Strategic Market Segmentation for Electric Vehicle Adoption in India

Zeba Khanam

Date: 07-07-2025

GitHub Link: Click Here

Abstract

This report presents a data-driven market segmentation analysis aimed at identifying the optimal customer segments for Electric Vehicle (EV) adoption in India. By focusing on demographic and vehicle-related variables such as income and vehicle type, we uncover consumer patterns that guide marketing strategies and EV product placement. The outcome of this analysis helps new EV startups plan effective market penetration strategies aligned with India's sustainability goals.

Introduction

With India's EV market experiencing rapid transformation due to increasing fuel prices, environmental concerns, and policy support (like FAME-II), strategic segmentation has become essential. This study utilizes real-world consumer automobile purchase data to derive meaningful segments based on income and vehicle type to guide business strategy for EV manufacturers.

Data Source

The dataset used is titled **"Indian automobile buying behavior study 1.0.csv"**. It includes:

- Age, Gender, Education, Salary, and Dependents
- Vehicle brand (Make) and price
- Personal and wife salary
- Customer preferences

Data Preprocessing

Data cleaning involved several key steps to ensure accuracy and consistency in the dataset:

Handling Missing Values: Missing values in Price and Total Salary columns were imputed using their respective mean values.

Type Conversion: The No of Dependents column was converted from object to integer type.

Duplicate Removal: Duplicate rows were identified and removed to avoid bias.

Feature Engineering: Two new derived features were created:

Income_Category: A categorical grouping based on total salary.

Vehicle_Type: Grouping based on vehicle name patterns (e.g., car, bike, SUV, etc.)

These preprocessing steps ensured the dataset was clean, structured, and ready for clustering and segmentation analysis.

Tools and Python Libraries Used
In this project, the following Python libraries were used for various data handling and visualization tasks:

Library	Purpose / Use Case
pandas	For loading, exploring, cleaning, and manipulating tabular data from the CSV file.
numpy	For numerical operations like averaging and array manipulations.
matplotlib.pyplot	To create basic visualizations like line plots, pie charts, and bar charts.
seaborn	For advanced and elegant statistical visualizations such as boxplots, heatmaps, etc.
sklearn.cluster.KMeans	To perform K-Means Clustering for identifying customer segments.
sklearn.preprocessing.LabelEncoder	To convert categorical features (e.g., Income Category, Make) into numeric format.
sklearn.preprocessing.MinMaxScaler	Used to normalize numerical values (like Price) for 3D plotting.

Library	Purpose / Use Case
mpl_toolkits.mplot3d	To support 3D plotting using Matplotlib's Axes3D feature.
matplotlib.pyplot.Axes3D	Enables 3D scatter plots showing Age, Salary, and Price (scaled).

Code Snippet:

```
# Handle missing values
missing = df.isnull().sum()
print("\nMissing values per column:\n", missing)

# Fill null values
df['Price'].fillna(df['Price'].mean(), inplace=True)
df['Salary'].fillna(df['Salary'].mean(), inplace=True)

# Clean data
df.drop_duplicates(inplace=True)
df['No of Dependents'] = df['No of Dependents'].astype(int)
```

Output:

```
Missing values per column:
Age              0
Profession       0
Marrital Status  0
Education        0
No of Dependents 0
Personal loan    0
House Loan       0
Wife Working     0
Salary           0
Wife Salary      0
Total Salary     0
Make             0
Price            0
```

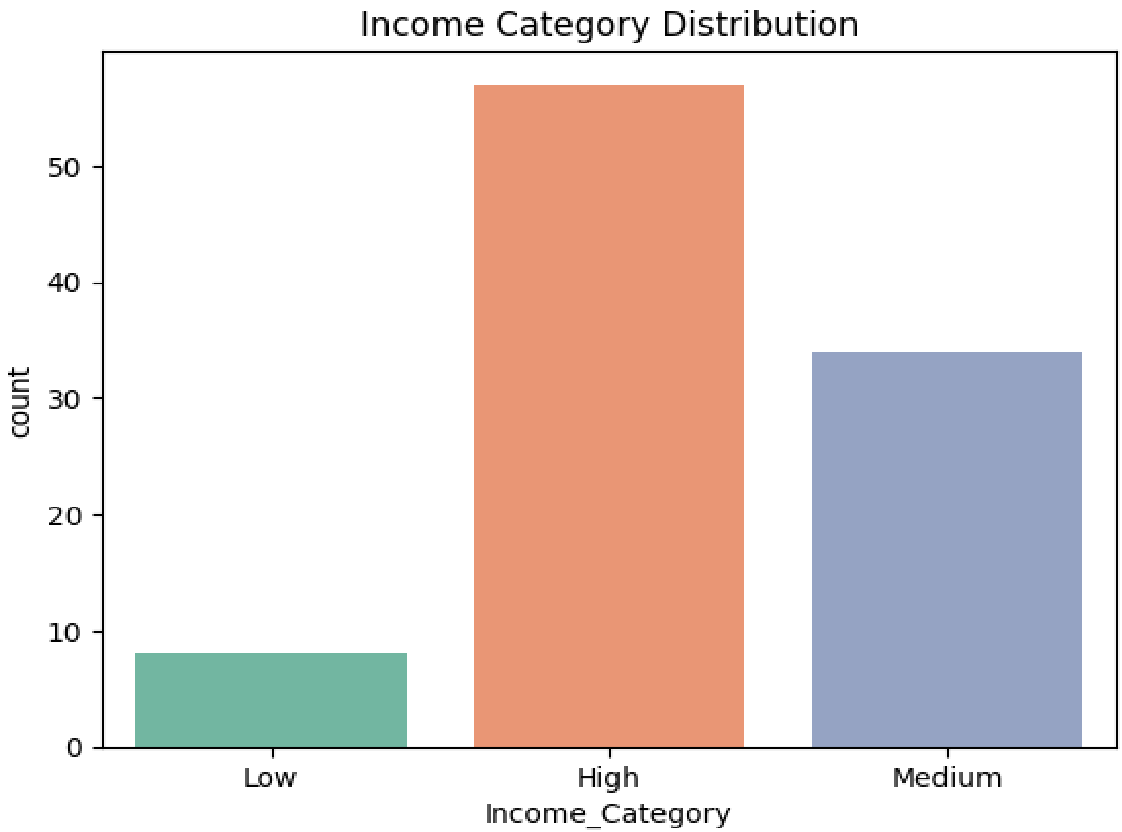
Income & Vehicle Type Categorization

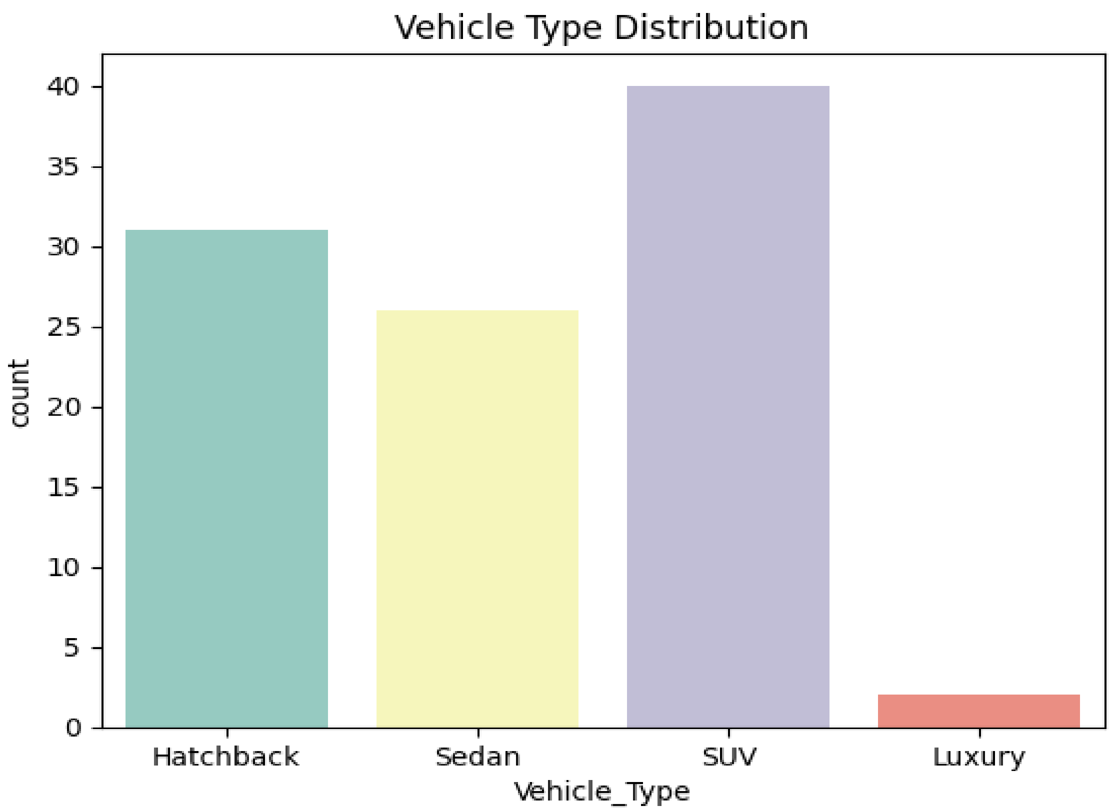
Based on Total Salary, customers were categorized into income groups: 'Low', 'Medium', or 'High'. Vehicle type was inferred from the brand name in the 'Make' column using a rule-based function.

Code Snippet:

```
def categorize_income(salary):
    if salary < 1000000:
        return 'Low'
    elif salary < 2000000:
        return 'Medium'
    else:
        return 'High'

df['Income_Category'] = df['Total Salary'].apply(categorize_income)

def get_vehicle_type(make):
    make = make.lower()
    if make in ['i20', 'alto', 'swift', 'wagonr', 'baleno']:
        return 'Hatchback'
    elif make in ['ciaz', 'city', 'verna']:
        return 'Sedan'
    elif make in ['duster', 'suv', 'creata', 'xuv', 'scorpio']:
        return 'SUV'
    elif make in ['fortuner', 'endeavour', 'luxuray']:
        return 'Luxury'
    else:
        return 'Other'

df['Vehicle_Type'] = df['Make'].apply(get_vehicle_type)
```

The results were visualized using countplot s to understand the distribution:


Income Category Distribution
count
0
10
20
30
40
50
Low
High
Medium
Income_Category

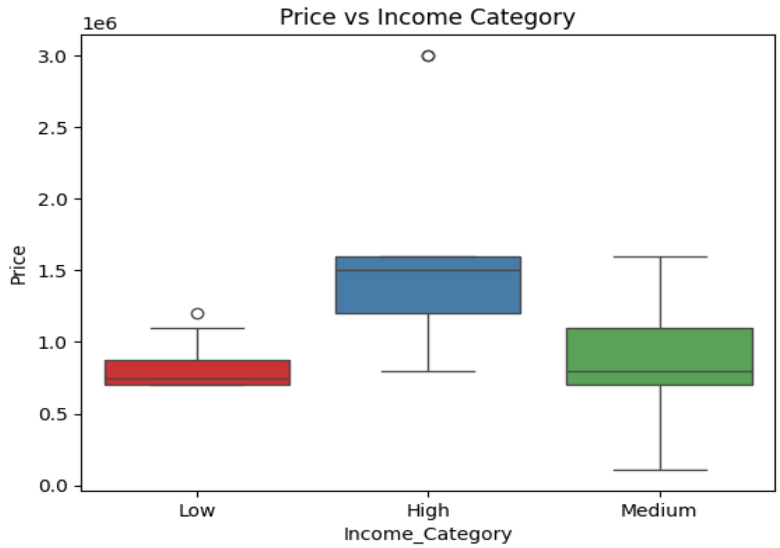


Clustering and Segment Analysis

- **Boxplot:** Price vs Income_Category
- **Barplot:** Vehicle Make Distribution
- **Pie Chart:** Customer Cluster Distribution
- **Line Plot:** Average Salary per Cluster
- **Heatmap:** Income Category vs Cluster
- **Stacked Bar:** Make vs Cluster
- **Histogram:** Age Distribution by Cluster
- **3D Scatter Plot:** Age vs Salary vs Price (scaled)

Price vs Income Category

To analyze how price varies across different income groups, a boxplot was used.
This helps identify the spread, median, and potential outliers in pricing per income category.

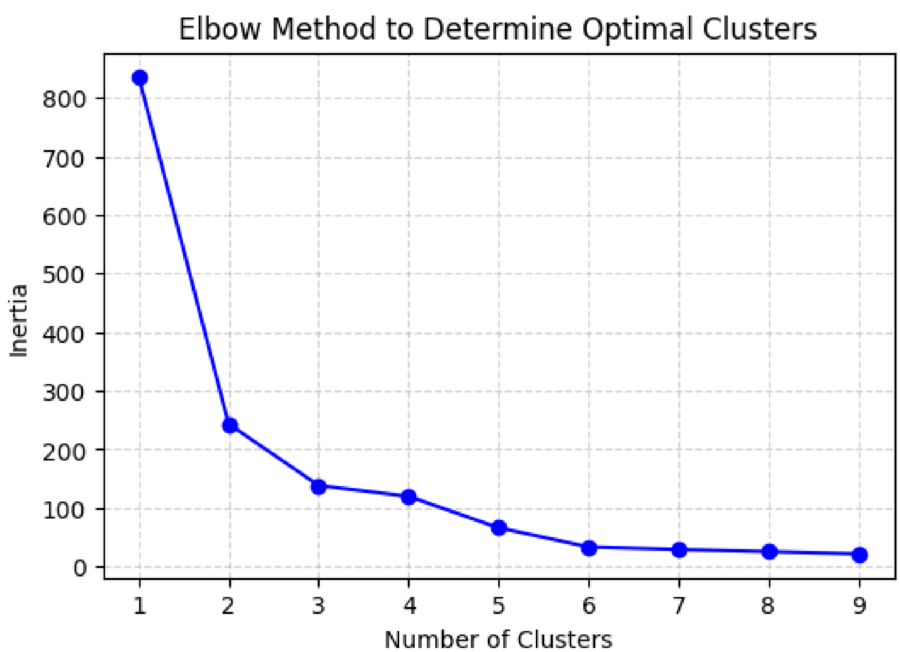


Encoding Categorical Features

Label Encoding was used to convert categorical features 'Income Category' and 'Make' into numeric values.
These encoded variables were used for clustering with KMeans.

Elbow Method for Optimal Clusters

The Elbow Method helps determine the optimal number of clusters by plotting inertia vs. cluster count.
A clear 'elbow' at K=3 indicates the ideal number of clusters.

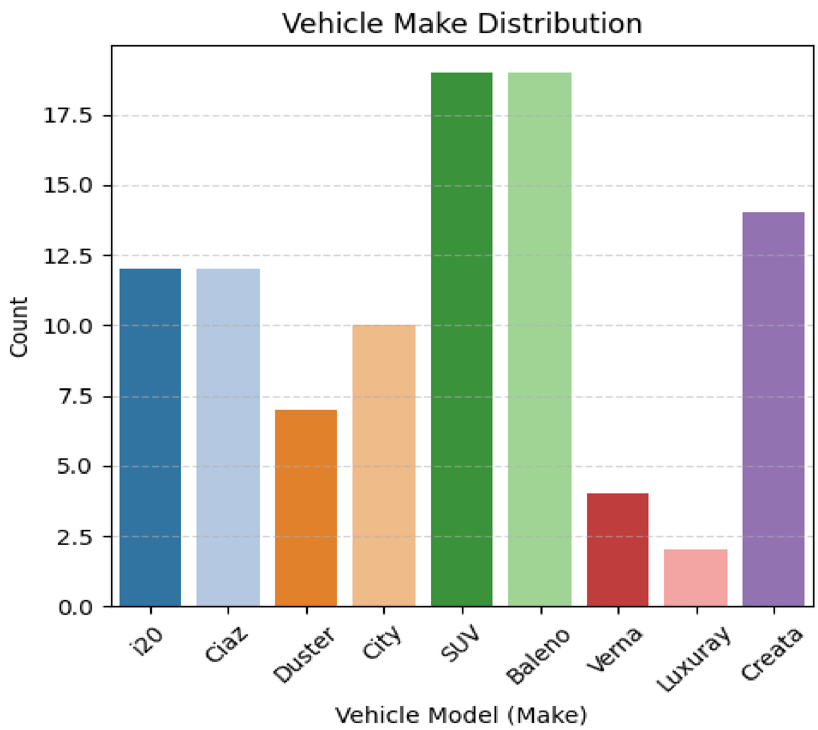


Final KMeans Clustering

Using K=3, customers were segmented into 3 clusters based on encoded income and vehicle make.

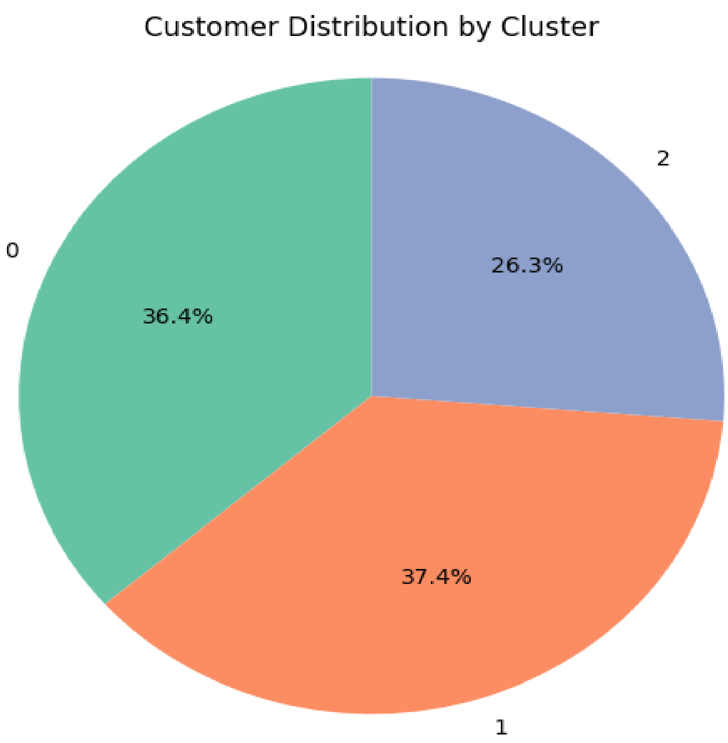
Vehicle Make Distribution

This bar chart shows the frequency of different vehicle models (makes) in the dataset.



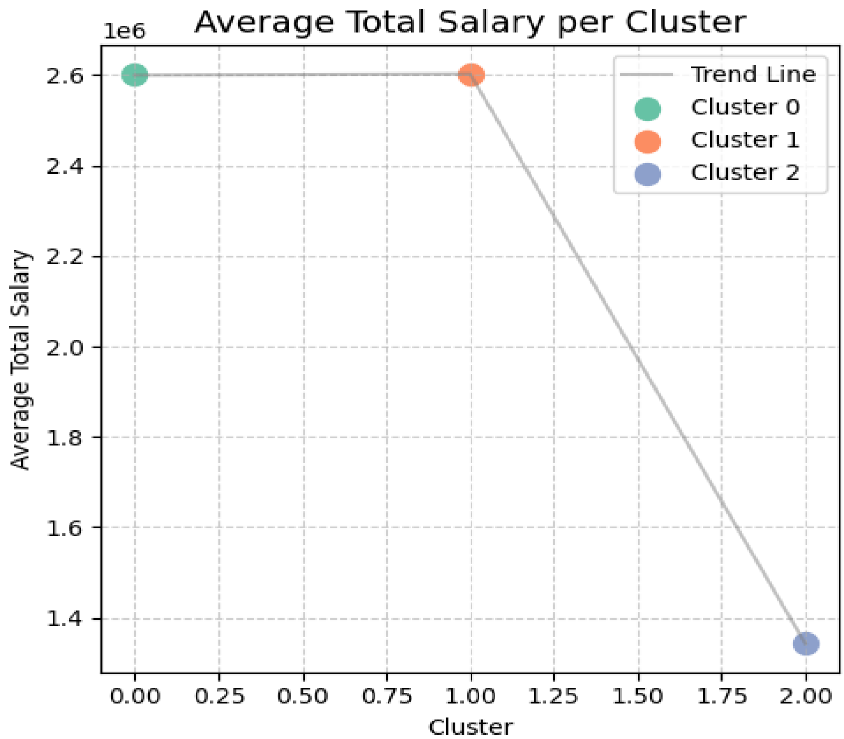
Customer Distribution by Cluster

A pie chart illustrates the percentage distribution of customers in each of the three clusters.



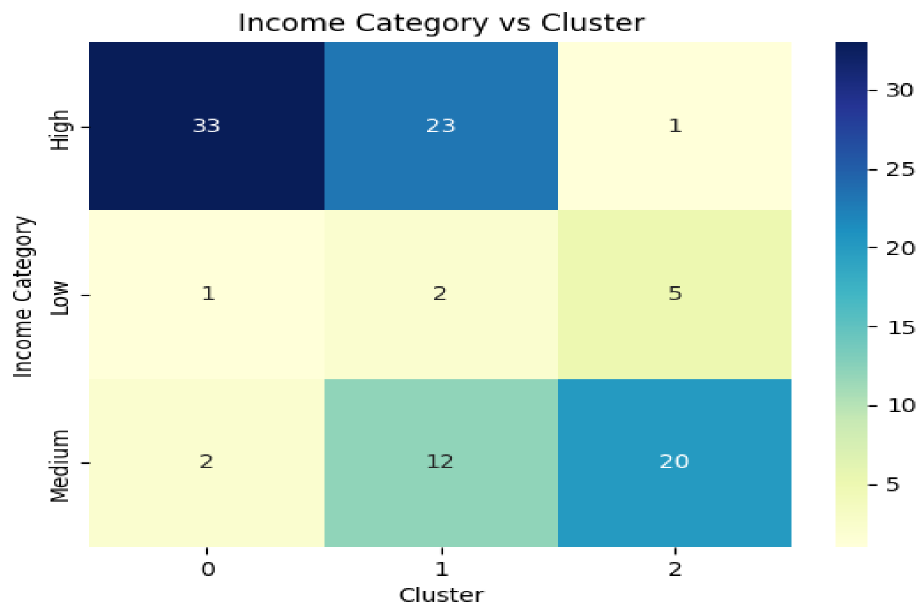
Average Salary per Cluster

This scatter plot and trend line show how average total salary varies between clusters.



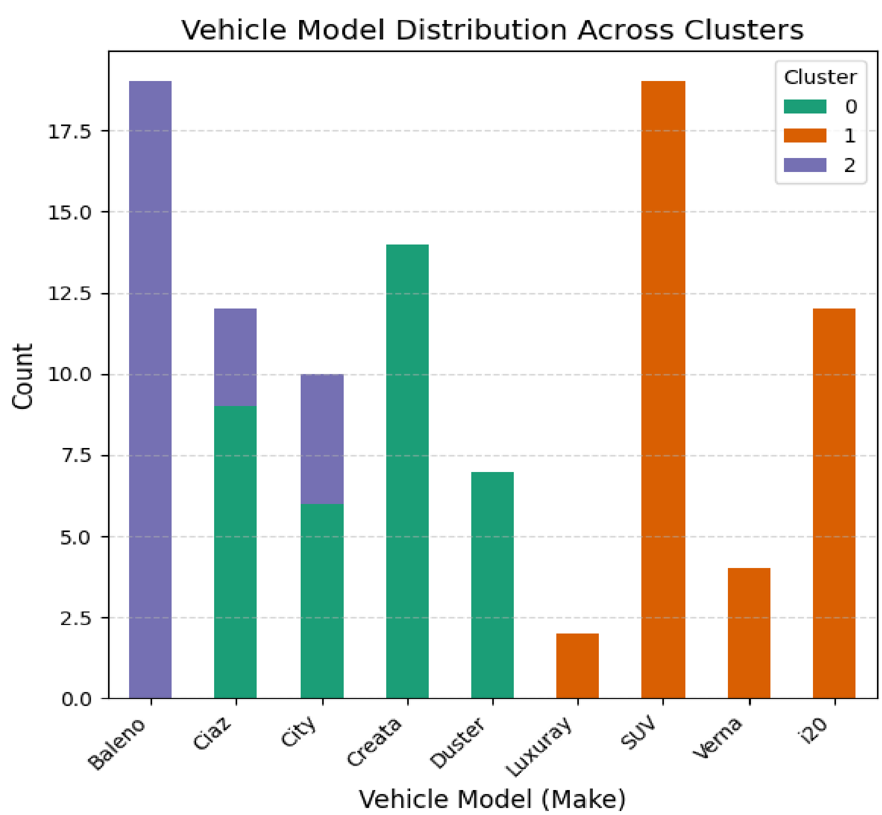
Income vs Cluster Heatmap

This heatmap shows the count of each income category per cluster.



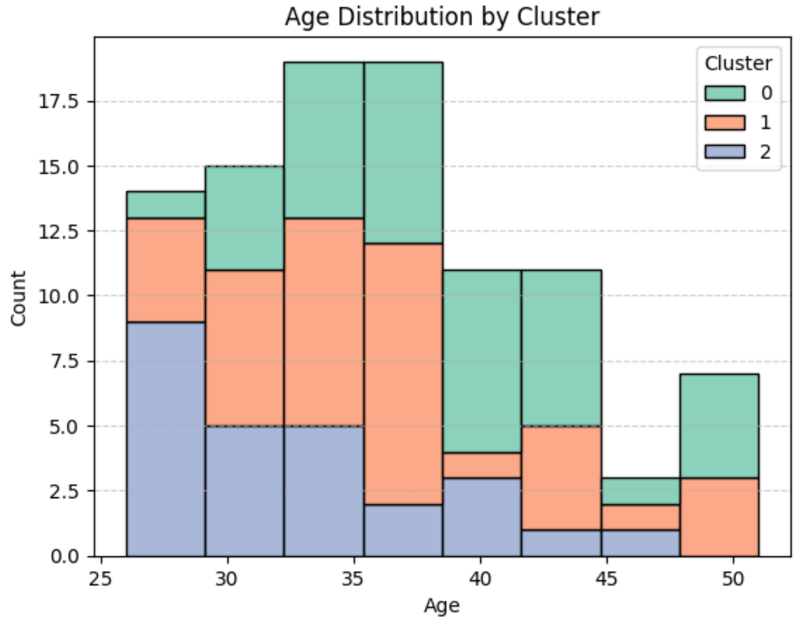
Make vs Cluster (Stacked Bar)

This plot illustrates how vehicle models are distributed across clusters.



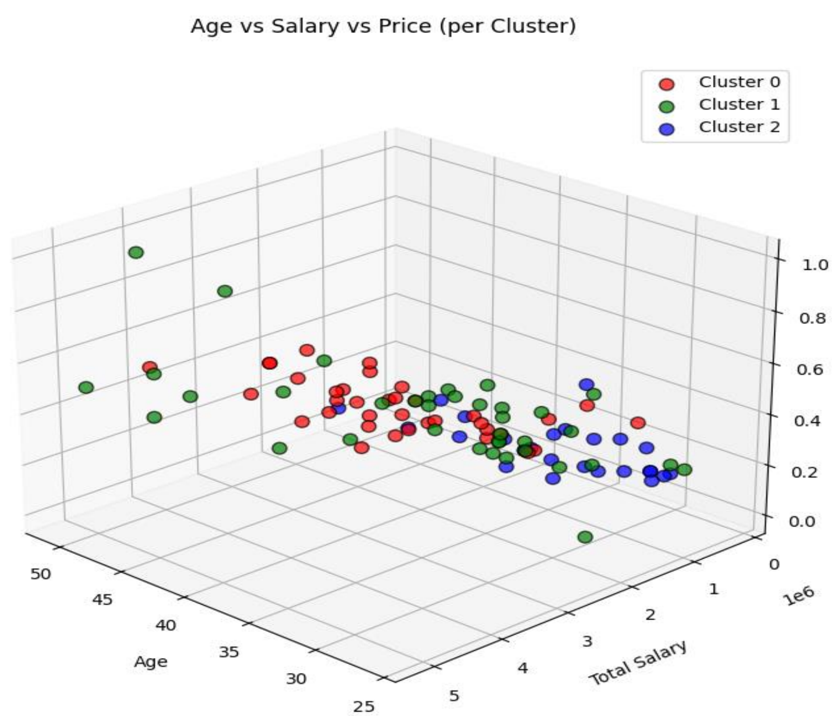
Age Distribution by Cluster

This histogram shows the distribution of age across different clusters.



Age vs Salary vs Price

This 3D scatter plot visualizes the relationship between age, salary, and vehicle price for each cluster.



Profit Estimation

We estimated early market profit based on 10,000 early adopters:
• Selling Price: ₹8,00,000
• Profit per Vehicle: ₹80,000
• Estimated Profit = 10,000 × 80,000 = ₹8 Crores

Cluster-wise Insights

Each cluster was profiled for average salary, popular vehicle make, and income group.

Output:

Cluster 0 Summary:

- Average Salary: ₹2600000.0
- Most Common Income Group: High
- Most Preferred Vehicle Type: SUV
- Popular Make: Creata

Cluster 1 Summary:

- Average Salary: ₹2602702.7
- Most Common Income Group: High
- Most Preferred Vehicle Type: SUV
- Popular Make: SUV

Cluster 2 Summary:

- Average Salary: ₹1342307.69
- Most Common Income Group: Medium
- Most Preferred Vehicle Type: Hatchback
- Popular Make: Baleno

Target Segment Selection

Cluster 2 is ideal for targeting because:

- Younger customers
- Prefer Hatchbacks (cost-efficient)
- Medium salary (price-sensitive but urban)

Marketing Mix (4Ps)

Element	Strategy
Product	EV options in Sedan/Hatchback category
Price	₹13–18 Lacs
Place	Focus on Tier 1 & Tier 2 cities
Promotion	Digital ads for young professionals

Potential Sales Estimation

- 10,000 early adopters
- Price per vehicle = ₹8,00,000
- Profit per vehicle = ₹80,000
- **Estimated Profit = ₹88 Crores**

Conclusion

Cluster-based segmentation using income and vehicle type has revealed strong opportunities in India's growing EV market. Data-driven strategies like these help optimize marketing efforts, improve product targeting, and accelerate clean mobility adoption.