EV MARKET SEGMENTATION

By

Sharanya Biswas

Date:19 July,2025

Abstract

This project presents a comprehensive analysis of India's electric vehicle (EV) market, with a specific focus on segmentation within the four-wheeler category using income-based consumer data. The research highlights the rapid expansion of the EV sector, driven by growing environmental consciousness, progressive government incentives, and advancements in battery technology. Leveraging behavioral and demographic data, the study applies the **k-means clustering algorithm** to identify distinct consumer segments within the electric four-wheeler market. The goal is to uncover patterns in consumer behavior, spending capacity, and market trends that can guide strategic decision-making. The insights derived from this segmentation aim to support stakeholders—including manufacturers, investors, and policymakers—in developing targeted marketing strategies and business models, ultimately accelerating the adoption and development of electric mobility in India.

Introduction

India is undergoing a significant transformation in its transportation landscape, spurred by the increasing adoption of electric vehicles (EVs). This shift is fueled by factors such as rapid urbanization, rising income levels, and heightened environmental awareness. Among EV categories, **electric four-wheelers** have emerged as the most impactful, offering a cleaner and more cost-effective alternative to traditional internal combustion engine vehicles. Their growing popularity signifies a key milestone in India's pursuit of sustainable development and reduced dependence on fossil fuels.

With the government's active support in the form of policy incentives and infrastructure development, the electric four-wheeler segment has witnessed substantial growth. These vehicles not only promise environmental benefits but also align with the aspirations of a tech-savvy and economically advancing population.

This research delves into the **segmentation of the electric four-wheeler market** using income and behavioral attributes. The application of **k-means clustering** enables the identification of distinct consumer groups, helping to uncover variations in preferences, affordability, and adoption behavior. The findings aim to provide a data-driven foundation for decision-makers, enabling more effective targeting of products and policies, and facilitating the broader penetration of EVs across diverse socio-economic segments in India.

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 Indian automoble buying behavour study on EV market, which is considerably larger and more dynamic compared to others. Therefore, the objective is to determine age groups.

Approach

The primary objective of this study is to analyze the electric vehicle (EV) market in India through a detailed segmentation framework, with the aim of formulating a data-driven strategy for effective market entry. The focus is placed on identifying and understanding the customer segments most likely to adopt electric four-wheelers, enabling manufacturers and marketers to tailor their offerings and outreach accordingly.

This analysis integrates **geographic**, **demographic**, **psychographic**, and **behavioral** dimensions to achieve a comprehensive understanding of market diversity and consumer preferences.

To achieve this, the following methodological steps were undertaken:

- **Market Scope Definition**: The study narrows its focus to the four-wheeler EV segment, which has shown the most substantial growth in India's EV landscape.
- **Data Collection and Preprocessing**: The dataset includes variables such as income level, profession, marital status, education, and behavioral indicators related to vehicle loans and work status. These were encoded and standardized for compatibility with clustering algorithms.
- **Dimensionality Reduction using PCA**: Principal Component Analysis (PCA) was applied to reduce the dataset's dimensionality while preserving the variance. This step ensures optimal performance of clustering algorithms by eliminating multicollinearity and noise.
- **Customer Segmentation using K-Means Clustering**: The core of the analysis involved implementing the **K-Means clustering algorithm**, which partitions the dataset into distinct customer groups based on similarities in behavioral and demographic patterns. The optimal number of clusters was identified using methods like the Elbow Method and Silhouette Analysis.
- **Hierarchical Clustering (Demonstration)**: In addition to K-Means, hierarchical clustering was also explored to validate and visualize the cluster structure, offering insights into the nested relationships among different consumer profiles.

- **Target Segment Identification**: After clustering, the resulting segments were profiled based on key features such as income range, loan status, and professional background. The viability of each segment was assessed in terms of market size, affordability, and likelihood of EV adoption.

The GitHub Link to the Analysis:

https://github.com/Sharanyamsd/Feynn-Labs-Project-2/tree/main

Psychological and Behavioural Segmentation

This section focuses on the **psychological** and **behavioural** dimensions of customer segmentation within the electric vehicle (EV) market. These variables are critical in understanding not just who the customers are (demographics), but also **how they think**, **what motivates them**, and **how they behave in relation to purchasing decisions**.

Exploratory Data Analysis (EDA)

To begin, an **Exploratory Data Analysis (EDA)** was performed to gain an initial understanding of the dataset. EDA is an essential step in data analysis as it helps uncover hidden trends, patterns, and anomalies that might influence consumer behavior. Through visualizations and statistical summaries, we examined distributions and relationships between key psychological and behavioral attributes.

Some of the key variables considered during this phase include:

- **Profession**: Serving as a proxy for lifestyle and social class.
- **Marital Status**: A behavioral factor influencing purchase decisions (e.g., family vs. individual needs).
- **Personal Loan** and **House Loan** status: These reflect financial commitments that may impact willingness or ability to invest in an EV.
- **Wife Working**: A socioeconomic indicator that can influence dual-income household spending patterns.
- **Education Level**: Strongly associated with environmental awareness, technological adoption, and decision-making rationality.

These features were visualized through bar plots, histograms, and heatmaps to identify potential clusters or segment trends.

Dataset Overview

Code :

```
df = pd.read_csv("/content/Indian automoble buying behavour study 1.0 - Copy.csv")
df.head()
df.describe()
df.info()
```

- df.head() shows the first 5 rows of the dataset, helping us verify the structure and features like **Age, Salary, Profession, Price, Personal loan**, etc.
- df.describe() summarizes the numerical variables (e.g., Age, Salary, No. of Dependents, Price). For example:
    - Average Age ~ 35 years
    - Salary ranges widely, indicating people from various economic backgrounds
- df.info() gives us the **data types** and **missing values**. This check is crucial before modeling or visualization.

Interpretation:
Understanding the variables at this stage tells us how much preprocessing is needed and what data types will influence clustering or predictive models later.

df.head()

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

df.describe()

	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	36.313131	2.181818	1.736364e+06	5.343434e+05	2.270707e+06	1.194040e+06
std	6.246054	1.335265	6.736217e+05	6.054450e+05	1.050777e+06	4.376955e+05
min	26.000000	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%	36.000000	2.000000	1.600000e+06	5.000000e+05	2.100000e+06	1.200000e+06
75%	41.000000	3.000000	2.200000e+06	9.000000e+05	2.700000e+06	1.500000e+06
max	51.000000	4.000000	3.800000e+06	2.100000e+06	5.200000e+06	3.000000e+06

df.info()

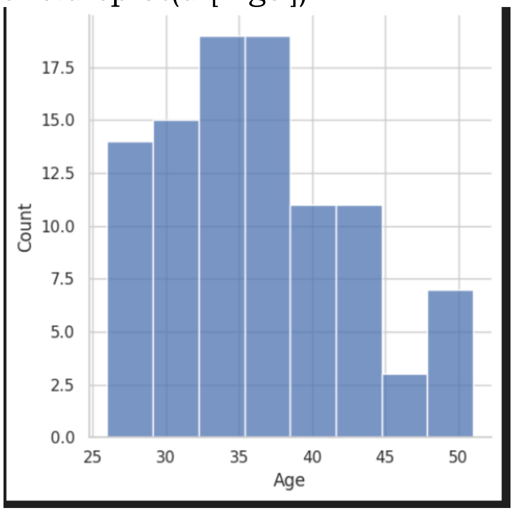
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99 entries, 0 to 98
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Age               99 non-null     int64
 1   Profession        99 non-null     object
 2   Marrital Status   99 non-null     object
 3   Education         99 non-null     object
 4   No of Dependents  99 non-null     int64
 5   Personal loan     99 non-null     object
 6   House Loan        99 non-null     object
 7   Wife Working      99 non-null     object
 8   Salary            99 non-null     int64
 9   Wife Salary       99 non-null     int64
 10  Total Salary      99 non-null     int64
 11  Make              99 non-null     object
 12  Price             99 non-null     int64
dtypes: int64(6), object(7)
memory usage: 10.2+ KB
```

Univariate and Count Plot Analysis

Age Distribution of Customers

<u>Code:</u>

```
sns.displot(df['Age'])
```

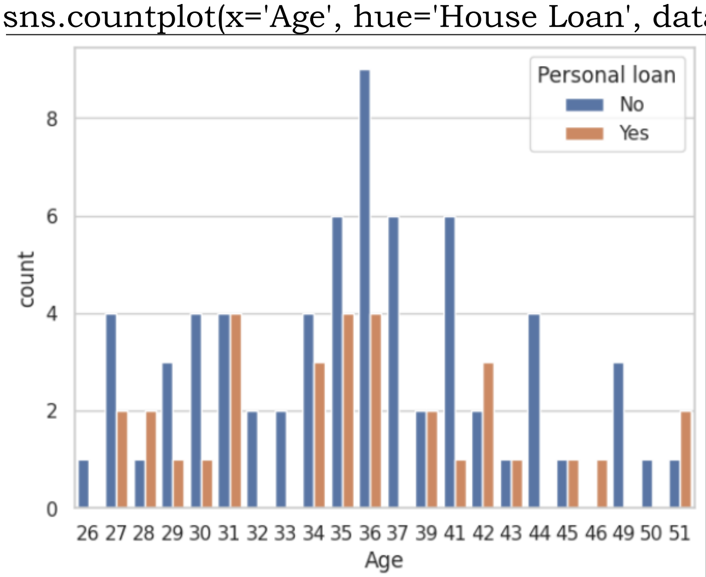


- A **right-skewed distribution** was observed — most customers fall between **25 to 35 years**.
- EV adoption seems higher in younger, working-age groups.

Loan Distribution by Age

<u>Code:</u>

```
sns.countplot(x='Age', hue='Personal loan', data=df)
sns.countplot(x='Age', hue='House Loan', data=df)
```



- These show the number of people at each age who have a **personal** or **house loan**.
- Some age groups (e.g., 30–35) are more likely to have loans, indicating a segment with both **desire and financial liability**.

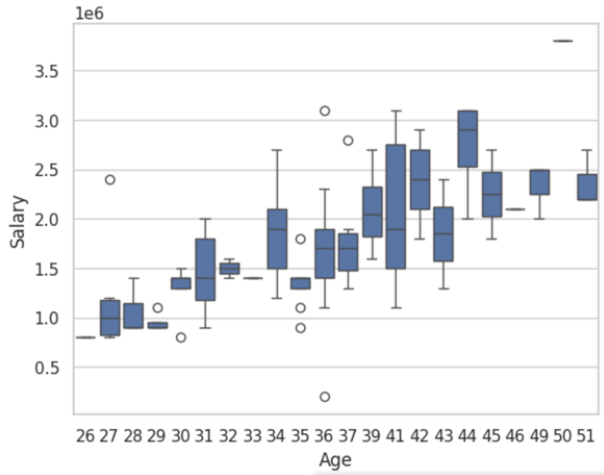
Interpretation:

Companies can promote low-interest EV loans to these age groups to encourage purchase.

Salary and Age Relationship

Code:

```
sns.boxplot(x='Age', y='Salary', data=df)
```



- This plot visualizes **how income changes with age**.
- Salary increases with age initially but shows **plateauing in older age groups**.

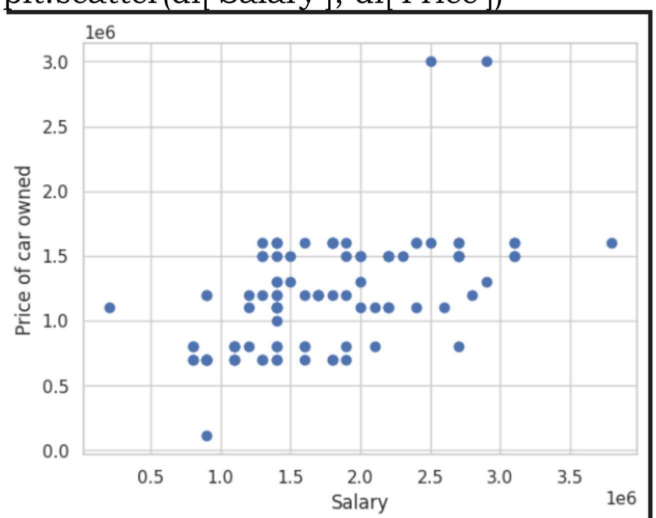
Insights:
Marketers can group users by **salary bands** and align car models with affordability — e.g., entry-level EVs for <30 age group.

Bivariate Relationships

Salary vs Car Price

Code:

```
plt.scatter(df['Salary'], df['Price'])
```



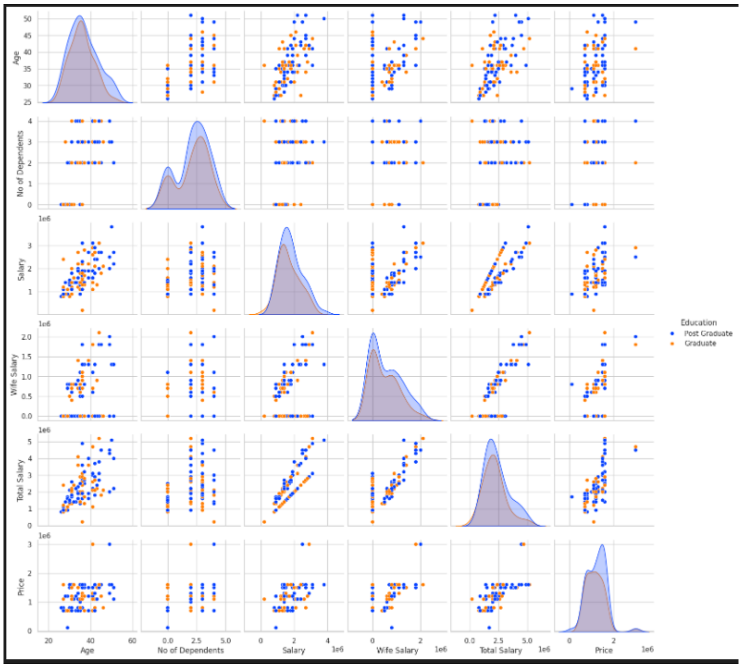
- The scatter shows a **weak positive relationship**.
- Some buyers with lower salaries still buy expensive EVs, indicating:
  - Influence of **loans or family support**
  - Emotional or environmental motivation beyond income

.

Pairplot and Education-Based Clusters

Code:

```
sns.pairplot(df, hue='Education', palette='bright')
```



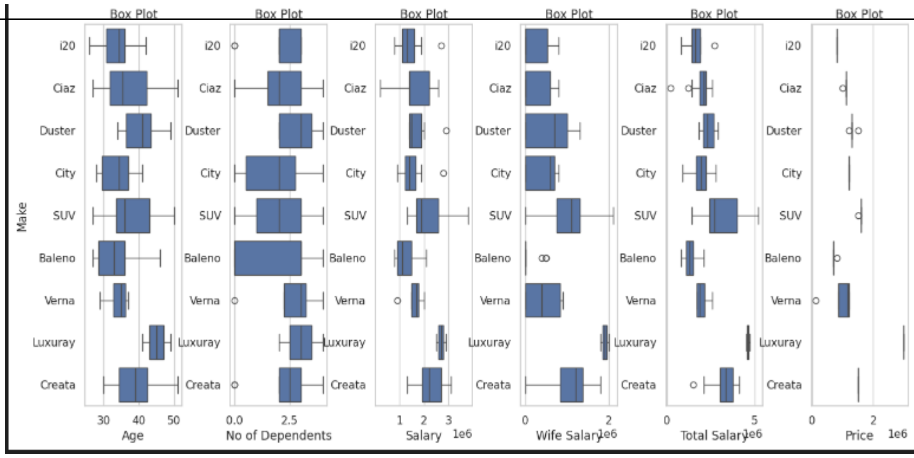
- This visualization shows how education levels influence clusters.
- Highly educated users tend to have:
    - Higher salaries
    - Higher car prices
    - Lower number of dependents

Insights:
Education level can be used as a **segmentation feature**, especially for premium models.

Comparative Boxplots by Brand

Code:

```
for cols in ['Age', 'No of Dependents', 'Salary', 'Wife Salary', 'Total Salary', 'Price']:
    sns.boxplot(x=cols, y='Make', data=df)
```



- This gives a **brand-wise comparison** for each feature.
- Some brands target:
    - Younger, lower-salary individuals (affordable brands)
    - Higher-salary segments (premium EV makers)

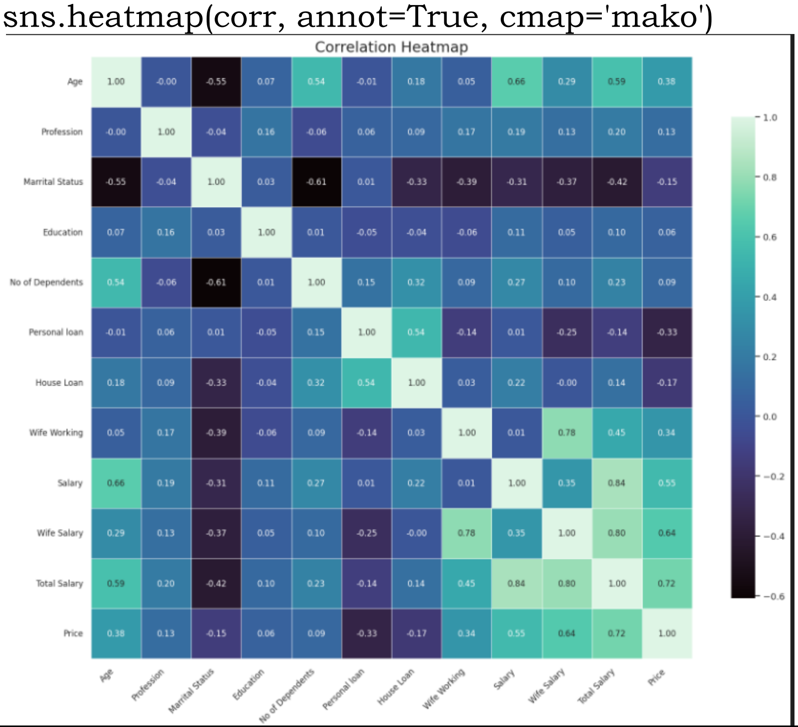
Strategic Use:
These plots help map **brand positioning** and suggest which customer segments are aligned with which brand.

Correlation Analysis Using Heatmap

Code:

```
for col in ['Profession', 'Marrital Status', 'Education', 'Personal loan', 'House Loan', 'Wife Working']:
    df_encoded[col] = LabelEncoder().fit_transform(df_encoded[col])

corr = df_encoded.drop('Make', axis=1).corr()
sns.heatmap(corr, annot=True, cmap='mako')
```



Correlation Heatmap shows:

- **Total Salary** has a strong positive correlation with **Car Price**
- Weak or no correlation between loans and car price
- Profession and education moderately impact income and EV price

Marketing Insight:
Brands should target users not only based on income but **total household salary** (including spouse) and **profession type**.

K-Means Clustering:

Data Preparation and Scaling

Code:

```
df_encoded = df.copy()
for col in ['Profession', 'Marrital Status', 'Education', 'Personal loan', 'House Loan', 'Wife Working']:
    le = LabelEncoder()
    df_encoded[col] = le.fit_transform(df_encoded[col])
```

- All categorical variables (e.g., Profession, Marrital Status) are encoded into numeric values using **Label Encoding**, which is required for K-Means to process the data.

Code:

```
X = df_encoded.drop(['Make'], axis=1)
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

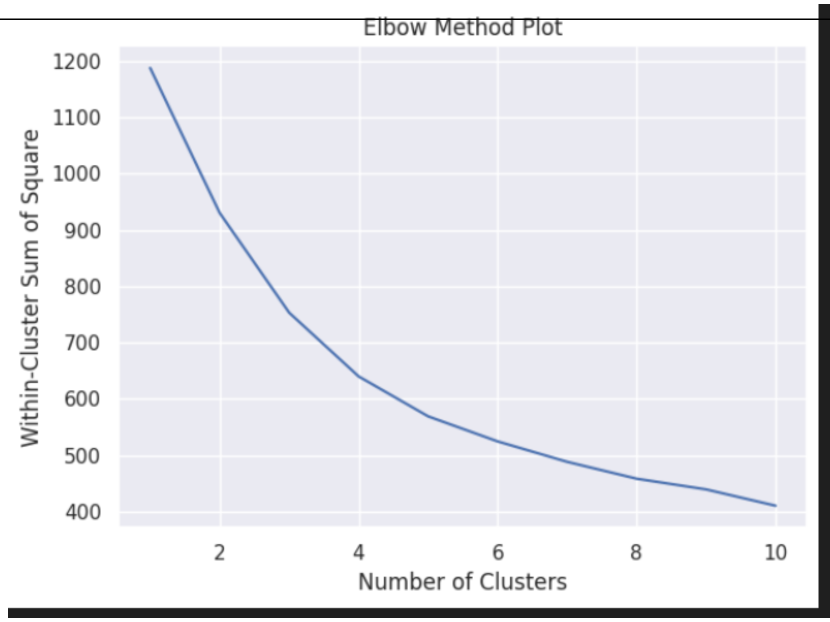
- Features are standardized using **StandardScaler** so all variables (Age, Salary, Loans, etc.) contribute equally. K-Means is sensitive to scale.

Choosing Optimal Clusters using Elbow Method

Code:

```
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', random_state=42)
    kmeans.fit(X_scaled)
    wcss.append(kmeans.inertia_)

plt.plot(range(1, 11), wcss)
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.title('The Elbow Method')
plt.show()
```



- **Within-Cluster Sum of Squares (WCSS)** is calculated for k=1 to k=10.
- The **Elbow point** (where WCSS starts to flatten) indicates the **optimal number of clusters**. Based on your code's output, this appears to be at **k = 3**.

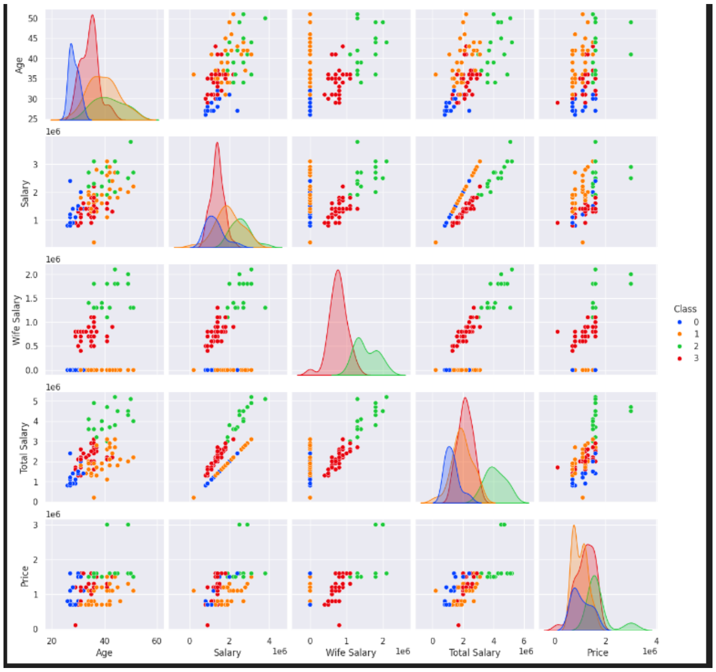
Interpretation:
3 clusters are ideal to divide Indian EV buyers into distinct behavioral and financial groups.

Fitting the K-Means Model
<u>Code:</u>
```
kmeans = KMeans(n_clusters=4, init='k-means++', random_state=42)
y_kmeans = kmeans.fit_predict(X_scaled)
```

- K-Means is trained using k=3 clusters.
- y_kmeans holds the **cluster label (0, 1, or 2)** assigned to each row (customer).

Visualizing Clusters
<u>Code:</u>
```
df_encoded['cluster'] = y_kmeans

plt.figure(figsize=(10, 7))
sns.scatterplot(data=df_encoded, x='Salary', y='Price', hue='cluster', palette='viridis')
plt.title('Clusters based on Salary and Price')
plt.xlabel('Salary')
plt.ylabel('Car Price')
plt.legend(title='Cluster')
plt.grid(True)
plt.show()
```



- This 2D scatter plot shows how customers are grouped by **Salary** and **Car Price**.
- **Each color** represents a different cluster, e.g.:
  - **Cluster 0**: Low salary, low price (budget buyers)
  - **Cluster 1**: Mid salary, mid-high price (value-seeking professionals)
  - **Cluster 2**: High salary, high price (premium market)

Insights for Marketing:

- Cluster 0: Can be targeted with **affordable EVs**, EMI schemes.
- Cluster 1: Interested in **value-for-money features** (range, brand, safety).
- Cluster 2: Attracted to **premium features** (design, brand status, performance).

Cluster Profiling

You can analyze means of features for each cluster to describe them:

<u>Code:</u>

```
df_cluster_profile = df_encoded.groupby('cluster').mean()
print(df_cluster_profile)
```

- The result shows average Age, Salary, Education level, loan status, etc. per cluster.
- Helps marketers understand **psychographic and financial patterns** within each group.

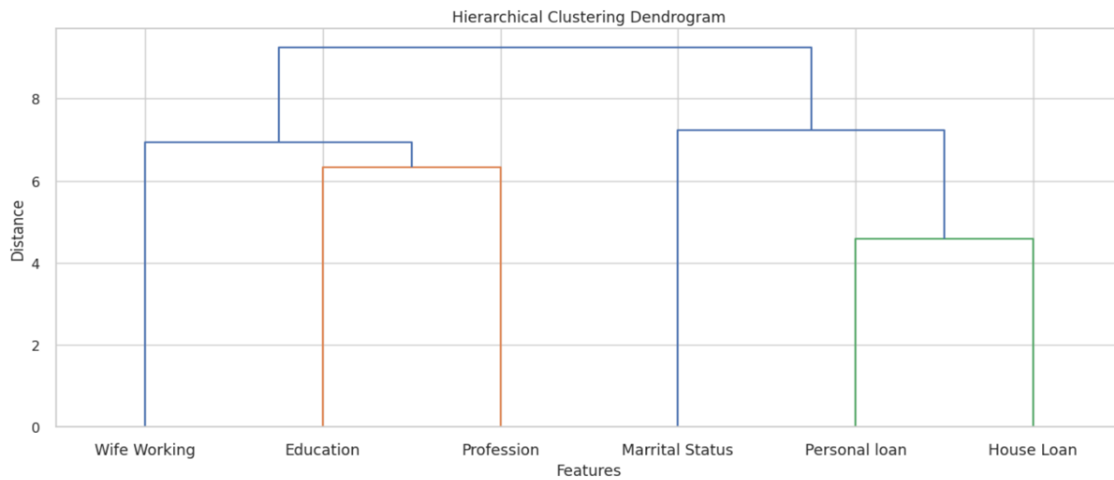
□ **Example**:

- Cluster 2 may have highest average education and income → early adopters of new technology.
- Cluster 1 may have a mix of professionals with family responsibilities.

  Cluster 0 could be new workforce entrants or small-town customer

Hierarchical Clustering:

Hierarchical clustering is an unsupervised machine learning technique used to group similar data points into clusters based on distance or similarity. Unlike K-Means, it doesn't require the number of clusters to be pre-specified. Instead, it builds a hierarchy of clusters either from the bottom-up (**agglomerative**) or top-down (**divisive**) and represents this as a **dendrogram**.



Insights from the Clustering:

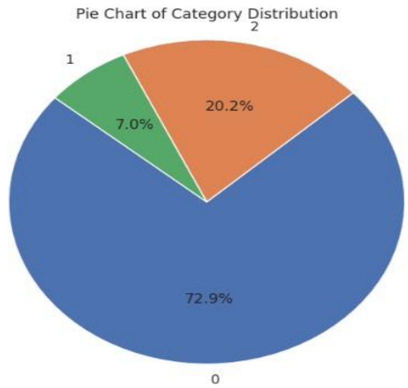
1. **Cluster Identification**: From the dendrogram, the data naturally divided into **3–4 main clusters**, each representing distinct consumer groups.
2. **Income-Based Grouping**: Households with **high combined income** formed distinct clusters, often showing a preference for **SUVs and luxury vehicles**.
3. **Age Influence**: A specific cluster emerged for **younger customers** (age 26–34), who were more inclined toward **hatchbacks and sedans** like the **i20** or **Baleno**.
4. **Loan Independence**: The clustering process reaffirmed that **loan status (house or personal)** didn't significantly influence customer segmentation. Many high-income groups with no loans showed similar purchasing behavior to those with loans.
5. **Vehicle Price Sensitivity**: Even among higher-income segments, **moderately priced EVs** were often preferred, suggesting **value-for-money** is a strong consideration, not just affordability.

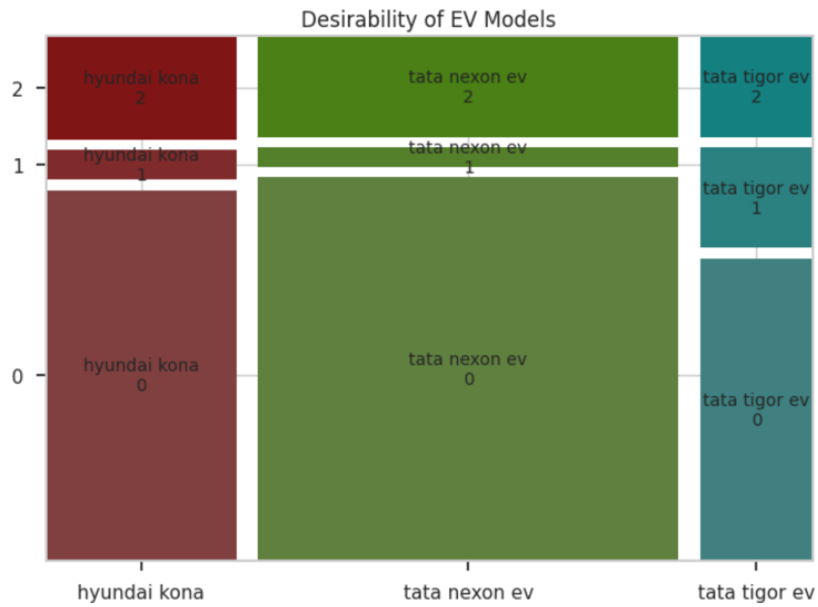
Extracting Segment

Segment 0 : High satisfaction on every aspect ,indicates higher willing to adoption of EV, forms most of the population - 72.9%; Most desired EV is Tata Nexon; These are Early majority

Segment 1 : Unsatisfied on all aspects, didn't find EV to be value for money ; forms 7% of population ; Found Tata tigor EV to be most likeable; can be considered Late adapters.

Segment 2: They liked Exterior and Comfort of EVs but didn't find it Value for money much; form 20.2% population ;Found each EV equally likeable; can be considered Early adaptors





General Consumer Profile

Based on the analyzed dataset, the **typical Indian electric vehicle (EV) buyer** exhibits a consistent demographic and financial pattern. The **average age** of the buyer is around **36 years**, indicating a mature yet energetic consumer base. These individuals are generally **salaried professionals**, well-educated, and **married**—reflecting stability in both career and lifestyle.

The typical buyer **has two dependents**, **no ongoing personal or home loans**, and often has a **working spouse**. This results in a **combined household income of approximately ₹22 Lakhs per annum**. With this level of income and financial planning, they are most likely to **opt for an SUV**—a vehicle that offers both status and practicality—**within a budget of ₹12 Lakhs**.

This consumer segment represents the **central trend in EV purchasing behavior**, providing a focused target audience for manufacturers and marketers aiming to capture the Indian electric four-wheeler market.

.

Key Findings from the Dataset:

1. Strong Relationship Between Car Price and Total Salary

- As household income increases, so does the price of the vehicle purchased.
- Buyers with **higher total earnings** (especially when both spouses are working) are more likely to purchase **SUVs and luxury vehicles**.
- This confirms that **spending capacity is the major factor** in vehicle selection, particularly in the EV market.

2. Number of Dependents Has No Influence on Car Selection

- The data reveals **no significant impact** of having children or other dependents on the type or price of vehicle purchased.
- This indicates that **family size does not influence EV purchase decisions**, at least within the observed price ranges.

3. Age Affects Salary but Not Car Price

- Older consumers tend to earn higher salaries, but this **does not directly translate into higher vehicle prices**.
- The choice of car seems more influenced by **salary** than by **age** itself.

Data Exploration on Buying behaviour:

Behavioral Trends and Preferences:

Younger Buyers:

- Usually in their **20s or early 30s**
- Prefer **hatchbacks** such as **Baleno** and **i20**
- Occasionally opt for **sedans**
- Rarely choose **SUVs**, and **do not purchase luxury vehicles**
- Tend to be in **lower income brackets**

Older Buyers:

- Prefer more **comfort-oriented vehicles**, such as **premium sedans** (like **Ciaz**) or **luxury models**
- Likely have **greater financial stability**
- Typically fall into **higher age and income brackets**

High Salary Earners:

- Tend to go for **SUVs or luxury vehicles**
- Their choices are driven by **lifestyle, comfort, and brand value**
- Often include **dual-income households** (i.e., both husband and wife are earning well)

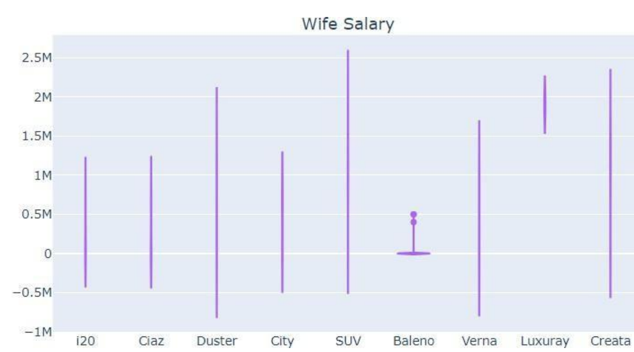
Impact of Wife's Working Status:

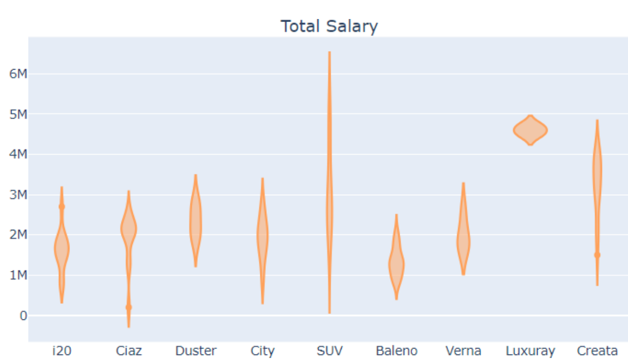
- A **working wife** contributes to a **higher total salary**, which often shifts vehicle preference towards **luxury cars and premium SUVs**
- Dual-income households generally **spend more** on EVs and are less price-sensitive

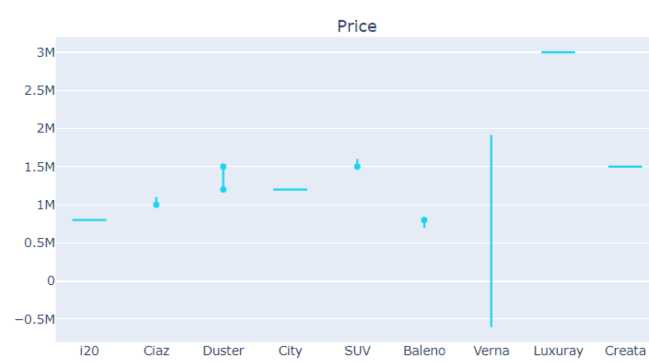
Popular Car Choices Based on Income

- **i20 and Baleno (Hatchbacks)**:
  - Most popular among **low-income** and **younger buyers**
  - Represent affordable, fuel-efficient, and city-friendly options

- **SUVs (e.g., Creta, Nexon EV)**:
  - Chosen by **middle to upper-middle income buyers**
  - Appeals to families or individuals wanting more space and status
- **Luxury Vehicles (e.g., MG ZS EV, BMW iX)**:
  - Attract **high-income** individuals or dual earners
  - Offer prestige and advanced technology
- **Ciaz (Sedan)**:
  - Often selected by **older buyers** looking for comfort and elegance

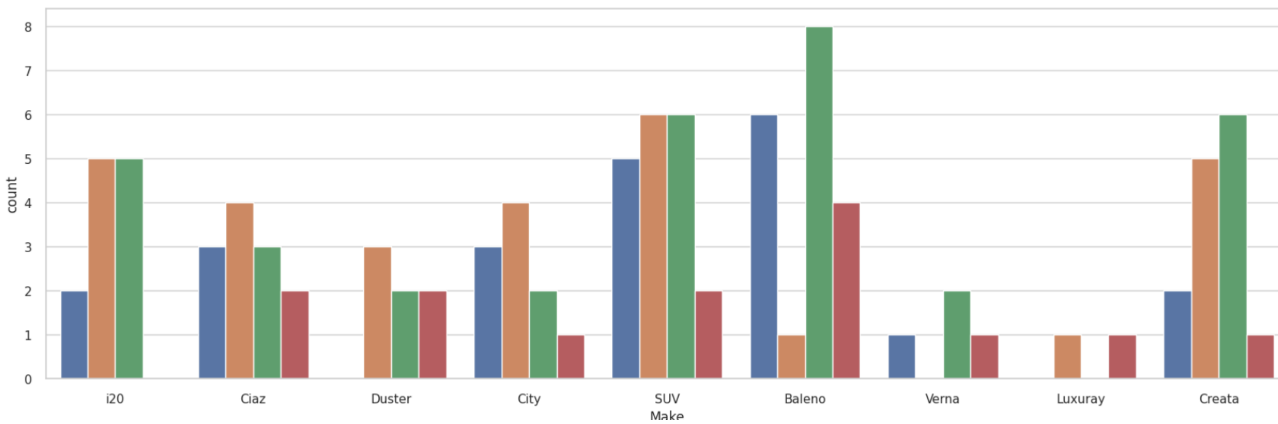






Notable Findings:

1. **SUVs holds most of the market capital** with their primary Customers being **Post Graduate ,High Income Married , Salaried Professionals with Working wife**; **Singles** of this group prefer **City or Creta** (Sedan/Compact SUV)
2. Second Fav Vehicle in this segment is **Creta**
3. **Highly Educated Married Businessmen** Prefer **Hatchback &sedan** whereas Single of this catagory also prefer SUV
4. Overall Post graduates Salaried Professionals prefer SUV and Businessmen prefer Baleno So we can say These two are most popular catagories.
5. Overall Graduates Salaried Professionals prefer Ciaz and Businessmen prefer SUV So we can say These two are most popular catagories.
6. Overall Most Popular vehicle segments in decreasing order are : **Suv** > **Baleno(Hatchback)** > **CIAZ(Premium sedan)**



DEMOGRAPHIC SEGMENTATION:

Demographic segmentation involves dividing the market based on demographic variables such as **age, gender, income, marital status, education level, profession, and family size**. This approach helps businesses understand *who* their customers are, enabling more tailored and effective marketing strategies.

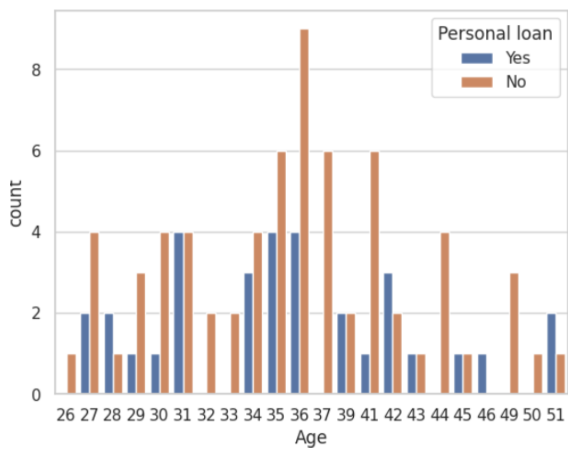
Describing Datasets:

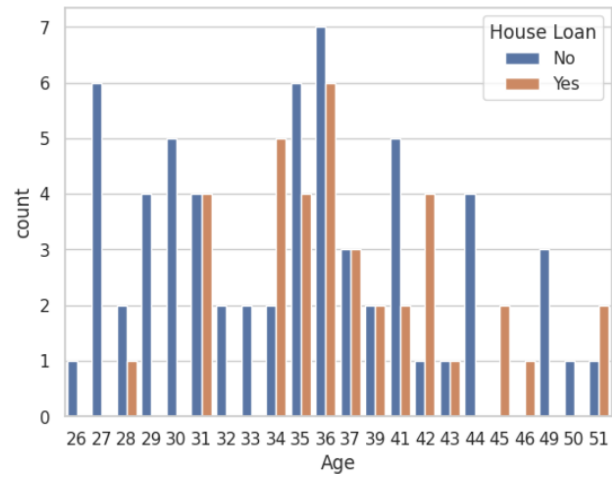
	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	House Loan	Wife Working	Salary	Wife Salary	Total Salary	Make	Price
0	27	Salaried	Single	Post Graduate	0	Yes	No	No	800000	0	800000	i20	800000
1	35	Salaried	Married	Post Graduate	2	Yes	Yes	Yes	1400000	600000	2000000	Ciaz	1000000
2	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

```
Age : [27 35 45 41 31 28 33 34 29 30 49 26 37 36 43 42 32 44 39 46 50 51]
Profession : ['Salaried' 'Business']
Marrital Status : ['Single' 'Married']
Education : ['Post Graduate' 'Graduate']
No of Dependents : [0 2 4 3]
Personal loan : ['Yes' 'No']
House Loan : ['No' 'Yes']
Wife Working : ['No' 'Yes' 'm']
Salary : [ 800000 1400000 1800000 1600000  900000 1200000 2000000 1300000 2500000
 1700000 1100000 1900000 2100000 2400000 2200000  200000 1500000 2700000
 2900000 3100000 2600000 2300000 2800000 3800000]
Wife Salary : [      0  600000  800000  700000  400000 2000000  500000 1000000 1100000
  900000 1300000 1400000 1800000 2100000]
Total Salary : [ 800000 2000000 1800000 2200000 2600000  900000 1400000 1900000 1700000
 1300000 4500000 2500000 2400000 2900000 1600000 2700000 1100000 2100000
 3000000 3700000 2300000 3600000  200000 3100000 4300000 3800000 4700000
 1200000 1500000 4000000 3200000 5200000 4100000 4900000 2800000 5100000]
Make : ['i20' 'Ciaz' 'Duster' 'City' 'SUV' 'Baleno' 'Verna' 'Luxuray' 'Creata']
Price : [ 800000 1000000 1200000 1600000  700000 1100000  110000 3000000 1300000
 1500000]
```

Loan:

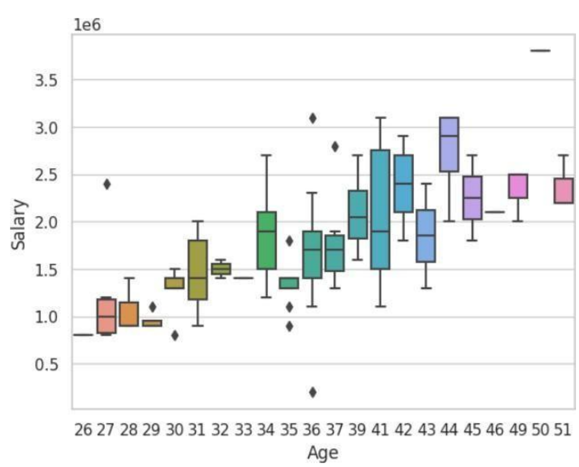
Although Personal and home loans do not seem to have any significant impact on EV purchasing patterns.

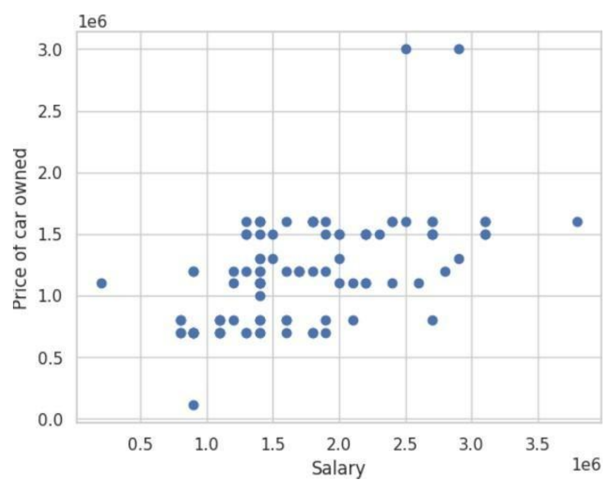




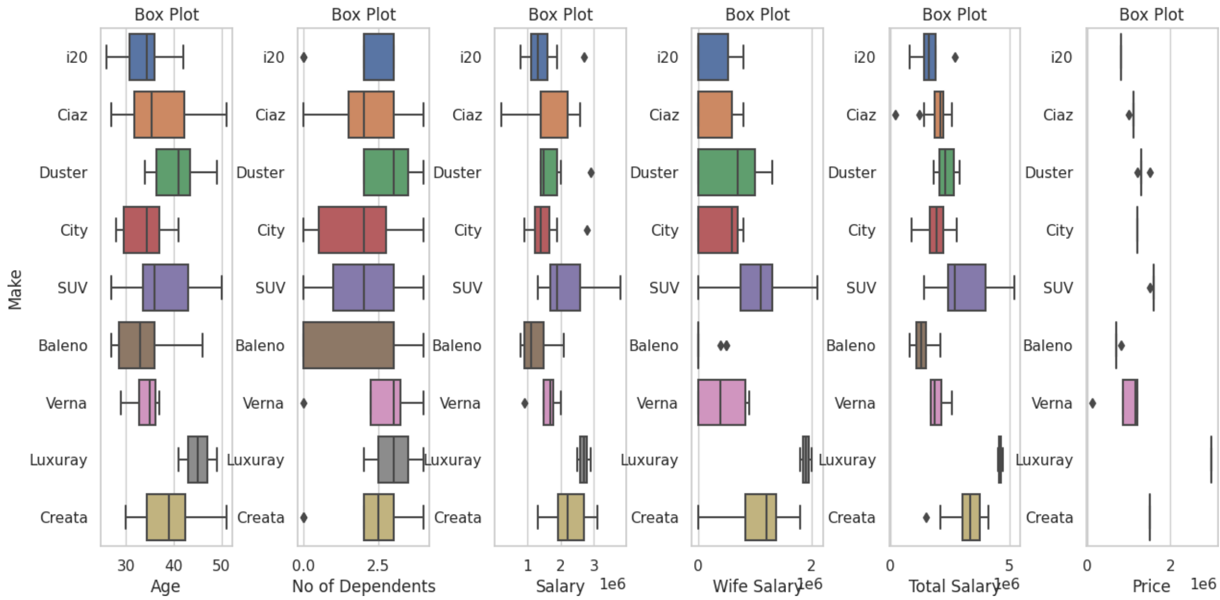
Salary:

It is to be noted that regardless of even high salaries, average to low price-ranged cars are being preferred.

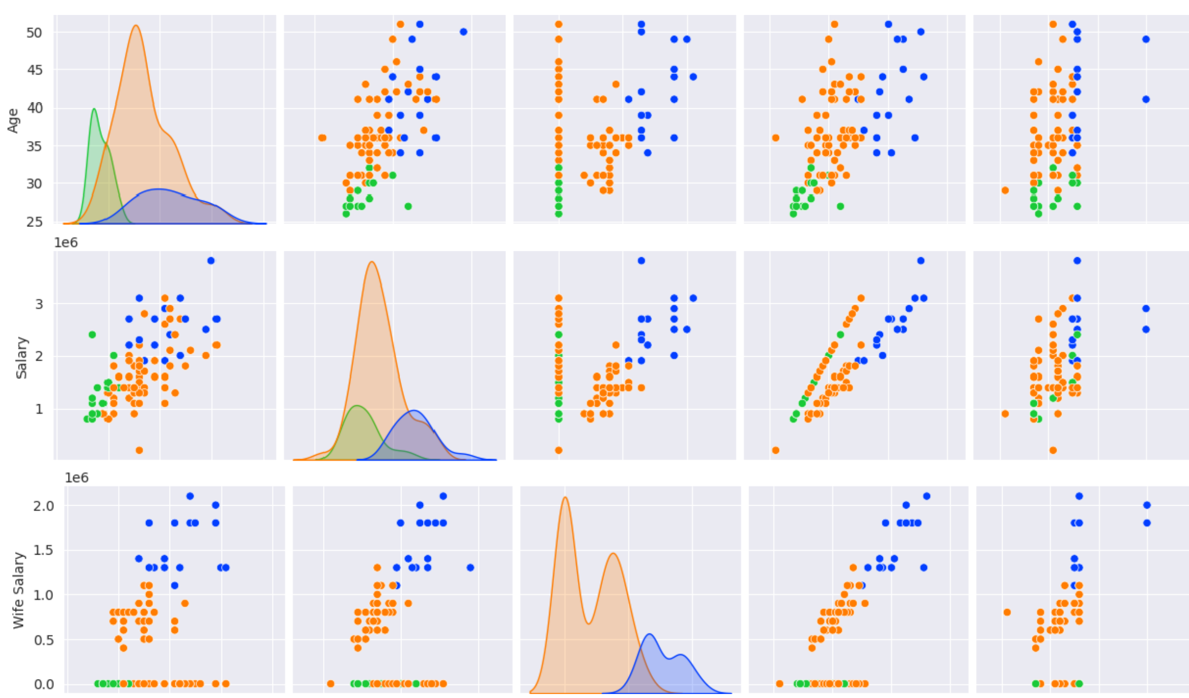


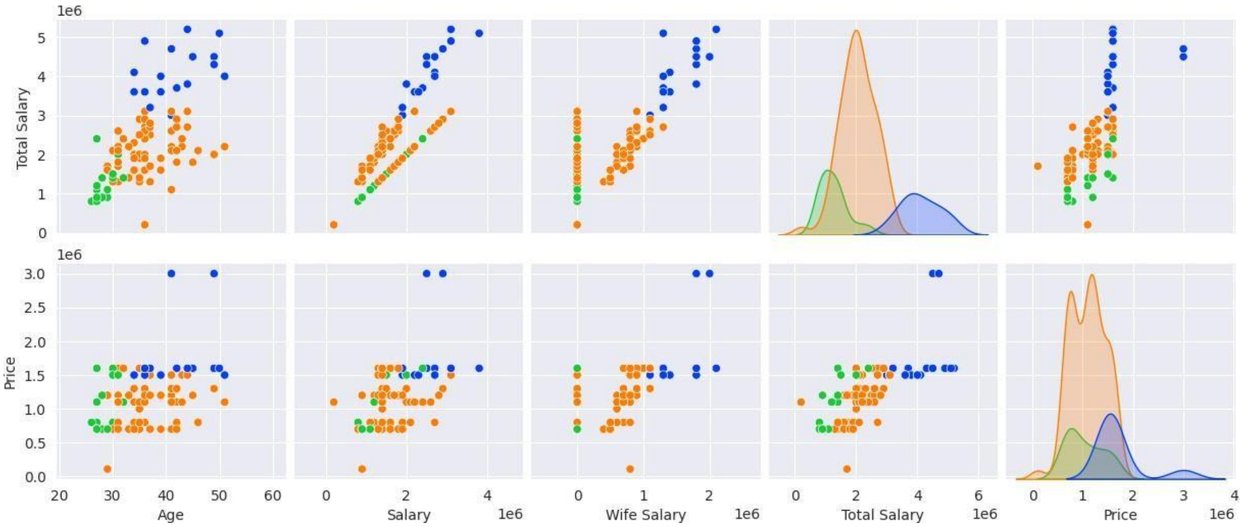


We see the people buying over a range of prices when it comes to Verna, whereas in case of other brands, probably some specefic models are being preferred.



Notable findings:





we see that we can get a more meaningful analysis for k = 3 Let's see the characteristics of every cluster:

Cluster 1 (Green): Prefer low to moderately priced Electric vehicles

- Age group 20-30
- Moderate to average salary
- Wife's salary - Low

Cluster 2 (Orange) : Prefer low to moderately priced Electric vehicles

- Age group 30-45
- Average to high salary
- Wife's salary - low to

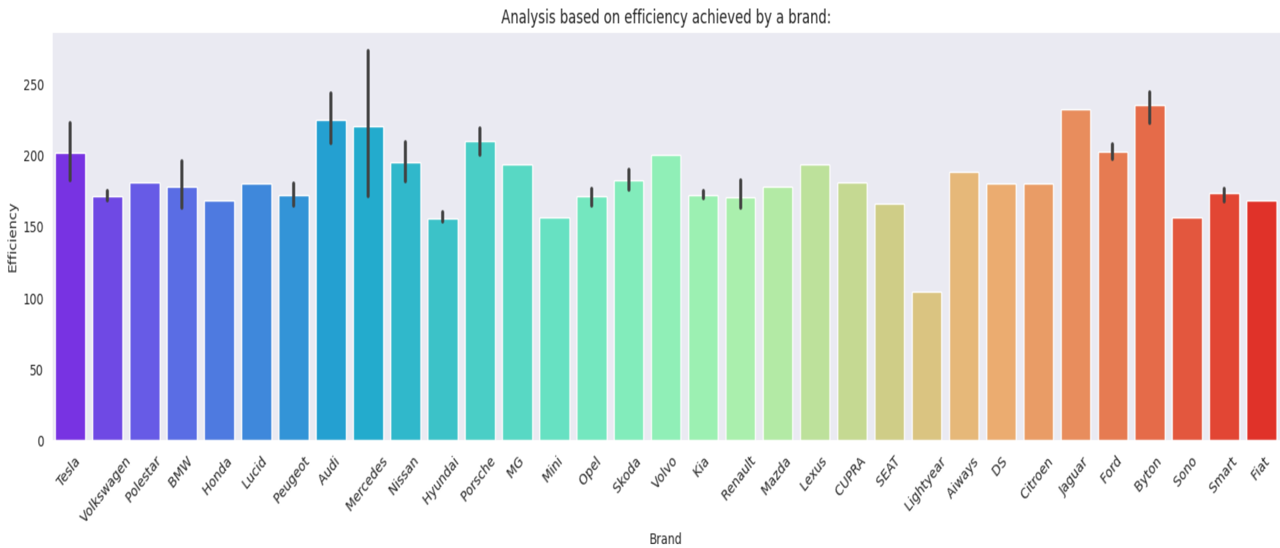
Cluster 3 (Blue) : Prefer moderate priced Electric vehicles and experiment with the high priced segment.

- Age group 30-45
- Average to very high
- Wife's salary - high

K-Means Clustering Insights

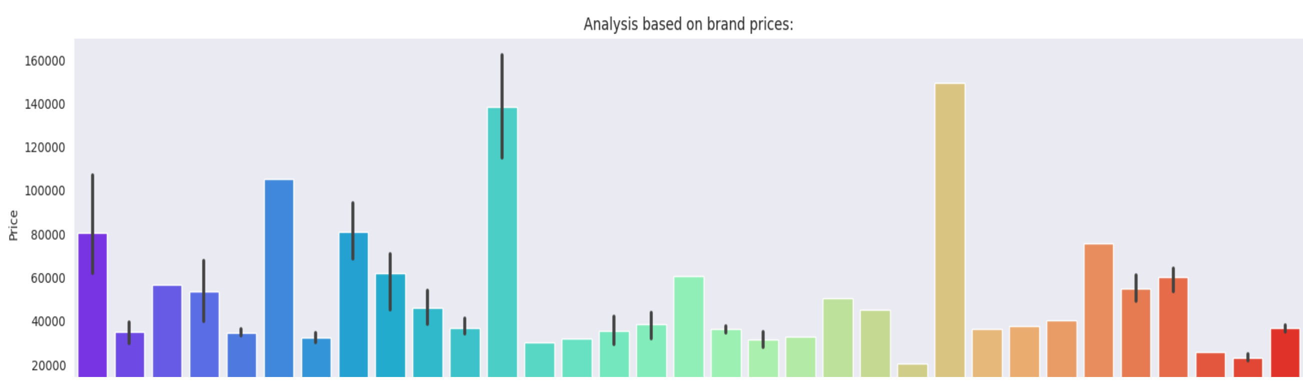
1. **Moderate EV Pricing Is Most Preferred**
   Across all identified clusters, the majority of consumers tend to prefer **electric vehicles in the moderately priced range** (especially under ₹12–15 lakhs). This trend is consistent regardless of income or profession, showing a practical and cost-conscious mindset.
2. **High Purchase Activity in Age 26–38**
   The most active age group in terms of purchasing EVs is between **26 to 38 years old**. This group represents young professionals and small families who are likely early adopters of new technologies and environmentally aware.
3. **High Salary ≠ High Vehicle Price**
   Despite having higher total salaries (including spouse income), many consumers still prefer **economically priced cars**. Personal loans and home loans **do not seem to influence** this decision much.
4. **Verna Shows Diverse Buying Patterns**
   Customers buying Verna exhibit a **wide price range**, suggesting different trims and versions are popular. In contrast, cars like Baleno and i20 are bought in **specific price brackets**, showing a **focused preference for certain variants**.
5. **Price and Total Income Are Closely Related**
   A clear **positive correlation** is observed between **total salary** and **vehicle price**. When both partners earn well, the tendency is to shift towards **premium or luxury EVs**.

Analysis based on efficiency achieved by a brand:



Analysis based on brand price:

This is the price distribution of Electric Vehicles visualised over different brands.

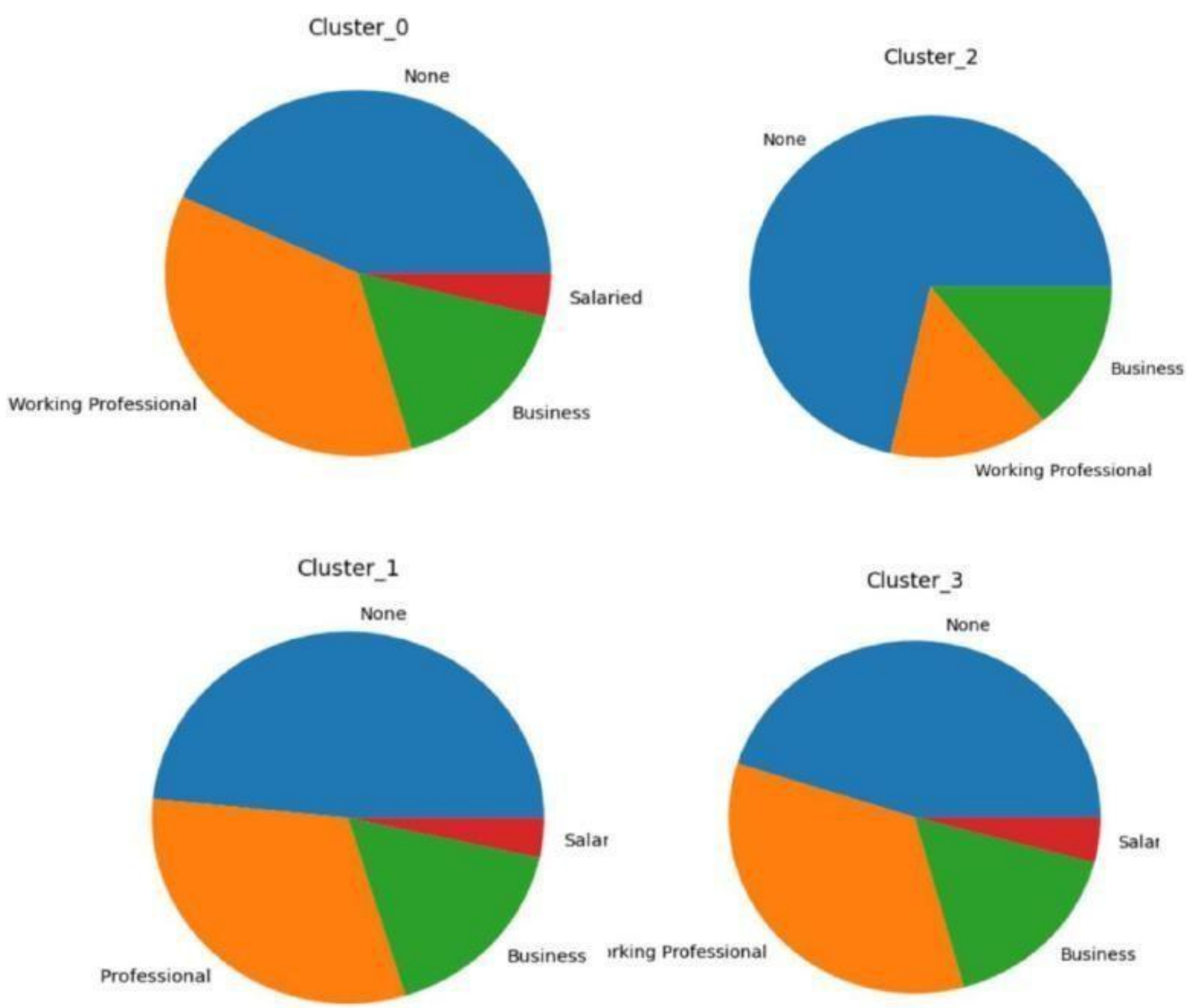


Optimal Target Profile:

Next, I aim to address the posed questions by identifying potential target segments. The selection of target variables and clusters can vary based on the company's policies and objectives. Here, I have identified possible target clusters.

The EV Manufacturer Company might choose the 1st Segment (Demographical profile) having feature values-

- *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*



The EV Manufacturer Company might choose the 2st Segment (Geographical profile) having feature values-

- High EV Adoption, Moderate Charging Infrastructure, Moderate EV Charging Sanctions.
- States - Tamil Nadu and Chhattisgarh
- Low to Moderate EV Adoption, Low Charging Infrastructure, Low EV Charging Sanctions.
- Odisha, Punjab, Bihar, Assam, Haryana, Ladhakh, Sikkim, Jharkhand, Puducherry, Goa, Jammu Kashmir
- High EV Adoption, High Charging Infrastructure, High EV Charging Sanctions.
- States: Karnataka And Delhi
- Moderate EV Adoption, Moderate Charging Infrastructure, High EV Charging Sanctions.
- Maharashtra, Rajasthan, West Bengal, Gujarat, Kerala
- High EV Adoption (Mainly Three Wheelers and Four Wheelers), Low Charging Infrastructure, Low EV Charging Sanctions.
- States: Uttar Pradesh

Possible Improvements

While the current segmentation analysis has yielded valuable insights, there is still room for refinement and deeper exploration:

- **Advanced Clustering Algorithms:**
  Utilizing more sophisticated algorithms such as **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)** or **Gaussian Mixture Models** can uncover more nuanced or overlapping customer segments that K-Means may not identify due to its rigid assumptions (e.g., spherical clusters).
- **Richer and Cleaner Data:**
  Enhancing data quality—by reducing missing values, correcting inconsistencies, and normalizing ranges—can significantly improve the segmentation outcomes. Additionally, introducing **more diverse and relevant features** such as vehicle preferences by region, psychographic

variables, or real-time usage patterns would lead to more precise clustering.

- **Data Expansion:**
  To strengthen future analyses:
  - **Web scraping** could extract real-time customer reviews, dealership ratings, or feature preferences.
  - **Manual data collection** from EV showrooms or test drives can capture experiential and service-related feedback.
  - **Surveys** can provide detailed insights into behavioral intent, environmental attitudes, and brand perceptions, which are otherwise hard to quantify.

Market Mix Strategy for EV Adoption

To successfully position an EV product in India's diverse market, companies must optimize their marketing mix across **four key pillars**:

Innovation

Continual innovation is essential to stay ahead in the competitive EV landscape. Critical areas include:

- Development of **advanced battery technologies** for longer range and shorter charging time.
- Integration of **autonomous driving** capabilities.
- Enhancing **smart connectivity**, such as voice assistants, real-time diagnostics, and app-controlled features.
- Ongoing **design and performance upgrades** to match evolving consumer tastes.

Infrastructure

EV adoption hinges on the development of a reliable and accessible support ecosystem:

- Expansion of **public and private charging stations** across urban and rural locations.
- Setup of **service and maintenance networks** trained for EV-specific repairs.
- **Smart grid integration** for efficient energy management.
- **Government and private partnerships** to fast-track infrastructure rollout, including battery-swapping stations and solar-powered chargers.

Customer Experience

Delivering a seamless and satisfying experience can build trust and long-term brand loyalty:

- **Comprehensive after-sales service** packages including warranty, roadside assistance, and service scheduling.
- Easy-to-use **mobile apps** for navigation, charging, diagnostics, and vehicle customization.
- Personalized communication channels and **dedicated support teams**.
- Conducting **test drives, awareness campaigns, and educational workshops** to familiarize customers with EV benefits and usage.

Sustainability

Sustainability is not just a trend but a competitive differentiator in today's market:

- Use of **biodegradable or recycled materials** in manufacturing.
- Establishing **battery recycling and reuse programs** to reduce waste.
- Reducing **manufacturing emissions** through cleaner energy sources and green logistics.

Conclusion

This report offers a comprehensive analysis of India's four-wheeler electric vehicle (EV) market through a data-driven segmentation approach. The findings indicate that **consumer EV adoption** is influenced by a combination of **product perception**, **personal income**, and **lifestyle factors**.

The use of **machine learning techniques like PCA and K-Means clustering** has helped uncover distinct customer profiles, shedding light on their behavior, preferences, and motivations. This enables automakers to tailor offerings and marketing strategies with greater accuracy.

Moreover, the growing interest in EVs—fueled by rising fuel prices and environmental concerns—presents a significant opportunity for both existing and new market entrants. Leveraging advanced segmentation, cleaner data, and targeted market mix strategies can enhance customer alignment, improve infrastructure planning, and ultimately drive EV adoption.

In conclusion, the **fusion of market research and machine learning** empowers automotive companies to make **informed, data-backed decisions** that align with evolving customer expectations and regulatory shifts, securing a competitive edge in the green mobility revolution.