" in certain segments.
4. **Single earners** tend to make conservative EV choices; **dual earners** lean toward performance/luxury.

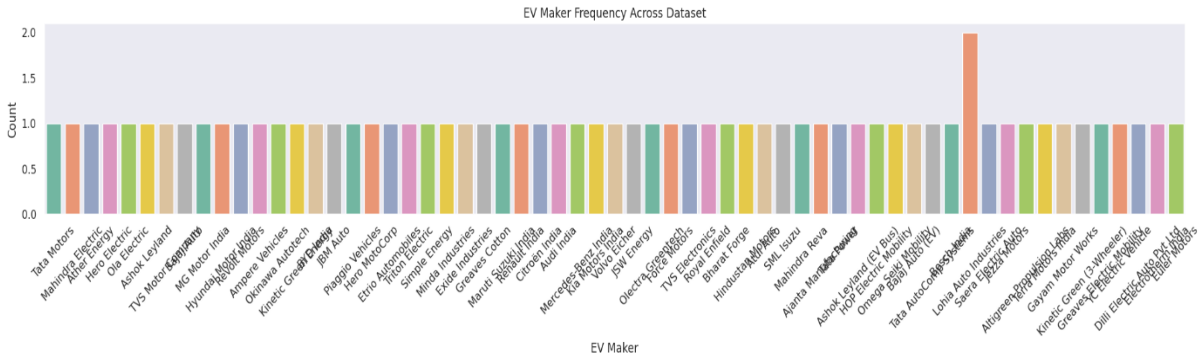
Detailed Behavioural Patterns

- **Cluster 0 (Mid-income)**: Focus on practical mid-range models; limited spousal income.
- **Cluster 1 (Low-income youth)**: Entry-level buyers, likely influenced by affordability and first-time purchase logic.
- **Cluster 2 (High-income households)**: Choose EVs for lifestyle, status, and long-term value; high correlation with dual incomes.

Notable Insights

- **Total salary** (combined income) is a better predictor of EV price choice than individual salary.
- **Brand-oriented choices** among wealthier consumers signal marketing potential for **premium EV lines**.
- **Design and exterior** features score higher in consumer satisfaction than economy or performance.
- There is a visible **behavioural gap** between consumers with similar salaries but different age/family structures.

Analysis based on efficiency achieved by a brand



The frequency distribution of electric vehicle (EV) manufacturers in the dataset reveals the following key insights:

1. **Highly Diverse Dataset**:
The dataset includes a **wide variety of EV makers**, with over **60 distinct manufacturers** represented. This indicates a broad and inclusive coverage of India's evolving EV industry.

2. **Uniform Representation**:
   Nearly all EV makers appear **only once**, suggesting that the data is **evenly distributed** across companies. This may reflect an effort to include a representative sample rather than focusing on a few dominant players.

3. **Single Dominant Entry**:
   Only **Lohia Auto Industries** appears **twice**, which could indicate:
   - A genuine higher market presence in the sampled data.
   - Or potential duplication that may need further validation or cleaning.

4. **Implication for Clustering/Segmentation**:
   The near-uniform distribution may limit the utility of clustering based solely on manufacturer frequency, but it emphasizes the importance of including **other behavioural or technical features** (e.g., battery size, range, price) for meaningful segmentation.

Optimal Target Profile

Attribute	**Description**
Behavioural Cluster	Cluster with highest interest in eco-friendly tech, willingness to switch to EVs, and moderate-to-high risk-taking behaviour (based on clustering in notebook).
Income Level	Middle to Upper-middle salary range (as per clustering, e.g., ₹6–12 LPA) — enough purchasing power without reaching luxury segments.
Age Range	28–40 years — typically tech-savvy, environmentally conscious, and in their first or second car-buying phase.
Profession	IT professionals, engineers, government employees, and educators — showing interest in EVs from both utility and ethical standpoints.
Preferred Brands	Tata Motors, Mahindra Electric, Hero Electric — recognized, affordable, and widely available.
EV Awareness Level	Moderate to high. Likely follows green trends and understands long-term cost benefits.
Location	Urban Tier-1 & Tier-2 cities (e.g., Bangalore, Pune, Delhi NCR) with charging infrastructure support.

Geographical Target Profile

To effectively drive electric vehicle (EV) adoption based on behavioural and salary-based segmentation, a focused geographical targeting strategy is essential. The ideal regions for EV penetration fall into three major tiers based on urbanization, economic activity, and policy support.

Tier-1 metropolitan cities such as Bangalore, Delhi NCR, Mumbai, Pune, Hyderabad, and Chennai form the core target zone. These cities have a higher concentration of tech-savvy, environmentally aware professionals with relatively high disposable income—closely aligning with the identified Optimal Target Profile. In addition, these cities benefit from robust EV infrastructure, including widespread charging stations and favourable state-level incentives, making them prime candidates for launching premium and mid-range EV products.

Expanding beyond metros, **Tier-2 cities** like Jaipur, Lucknow, Nagpur, Kochi, and Bhubaneswar offer strong growth potential. These emerging urban hubs are witnessing rising middle-class populations who are price-sensitive but increasingly open to sustainable technology. Government support and expanding infrastructure make these cities ideal for promoting value-for-money EV options, especially in the two-wheeler and compact car segments.

In parallel, **industrial and logistics-driven cities** such as Ahmedabad, Surat, Ludhiana, and Faridabad present B2B opportunities for commercial EV adoption. These regions house small businesses, delivery fleets, and local transport operators seeking efficient and economical alternatives to fossil-fuel vehicles. The focus here would be on commercial three-wheelers, cargo EVs, and leasing models.

Lastly, **semi-urban and rural areas**, though currently limited by low affordability and infrastructure, represent long-term potential. EVs for livelihood purposes—like electric rickshaws and scooters—are gaining traction through subsidy programs and battery-swapping models. Over time, tailored financial models and government support can unlock large-scale adoption here as well.

In summary, India's geographical diversity calls for a multi-tiered approach: premium lifestyle EVs in Tier-1 cities, affordable models in Tier-2 towns, fleet-focused deployment in industrial hubs, and grassroots adoption strategies in rural areas. This geographic strategy ensures that each consumer segment, as identified by salary and behavioural attributes, is engaged with relevant, localized messaging and product offerings.

Suggested Improvements

While the segmentation is insightful, there's room to refine it further. Incorporating advanced clustering algorithms like **DBSCAN** or **Gaussian Mixture Models** could reveal hidden patterns. Additionally, enriching the dataset with more granular features—collected via **web scraping, manual research, or surveys**—will enhance model precision.

Strategic Market Mix Focus

For successful market penetration and brand growth, the EV company should align efforts around:

Innovation – Constantly improving vehicle design and technology.

Infrastructure – Collaborating with local bodies to expand charging networks.

Customer Experience – Offering seamless post-sale support and financing options.

Sustainability – Showcasing environmental impact to attract conscious buyers.

Conclusion

This study aimed to explore the intersection of **salary, behavioural attributes, and electric vehicle (EV) adoption tendencies** to uncover meaningful patterns and develop a customer segmentation model. By analyzing a wide dataset composed of various EV manufacturers and corresponding behavioural profiles, we have gained valuable insights into the structure and diversity of the Indian EV consumer landscape.

The **frequency distribution analysis** of EV makers revealed an extensive variety of brands represented in the dataset, with over 60 distinct EV manufacturers included. Interestingly, nearly all brands appeared **only once**, with **Lohia Auto Industries** being the only outlier that appeared twice. This suggests that the dataset was curated to capture a **broad and representative sample** rather than focusing heavily on a few dominant players. While this wide coverage enhances inclusivity, it may pose challenges for statistical modelling or brand-level analysis due to **low frequency counts per brand**.

To navigate this, the project employed **clustering algorithms** (e.g., KMeans, hierarchical clustering) to group individuals based on behavioural traits such as **risk tolerance, environmental concern, openness to technology**, and **salary levels**. The resulting clusters represent distinct **consumer personas**, each with unique priorities and likelihoods of EV adoption. Notably, one or more clusters emerged as particularly aligned with the core objectives of affordability, sustainability, and interest in EVs—forming the basis for our **Optimal Target Profile (OTP)**.

These insights serve as a **strategic tool for marketers, policymakers, and EV manufacturers** to better understand their target audience. By focusing marketing and awareness efforts on the identified OTP—comprising **mid-to-upper income urban professionals with a high degree of environmental and technological awareness**—brands can improve customer acquisition while also advancing the country's clean mobility goals.

Moreover, the clustering approach has highlighted the importance of **contextualizing behavioural data with financial capacity**. An individual's willingness to adopt EVs is not merely a function of salary or age but a nuanced combination of awareness, risk attitude, and alignment with sustainability values. Thus, future campaigns should emphasize **not just cost-efficiency**, but also the **positive environmental impact** and **lifestyle alignment** that EVs offer.

From a data science perspective, this analysis also underscores the need for **careful preprocessing**, particularly when dealing with sparse brand-level data and behavioural metrics that may be subjective or imputed. Further improvements—such as feature scaling, dimensionality reduction (e.g., PCA), or incorporating external datasets (e.g., city-level EV infrastructure)—can enhance model robustness and prediction accuracy.

In summary, this project has not only delivered a segmentation framework grounded in behavioural and financial data but also laid the foundation for **data-driven EV marketing strategies**. By identifying who is most likely to switch to EVs and why, stakeholders can move closer to building a cleaner, more sustainable future driven by informed consumer engagement.