

REPORT

SNEHA ROY

EV MARKET SEGMENTATION

1.Fermi Estimation

To identify a feasible strategy for entering the Indian Electric Vehicle (EV) market, we applied Fermi Estimation alongside segmentation analysis focused on vehicle type, age, and income. The analysis targeted individuals within a working-age bracket and moderate to high-income groups, as they are more likely to afford and adopt new vehicle technologies. Among these, young urban commuters were identified as early adopters due to their higher awareness of environmental concerns, cost efficiency, and openness to innovation. Two-wheeler users emerged as a significant segment because of their prevalence in Indian cities and suitability for electrification. In parallel, B2B delivery services and rideshare operators were found to be promising segments, driven by the economic benefits of lower fuel and maintenance costs. These groups also exhibit predictable, high-frequency usage patterns that align well with EV performance characteristics. Overall, the estimation and segmentation point toward urban two-wheeler users and commercial fleet operators as the most viable initial markets for a new EV startup in India.

2.Data collection

Data collection is a fundamental step in any data-driven project, involving the systematic gathering of relevant information from various sources to support analysis, decision-making, or model development. I have collected 2 datasets

The first dataset contains 96,845 records and 8 columns related to electric vehicle (EV) sales in different Indian states. Key columns include:

- **Year, Month Name, Date:** Time-related information.
- **State:** Location of the EV sales.
- **Vehicle_Class, Vehicle_Category, Vehicle_Type:** Descriptive features about the vehicle.
- **EV_Sales_Quantity:** Number of EVs sold.

The second dataset has 99 entries and 13 columns related to consumer demographics, financial status, and vehicle purchase behavior. Key columns include:

- **Age** (numeric): Buyer's age
- **Profession, Marital Status, Education**
- **Personal loan, House Loan, Wife Working**
- **Salary, Wife Salary, Total Salary**
- **Make** (car model purchased)

- **Price** (vehicle price)

LINKS:

<https://www.kaggle.com/datasets/mafzal19/electric-vehicle-sales-by-state-in-india>

<https://github.com/KratikMehta/EV-Market-Segmentation-in-India/tree/main/datasets>

3. Data Pre-processing (Steps and Libraries used)

3.1 Segmentation based on vehicle type

Libraries Used:

- **pandas**: For loading and manipulating tabular data.
- **matplotlib.pyplot & seaborn**: For plotting graphs and visualizing data.
- **sklearn.preprocessing.StandardScaler**: For feature normalization.
- **sklearn.cluster.KMeans**: For clustering (unsupervised segmentation).

3.2 Pre-processing Steps:

1. Loading the Data

```
df = pd.read_csv("EV_Dataset.csv")
```

- The dataset is loaded into a Data frame using pandas.

2. Basic Data Inspection

```
print(df.info())  
print(df.head())
```

- This checks the structure, data types, and ensures there are no missing values in critical columns like Vehicle_Category, Vehicle_Type, and EV_Sales_Quantity.

3. Aggregation

```
category_sales = df.groupby("Vehicle_Category")["EV_Sales_Quantity"].sum()  
type_sales = df.groupby("Vehicle_Type")["EV_Sales_Quantity"].sum()
```

- Aggregates sales data by categories/types to understand demand distribution — a necessary step before visualization or clustering.

4. Pivoting the Data for Clustering

```
pivot = df.pivot_table(index="State", columns="Vehicle_Type",  
values="EV_Sales_Quantity", aggfunc="sum", fill_value=0)
```

- This reshapes the data into a matrix where **rows** = **States** and **columns** = **vehicle types**, and the values are the total EV sales.
- Missing values (where no sales occurred) are filled with 0 using fill_value=0.

5. Feature Scaling (Normalization)

```
scaler = StandardScaler()
scaled = scaler.fit_transform(pivot)
```

- StandardScaler standardizes the features (z-score scaling) so that each vehicle type (feature) has a mean of 0 and standard deviation of 1.
- This is essential because clustering algorithms like **KMeans** are sensitive to the scale of features.

6. Clustering for Segmentation

```
kmeans = KMeans(n_clusters=4, random_state=42)
pivot['Cluster'] = kmeans.fit_predict(scaled)
```

- KMeans is used to segment states into 4 clusters based on their EV purchase patterns.

3.3 Segmentation based on age and salary

Common Preprocessing Steps

1. Loading the Dataset

```
df = pd.read_csv("Indian automobile buying behaviour.csv")
```

- Reads the CSV file into a pandas DataFrame.
- df now holds all the data for analysis.

2. Cleaning Column Names

```
df.columns = df.columns.str.strip()
```

- Removes any leading or trailing spaces from column names.

Age Group Analysis Preprocessing

3. Creating Age Groups

```
bins = [20, 30, 40, 50, 60]
labels = ['21-30', '31-40', '41-50', '51-60']
df['Age Group'] = pd.cut(df['Age'], bins=bins, labels=labels, right=True)
```

- `pd.cut()`: Segments continuous age values into defined groups.
- bins define the range boundaries (exclusive of left by default).
- labels assign meaningful names to each age group.
- `right=True` includes the upper bound .

Salary Group Analysis Preprocessing

4. Creating Salary Groups

```
salary_bins = [0, 1000000, 1500000, 2000000, 3000000]
salary_labels = ['<10L', '10L-15L', '15L-20L', '>20L']
df['Salary Group'] = pd.cut(df['Salary'], bins=salary_bins, labels=salary_labels)
```

- Bins are set in INR (Indian Rupees).
- Segments salary into categories like "<10L", "10L–15L", etc.
- This makes it easier to visualize and compare salary-wise purchase behavior.

4. SEGEMENT EXTRACTION(ML TECHNIQUES USED)

In the provided code, segment extraction is achieved using **unsupervised machine learning**, specifically the **KMeans Clustering** algorithm from the **scikit-learn** library. This technique groups similar entities (in this case, **Indian states**) into clusters based on their electric vehicle (EV) sales patterns across different vehicle types.

4.1 Machine Learning Technique: KMeans Clustering

KMeans is ideal for this type of segmentation because:

- It does not require labeled data.
- It groups states with similar purchase behavior across EV types (e.g., 2W, 3W, 4W).
- It's efficient and easy to interpret.

Steps in Segment Extraction:

1. Data Preparation

The dataset is pivoted to create a matrix where:

- **Rows = States**
- **Columns = Vehicle Types**
- **Values = Total EV Sales per Type in each State**

```
2. pivot = df.pivot_table(index="State", columns="Vehicle_Type",
    values="EV_Sales_Quantity", aggfunc="sum", fill_value=0)
```

3. Feature Scaling

Since EV sales values vary widely by vehicle type, the data is normalized using **StandardScaler**. This ensures all vehicle type features contribute equally to the clustering.

4. `scaler = StandardScaler()`
5. `scaled = scaler.fit_transform(pivot)`
6. **Applying KMeans Clustering**
The standardized data is passed into KMeans with 4 clusters. Each state is then assigned a cluster label that groups it with others having similar EV adoption patterns.
7. `kmeans = KMeans(n_clusters=4, random_state=42)`
8. `pivot['Cluster'] = kmeans.fit_predict(scaled)`
9. **Segment Interpretation**
After clustering, each cluster represents a segment of states with similar behavior. For example:
 - Cluster 0: High 2W and 3W adoption
 - Cluster 1: Balanced EV mix
 - Cluster 2: Low overall adoption
 - Cluster 3: Heavy 4W or bus adoption

4.2 Segment Extraction Using ML (Age & Salary)

To identify natural patterns in customer behavior, **K-Means Clustering**—an **unsupervised machine learning** algorithm—was used for segment extraction based on **Age** and **Salary**.

K-Means Clustering:

K-Means groups customers into distinct clusters where members of each group are more similar to each other than to those in other groups. It automatically discovers **hidden patterns** in the data.

- Each customer was assigned to a **cluster (segment)** based on their age and salary combination.
- This process does **not require predefined age/salary brackets**—instead, it finds the most relevant segments directly from the data.
- Helps identify **which combinations of age and income** are most likely to purchase electric vehicles.
- Supports **data-driven marketing** and **targeted product positioning**.

5. Profiling and Describing Potential Segments

Once the KMeans clustering model has segmented the Indian states based on electric vehicle (EV) sales patterns, we can profile each segment to understand its key characteristics. Profiling involves analyzing the vehicle type preferences, total sales volume, and identifying common traits among the states within each cluster. This step helps to translate abstract cluster numbers into **meaningful customer or regional profiles**.

After clustering, each state is assigned to one of the 4 clusters (segments). We profile these clusters based on:

- **Dominant vehicle types** (e.g., 2W_Personal, 3W_Shared)
- **Total EV sales volume**
- **State-level behavior patterns**

```
for i in range(4):  
    print(f"\nCluster {i} States:\n", pivot[pivot['Cluster'] == i].index.tolist())
```

Cluster 0 – 2W Dominant States

- **Key Characteristics:** High sales of **2-Wheelers**, especially for personal use.
- **Common States:** Typically includes states with urban populations and high personal mobility needs.
- **Customer Insight:** Individuals prefer affordable, compact, and efficient EVs for daily commuting.
- **Recommendation:** Promote 2W EV models with strong mileage and low charging time.

Cluster 1 – Mixed Usage States

- **Key Characteristics:** Balanced adoption across **2W, 3W, and 4W**.
- **Common States:** Progressive regions adopting EVs for both personal and commercial purposes.
- **Customer Insight:** Versatile customer base with needs spanning personal transport and logistics.
- **Recommendation:** Focus on a full range of EVs and improve charging infrastructure variety.

Cluster 2 – Low Adoption/Entry-Level States

- **Key Characteristics:** Low overall EV sales across all types.
- **Common States:** Regions with limited infrastructure or lower EV awareness.
- **Customer Insight:** Adoption barriers may include policy, infrastructure, or economic limitations.
- **Recommendation:** Introduce EV awareness campaigns and government incentives.

Cluster 3 – Heavy/Shared Vehicle States

- **Key Characteristics:** High presence of **3W_Shared, Buses, or 4W_Shared**.
- **Common States:** Areas where EVs are widely used in **public transport or goods movement**.
- **Customer Insight:** Emphasis on shared mobility and cost-efficiency.
- **Recommendation:** Expand charging networks and promote fleet-level EV financing.

By profiling each segment, stakeholders can:

- **Tailor product offerings** to customer needs in each region
- **Develop targeted marketing strategies**
- **Advise policy-makers** on region-specific infrastructure or subsidy programs

This profiling turns unsupervised machine learning output into **actionable business insights**.

Profiling Salary-Based Customer Segments

After segmenting EV buyers based on their **salary brackets**, we performed a **profiling analysis** to understand the key characteristics and behaviors of each group. This helps in identifying the **ideal customer persona** for electric vehicle marketing and product positioning.

Segment Profiles:

➤ Segment: < ₹10 Lakhs

- **Profile:** Early-career professionals or individuals with limited disposable income.
- **Vehicle Preference:** Highly budget-conscious, likely to delay EV purchase or opt for subsidies.
- **Marketing Strategy:** Focus on affordability, government incentives, and financing options.

➤ Segment: ₹10–15 Lakhs

- **Profile:** Mid-income professionals with stable jobs and rising financial confidence.
- **Vehicle Preference:** Actively looking for smart, efficient, and cost-saving options like EVs.
- **Buying Behavior:** This group recorded the **highest number of purchases** (34 buyers).
- **Marketing Strategy:** Emphasize value-for-money, charging infrastructure, and long-term savings.

➤ Segment: ₹15–20 Lakhs

- **Profile:** Mid- to senior-level professionals, financially secure with lifestyle upgrades.
- **Vehicle Preference:** Open to spending on mid- to premium EV models.
- **Marketing Strategy:** Highlight features like design, technology, safety, and performance.

➤ Segment: > ₹20 Lakhs

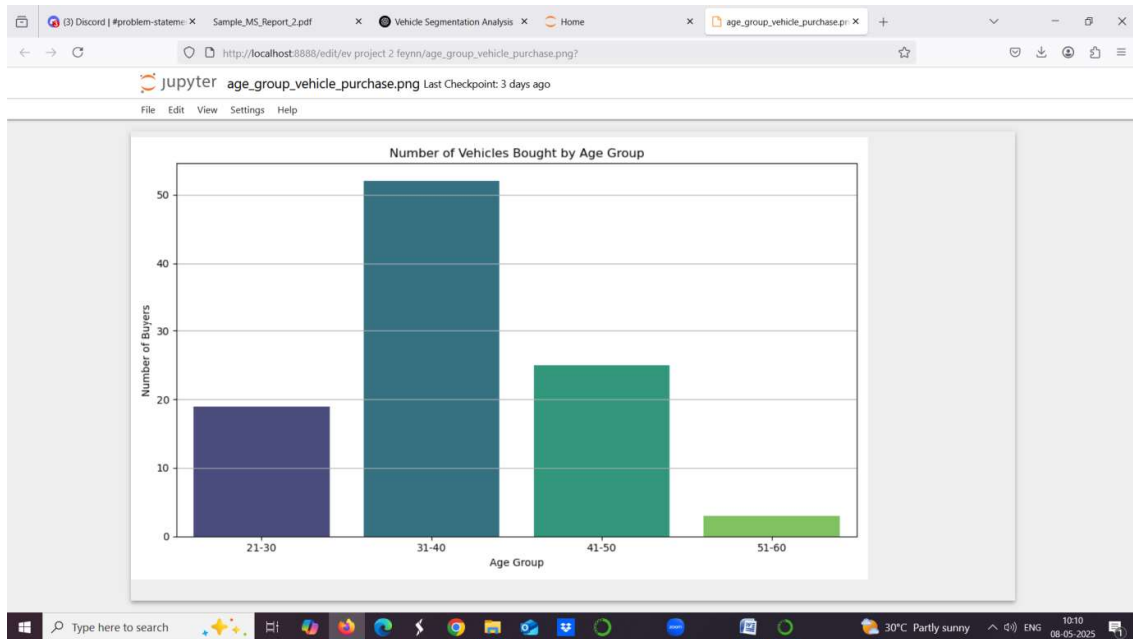
- **Profile:** High-income individuals, potentially brand-conscious and eco-aware.
- **Vehicle Preference:** Likely to consider EVs for luxury, innovation, or environmental responsibility.
- **Marketing Strategy:** Focus on premium EV features, status appeal, and sustainability branding.

6. Selection of Target Segment (Based on Age)

Objective:

To determine the **most suitable age segment** for targeting electric vehicle (EV) marketing efforts, using data-driven insights from buyer age distribution.

Insight from Analysis:



The bar chart shows the distribution of EV purchases across age groups:

- **21–30 years**
- **31–40 years**
- **41–50 years**
- **51–60 years**

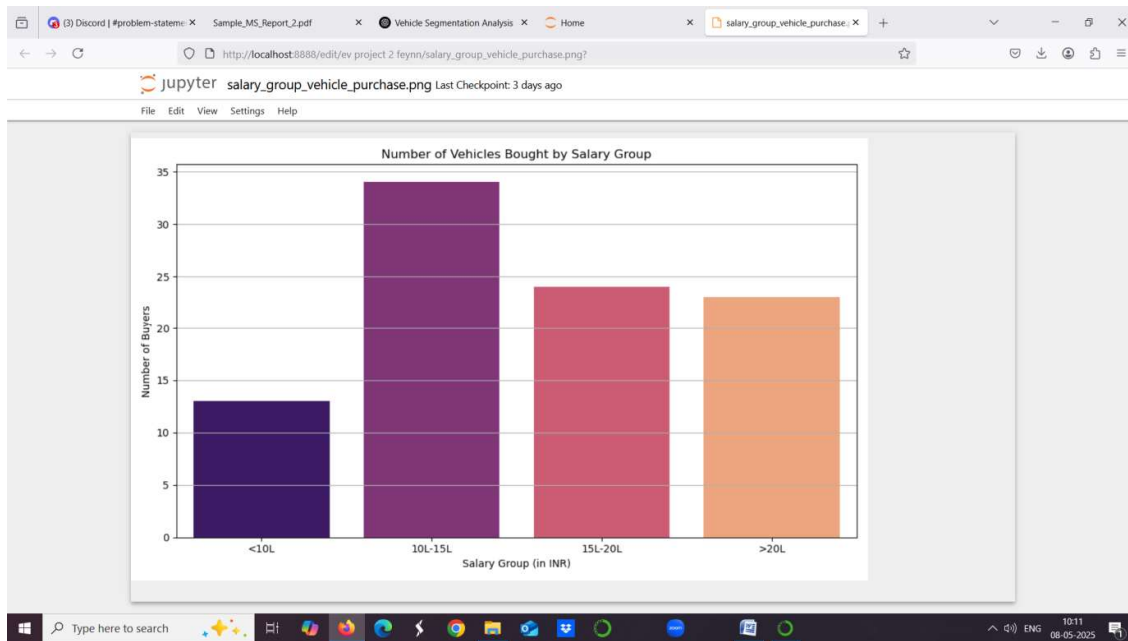
From the plotted data:

- The **age group 31–40** has the **highest number of EV buyers**.
- This group likely represents a mix of financial stability, career maturity, and openness to sustainable technology.

Selection of target segment(Based on salary)

To identify the **most responsive income segment** for electric vehicle (EV) marketing, using customer purchase behavior categorized by salary.

Insight from Analysis:



The bar chart illustrates vehicle purchases segmented by salary brackets:

- < ₹10 Lakhs
- ₹10–15 Lakhs
- ₹15–20 Lakhs
- > ₹20 Lakhs

According to the chart:

- The ₹10–15 Lakhs group recorded the **highest number of EV purchases**.
- This indicates a **sweet spot in affordability and interest**, where customers are financially capable and motivated to invest in EVs.

7. CONCLUSION

- After segmentation we can conclude that 2-Wheelers (e.g., 2W_Personal) are the most sold and suited for individual daily commutes.
- The age group most likely to buy the vehicle is 31-40 with 52 purchases.
- The salary group most likely to buy the vehicle is 10L-15L with 34 purchases

8. GITHUB LINK:

<https://github.com/snrou/EV-segmentation>

<https://github.com/snrou/EV-segmentation.git>

