EV Market Segmentation Analysis Report

Joyashree Mandal

28/04/2025

Analysing the Electric Vehicle market in India using Segmentation analysis for an Electric Vehicles Startup and coming up with a feasible strategy to enter the market, targeting the segments most likely to use Electric vehicles.

Github Link
https://github.com/joyashre/EV-Market-Segmentation

SETUP REQUIREMENTS :

This project requires Python and the following Python libraries installed:

- **NumPy:** Scientific Computing Library
- **Pandas:** Data Analysis Library
- **Matplotlib:** Data Visualization Library
- **Seaborn :** Data Visualization Library
- **Scikit-learn:** Machine Learning Library

DATA SOURCES :

- **EVStats.csv:** For region based analysis of different types of Electric Vehicle sales in India for the year 2018.

FOR DATASETS: https://electricvehicles.in/electric-vehicles-sales-report-in-india-2018/

- **Indian automoble buying behavour study 1.0.csv:** For behavioral, psychographic and demographic analysis of Indian automobile market.

FOR DATASETS: https://www.kaggle.com/karivedha/indian-consumers-cars-purchasing-behaviour

- # 🔍 Exploratory Data Analysis (EDA)

🔍 Exploring the Dataset

The dataset contains the following features:

1. 🎂 **Age** – Age of the buyer
2. 👔 **Profession** – Occupation of the buyer
3. 💍 **Marital Status** – Marital status (Married/Single)
4. 🎓 **Education** – Highest education level
5. 👶 **No. of Dependents** – Number of dependents
6. 💳 **Personal Loan** – Has a personal loan (Yes/No)
7. 🏠 **House Loan** – Has a house loan (Yes/No)
8. 👩 **Wife Working** – Is the wife working? (Yes/No)
9. 💰 **Salary** – Buyer's salary
10. 💵 **Wife Salary** – Wife's salary
11. 🏦 **Total Salary** – Combined household salary
12. 🚗 **Make** – Car brand purchased
13. 🛒 **Price** – Price of the car purchased

```
data.head()
```

	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

🧹 Data Cleaning & Encoding

Now we clean the data. We handled missing values, fixed inconsistent entries(like 'm' → 'No'), and converted categorical variables into numeric format for analysis 🔄📊.

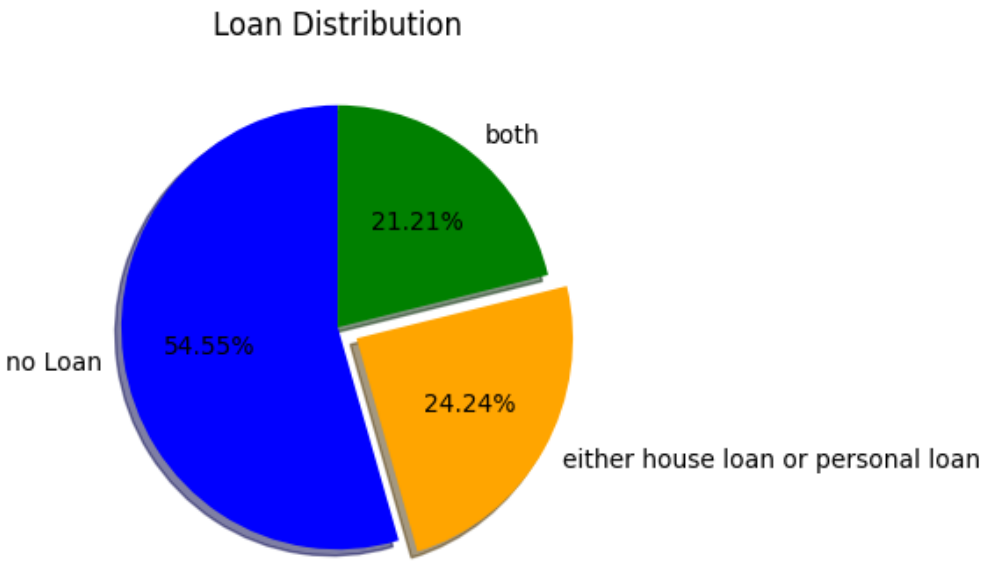
```
data.loc[data['Wife Working']=='m']
```

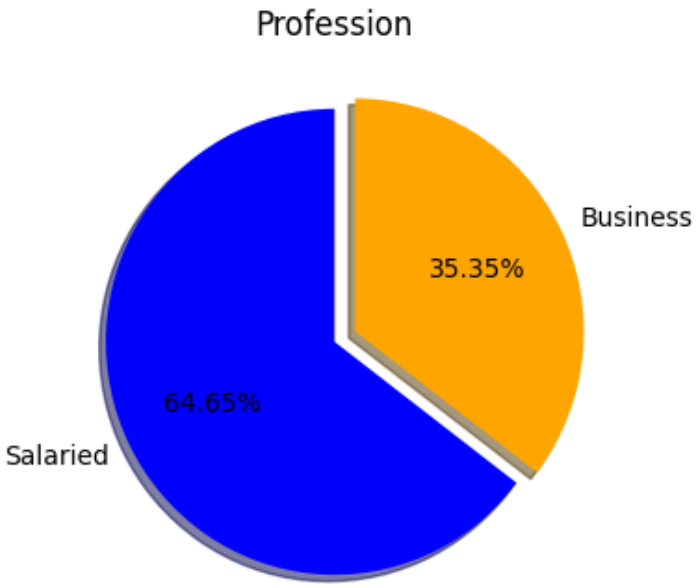
	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	House Loan	Wife Working	Salary	Wife Salary	Total Salary	Make	Price
11	35	Salaried	Married	Graduate	4	Yes	Yes	m	1400000	0	1400000	Baleno	700000

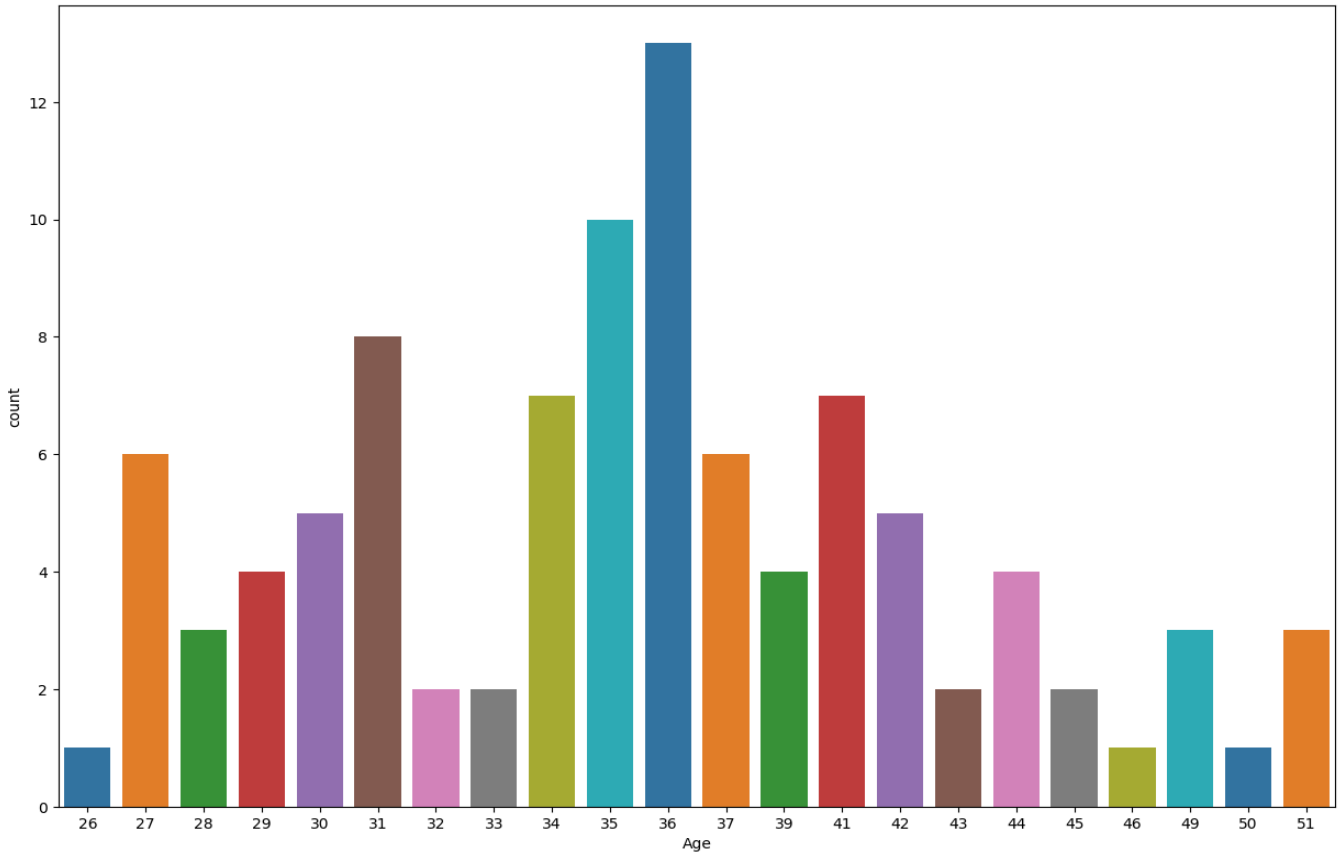
🧹 Encoding

	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	House Loan	Wife Working	Salary	Wife Salary	Total Salary	Make	Price
0	27	1	1	1	0	1	0	0	800000	0	800000	8	800000
1	35	1	0	1	2	1	1	1	1400000	600000	2000000	1	1000000
2	45	0	0	0	4	1	1	0	1800000	0	1800000	4	1200000
3	41	0	0	1	3	0	0	1	1600000	600000	2200000	2	1200000
4	31	1	0	1	2	1	0	1	1800000	800000	2600000	6	1600000

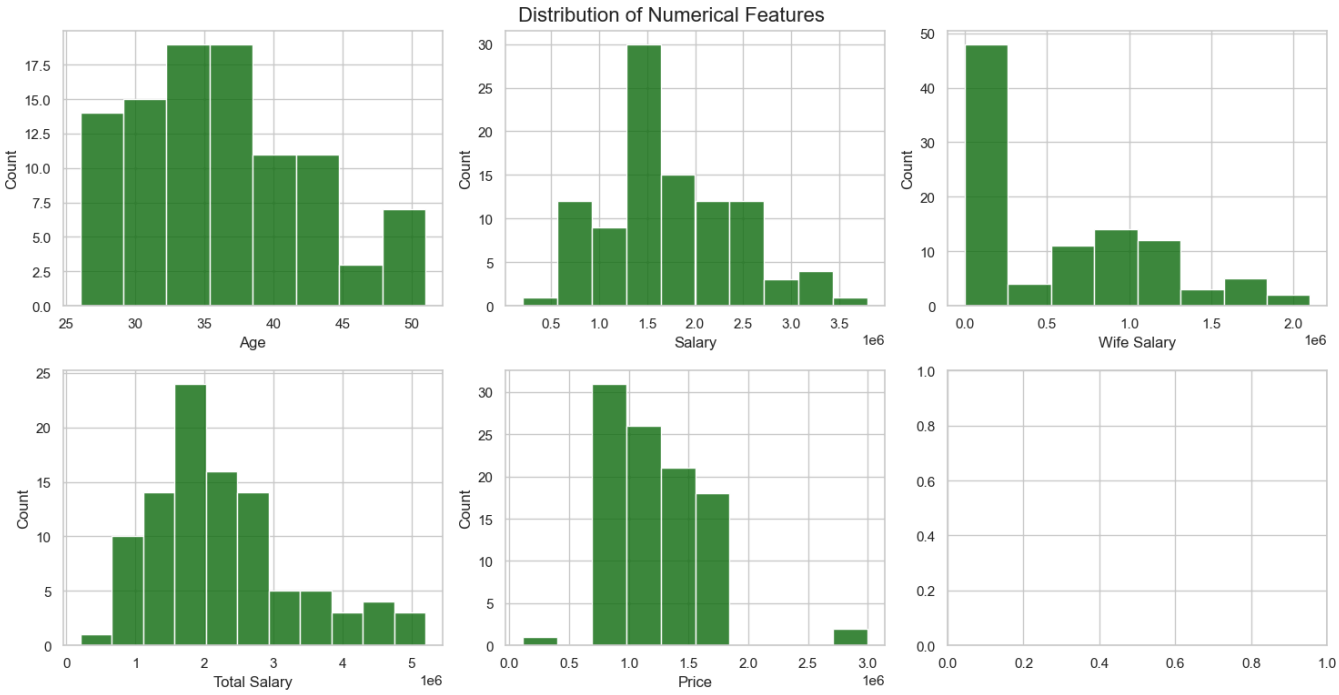
Now , We explored loan patterns, profession types, and consumer demographics through insightful visualizations 📊🎯 — revealing trends, correlations, and behaviors in vehicle purchasing habits 🚗💡 .







- From the above graph we can see Age 36 mostly buy vehicles.



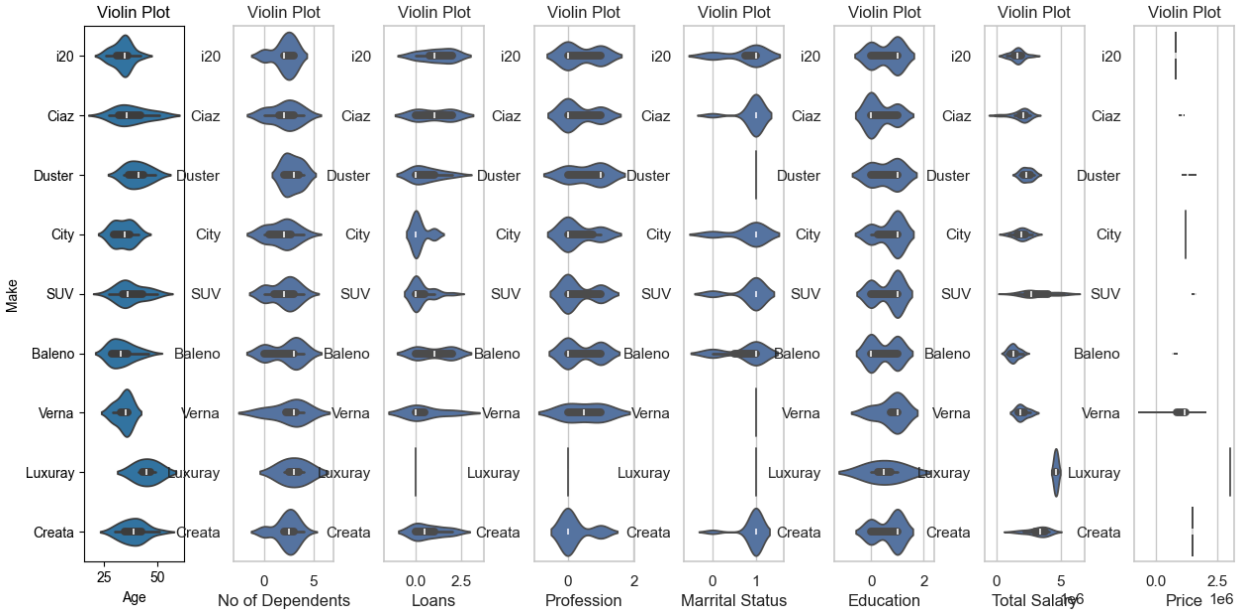
🔍 Key Observations:

- ★ Average Age of car buyer is around 36 and 92% of customers fall between the age of 25 - 45.
- ★ Salaries are predominantly found between 1 million and 2.5 million, with 72% of customers falling within this range.
- ★ 79% of Customers buy cars with Price range of 7 Lakhs to 16 lakhs.
- ★ A majority of customers have working spouses which is around 52%.
- ★ The household income for 73% of customers falls between 1 and 3 million.

🎯 Behavioral & Psychographic Analysis

Using violin plots, we explored how features like age, salary, and dependents etc. relate to the type of vehicle consumers purchase.

These insights reveal patterns in buying behavior and lifestyle choices 🚗🧠.

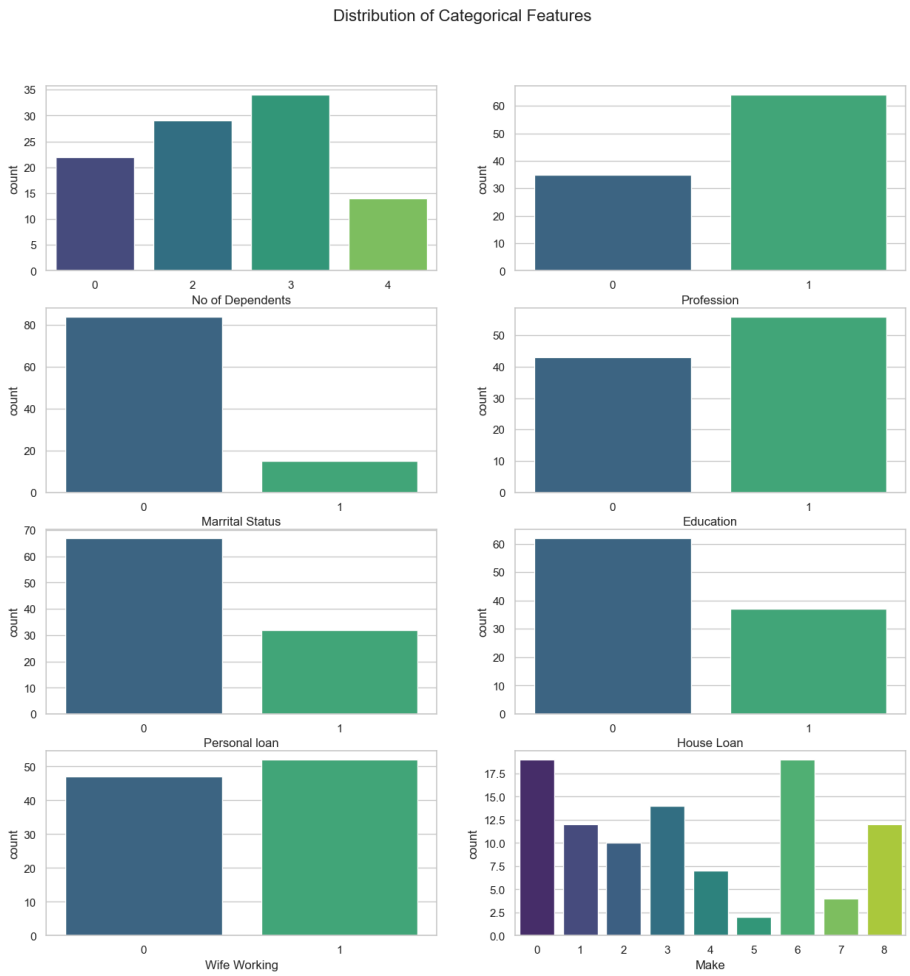


- 🎻📊 **Violin Plot Observations:**

1. Age : Younger Consumer Purchase Less Expensive Vehicles.

2. Number of Dependents: greater the number Of Dependents Makes the consumer buy a vehicle with more seats.

3. Salary: If you overlap the normalised salary plots with price plot, you would observe the median of salary violin plot matches that of the price of the vehicle indicating a very direct relationship.



- 🔍 **Key Observations:**

★ Most of the customers have 2 to 3 dependents

★ Approximately 65% of the customer base is employed in salaried positions, while 35% are self-employed.

★ Approximately 15% of the customers are single, while the remaining 85% are married.

★ Approximately 55% customers hold a Post Graduate degree while 35% have a Graduate degree.

★ Approximately 68% of the customers do not have a personal loan.

★ Approximately 62% of the customers do not have a house loan.

★ A majority of customers have working spouses which is around 52%.

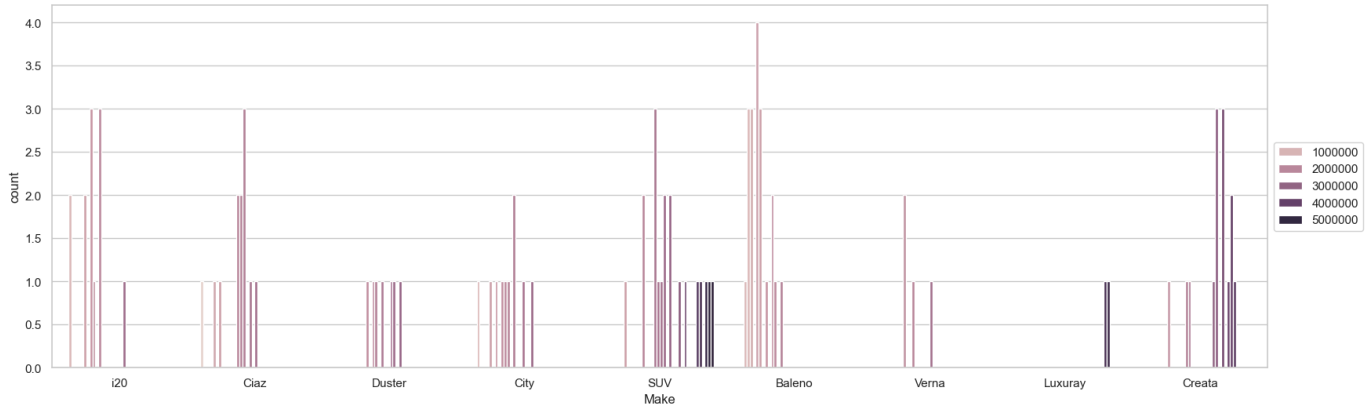
★ The most popular vehicle purchases among customers are SUVs and Balenos followed by Cretas.

- 💸 **Salary vs. Vehicle Purchase Behaviour**

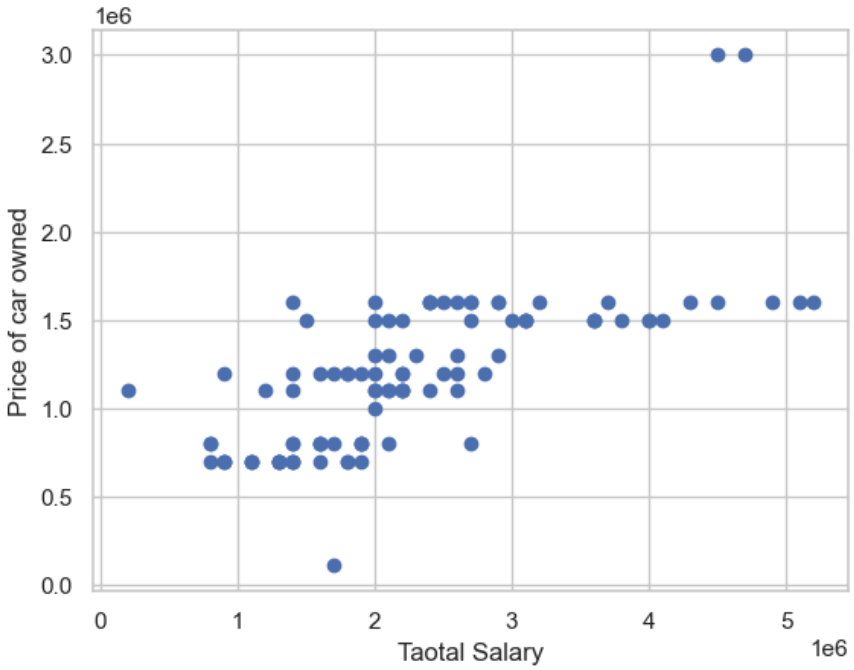
We examined how total salary influences vehicle make and price.

Turns out, higher salaries often correspond to more expensive cars—makes sense, right? 💼➡️🚗

>> make vehicles they tend to purchase



>> Price of vehicle owned



This scatter plot explores how a customer’s total salary influences the price of the car they own.

- 🔍 **Key Observations:**

🔄 There's a positive correlation — higher salary individuals generally own more expensive cars.

💡 However, the relationship appears non-linear beyond a certain point (~1.5M), as many car prices plateau around 1.5M despite rising salaries.

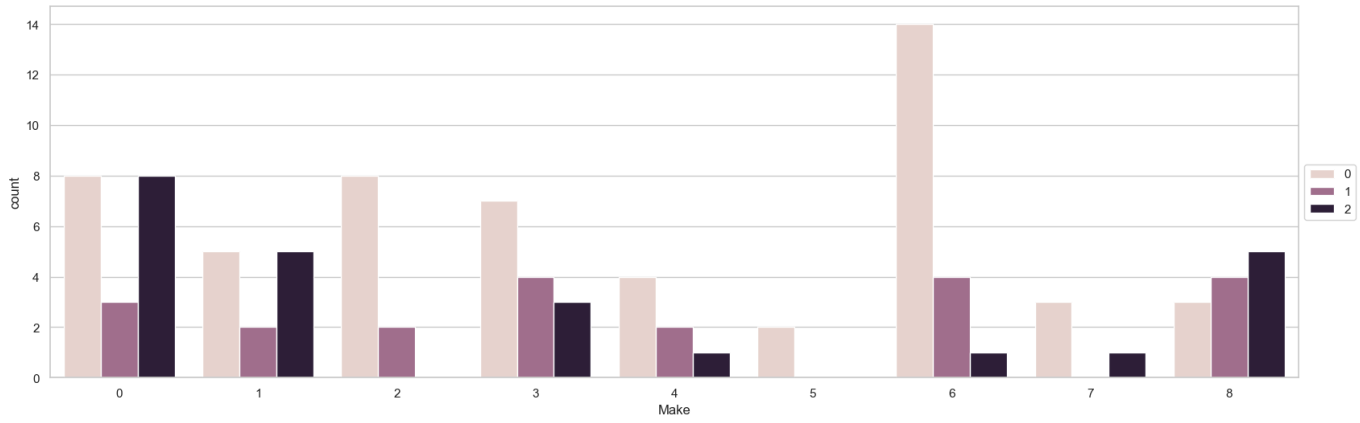
⚠️ A few outliers are visible — people with mid-range salaries (~2M) owning relatively cheaper cars, and vice versa.

🚗 Most owned cars cluster between 1M and 1.5M, suggesting this is the sweet spot for affordability across income groups.

✅ Insight: While salary impacts car choice, other factors like savings, loans, lifestyle, or financial goals likely moderate the final decision. 💳🚗

- ## 💳 Consumer Loan Status vs Vehicle Preference

> 🚗 *Vehicle Purchase Trend by Loan Status*



(0 = No Loans, 1 = Either Home or Personal Loan, 2 = Both Loans)
This bar chart highlights the distribution of car models purchased based on customers' loan status. It's an insightful way to understand how financial obligations shape vehicle preferences.

- 🔍 **Key Observations:**

🚙 SUVs are overwhelmingly favored by individuals with no loans (Cluster 0) — suggesting that financially unburdened buyers may aim for larger, more expensive vehicles.

🚘 City, Baleno, and Creta are also common choices for Cluster 0, showing a general tendency for wider vehicle variety among those without loans.

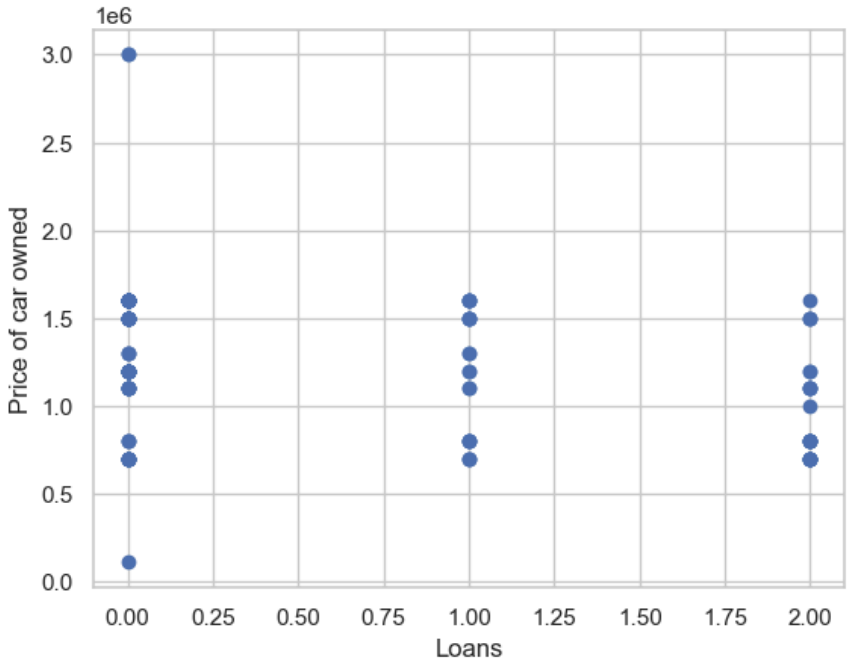
💸 People in Cluster 1 (with either a home or personal loan) show moderate interest in i20, Creta, and Baleno, possibly opting for more economical or fuel-efficient models due to loan constraints.

🏦 Consumers in Cluster 2 (those juggling both types of loans) mostly go for i20, Ciaz, and Baleno, which are relatively budget-friendly and practical options — reflecting a more cautious or necessity-driven buying approach.

💼 Luxury models and Verna remain the least chosen across all loan categories — likely due to affordability and priority shifts.

✅ Insight: As financial burden increases (from Cluster 0 to 2), people tend to move from high-end and spacious vehicles (like SUVs) toward compact, practical cars.

> price of vehicle owned (based on loans



- 🔍 **Interpretation (Quick Summary):**

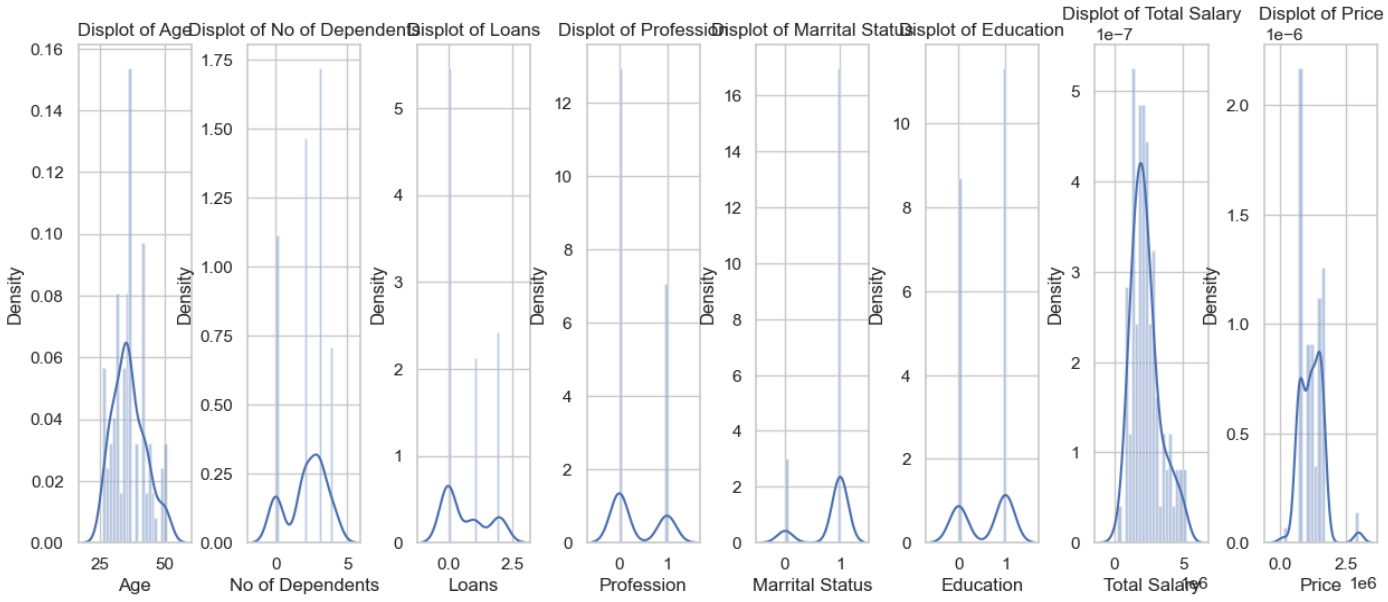
🟢 No Loans (0)
💸 Car prices range from very low to very high — includes both budget buyers and wealthy spenders.

🟠 One Loan (1)
🚗 Mostly mid-range cars - generally own cars in the moderate price range — likely being careful with spending due to financial commitments.

🔴 Both Loans (2)
🚘 Also in the mid-range zone — similar to group 1, but with slightly lower median, possibly due to higher financial burden.

- 🧬 **Demographic Analysis**

Distributions and pair plots helped us explore relationships and correlations among all major features.
Perfect for spotting trends and potential predictors in vehicle preferences 👥📉



- > 🔍 **Interpretation (Quick Summary):**

📊 This set of displots provides insights into the distribution patterns of key variables relevant to EV market segmentation.

- 🔍 **Key Observations:**

👥 Age is slightly right-skewed, with most consumers between 25 and 40 years old — a prime target age group.

👶 Number of Dependents shows a bimodal distribution, indicating distinct family-size segments.

💸 Loans, both personal and house, are sparse — suggesting a majority of users have no or limited active loans.

💼 Profession, 🧑‍🎓 Education, and 💍 Marital Status are binary-coded and show class imbalance — useful for group profiling.

💵 Total Salary is right-skewed, meaning a few consumers have significantly higher income.

🚗 Price also shows a skew toward lower values, indicating most customers are likely to prefer budget to mid-range EVs.

✅ These distributions help in understanding consumer profiles and refining clustering or marketing strategies.

- **Observations:**

Customer Demographics:

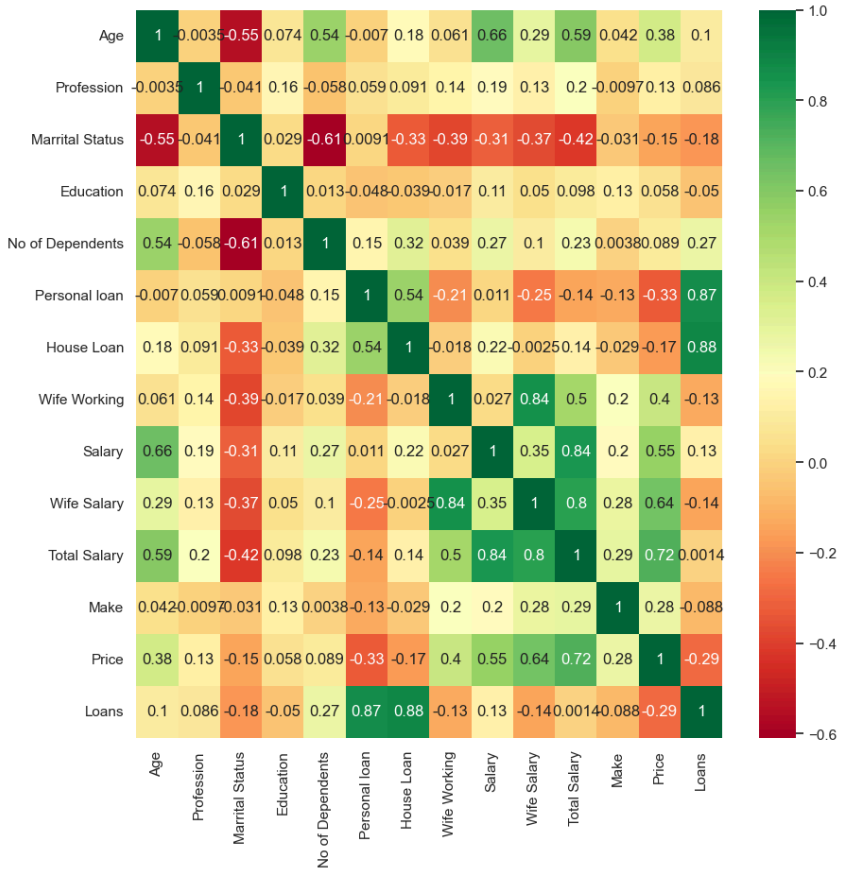
Field	Description
Average Age	36 years old
Age Distribution	92% of customers are between 25 and 45 years old
Marital Status	85% married, 15% single
Dependents	Most customers have 2 or 3 dependents
Employment	65% salaried, 35% self-employed
Education	55% hold a postgraduate degree, 35% hold a graduate degree

Financial Information:

Field	Description
Salary Range	72% of customers earn between ₹1 - ₹2.5 million annually
Household Income	73% of customers have income between ₹1 - ₹3 million
Personal Loans	68% of customers do not have a personal loan
House Loans	62% of customers do not have a house loan
Working Wife	52% of customers have working spouses

Vehicle Purchases :

Field	Description
Popular Models	SUVs, Baleno, Creta are the most popular vehicle choices
Price Range	79% of customers purchase vehicles within range of ₹7 - ₹16 lakh
Luxury Vehicles	Very few customers purchase luxury vehicles



- 🔍 **Heatmap Insights (Correlation Matrix)**

🧠 The heat map above shows the correlation between different features related to customer financial and demographic profiles. Key observations:

💰 Salary, Wife Salary, and Total Salary are strongly correlated — highlighting the combined household income influence.

🏦 Personal Loan and House Loan have a high correlation with the number of dependents and each other — reflecting linked financial obligations and borrowing behavior.

📈 Price is positively correlated with salary-related features, helping to segment customers based on purchasing power.

👨‍👩‍👧‍👦 Marital Status, Education, and Profession show weak correlations with most variables — suggesting a smaller role in financial decisions and loan patterns.

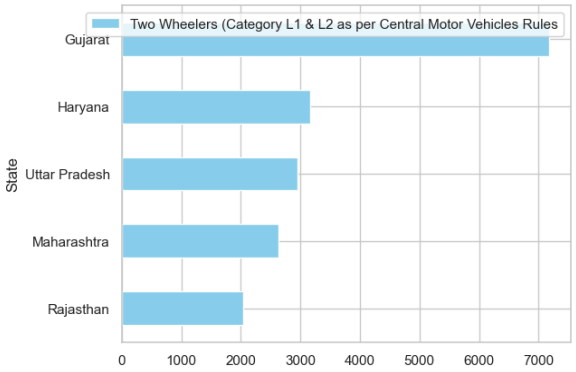
✅ These correlations guide us in selecting key features for effective clustering, segmentation, and predictive modeling.

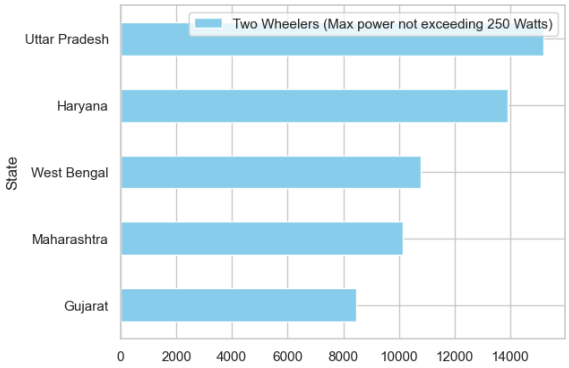
🌏 Geographic Analysis

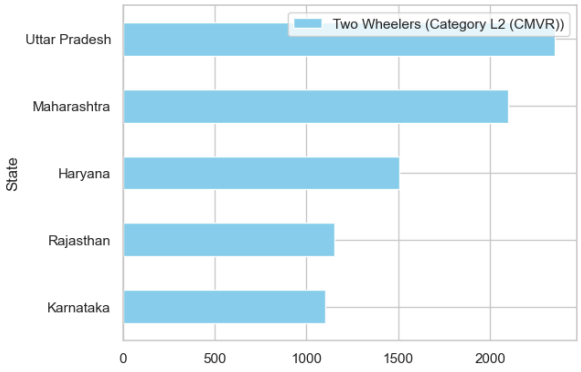
For geographic analysis we used state-wise sales of different types of **Electric Vehicles dataset**

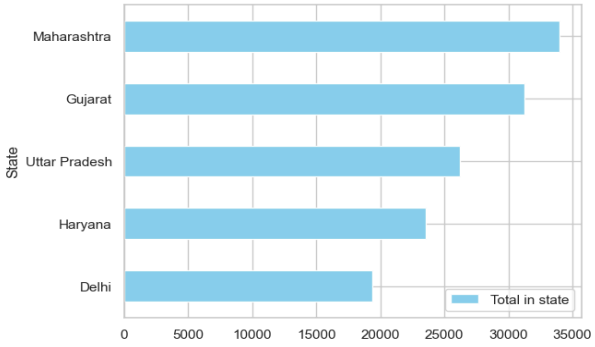
🔍 Objective:

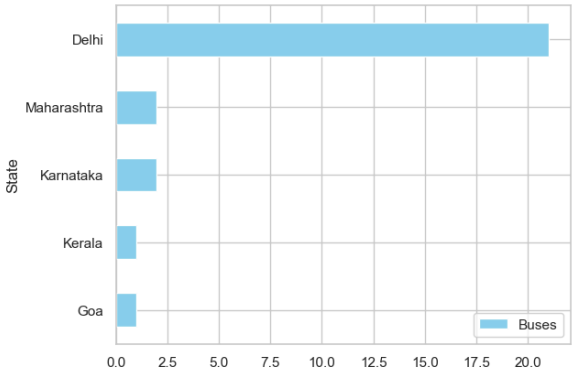
Identify the most suitable regions in India to develop an early market for electric vehicles (EVs)

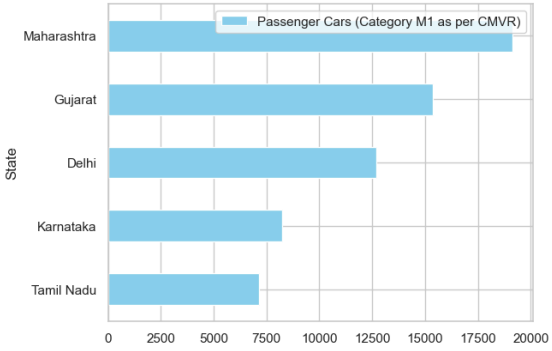


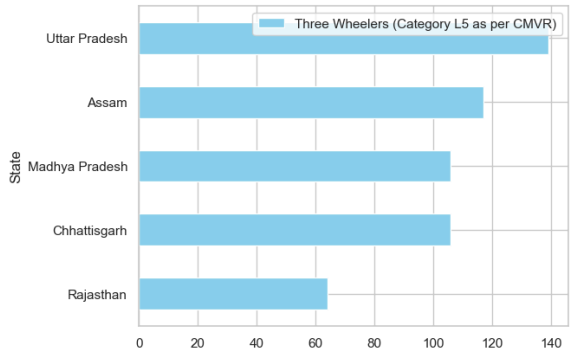


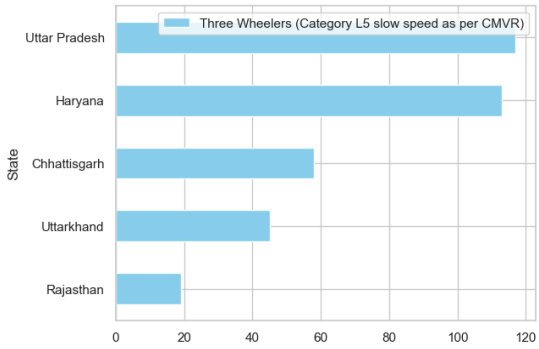












- 🔍 **Interpretation (Quick Summary):**

🛵 Two Wheelers (Category L1 & L2): Haryana leads , followed by Uttar Pradesh and Maharashtra.

🛵 Two Wheelers (Category L2 only): Maharashtra tops , with Haryana and Rajasthan close behind.

🔋 Low Power Two Wheelers (Max 250 Watts): Massive numbers in Uttar Pradesh and Haryana.

🚗 Three Wheelers (Slow speed): Uttar Pradesh and Haryana are major hubs .

🚗 Three Wheelers (Category L5): Assam and Madhya Pradesh show strong presence.

🚘 Passenger Cars: Maharashtra is a clear leader , followed by Gujarat and Delhi.

🚌 Buses: Delhi dominates , Maharashtra slightly behind.

🚚 Goods Vehicles (likely Category N): Maharashtra and Gujarat lead in goods transport.

- 🚗⚡ **Insight:** Based on the type of electric vehicle, states with a higher concentration of EV registrations emerge as prime targets. These regions likely consist of early adopters and innovators who are more open to purchasing new technologies.

- 🎯 **Strategy**: Focus marketing and expansion efforts in these high-potential states to accelerate early market penetration and build momentum for broader adoption.

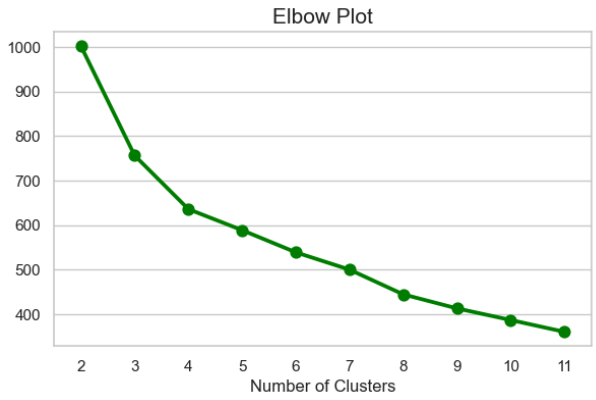
- **Market Segmentation**

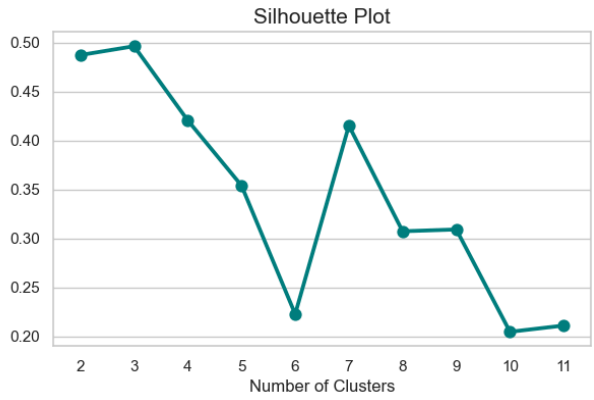
🤖 K-Means Clustering Setup

We're diving into K-Means Clustering, one of the most widely used unsupervised learning techniques for identifying hidden patterns in data.

📦 First up: scaling the data to ensure fair comparisons across features.

🔍 Then, we'll group similar consumer profiles into meaningful clusters 📊.



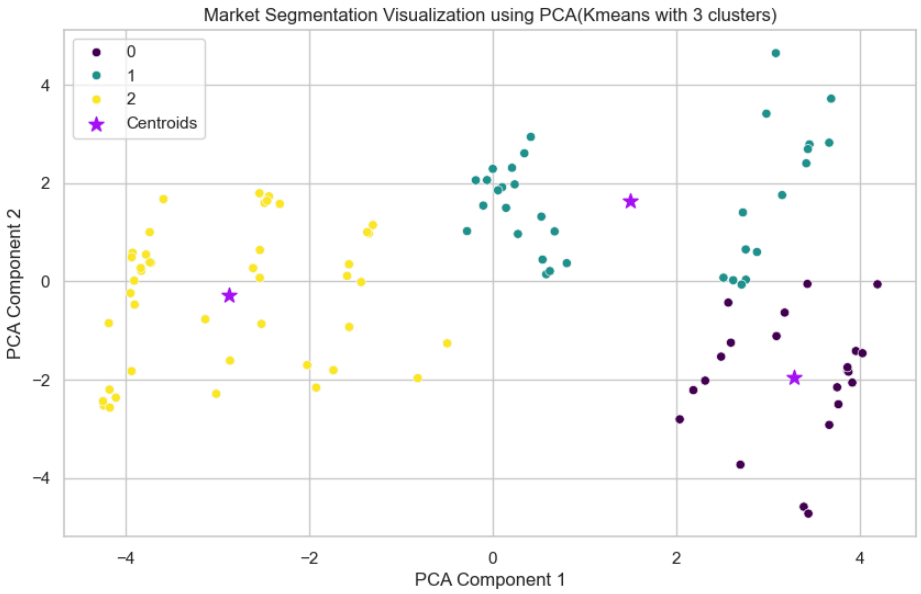


- 🔍 **Interpretation (Quick Summary):**

  📊 Elbow Method for Optimal Clusters : The plot above shows the inertia (within-cluster sum of squares) for different numbers of clusters. The "elbow" appears around 3 to 4 clusters, suggesting this as the optimal range for segmenting the EV market based on income levels🚗💰.

  * Both the elbow and silhouette methods indicate that 3 clusters are optimal for the K-means algorithm.

- **Apply k means with 3 clusters**



🧠 ****PCA Market Segmentation Interpretation****

🎯 Clear Clustering:

The PCA plot shows three well-separated clusters (Cluster 0, 1, 2).
This indicates that your income-based features are effectively segmenting customers into distinct market groups.

💰 Income-Based Differences:

Since PCA was applied on income-related data (Salary, Wife Salary, Total Salary, etc.), the clusters likely represent different income levels or spending capacities.
Example:
Cluster 2 (💛 Yellow) may represent low-income buyers.
Cluster 0 (🟣 Purple) appears as a moderate-income group.
Cluster 1 (🟢 Teal) might indicate high-income consumers with greater EV purchasing power.

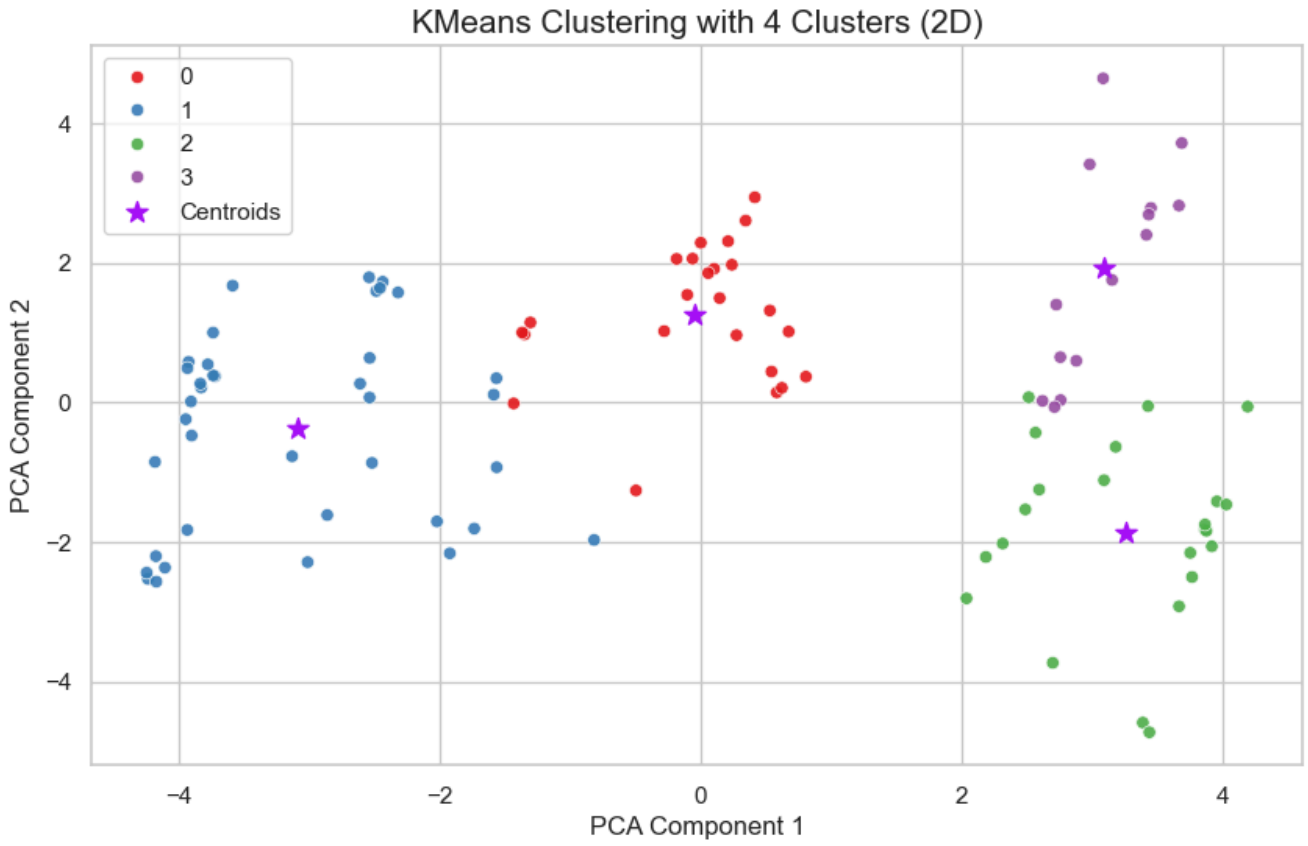
🔄 Low Overlap, High Separability:

The clusters are not overlapping much, meaning the model is doing a good job at distinguishing customer groups based on income.

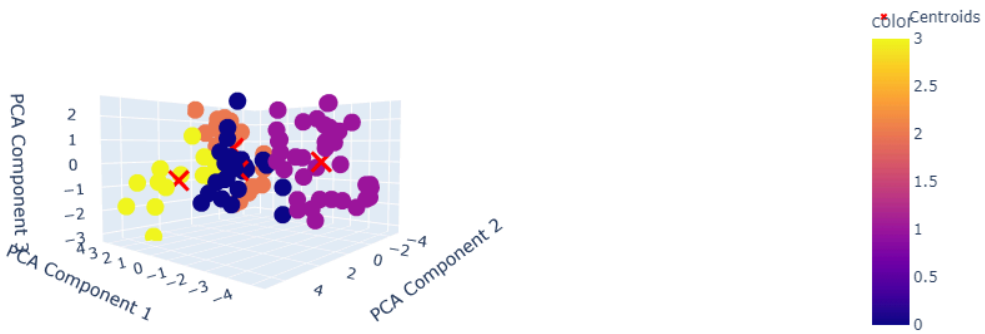
🚘 Strategic Insights:

We could use this segmentation to target each group differently.
💛 Cluster 2: Offer affordable EVs, subsidies, or financing options.
🟣 Cluster 0: Mid-range EVs with value-for-money features.
🟢 Cluster 1: Premium EVs, luxury features, or performance-based models.

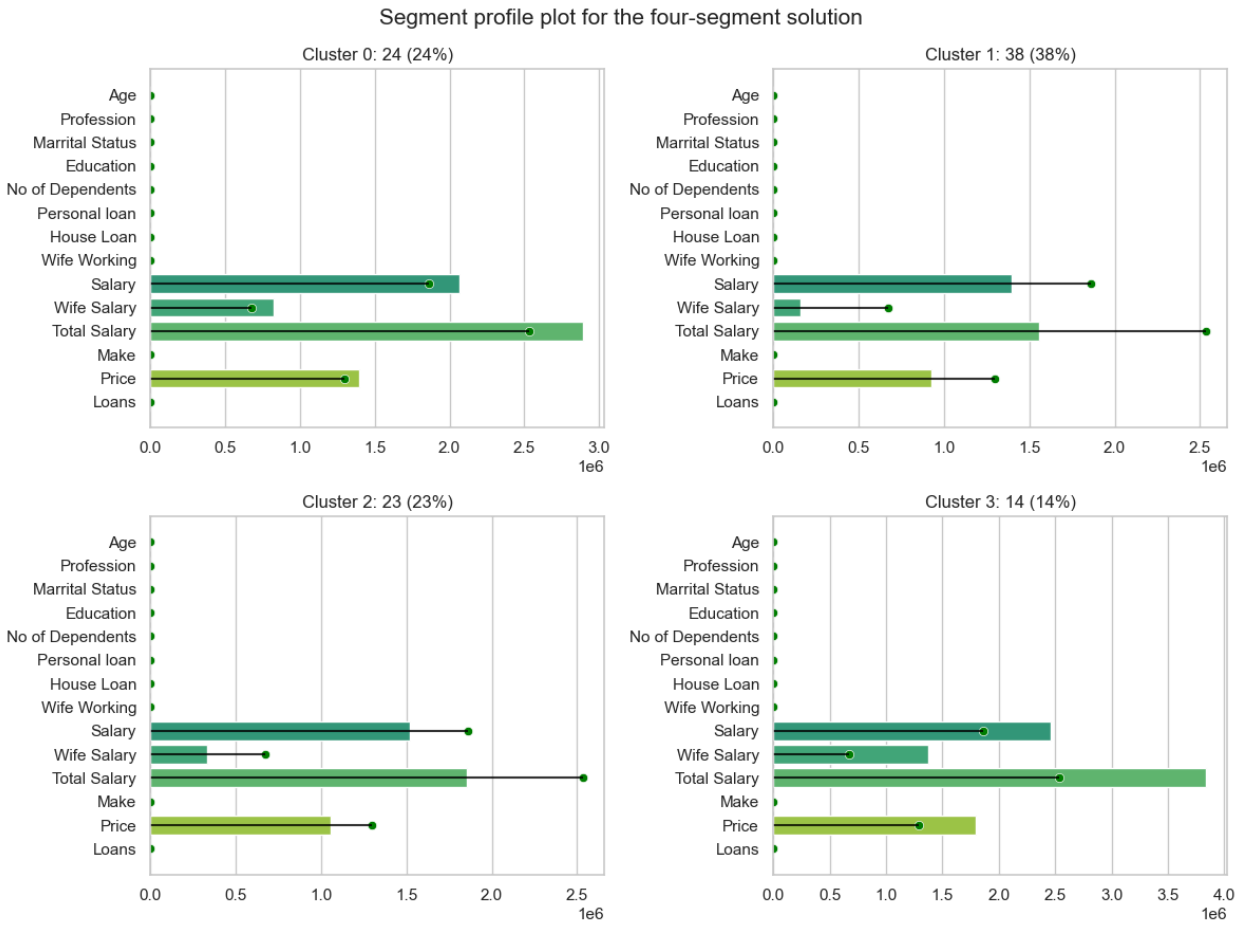
- **Apply K Means with 4 clusters and plotting Clusters using PCA**



KMeans Clustering with 4 Clusters (3D)



- # 🛍️ Market Segment Profiling (Four-Segment Solution)



This section presents a comprehensive analysis of the consumer segments based on their mean characteristics, as detailed in the provided data. The following insights capture the unique profiles of each segment:

- 🔵 **Cluster 0:**

Size: 24% of customers

Profile:

* High salary and total salary 💰 . * Wife's salary is also considerable 👩‍💼 . * Make value is decent 🚗. * Less significant in loans or personal loans 🤷‍♂️.

Summary:

→ Well-earning households with working spouses, financially stable, low loan dependency. 🏡

- 🟢 **Cluster 1:**

Size: 38% of customers (largest group)

Profile:

*Moderate salary, but total salary varies a lot 📉📈. *Some members have much lower wife salary 👩‍🦰 . * Significant spread in loan amounts 💳. * Moderate Make/price spending 🚙

Summary:

→ Middle-class segment, mixed financial background, moderate asset acquisition, higher loan exposure. 🏦

- 🟡 **Cluster 2:**

Size: 23% of customers

Profile:

* Good salary and total salary 💼. *Moderate wife salary too 👩‍💼. *Balanced Make/Price spending 🚘. *Loans are present but not extreme 💳

Summary:

→ Financially sound families, dual-income, balanced spending and loan behavior. ⚖️

- 🟣 **Cluster 3:**

Size: 14% of customers (smallest group)

Profile:

*Highest total salary among all clusters 🏆💸. *High wife salary contribution 👩‍💼. *Very strong in Make/Price spending 🏎️. *Loan values moderate but manageable 💵

Summary:

→ Affluent segment with luxury purchases, strong dual-income, financially powerful. 👑

In summary, each segment reflects distinct financial profiles and priorities, with varying emphasis on salary, dependents, loans, and vehicle make. This segmentation helps in tailoring marketing strategies to address the specific needs and preferences of each group.

- # 🎯 Describing Segmentation

Cluster	Age	No of Dependents	Salary	Wife Salary	Total Salary	Price	Size	Profession	Marrital Status	Education	Personal loan	Make	House Loan	Wife Working
0	39.0	3.0	2000000.0	900000.0	2850000.0	1500000.0	24	Salaried	Married	Post Graduate	No	Creata	No	Yes
1	33.5	2.0	1400000.0	0.0	1450000.0	900000.0	38	Salaried	Married	Graduate	No	Baleno	No	No
2	34.0	2.0	1400000.0	0.0	1900000.0	800000.0	23	Salaried	Married	Post Graduate	No	i20	No	No
3	41.5	3.0	2500000.0	1300000.0	4000000.0	1600000.0	14	Salaried	Married	Post Graduate	No	SUV	No	Yes

🚗 Cluster 0: Mid-Aged Dual Income Professionals

- **Average Age:** 39 years 👨‍🦳
- **Number of Dependents:** 3 👨‍👩‍👧
- **Average Salary:** ₹2,000,000 💰
- **Average Wife's Salary:** ₹900,000 👩‍💼
- **Total Household Salary:** ₹2,850,000
- **Average Price of Car Purchased:** ₹1,500,000 🚙
- **Segment Size:** 24 customers 📊
- **Most Common Profession:** Salaried
- **Most Common Marital Status:** Married
- **Most Common Education Level:** Post Graduate 🎓
- **Personal Loan Status:** No 🚫
- **House Loan Status:** No 🏠❌
- **Wife Working Status:** Yes 👩‍💼✅
- **Most Common Car Make:** Creta 🚗

👉 **These are mid-aged, highly educated married couples with double income streams and 3 dependents. They prefer premium compact SUVs like Creta.**

🚀 Cluster 1: Young Married Professionals

- **Average Age:** 33.5 years 👨‍💼
- **Number of Dependents:** 2 👶👶
- **Average Salary:** ₹1,400,000 💵
- **Average Wife's Salary:** ₹0 🚫
- **Total Household Salary:** ₹1,450,000
- **Average Price of Car Purchased:** ₹900,000 🚘
- **Segment Size:** 38 customers 📈
- **Most Common Profession:** Salaried
- **Most Common Marital Status:** Married
- **Most Common Education Level:** Graduate 🎓
- **Personal Loan Status:** No ❌
- **House Loan Status:** No 🏠❌
- **Wife Working Status:** No 👩‍🦰🚫
- **Most Common Car Make:** Baleno

👉 **This group consists of younger married salaried professionals, often with a single income and small families. They prefer budget-friendly vehicles like the Baleno.**

Cluster 2: Mid-Aged Single Income Families

- **Average Age:** 34 years 🧔
- **Number of Dependents:** 2 👨‍👩‍👧
- **Average Salary:** ₹1,400,000 💸
- **Average Wife's Salary:** ₹0 🚫
- **Total Household Salary:** ₹1,900,000
- **Average Price of Car Purchased:** ₹800,000 🚗
- **Segment Size:** 23 customers 📊
- **Most Common Profession:** Salaried 👔
- **Most Common Marital Status:** Married 💍
- **Most Common Education Level:** Post Graduate 🎓
- **Personal Loan Status:** No ❌
- **House Loan Status:** No 🏠❌
- **Wife Working Status:** No 👩🚫
- **Most Common Car Make:** i20 🚗

👉 **These are mid-aged married individuals with no working spouse, slightly better salary but still prefer relatively budget-conscious cars like Hyundai i20.**

🏆 Cluster 3: Affluent Dual Income Executives

- **Average Age:** 41.5 years 👨‍🦳
- **Number of Dependents:** 3 👨‍👩‍👦‍👦
- **Average Salary:** ₹2,500,000 💵
- **Average Wife's Salary:** ₹1,300,000 👩‍💼
- **Total Household Salary:** ₹4,000,000
- **Average Price of Car Purchased:** ₹1,600,000 🚘
- **Segment Size:** 14 customers 📉
- **Most Common Profession:** Salaried 👔
- **Most Common Marital Status:** Married 💍
- **Most Common Education Level:** Post Graduate 🎓
- **Personal Loan Status:** No 🚫
- **House Loan Status:** No 🏠❌
- **Wife Working Status:** Yes 👩‍💼✅
- **Most Common Car Make:** SUV 🚙

👉 **This is a smaller, affluent segment of experienced dual-income couples, highly educated, high earners, and strong preference for SUVs.**

📝 Summary:

✅ All clusters consist of **married salaried professionals**.
✅ Higher total salaries usually correspond with choosing **bigger cars like SUVs**.
✅ **Wife working status** has a major impact on total income.
✅ No cluster appears to have major debts (no personal or house loans).

📋 Conclusion

- **Cluster 0 (Mid-Aged Dual Income Professionals):**

This segment is composed of middle-aged married couples with multiple dependents. They enjoy a strong dual-income household, with both partners contributing significantly. Their total income is quite high, and they prefer investing in moderately high-priced vehicles like the Creta. They are highly educated (mostly postgraduates), professionally stable, and financially self-sufficient, with minimal reliance on personal or house loans.

- **Cluster 1 (Young Married Professionals):**

This group mainly consists of younger married individuals with small families. Typically, there is only one working partner, resulting in a moderate total household income. They prefer budget-friendly vehicles like the Baleno, reflecting practical purchasing behavior. While well-educated (mostly graduates), they are earlier in their career trajectories and tend to avoid taking loans or financial liabilities.

- **Cluster 2 (Mid-Aged Single Income Families):**

These are mid-aged married individuals with two dependents but only a single income. Their earnings are modest, and they opt for cost-effective car options like the Hyundai i20. They are highly educated (postgraduates) but demonstrate cautious financial behavior, showing no significant dependence on loans. Notably, the wife is generally not working, which slightly impacts the household's financial capacity.

- **Cluster 3 (Affluent Dual Income Executives):**

This cluster is characterized by older, affluent families with multiple dependents and high total household incomes. Both partners are working, contributing to an impressive financial profile. They prefer higher-end vehicles such as SUVs, reflecting their premium lifestyle. With high educational qualifications and strong salaried careers, they show no reliance on personal or house loans, indicating strong financial independence.