ELECTRIC VEHICLE MARKET SEGMENTATION

R.Ramanidevi

22/11/2024

GitHub : https://github.com/gitramani/rramanidevi.git

Exploring Market Segmentation and Buyer Behavior for EV Adoption in India

The rapid evolution of transportation technology, particularly electric vehicles (EVs), highlights the importance of market segmentation to accelerate their adoption in emerging markets like India. EVs, with their low emissions and reduced operating costs, are poised to transform the automotive landscape, sparking significant academic interest in understanding consumer behavior and preferences. This study aims to analyze and identify distinct consumer segments for EVs by integrating psychographic, behavioral, and socio-economic data.

To achieve this, two datasets have been utilized: one focusing on **Electric Vehicle (EV) buyers** and the other capturing **Indian automobile consumer behavior**. By employing an integrated research framework of *'perceived benefits-attitude-intention'*, we aim to uncover the drivers of EV adoption and how various consumer profiles influence purchase decisions. The insights derived from this study will provide actionable intelligence to policymakers, manufacturers, and marketers to effectively target potential EV buyers and facilitate the transition to sustainable mobility solutions in India.

Data Source Summary:

The dataset provides an overview of electric vehicle (EV) sales across various Indian states in 2018. It categorizes EVs based on their types, such as two-wheelers, three-wheelers, passenger cars, and buses, in accordance with the **Central Motor Vehicles Rules (CMVR)**. The data emphasizes the spread and adoption of electric vehicles, particularly focusing on regions and categories where growth is noticeable. The original source is from the website "electricvehicles.in", a prominent platform for EV news and market insights.

Key Touch Points from the Dataset:

1. **EV Classification by Category**:
    - Two-wheelers are subdivided into:
        - **Category L1 & L2**: Includes basic two-wheelers.
        - **Category L2 (CMVR)**: A specific CMVR-defined subset.
        - **Two-wheelers with max power ≤250 Watts**: Reflecting low-speed electric scooters.
    - Three-wheelers are classified under:
        - **Category L5 (slow speed and standard CMVR types)**.
    - **Passenger cars (Category M1)** and **buses** are also detailed.
2. **Top States with High EV Sales**:
    - **Maharashtra** leads the chart with **34,013 EVs**, indicating a strong EV adoption rate.
    - **Gujarat** and **Uttar Pradesh** follow with **31,267** and **26,209 EVs**, respectively.
    - High adoption in urbanized and industrially significant states reflects progressive EV policies.
3. **Prominent Contributions by Vehicle Type**:
    - Two-wheelers dominate in terms of numbers, showcasing their affordability and practicality.
    - Passenger cars show significant numbers in states like **Delhi (12,695)** and **Karnataka (8,242)**, reflecting urban adoption.
    - Buses remain minimal, indicating potential areas for future policy-driven adoption.
4. **Regional Patterns**:
    - **North India**: Haryana and Uttar Pradesh exhibit a robust EV adoption, particularly in two-wheelers and passenger cars.
    - **South India**: Tamil Nadu and Karnataka highlight significant numbers in both two-wheelers and passenger cars.
    - **Western India**: Maharashtra and Gujarat lead in all EV categories.
    - **Northeast**: States like Assam and Tripura show potential with consistent numbers despite smaller populations.

5. **Data Insights**:
   - Urbanization and industrialization strongly correlate with EV adoption.
   - Central and state government subsidies, along with policies like **Faster Adoption and Manufacturing of Electric Vehicles (FAME)**, have likely influenced these numbers.
   - The dataset is instrumental for analyzing regional EV market trends and identifying underperforming areas for targeted interventions.

By integrating this data with ongoing government efforts and market trends, policymakers and businesses can strategize for enhancing EV adoption.

The dataset provided appears to contain information about individuals and their car purchase details. The original source is from the website "Indian automobile consumers behaviour". Below is a summary of the data:

Dataset Columns:

1. **Age**: Age of the individual.
2. **Profession**: Type of profession (Salaried or Business).
3. **Marital Status**: Indicates if the individual is Single or Married.
4. **Education**: Level of education (Graduate, Post Graduate).
5. **No of Dependents**: Number of dependents in the family.
6. **Personal Loan**: Indicates if the individual has a personal loan (Yes/No).
7. **House Loan**: Indicates if the individual has a house loan (Yes/No).
8. **Wife Working**: Indicates if the wife is working (Yes/No).
9. **Salary**: Salary of the individual.
10. **Wife Salary**: Salary of the wife.
11. **Total Salary**: Combined salary of the individual and their wife.
12. **Make**: The make/model of the car purchased.
13. **Price**: Price of the car.

Data Insights:

- **Profession**: The dataset includes two main categories: "Salaried" and "Business."
- **Marital Status**: Both married and single individuals are represented.
- **Car Price Range**: Cars vary in price, from budget-friendly options (e.g., Baleno at €700,000) to luxury cars (e.g., Luxury at €3,000,000).
- **Loan Information**: Data includes whether the individual has personal or house loans.
- **Family Dependents**: Number of dependents ranges from 0 to 4.
- **Salary Distribution**: Salary data is split into individual and combined family income.
- **Car Preferences**: Popular car models include i20, Baleno, City, SUV, and Creta.

Data Preprocessing and Model Deployment for K-Means Clustering

Data Preprocessing:

In the initial stage of data analysis, we observed that some categories in the dataset were incorrectly classified under certain attributes. After reassessing and correctly categorizing the data, we proceeded to clean and preprocess it further. One key step in preprocessing involved checking for null or missing entries in the dataset. Fortunately, our dataset was free of such entries, with 0% missing values in all columns as verified by the df.isnull().sum() check.

Dataset Overview:

- The dataset consists of 99 rows and 13 columns.
- The columns contain a mix of numerical and categorical data.
- Numerical columns include features like Age, No of Dependents, Salary, Wife Salary, Total Salary, and Price.
- Categorical columns include Profession, Marital Status, Education, Personal loan, House Loan, Wife Working, and Make.

Column-wise Summary:

- The dataset has no missing values (all columns have 0.0% empty entries).
- The numerical columns have values ranging from 26 to 51 for Age, with salaries ranging from ₹2,00,000 to ₹38,00,000.
- Categorical columns such as Profession have two unique values: 'Salaried' and 'Business'. Other categorical columns also have well-defined sets of values.

Since **K-Means Clustering** requires numerical input and cannot process categorical variables directly, we needed to transform these categorical attributes into numerical formats. This transformation was accomplished by encoding various categorical features. We performed this encoding on a copy of the original dataset, preserving the integrity of the original data while preparing it for model training.

Model Deployment:

The first step in deploying the **K-Means Clustering** model is to determine the optimal number of clusters, denoted as **K**. To achieve this, we used the **Elbow Method**, a popular technique for selecting the optimal value of K.

In the **Elbow Method**, we calculate the **Within-Cluster Sum of Squares (WCSS)** for various values of K. WCSS measures the sum of squared distances between data points and their respective cluster centroids. We then plot the WCSS against the number of clusters (K). The optimal K is identified at the point where the WCSS begins to decrease at a slower rate, creating an "elbow" shape in the graph. This point indicates the best balance between minimizing the sum of squared distances and avoiding overfitting by using too many clusters.

Once the optimal K is identified, the K-Means Clustering algorithm can be applied to segment the data into K clusters, enabling us to better understand and target specific customer segments.

This process of data preprocessing and applying the K-Means algorithm is essential for building a more accurate and effective segmentation model that can provide valuable insights into consumer behavior based on the dataset provided.

Approaches for Market Segmentation: K-Means Clustering

Market segmentation is a critical step in understanding and targeting specific consumer groups effectively. To identify the ideal target segment for market penetration, we use a population behavioral study, where a subset of 100 individuals is selected from the broader population. This sample provides valuable insights into the automobile purchase behavior and capabilities, which forms the foundation of our segmentation analysis.

In general, there are two primary approaches to classification: **common sense classification** and **data-driven classification**. For our analysis, we will implement a **data-driven classification** method using the **K-Means Clustering** algorithm. This approach ensures that the segmentation is based on actual data patterns rather than subjective assumptions.

K-Means Clustering Algorithm:

K-Means is a widely used unsupervised learning algorithm, designed to identify groups or clusters within unlabeled data (i.e., data that does not have predefined categories or labels). The primary objective of K-Means is to partition the data into K distinct groups, where each group is formed by grouping similar data points based on their features.

How K-Means Works:

The K-Means algorithm follows a series of steps to assign data points to clusters:

1. **Specify the number of clusters (K):** First, we define the number of clusters that we want to create. This is typically based on domain knowledge or trial and error.
2. **Initialize centroids:** Randomly select K data points (representing consumers) from the dataset as the initial cluster centroids, denoted as C=c1,c2,...,ckC = {c1, c2, ..., ck}.
3. **Assign points to clusters:** For each data point (consumer), calculate the distance to each of the K centroids. Assign each point to the nearest centroid, forming K market segments (S1, S2, ..., Sk).

4. **Recompute centroids:** After each assignment, the centroids are recalculated by averaging the positions of the data points in each cluster. The new centroids represent the central point of each segment.
5. **Repeat the process:** Steps 3 and 4 are repeated iteratively, with each update moving the centroids closer to their optimal positions. The algorithm stops when the centroids no longer change significantly (convergence), or when a pre-set maximum number of iterations is reached.

Key Insights:

- **K-Means Clustering** provides a data-driven method for grouping consumers based on their purchasing behavior, ensuring that segments are defined by actual consumer patterns rather than arbitrary classification.
- The centroids represent the central characteristics of each segment, which can be used to label new data points.
- This iterative process ensures that the final segmentation is the most optimal, with minimal distance between the data points and their assigned centroids.

By applying this algorithm, we can identify distinct market segments based on purchasing behavior, enabling more precise and targeted marketing strategies.

Prerequisites for Running the Project

This project leverages Python and a collection of powerful libraries to handle data analysis, visualization, and machine learning tasks effectively. Below is an elaboration of the libraries used and their specific roles in the project:

1. **NumPy**:
    - Provides support for large, multi-dimensional arrays and matrices.
    - Includes a vast collection of mathematical functions to operate on these arrays.
    - Used for numerical computations such as array manipulations and linear algebra operations, which form the foundation for data preprocessing.
2. **Pandas**:
    - Offers high-level data structures like DataFrames for organizing and analyzing data.

- Simplifies tasks such as reading datasets from various formats (CSV, Excel, etc.), cleaning, and reshaping data.
- Enables grouping, filtering, merging, and aggregation of data for detailed insights.

3. **matplotlib**:
   - A fundamental library for creating static, animated, and interactive visualizations.
   - Useful for plotting basic charts like histograms, line plots, scatter plots, and bar graphs to understand trends and distributions in data.
   - Provides fine-grained control over plot appearance, ensuring clarity and customization.
4. **seaborn**:
   - Built on top of matplotlib, it simplifies the creation of attractive and informative statistical graphics.
   - Ideal for advanced visualizations like heatmaps, pair plots, and violin plots that reveal relationships and distributions in datasets.
   - Integrates seamlessly with Pandas for quick plotting of DataFrame data.
5. **scikit-learn**:
   - A comprehensive machine learning library for building and evaluating models.
   - Provides tools for preprocessing data, such as scaling and encoding, to prepare it for machine learning algorithms.
   - Includes a wide range of supervised and unsupervised learning algorithms, like linear regression, decision trees, clustering, and more.
   - Offers metrics for model evaluation, such as accuracy, precision, recall, and confusion matrices.

Software Requirements

To run this project, ensure the following software is installed:

- **Jupyter Notebook**: Allows interactive coding and documentation, making it ideal for exploring datasets and presenting results.
- **Python Environment**: Install and manage the libraries above through tools like **Anaconda** or **pip**.

With these tools and libraries, the project ensures a robust framework for data analysis, visualization, and predictive modeling.

Data Sources

- Electric Vehicle Dataset - For region based analysis of different types of Electric Vehicle sales in India for the year 2018.
- Indian EV Consumers Behaviour - For behavioral, psychographic and demographic analysis of Indian automobile market.

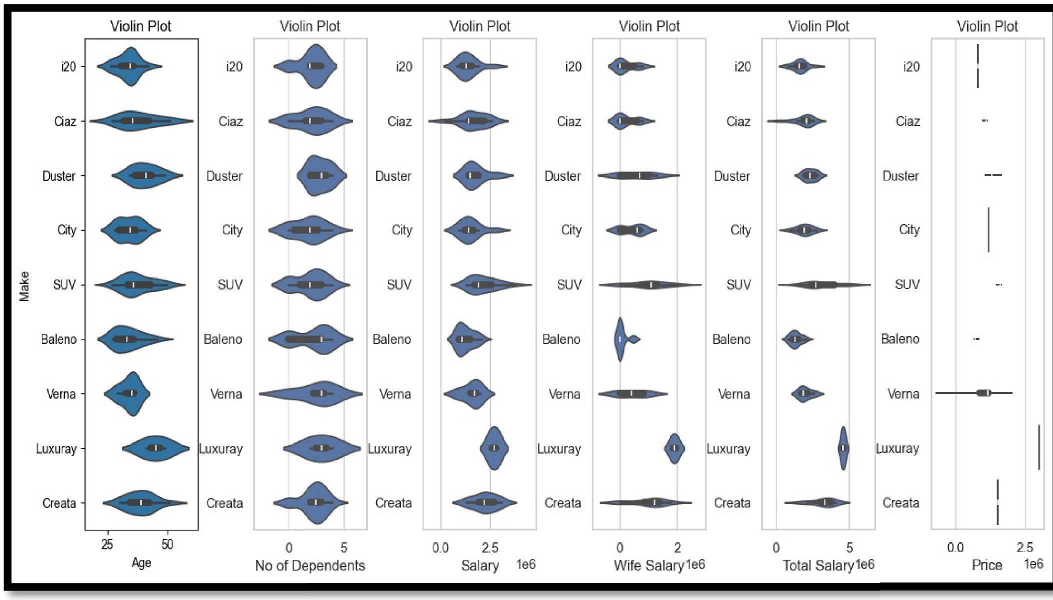
Behavioral and Psychographic Analysis

Behavioral segmentation is a powerful method for classifying customers based on their behavior patterns when interacting with a brand or making purchasing decisions. This form of segmentation enables businesses to group customers based on their awareness, attitudes, usage, or response to products, services, or brands. On the other hand, psychographic segmentation goes deeper into understanding a consumer's lifestyle, interests, and opinions, offering valuable insights into their motivations and preferences.

In this study, we integrate both behavioral and psychographic analysis, as a consumer's lifestyle and interests often directly influence their purchasing behavior. For instance, consumers who value sustainability might be more likely to purchase eco-friendly products. Similarly, consumers' preferences for certain brands or types of vehicles can be strongly linked to their personal values, lifestyle, and experiences.

The dataset utilized in this project is based on a survey of individuals who own specific brands of fuel-based vehicles. It includes key demographic and personal information such as age, salary, loan status, marital status, number of dependents, education, occupation, and the make and price of their car. By analyzing these variables, we aim to identify patterns in purchasing decisions and consumer preferences.

The **violin plot** below provides a visual representation of the relationships between the segmentation and descriptive variables in the data, shedding light on important trends and insights.



Observations:

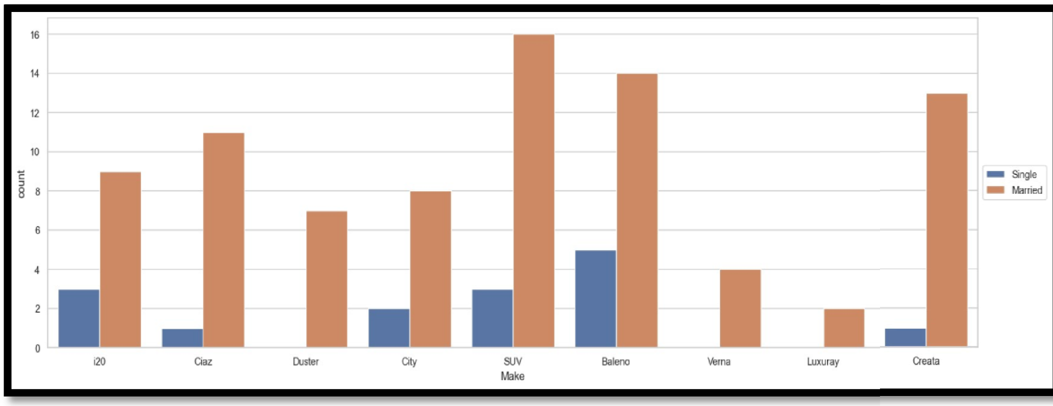
- **Age**: Younger consumers tend to purchase less expensive vehicles. This can be attributed to factors such as lower income, fewer dependents, and the likelihood of being single. As a result, they have limited purchasing power and fewer immediate needs for expensive vehicles.
- **Number of Dependents**: Consumers with more dependents often prefer larger vehicles, such as SUVs, to accommodate their families. These consumers prioritize space and comfort, which influences their vehicle choices.
- **Salary**: A clear relationship exists between salary and vehicle price. When comparing the normalized salary plot with the vehicle price plot, we observe that the median salary aligns closely with the median vehicle price. This correlation highlights that consumers tend to purchase vehicles within their financial means, reinforcing the notion that people are more likely to buy cars they can afford.

By combining both behavioral and psychographic insights, this analysis helps in understanding how various factors, such as lifestyle and personal circumstances, influence consumer purchasing behavior. These findings are crucial for tailoring marketing strategies and product offerings to meet the needs of different consumer segments.

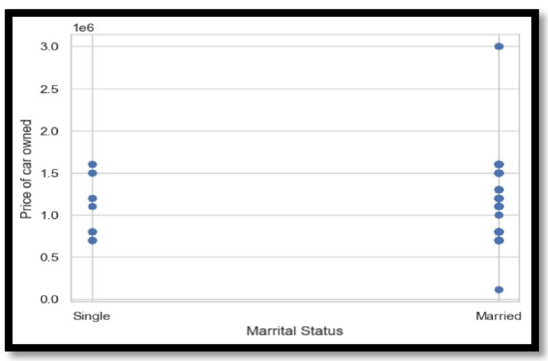
Dependency of make and price of vehicles on other descriptor variables:

- **Marital Status:**

  - Make of vehicles they tend to purchase:

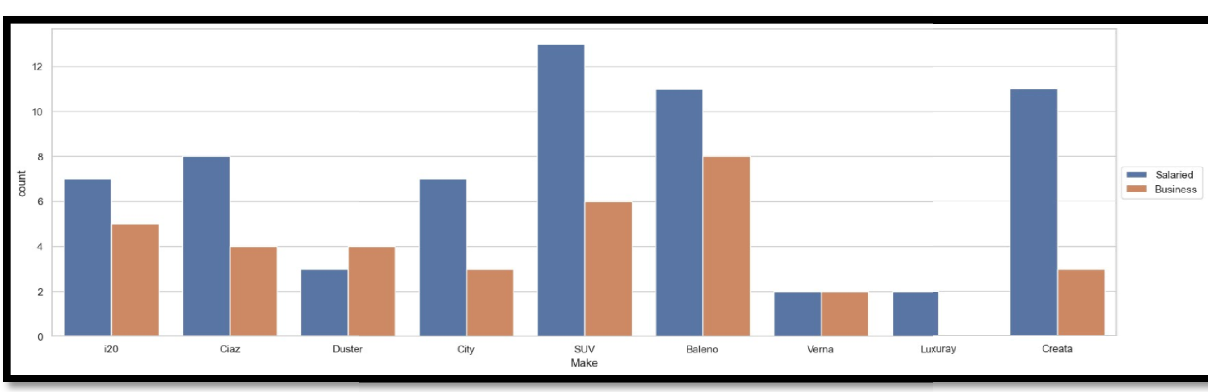


  - Price of vehicle owned:



- **Profession:**

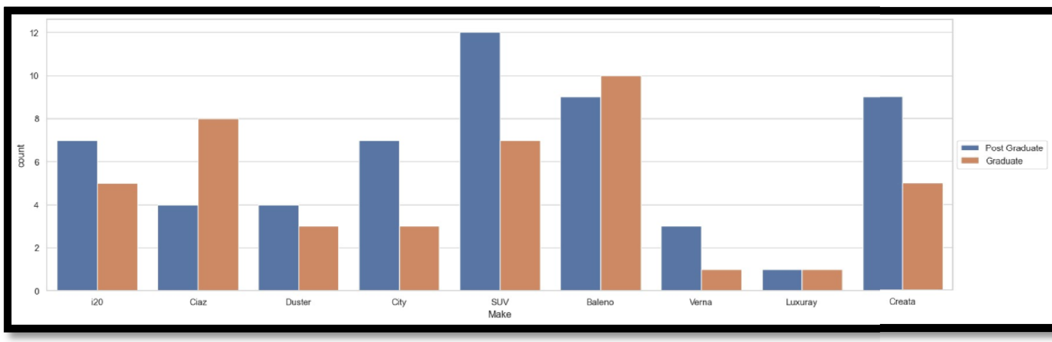
  - Make of vehicles they tend to purchase:



- Price of vehicle owned:

- **Education:**

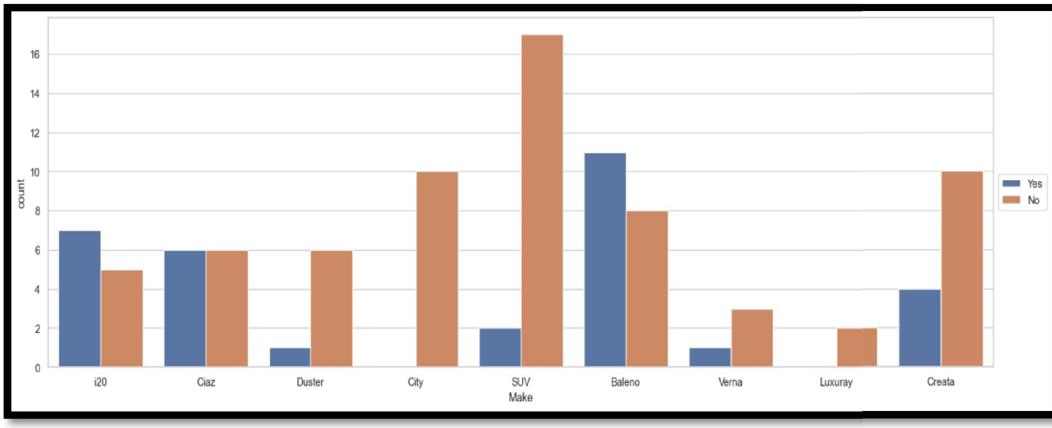
- Make of vehicles they tend to purchase:



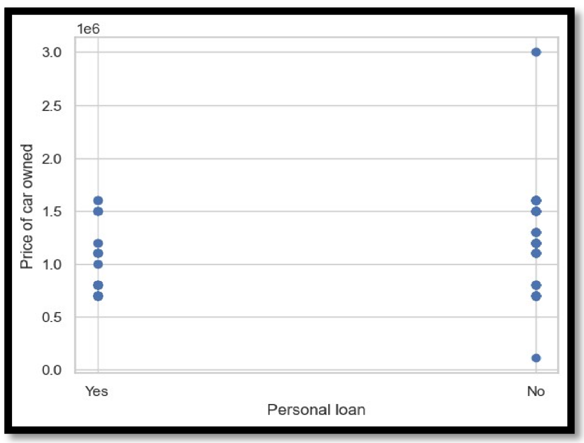
- Price of vehicle owned:

- **Personal Loan:**

  - Make of vehicles they tend to purchase:

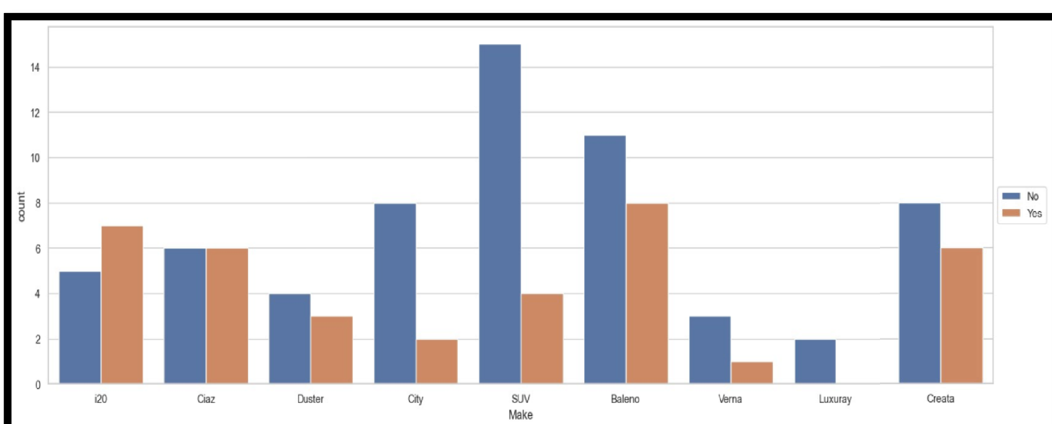


  - Price of vehicle owned:

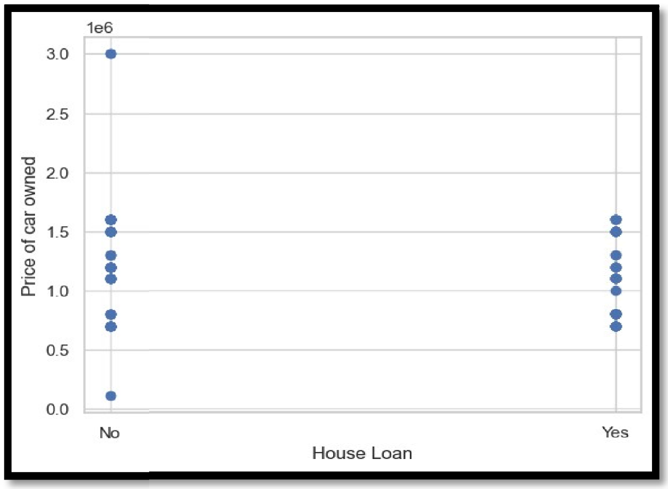


- **House Loan:**

  - Make of vehicles they tend to purchase:



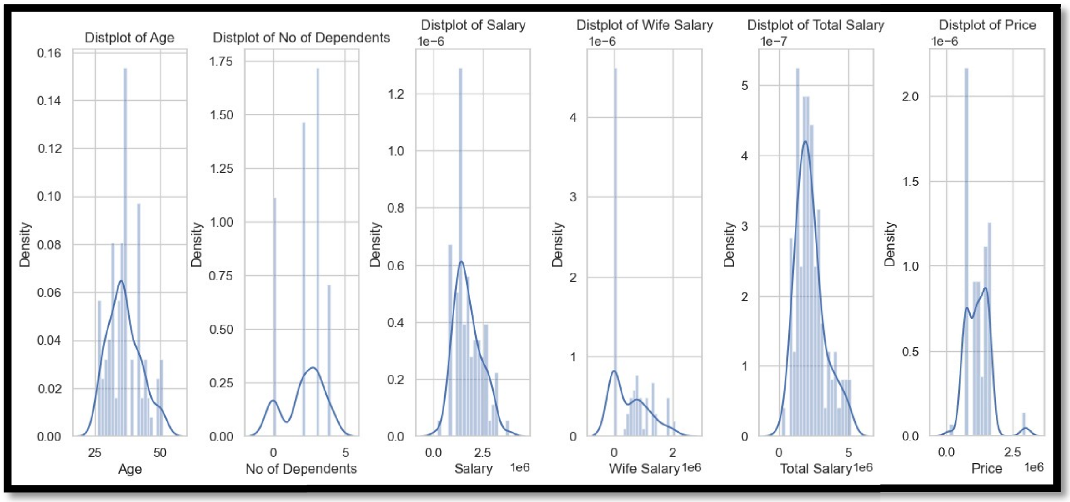
- Price of vehicle owned:



Demographic Segmentation and Analysis

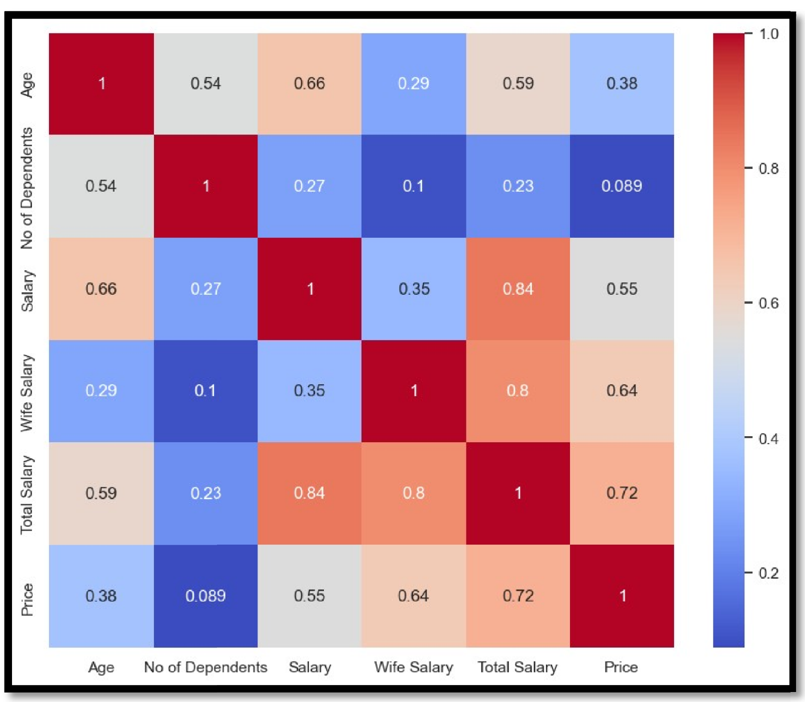
Demographic segmentation is a marketing strategy that categorizes customers based on specific traits such as age, gender, income, education, occupation, and family status. This approach operates on the premise that individuals within the same demographic group will likely exhibit similar behaviors, preferences, and needs. By leveraging demographic data, organizations can craft more targeted marketing strategies and outreach programs that resonate with their most probable customers, ultimately refining the targeting of products and services.

For this analysis, we have utilized the same dataset from the behavioral and psychographic analysis, which provides insights into the socio-demographic structure of the market. The following visualizations offer a deeper understanding of demographic patterns within the customer base:



Key Observations from the Distribution Plot:

- A large portion of the consumer market falls within the age range of 25 to 50 years, indicating this group is the most active when it comes to vehicle purchases.
- Consumers with an average annual salary of approximately 30 lakh INR tend to have a higher inclination towards purchasing vehicles.
- The price range of vehicles purchased most frequently falls between 10 to 20 lakh INR, pointing to a preference for mid-range vehicles.

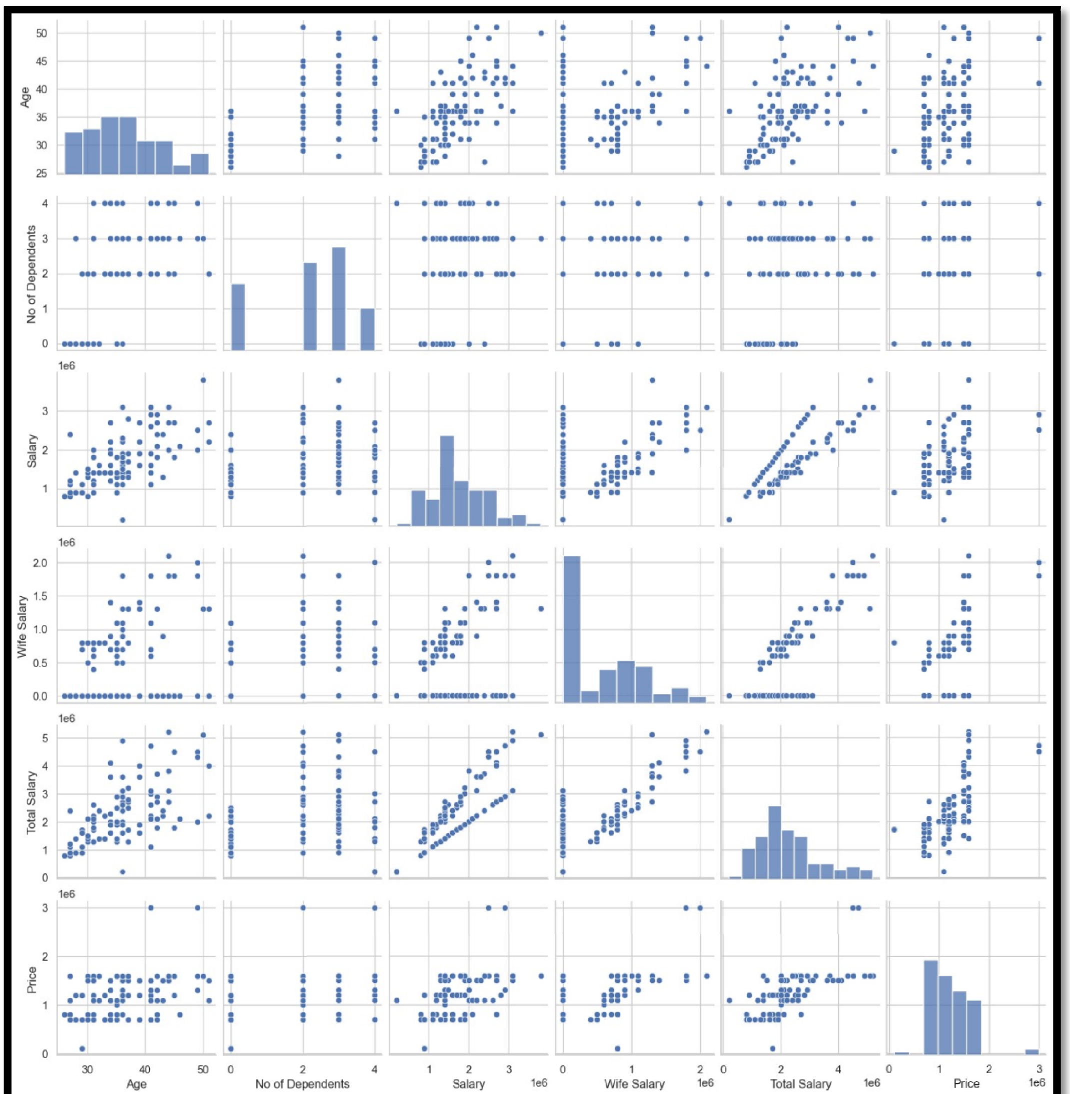


Insights from the Heatmap:

- While the heatmap doesn't reveal any dramatically new patterns, it reinforces the findings from the distribution plot, further validating the relationships between income, age, and vehicle purchase behavior.

Insights from the Demographic and Vehicle Price Scatterplot Matrix

The scatterplot matrix provides valuable insights into the relationships between demographic variables and vehicle price. Here's a summary of the key findings:



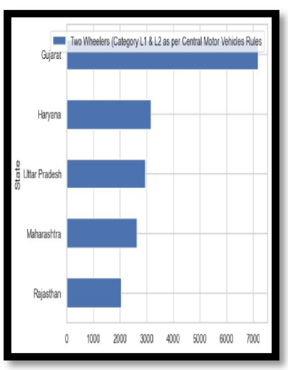
1. **Age vs. Price:**
   - There seems to be a scattered relationship between age and price, indicating no strong correlation between these two variables. This suggests that age does not significantly influence the price of the vehicle purchased.
2. **No. of Dependents vs. Price**:
   - There is a weak correlation between the number of dependents and the price of the vehicle. This might suggest that the number of dependents doesn't directly affect vehicle price choice, or the relationship is complex.
3. **Salary vs. Price**:
   - There appears to be a positive correlation between salary and vehicle price. Consumers with higher salaries tend to purchase higher-priced vehicles, which is consistent with common consumer behavior patterns.
4. **Wife Salary vs. Price**:
   - The relationship between wife's salary and vehicle price is not immediately clear from the scatter plot. However, it can be inferred that the wife's salary may have a minor role in vehicle purchasing decisions compared to the primary wage earner.
5. **Total Salary vs. Price**:
   - There is a clearer positive correlation between total salary (combined salary of both partners) and vehicle price. Higher total salaries tend to correlate with higher vehicle prices, supporting the notion that consumers with higher combined income are more likely to afford higher-priced vehicles.
6. **Overall Trends**:
   - **Salary** and **Total Salary** exhibit a clearer relationship with vehicle **Price**, whereas **Age** and **No. of Dependents** have relatively weaker or inconsistent relationships with vehicle pricing.
   - **Wife Salary** doesn't show a strong direct relationship with the price of the vehicle, possibly suggesting that the primary salary (from the main consumer) plays a more significant role in vehicle purchasing decisions.

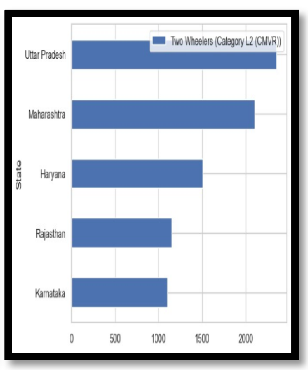
These insights indicate the importance of income levels, both individual and combined, in determining the vehicle price, while demographic factors such as age and dependents have a lesser impact.

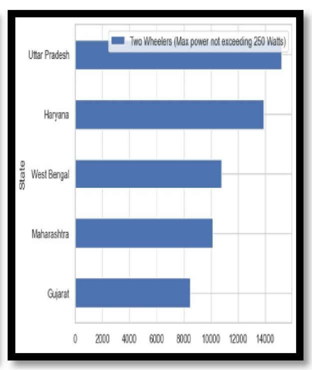
Geographic Analysis of Electric Vehicle Sales in India

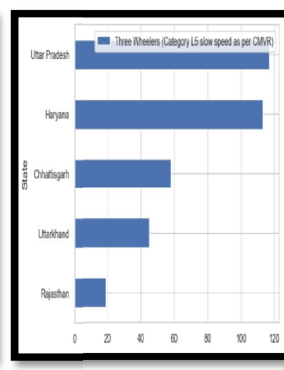
Geographic segmentation plays a crucial role in crafting effective marketing strategies by targeting consumers based on their location. This segmentation divides markets into regions, states, cities, or areas, allowing businesses to tailor their product offerings accordingly. In the context of electric vehicles (EVs), understanding the geographic distribution of consumers is essential for optimizing marketing efforts and enhancing sales.

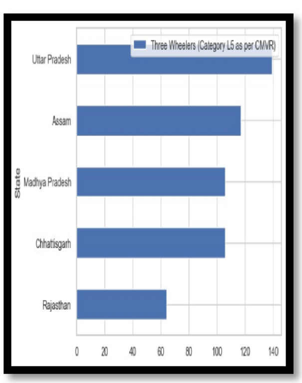
For this analysis, we utilized a state-wise dataset showcasing sales of various types of Electric Vehicles (EVs) across India. By examining the sales data, we can identify regions with the highest sales of specific EV types, enabling us to target these areas for future marketing campaigns.

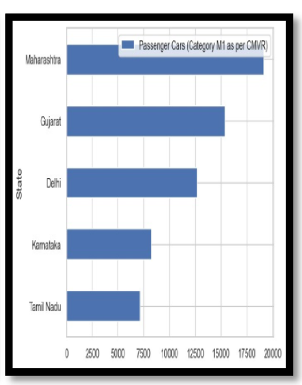


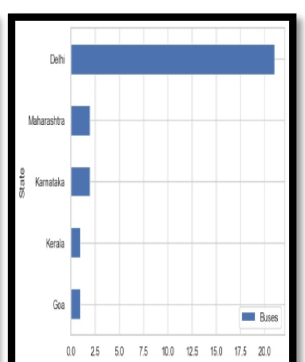


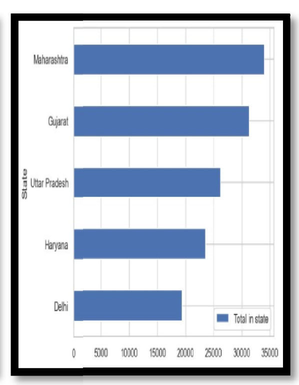












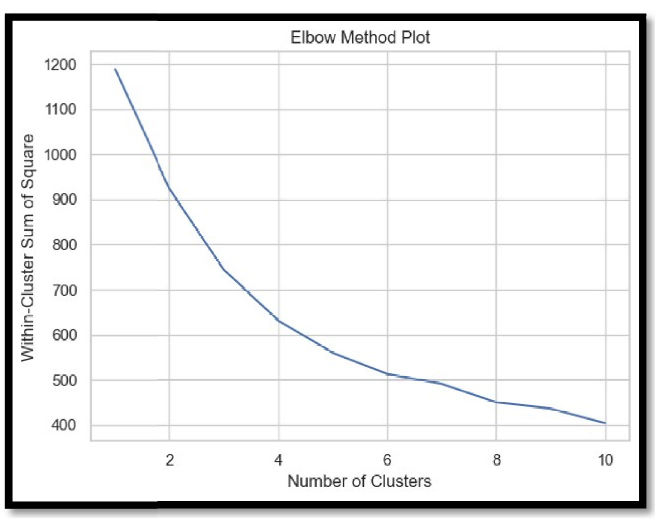
Key Insights:

- States with higher sales of specific EV types present an opportunity for businesses to concentrate their marketing efforts, as consumers in these regions are more inclined towards purchasing EVs.
- Targeting developed urban areas within these states can prove more effective, as these cities are likely to have a higher number of consumers who are willing to adopt electric vehicles.
- Factors such as the cost of the EV, average consumer income, availability of charging infrastructure, and the ability to maintain the vehicle are critical when considering the viability of targeting a particular region.

The following bar charts represent the top 5 states with the highest sales for different types of Electric Vehicles, offering valuable insights into potential target regions. By focusing on these states, businesses can align their product offerings and marketing strategies with consumer demand.

Determining the Optimal Number of Clusters

Upon analyzing the elbow plot, we observe two significant points where the curve shows a noticeable bend, indicating potential optimal values for **K**. These bends appear at **K=3** and **K=5**, suggesting that either of these values might provide the best clustering results. To further validate the ideal number of clusters, we will proceed with training the **K-Means Clustering** algorithm using both **K=3** and **K=5** values.

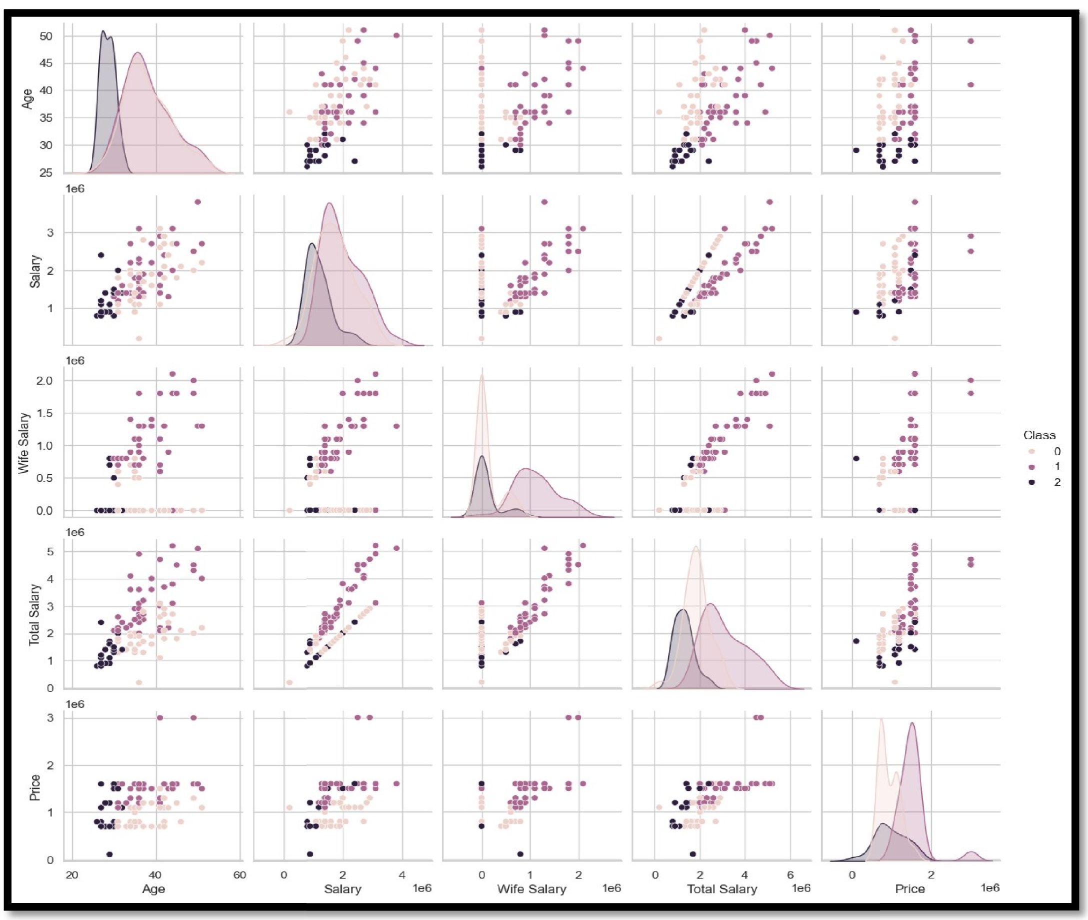


Impactful Attributes for Clustering

Through comprehensive analysis of the **behavioral, psychographic, geographic, and demographic** factors, we identified several attributes that significantly influence consumer grouping. Upon reviewing the clustered dataset, we found that the following five attributes played the most pivotal roles in determining the clusters:

1. **Age**
2. **Salary**
3. **Wife's Salary**
4. **Total Salary**
5. **Price**

These influential attributes were further visualized in the pairplot, created for both **K=3** and **K=5**, providing a clearer understanding of how these features contributed to the clustering patterns.

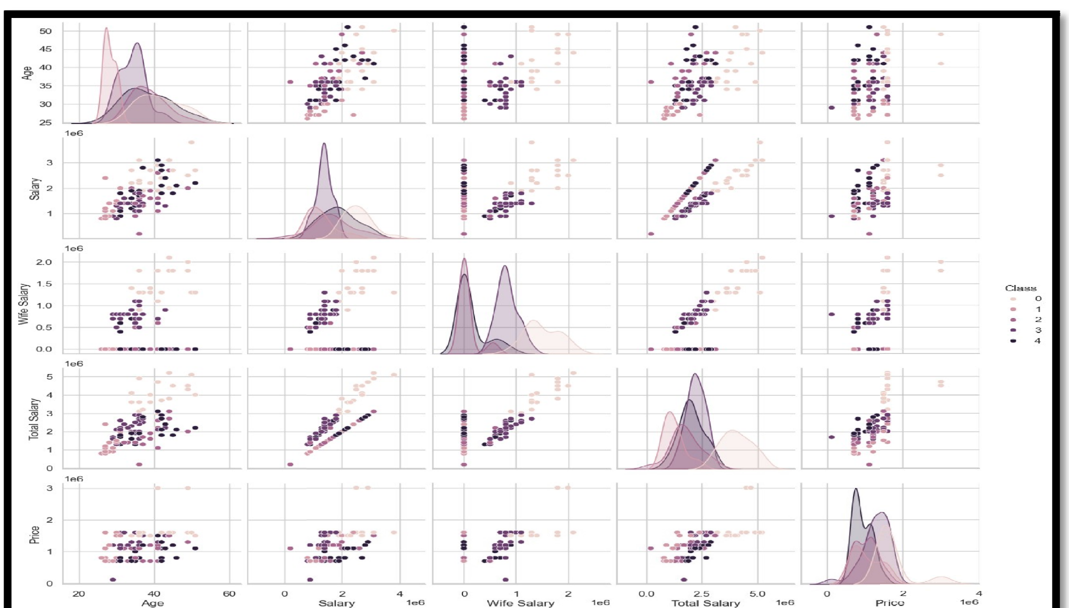


Clustering Insights Based on Total Income

Upon examining the clustered dataset, we observe that the **K-Means Clustering** algorithm has naturally grouped individuals based on their total income. Specifically, the model has classified the individuals into three distinct groups based on the relationship between their **husband's salary** and **total salary**:

- **Class 1**: This group consists of individuals whose **Total Salary** is approximately equal to their **husband's salary**. These individuals fall into a similar income bracket, indicating less variation in household income.
- **Class 2**: This group includes individuals whose **Total Salary** is notably higher than the **husband's salary**, suggesting a higher combined household income and possibly different financial characteristics compared to Class 1.
- **Class 0**: This class represents individuals whose **Total Salary** is close to their **husband's salary**, but their total income is relatively lower compared to others. These individuals may have lower combined household income, which sets them apart from the other groups.

This analysis highlights the underlying income distribution and demonstrates how clustering can reveal meaningful insights about consumer groups based on financial factors.



Optimal Clustering Approach: Maintaining Segment Homogeneity:

Upon analyzing the clustering results, we observe that the dataset is being divided into very small groups. While the model identifies these groups as trends, this segmentation does not reflect the true underlying structure of the data. To ensure that we maintain homogeneity within our segments and avoid overfitting, we conclude that **K=3** provides the most meaningful and effective clustering. This choice ensures that the segments remain sufficiently broad while accurately reflecting the key patterns in the data. Thus, using **K=3** will yield the best results for this clustering analysis.

Target Segment Analysis and Marketing Strategy for Electric Vehicles

Target Segment

1. Age Group and Affordability

- **Younger Population**: While younger individuals are environmentally conscious and eager to embrace new technologies, their preference for less expensive vehicles makes Electric Vehicles (EVs) less affordable for this group.
- **Recommended Target**: Focus on financially capable individuals aged **30 to 40 years**. This demographic is likely to adopt new technologies while having the financial stability to afford EVs.

2. Geographic and Urban Focus

- **Urban Cities**: Major urban areas with established infrastructure and greater awareness of technology benefits offer a promising market for EVs.
- **Education and Awareness**: Educated urban populations are more likely to appreciate and invest in EVs due to environmental benefits.

3. Lifestyle and Family Dynamics

- **Married Individuals with Dependents**: This segment is more inclined to purchase vehicles, making them a viable target group.

4. Income Consideration

- **Income Range**: Buyers typically earn an average salary of around **30 lakh per annum**, with most automobile purchases falling in the **10-20 lakh range**.
- **Two-Wheeler Affordability**: Two-wheelers dominate the market as they are more cost-effective for the general population.

Marketing Mix

The marketing mix plays a crucial role in effectively positioning EV products in the market by addressing **Product**, **Price**, **Place**, and **Promotion** aspects.

1. Product Strategy

- **Two-Wheelers**: Given that two-wheelers hold the largest share of the Indian automobile market, an EV startup should prioritize this segment. They are cost-effective and align with the existing infrastructure.
- **Public Transport Vehicles**: With supportive government policies for electric-based engines, public transport vehicles present a secondary opportunity.

2. Pricing Strategy

- **Affordability**: Pricing must balance cost-effectiveness with profitability. The ideal range for EVs should be between **10-20 lakh**, catering to the majority of buyers.
- **Maintenance Costs**: Beyond purchase affordability, maintaining low operational costs will enhance appeal.

3. Place Strategy

- **Urban Focus**: Launching in major urban cities ensures alignment with infrastructure readiness and an educated customer base.
- **Geographic Analysis**: Identify and target states with the most promising markets based on infrastructure and demand potential.

4. Promotion Strategy

- **Education and Awareness**: The key promotional strategy should highlight the benefits of EVs, including reduced environmental impact and long-term cost savings compared to fuel-based vehicles.
- **Affordable Options**: If the company develops an affordable product, emphasizing this in marketing campaigns will attract more buyers.

By focusing on these elements, an EV startup can develop and execute effective strategies to penetrate and thrive in the competitive Indian market.