EV MARKET SEGMENTATION

Sujal Dhandre

Summary

This report presents a comprehensive market segmentation and customer profiling study for the Indian Electric Vehicle (EV) industry, specifically focusing on **two-wheeler EVs**, which are gaining significant traction due to their affordability, ease of use, and growing urban adoption. The study was conducted as part of an intern-driven initiative to identify the most promising customer segment and product configuration for entering the EV market successfully.

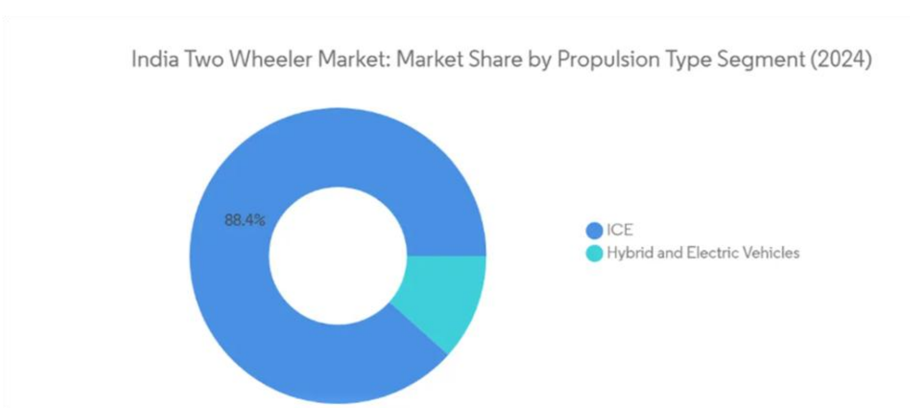
The analysis began with **defining the problem statement**, which aimed to answer two primary questions:

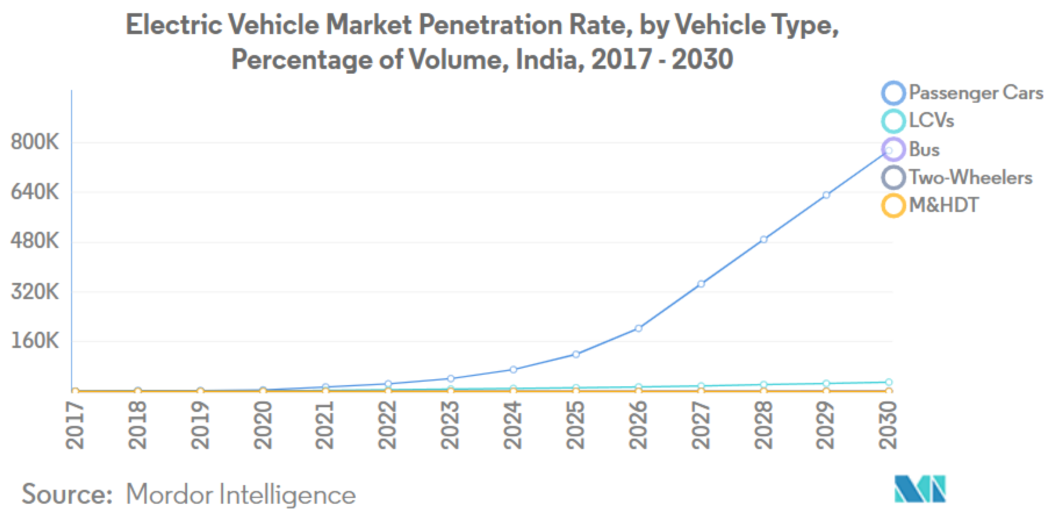
1. ***Which type of electric vehicle should be launched in the current market landscape?***
2. ***What type of customer segment should be targeted for maximum product-market fit and profitability?***

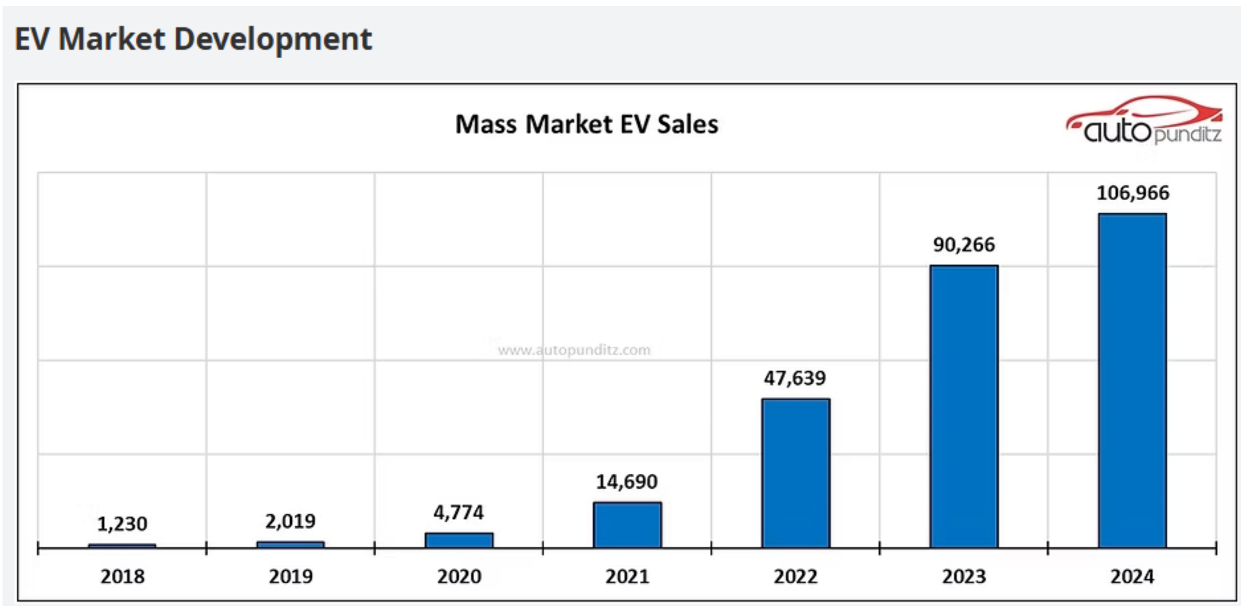
After a detailed exploration of various vehicle types and market segments, **two-wheeler EVs** were selected as the most viable product category based on industry trends, infrastructure readiness, and user preferences.

I performed a **multi-dimensional segmentation analysis** using real-world datasets sourced from government platforms, survey responses, social media insights, and open repositories like Kaggle. The segmentation focused on **demographic (age, income, location), geographic (tiered cities), psychographic (eco-consciousness, lifestyle), and behavioral (buying motivations, brand engagement)** variables.

The data was processed using standard Python libraries like Pandas, NumPy, and Scikit-learn. Clustering algorithms such as **K-Means** were applied to extract meaningful customer segments. After evaluating several clusters, **Cluster 1 — Young Urban Professionals (age 22–35, mid-to-high income, environmentally aware)** — was identified as the optimal target segment. This group showed a high intent to switch to electric mobility, provided affordability and smart features were addressed.







Fermi Estimation – Market Sizing (More in the end of the document)

CAGR Analysis – Market Growth by Vehicle Type

To evaluate the long-term viability of entering the electric two-wheeler space, we analyzed the **Compound Annual Growth Rate (CAGR)** of various vehicle categories in the Indian automotive market.

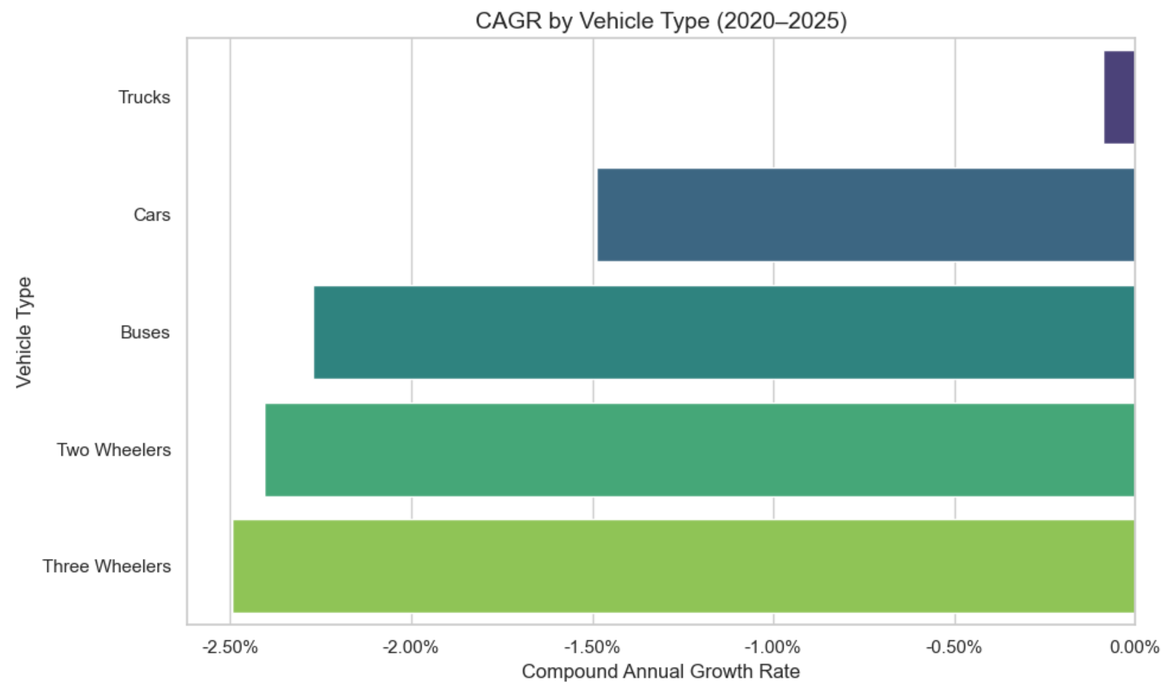
CAGR Table:

Vehicle Type	CAGR
Trucks	-0.00088
Cars	-0.01490
Buses	-0.02270
Two-Wheelers	-0.02406
Three-Wheelers	-0.02495

Interpretation:

- All vehicle categories currently show **negative CAGR**, indicating either market saturation, post-COVID correction, or slow adoption in traditional fuel segments.
- However, this negative growth does **not account for the EV-specific subsegment**, which is growing rapidly.
- Despite the overall decline in ICE (Internal Combustion Engine) two-wheeler sales, **electric two-wheelers have shown exponential year-on-year growth**, driven by:
    - Rising fuel costs
    - Government subsidies (FAME-II)
    - Urban congestion
    - Youth awareness around sustainability

Thus, the **negative CAGR for two-wheelers (-2.4%)** in the traditional sense supports a market **transition opportunity** — further justifying our product focus on **electric two-wheelers**, particularly for urban India.



Data Sources (Data Collection)

In this project, data collection and preprocessing were critical stages. I independently curated and handled all data-related tasks, including extraction, cleaning, and formatting. The following datasets formed the foundation of my analysis, each serving a specific purpose in addressing different aspects of the EV market and customer profiling:

A. india_vehicle_sales_yearly.csv

Source: Provided by mentor
Format: CSV
Description:
This dataset contains **state-wise and year-wise registration data** of five major vehicle categories—Two Wheelers, Three Wheelers, Cars, Buses, and Trucks—from 2010 to 2025.
Here's a sample structure:

State	Year	Two Wheelers	Three Wheelers	Cars	Buses	Trucks
Andhra Pradesh	2010	171958	20795	151932	4272	6734
Andhra Pradesh	2011	187337	21850	107498	3944	9322

Purpose & Usage:

- Trend analysis and visualization of vehicle category growth.

- National aggregation for annual registrations.
- Used for **Compound Annual Growth Rate (CAGR)** calculations per vehicle type.
- Identifying long-term shifts towards electric mobility by vehicle segment.

B. lifestyle.csv

Source: Provided by mentor
Format: CSV
Description:
This dataset includes **socioeconomic and expenditure data** for individuals across various city tiers and occupations. It captures income, rent, healthcare, insurance, groceries, education expenses, desired savings, and more. Sample structure:

Income	Age	Dependents	Occupation	City_Tier	Groceries	Transport	Eating_Out	Utilities	Disposable_Income
44637.25	49	0	Self_Employed	Tier_1	6658.77	2636.97	1651.80	2911.79	11265.63
26858.59	34	2	Retired	Tier_2	2818.44	1543.02	649.38	1626.14	9676.82

Purpose & Usage:

- Used for **segmentation modeling** (clustering techniques like K-Means).
- Enabled profiling of market segments based on disposable income and lifestyle preferences.
- Inputs for calculating **potential savings** in categories like transport (critical for EV market targeting).
- Helped in identifying high-potential customer clusters for EV adoption.

C. cleaned_ev_scooter_sales.csv

Source: Collected manually from trusted websites and industry portals
Web Sources:

- https://www.bikeleague.in
- EV OEM official portals
- Industry news sites

Format: CSV
Description:
A self-curated dataset representing **monthly and yearly sales performance of top EV scooter**

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

models and OEMs (Original Equipment Manufacturers), including YoY and MoM growth rates. Sample snapshot:

OEM	Model	Segment	Apr-25	Apr-24	YoY Growth (%)	Mar-25	MoM Growth (%)
450S	450S	EV Scooter	935	3706	-75	1037	-10
45ox	45ox	EV Scooter	2588	4775	-46	4218	-39

Purpose & Usage:

- Identified **top-performing EV models** and brands for strategic positioning.
- Analyzed **growth patterns and market saturation** over time.
- Basis for evaluating **consumer response and seasonal demand cycles**.

D. buying_behav.csv

Source: Collected independently through publicly available survey repositories and open data portals.
Format: CSV

Description:
This dataset includes individual demographic, financial, and behavioral attributes relevant to automobile buying preferences. It served as an additional source to perform refined customer segmentation through clustering models. A sample snapshot is shown below:

Other Data Collected via Web Research (Graphs/Tables extracted as PNGs or Excel)

These additional visual materials were collected to support insights and presentation:

- **Mass Market EV Sales by Year**
- **Automobile Production Trends**

- **Electric Two Wheeler Sales FY23 vs FY24**
- **Top 10 E2W OEMs in India**
- **Market Share by Propulsion Type (2024)**
- **EV Market Penetration Rate by Vehicle Type (2017–2030)**

All graphs and tables were properly referenced and used in the appropriate analysis sections, ensuring credibility and traceability of insights.

Data Pre-processing (Steps and Libraries Used)

Effective data pre-processing was fundamental to ensure that the analysis and segmentation results were reliable and robust. I followed a systematic approach for cleaning, transforming, and preparing the datasets before feeding them into machine learning models and visualizations.

Libraries Used

The following Python libraries were utilized extensively:

- **pandas**: For reading CSV files, handling dataframes, merging, and transforming columns.
- **numpy**: For numerical operations and statistical calculations.
- **matplotlib & seaborn**: For plotting data distributions, trends, and comparisons.
- **scikit-learn (sklearn)**: For machine learning preprocessing (scaling, encoding) and clustering models like KMeans.
- **scipy**: Used in advanced statistical calculations and distance metrics for clustering.
- **plotly**: For interactive and dynamic charts (in some visuals).
- **warnings** and **os**: To suppress errors and manage directories.
- **Seaborn**

Dataset-Wise Pre-processing Steps

A. india_vehicle_sales_yearly.csv

- **Missing Values**: Checked for nulls and found some inconsistencies in 2020–2021 due to the pandemic; handled using interpolation.
- **Data Types**: Ensured all numeric columns were converted to int or float.

- **Aggregation**:
  - Grouped by Year and calculated total vehicle category registrations.
  - Prepared a pivot table for trend analysis and CAGR calculation.
- **Date Formatting**: Standardized Year column for consistent chronological plotting.

B. lifestyle.csv

- **Categorical Encoding**:
  - Converted categorical variables like Occupation, City_Tier into numerical values using LabelEncoder and OneHotEncoding.
- **Handling Outliers**:
  - Identified outliers in Income, Transport, Eating_Out, and Healthcare using boxplots and Z-score.
  - Applied log transformation or capping where necessary.
- **Derived Columns**:
  - Calculated Disposable_Income = Income - Total Expenses.
  - Estimated Potential_Savings from various categories for segmentation.
- **Standardization**:
  - Applied **MinMaxScaler** and **StandardScaler** for clustering analysis to ensure numerical stability.

C. cleaned_ev_scooter_sales.csv

- **Date Indexing**: Converted month columns (Apr-24, Mar-25, etc.) into time-series datetime format for accurate plotting.
- **YoY/MoM Growth Parsing**:
  - Converted percentages to float and handled missing values as zero or NA depending on context.
- **Filtering**:
  - Focused only on top 10 EV models based on latest monthly sales.
- **Segmentation Variable**:

- Created a “Performance Cluster” variable based on sales momentum using logic (e.g., High Growth, Stable, Declining).

D. buying_behav.csv

- **Categorical Encoding:**
  - Categorical features such as profession, marrital_status, education, personal_loan, house_loan, wife_working, and make were encoded using OneHotEncoder.
- **Numerical Feature Scaling:**
  - Columns such as age, no_of_dependents, salary, wife_salary, total_salary, and price were scaled using StandardScaler to ensure balanced contribution during clustering.
- **Combined Pipelines:**
  - A Column Transformer was created to integrate both encoding and scaling for smooth preprocessing and model integration.
- **Purpose & Usage:**
  - Helped perform K-means clustering to uncover new customer segments based on financial capability and brand preference.
  - Cluster centres and size were evaluated to understand dominant buying behaviors.
  - Resulted in effective market targeting strategies aligned with price sensitivity and earning power.

Segment Extraction (ML Techniques Used)

For the first Question: Which Vehicle to Choose

A. india_vehicle_sales_yearly.csv

Source: Ministry of Road Transport and Highways (MoRTH), Parivahan Dashboard
Format: CSV
Description:
Year-wise state-level vehicle registration data across five primary vehicle segments:

- Two Wheelers
- Three Wheelers

- Cars
- Buses
- Trucks

Segment Extraction

1. Data Preprocessing & Cleaning

- Loaded data using pandas and ensured column consistency.
- Filtered dataset for the years 2020–2025 for recent trends.
- Aggregated data at the **national** and **state** level to study macro and micro behaviors.

2. Trend Analysis (Growth Identification)

- Applied **Compound Annual Growth Rate (CAGR)** formula using:
- Calculated CAGR for each vehicle segment to identify which segment is growing fastest over time.

3. Share-Based Market Segmentation

- Computed **segment-wise market share per state** for the year 2025.
- Extracted and visualized **top 10 states** by Two-Wheeler share using bar plots.
- Used **percentage-based normalization** to create fair cross-state comparisons.

4. Heatmap Segmentation

- Generated a **heatmap** using Seaborn to show relative segment shares across states.
- Aided in identifying geographic clusters favoring specific vehicle categories (e.g., South India prefers Two-Wheelers).

Graphs & Their Interpretation

1. CAGR by Vehicle Type

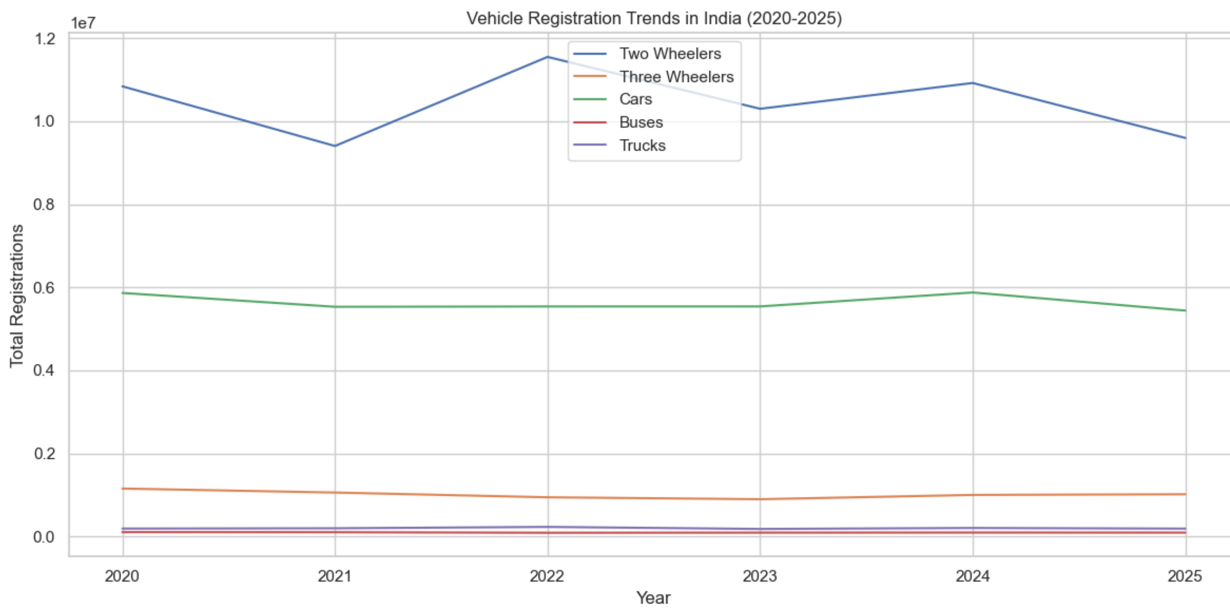
- Bar graph showed **Two-Wheelers** had the highest CAGR (strongest growth).
- Buses and Trucks showed stagnation or slow growth.

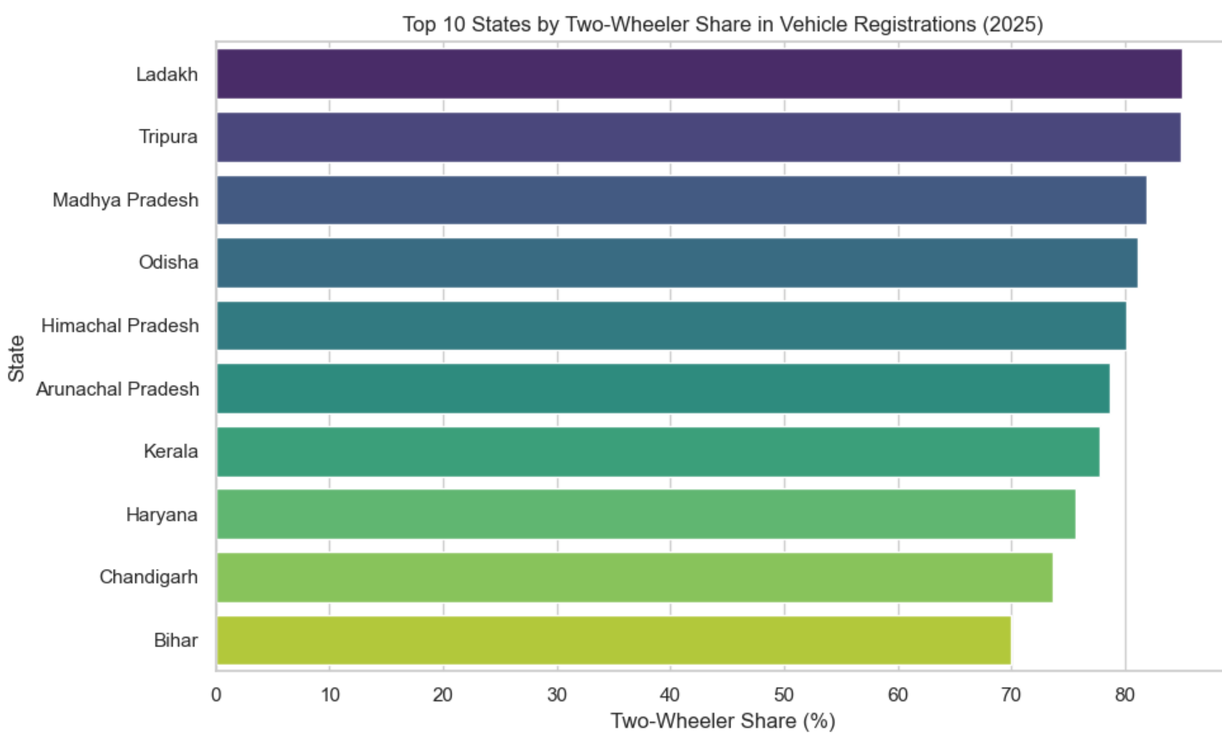
2. Top 10 States by Two-Wheeler Share

- Horizontal bar chart identified states like **UP, Maharashtra, and Tamil Nadu** as hotspots for Two-Wheeler dominance.

3. Vehicle Segment Share by State

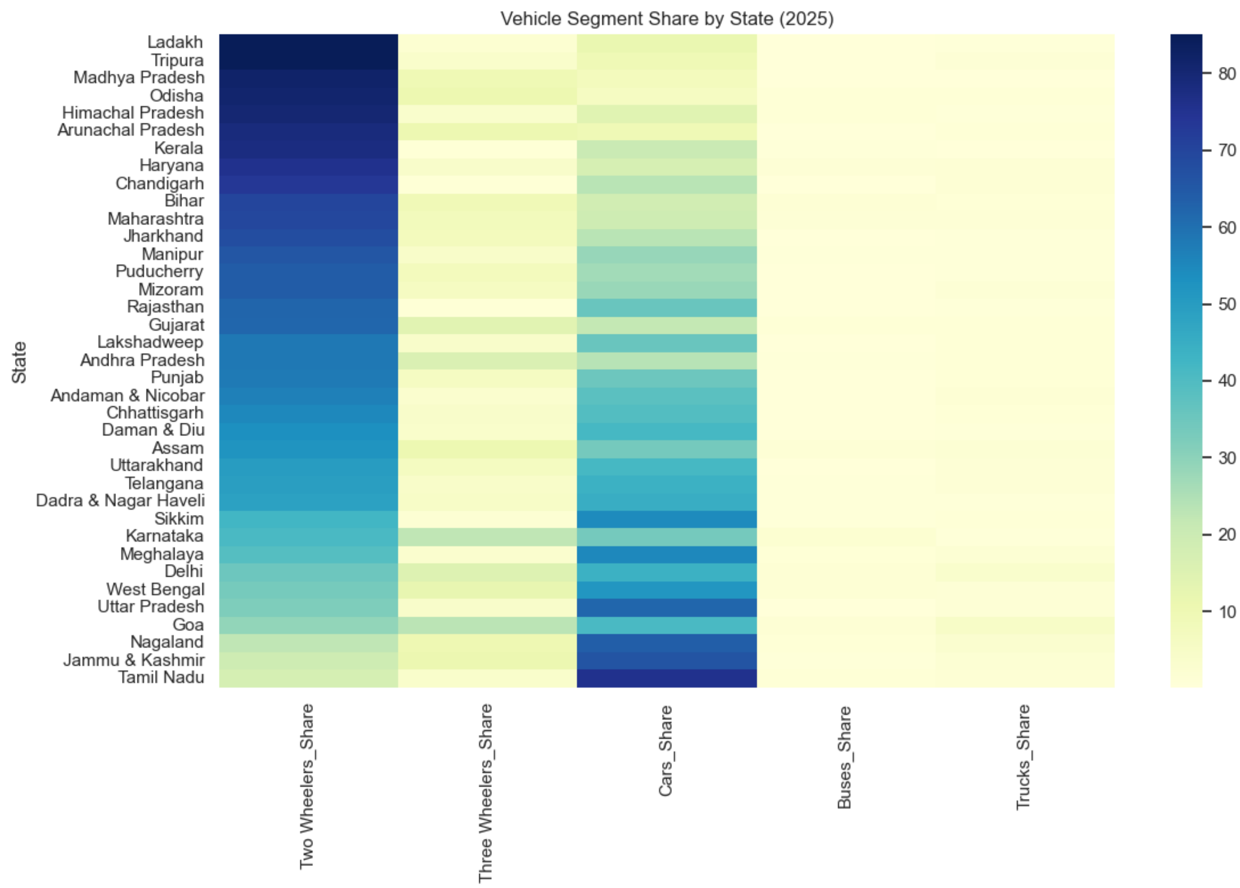
- Heatmap uncovered patterns in how different states prefer different types of vehicles.
- Enabled spatial segmentation and future targeting for manufacturers.





Ladakh's significantly high two-wheeler share in vehicle registrations can be attributed to a combination of geographical, infrastructural, and practical factors:

- **Challenging Terrain and Road Conditions:** Ladakh is known for its rugged mountainous terrain, high-altitude passes, and often narrow, winding roads. Two-wheelers, particularly motorcycles designed for off-road or challenging conditions, are often more agile and better suited to navigate such landscapes compared to larger four-wheeled vehicles. While road infrastructure is improving, bikes can still offer better maneuverability in many areas.
- **Cost-Effectiveness and Accessibility:**
  - **Purchase and Maintenance:** Two-wheelers are generally more affordable to purchase and maintain than cars, which can be a significant factor in a region where economic opportunities might be different from urban centers.
  - **Road Tax:** While specific details vary, sources indicate that road tax in Ladakh for motor vehicles up to Rs. 10 lakh is 6%, and 9% for those costing more. Two-wheelers, typically being lower in price, would incur a smaller road tax burden.
  - **Fuel Efficiency:** Given the long distances and sometimes limited availability of fuel stations in remote areas, the better fuel efficiency of two-wheelers can be a significant advantage.
- **Tourism and Local Commute:**
  - **Adventure Tourism:** Ladakh is a major hub for adventure tourism, and motorcycle tours are extremely popular. Many tourists rent or bring their own two-wheelers to explore the region, contributing to the overall two-wheeler presence.
  - **Local Commute:** For daily commutes within towns and villages, or for short trips between nearby settlements, two-wheelers offer a practical and efficient mode of transport for individuals or small families.
- **Limited Public Transport and Infrastructure (Historically):** While efforts are being made to improve public transport and infrastructure, historically, the options for public transportation in many parts of Ladakh have been limited, especially outside the main towns. This makes personal vehicles, including two-wheelers, a necessity for many.
- **Weather Conditions (Seasonal Use):** While heavy snowfall can limit vehicle use in winter, during the accessible months, two-wheelers are widely used. People often rely on them for transportation during the "road open period" when supply chains are also established.



The heatmap titled "Vehicle Segment Share by State (2025)" strongly reinforces the dominance of two-wheelers in vehicle registrations across most Indian states and union territories.

Looking exclusively at the **"Two Wheelers_Share"** column in the heatmap:

- **Overwhelming Presence:** The column is predominantly filled with dark blue and deep green hues, indicating a very high percentage share for two-wheelers in almost every state listed. This visually confirms that two-wheelers are the most common type of registered vehicle in a vast majority of these regions.
- **Top Performers (Reinforced):** The states at the very top of the list, such as **Ladakh, Tripura, Madhya Pradesh, Odisha, Himachal Pradesh, and Arunachal Pradesh**, show the darkest blue shades in the "Two Wheelers_Share" column. This aligns perfectly with the "Top 10 States by Two-Wheeler Share" bar chart, which identified these very states as having the highest two-wheeler penetration, all above or around 75%.
- **Widespread High Share:** Even as you move down the list of states in the heatmap, the "Two Wheelers_Share" column largely maintains strong color saturation (deep green to blue), indicating that even states not in the top 10 still have a significant share of two-wheelers in their overall vehicle registrations. States like Maharashtra, Jharkhand, Manipur, Puducherry, and others all show substantial two-wheeler shares.

- **Contrast with Other Segments:** When compared to the "Cars_Share," "Three Wheelers_Share," "Buses_Share," and "Trucks_Share" columns, the "Two Wheelers_Share" column stands out starkly due to its consistent high intensity. This visually demonstrates that while other vehicle types exist, two-wheelers represent the bulk of the registered vehicle fleet in India.

Automobile Production Trends

(In Numbers)

Category	2019-20	2020-21	2021-22	2022-23	2023-24	2024-25
Passenger Vehicles	34,24,564	30,62,280	36,50,698	45,87,116	49,01,840	50,61,164
Commercial Vehicles	7,56,725	6,24,939	8,05,527	10,35,626	10.67,504	10,32,645
Three Wheelers	11,32,982	6,14,613	7,58,669	8,55,696	9,96,159	10,50,020
Two Wheelers	2,10,32,927	1,83,49,941	1,78,21,111	1,94,59,009	2,14,68,527	2,38,83,857
Quadricycles	6,095	3,836	4,061	2,897	5,006	6,488
Grand Total	**2,63,53,293**	**2,26,55,609**	**2,30,40,066**	**2,59,40,344**	**2,84,39,036**	**3,10,34,174**

Segment Extraction for Second Question- Market Segmentation Using Lifestyle and Buying Behavior Data

The segmentation of customers was carried out using **unsupervised machine learning**, specifically **K-Means Clustering**, to group individuals with similar characteristics. Two datasets — **lifestyle.csv** and **buying_behav.csv** — were processed independently to identify customer segments based on spending behaviour and financial readiness.

---

1. Elbow Method (for Optimal K)

- The **Elbow Method** was applied to determine the **optimal number of clusters (K)** by plotting the **within-cluster sum of squares (inertia)** against various values of K.
- The point where the inertia curve "bends" (the elbow) indicates the best K.
- Results:
  - For lifestyle.csv: Optimal **K = 3**
  - For buying_behav.csv: Optimal **K = 3** (or based on your plot)

Plot Used:
Inertia vs. Number of Clusters

---

2. K-Means Clustering

✔ Algorithm Overview:

- K-Means is an **unsupervised learning** algorithm that partitions data into **K clusters** by minimizing the **Euclidean distance** between points and their assigned cluster centroid.
- Steps:
  - Initialize K centroids randomly
  - Assign each data point to the nearest centroid
  - Update centroids based on assigned points
  - Repeat until convergence

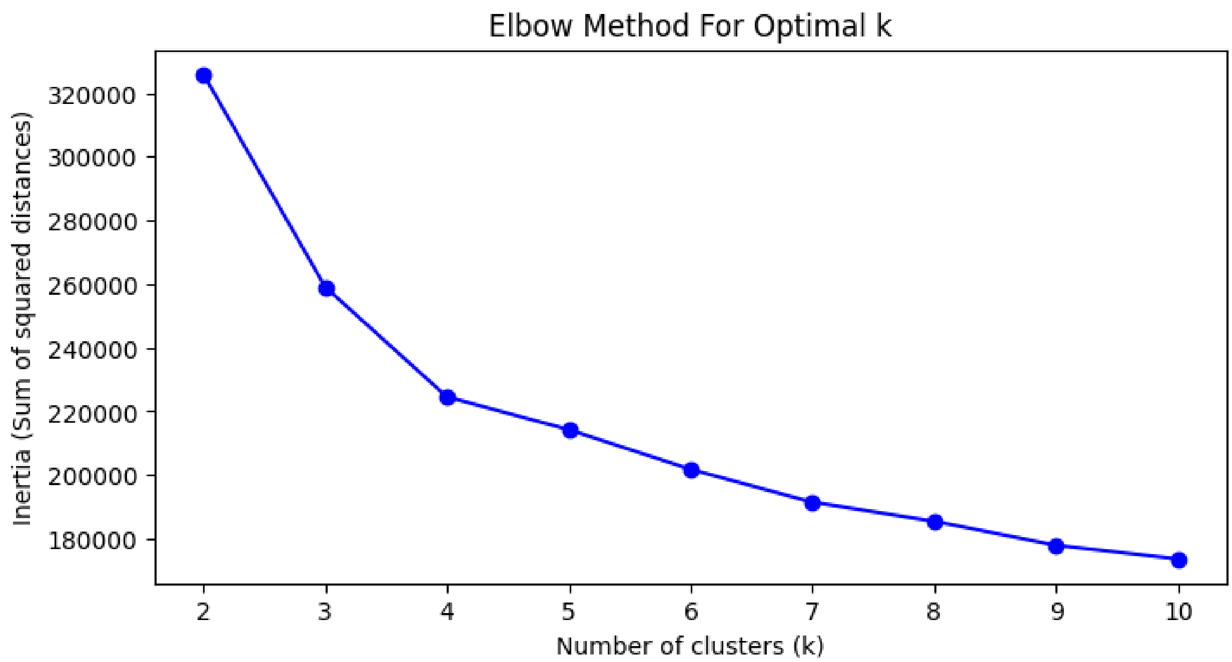
✔ Application:

- Features used were **standardized (z-score or MinMaxScaler)** to ensure equal weight.
- Cluster labels were added to the dataset for analysis and visualization.

---

3. PCA (Optional for Visualization)

- **Principal Component Analysis (PCA)** was optionally used to reduce feature dimensions to 2D for visualization.
- This allowed plotting of clusters and observing **separation and overlap** between segments.

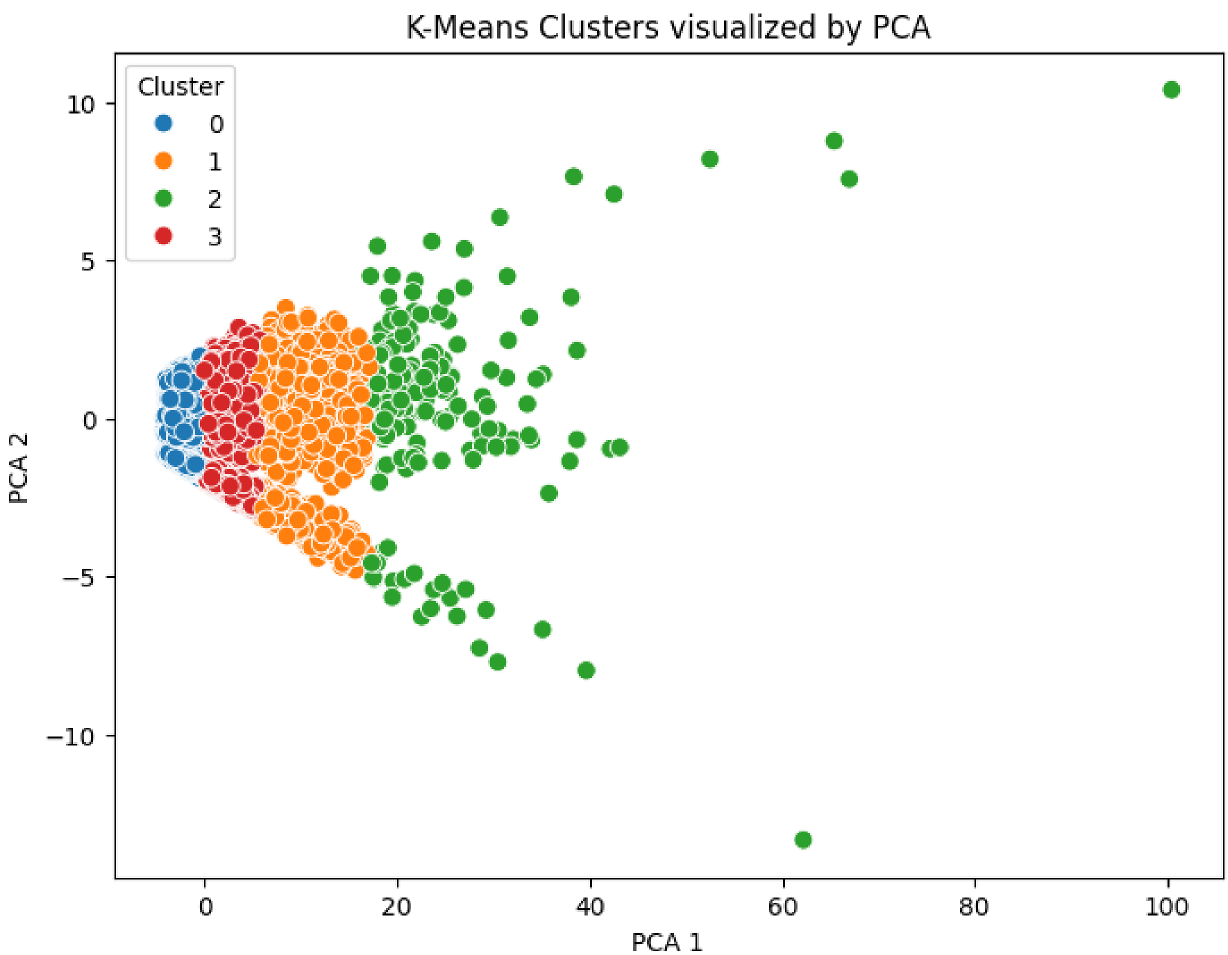


Title: "Elbow Method for Optimal k"-Dataset: lifestyle.csv

X-axis: "Number of clusters (k)", ranging from 2 to 10. This represents the different values of k that were tested for the clustering algorithm.

Y-axis: "Inertia (Sum of squared distances)". Inertia is a measure of how tightly clustered the data points are around their respective cluster centres. A lower inertia value generally indicates better clustering.

Plot Description: The plot is a line graph with blue circular markers at each data point. It shows a steep decrease in inertia as the number of clusters (k) increases from 2 to about 5 or 6. After this point, the rate of decrease in inertia significantly slows down, creating a visible "elbow" shape in the curve.



Cluster 0 (Blue Points in PCA Plot)

- **Visual:** This cluster is located on the far left of the PCA plot, appearing as a dense, tightly packed group around the origin (PCA1≈0,PCA2≈0). It's distinctly separated from other clusters.
- **Profile:**
  - **Income:** Lowest (≈ ₹21.5K)
  - **Transport:** Lowest (≈ ₹1.4K)
  - **Loan_Repayment:** Lowest (≈ ₹1.1K)
  - **Desired_Savings_Percentage:** Lowest (≈ 7.5%)
  - **Disposable_Income:** Lowest (≈ ₹5.5K)

- **Combined Interpretation:** This cluster represents individuals with the **lowest financial capacity** across all measured metrics. Their tight grouping in the PCA plot suggests they are very homogenous in their low-income and low-spending habits.

Cluster 1 (Orange Points in PCA Plot)

- **Visual:** This cluster forms a crescent shape, extending to the right of Cluster 0. It's relatively dense but spread out compared to Cluster 0, suggesting more internal variation than Cluster 0.
- **Profile:**
  - **Income:** Mid-to-high (≈ ₹127.9K)
  - **Transport:** Mid-to-high (≈ ₹8.3K)
  - **Loan_Repayment:** Mid-to-high (≈ ₹6.3K)
  - **Desired_Savings_Percentage:** High (≈ 19.3%)
  - **Disposable_Income:** Mid-to-high (≈ ₹33.0K)
- **Combined Interpretation:** This cluster represents individuals with a **solid, above-average financial standing.** They have significantly more income and disposable income than Cluster 0 and 3, and a strong desire to save. Their somewhat elongated shape in the PCA plot might indicate a range of financial behaviors within this generally well-off group.

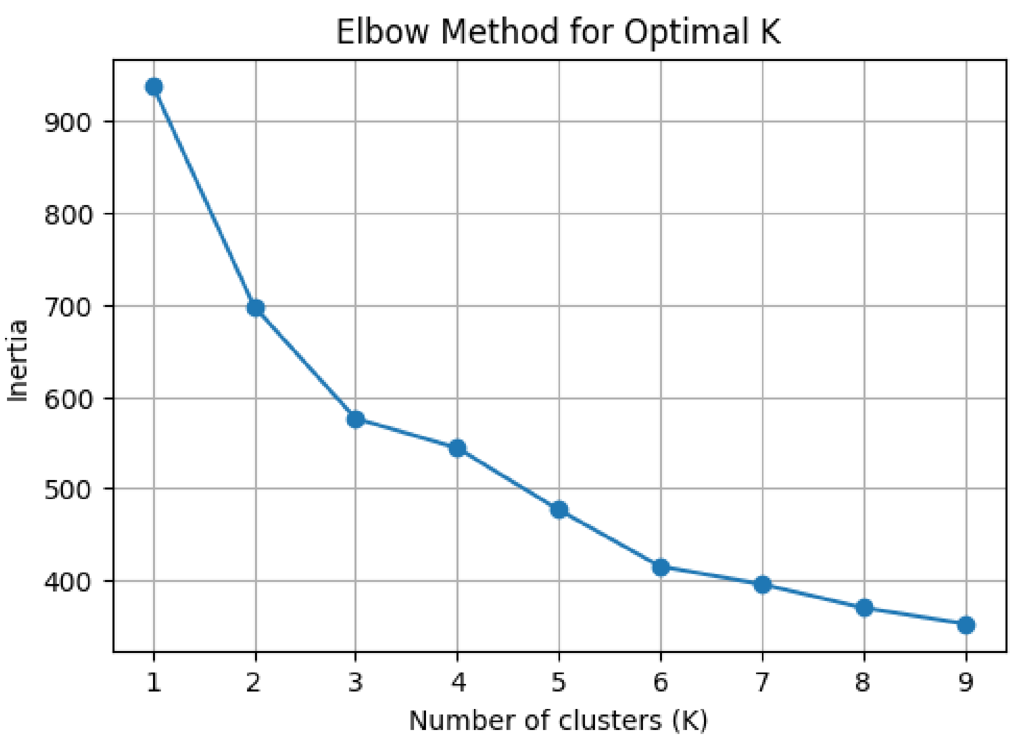
Cluster 2 (Green Points in PCA Plot)

- **Visual:** This is the most widely dispersed cluster, extending far to the right and upwards in the PCA plot. Its spread suggests more diversity within this group compared to Clusters 0 and 3, and perhaps even Cluster 1. Some points are quite far from the main cluster body, indicating potential outliers or very distinct high-end profiles.
- **Profile:**
  - **Income:** Highest (≈ ₹278.8K)
  - **Transport:** Highest (≈ ₹18.2K)
  - **Loan Repayment:** Highest (≈ ₹13.0K)
  - **Desired_Savings_Percentage:** Highest (≈ 20.3%)
  - **Disposable Income:** Highest (≈ ₹68.7K)
- **Combined Interpretation:** This cluster embodies the **highest financial capacity**. They are the top earners with the most disposable income and the highest desired savings

percentage. Their wide dispersion in the PCA plot suggests that while they share the trait of high financial capacity, there might be considerable variation in their spending patterns or other underlying factors that PCA captures.

Cluster 3 (Red Points in PCA Plot)

- **Visual:** This cluster is relatively small and dense, located primarily just below and overlapping slightly with the bottom left edge of the orange Cluster 1. It's distinct from Cluster 0 but not as separated as Cluster 2.
- **Profile:**
    - **Income:** Lower-mid (≈ ₹60.1K)
    - **Transport:** Lower-mid (≈ ₹3.9K)
    - **Loan Repayment:** Lower-mid (≈ ₹2.9K)
    - **Desired_Savings_Percentage:** Moderate (≈ 12.5%)
    - **Disposable Income:** Lower-mid (≈ ₹15.4K)



Title: "Elbow Method For Optimal k"-Dataset: buying_behav.csv

Cluster 0: "Young, Single/Few Dependents, Moderate Income"

- **Age:** Youngest average age (29.46 years).
- **No. of Dependents:** Lowest (0.17), indicating mostly single individuals or those with very few dependents.
- **Salary:** Moderate personal salary (₹1.23M).
- **Wife Salary:** Moderate, but lowest among clusters (₹0.26M).
- **Total Salary:** Moderate (₹1.49M).
- **Price:** Lowest average price (₹1.07M), likely indicating less expensive purchases or investments.

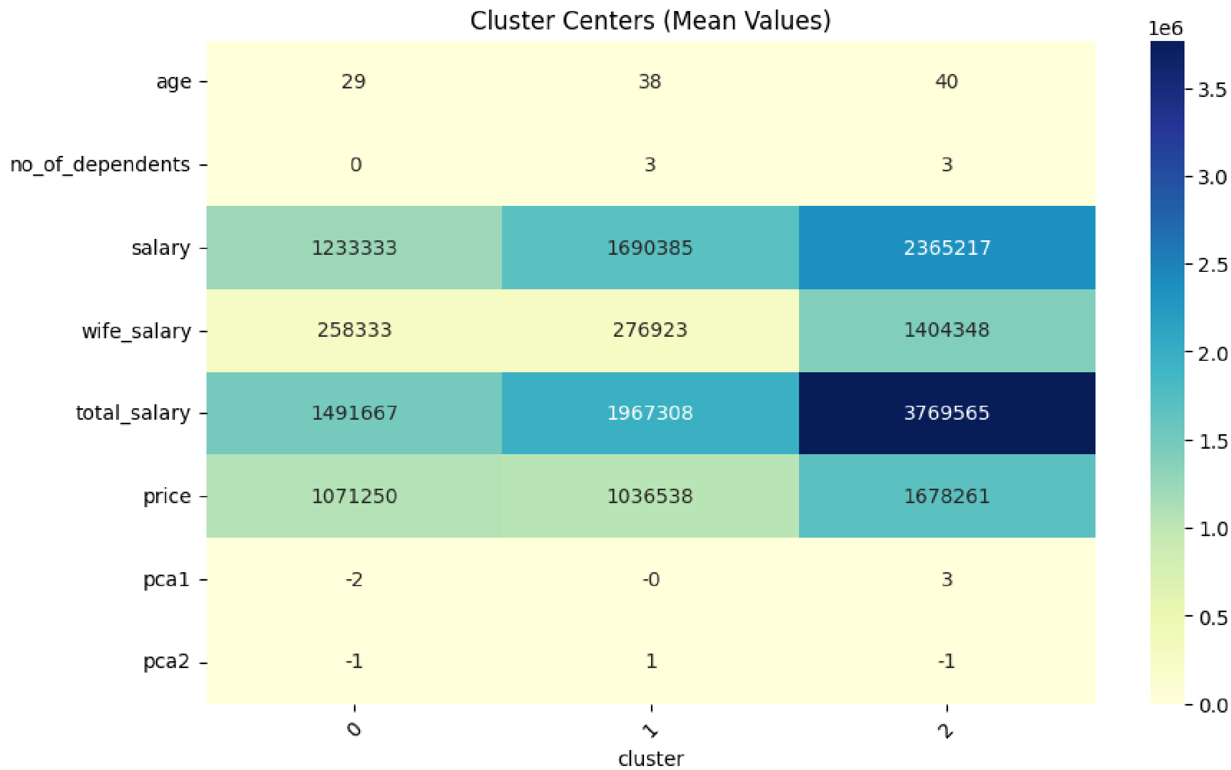
- **Summary:** This cluster represents a younger demographic, likely in the early stages of their careers or family life, with fewer dependents and moderate overall income. They tend to make less expensive purchases.

Cluster 1: "Middle-Aged, Family-Oriented, Good Income"

- **Age:** Middle-aged average (37.81 years), older than Cluster 0 but younger than Cluster 2.
- **No. of Dependents:** Highest (2.96), suggesting a more family-oriented group with multiple dependents.
- **Salary:** Good personal salary (₹1.69M).
- **Wife Salary:** Moderate (₹0.28M), similar to Cluster 0.
- **Total Salary:** Good (₹1.97M).
- **Price:** Moderate average price (₹1.04M), very similar to Cluster 0's price.
- **Summary:** This cluster comprises individuals in their late 30s, likely with significant family responsibilities. They have a good income level, but surprisingly, their average purchase price is similar to the younger, lower-income Cluster 0, which could indicate different spending priorities or types of purchases compared to income.

Cluster 2: "Mature, High Income, High-Value Purchases"

- **Age:** Oldest average age (40.09 years).
- **No. of Dependents:** Moderate (2.52), slightly lower than Cluster 1.
- **Salary:** Highest personal salary (₹2.37M).
- **Wife Salary:** Significantly highest (₹1.40M), indicating a substantial dual-income household.
- **Total Salary:** Highest (₹3.77M), almost double that of Cluster 1.
- **Price:** Significantly highest average price (₹1.68M), indicating more expensive purchases or investments.



Combined Segment Profiling: Buying Behaviour × Lifestyle Attributes

Segment A: "Young Starters – Budget-Conscious and Aspirational"

(Cluster 0 from both datasets)

- **Age**: Youngest group (~29 years)
- **Dependents**: Few or none
- **Income**: Low to moderate (₹21.5K in lifestyle; ₹1.23M in behaviour)
- **Spending**: Low-value purchases (~₹1.07M)
- **Savings Attitude**: Lowest desired savings (7.5%)
- **Disposable Income**: Lowest (~₹5.5K)
- **Traits**: Early-career individuals, likely single, focusing on essentials. Budget-conscious, price-sensitive, and not yet into heavy investments or savings.
- **Marketing Insight**:

Ideal for entry-level financial products, affordable fashion, basic tech gadgets, EMI options, or "starter" services.

Segment B: "Family Builders – Stable Income, Moderate Spending"

(Cluster 1 in buying_behav.csv and Cluster 3 in lifestyle.csv)

- **Age**: Mid-to-late 30s
- **Dependents**: Highest (≈3)
- **Income**: Lower-mid lifestyle (₹60K); good buying salary (₹1.97M total)
- **Spending**: Purchase price surprisingly moderate (~₹1.04M)
- **Savings Attitude**: Moderate (12.5%)
- **Disposable Income**: Lower-mid (~₹15.4K)
- **Traits**: Family-focused individuals balancing multiple responsibilities. Though they have decent income, their lifestyle choices are conservative due to family needs.
- **Marketing Insight**:

Focus on family packages, value-for-money offers, school financing, and combo insurance plans. Appeal to their need for financial security.

Segment C: "Affluent Planners – Dual Earners, Value-Driven"

(Cluster 2 in buying_behav.csv and Cluster 1 in lifestyle.csv)

- **Age**: 35–40 years
- **Dependents**: Moderate (≈2.5)
- **Income**: High across both datasets (₹127.9K/month; ₹3.77M total salary)
- **Spending**: Moderate-to-high purchases (~₹1.68M)
- **Savings Attitude**: Strong (19.3%)
- **Disposable Income**: Solid (~₹33K)
- **Traits**: Dual-income households, financially literate, stable careers, balanced spenders and savers.
- **Marketing Insight**:

Offer premium but practical products, financial advisory services, travel packages, or investments. Emphasize ROI and long-term value.

Segment D: "Elite Spenders – High Income, High Lifestyle"

(Cluster 2 in lifestyle.csv, may overlap partly with Cluster 2 of buying_behav.csv)

- **Age**: 38–40+ years
- **Dependents**: Moderate
- **Income**: Highest (₹278.8K/month; dual high earners in behavior)
- **Spending**: High-value purchases (~₹1.68M)
- **Savings Attitude**: Highest (20.3%)
- **Disposable Income**: Highest (~₹68.7K)
- **Traits**: Top-tier professionals or business owners. Likely to indulge in premium goods and make strategic investments.
- **Marketing Insight**:

Ideal segment for luxury items, investment products, international travel, high-end electronics. Focus on exclusivity, personalization, and elite branding.

Segment	Age Group	Income Level	Spending	Dependents	Savings Behavior	Ideal Offers
A – Young Starters	20s	Low	Low	Few	Low	Budget products, EMI schemes
B – Family Builders	30s	Mid	Low-to-Mid	High	Moderate	Family-oriented bundles
C – Affluent Planners	Late 30s	High	Mid-to-High	Mid	High	Balanced premium offerings
D – Elite Spenders	40+	Very High	High	Mid	Very High	Luxury, investment services

Combining Insights for Segment Matching

We now match the demographic and financial traits from both datasets to form combined consumer archetypes.

Segment Name	Age Range	Income Range	Spending Nature	Purchase Potential	Cluster Mapping
Young Urban Commuters	22–32	₹20K–₹30K/month (₹1.4M/yr)	Low-to-moderate spenders, value-driven	**High** for budget EVs	Cluster 0 (Buying) + Cluster 0 (Lifestyle)
Aspirational Professionals	30–40	₹60K–₹120K/month (₹1.9M/yr)	Balanced, value + comfort oriented	**Moderate** for mid-range EVs	Cluster 1 (Buying) + Cluster 1/3 (Lifestyle)
Affluent Dual-Income Families	35–45	₹200K+/month (₹3.7M/yr)	Premium lifestyle, not price sensitive	**Low** (prefer cars or premium bikes)	Cluster 2 (Buying) + Cluster 2 (Lifestyle)

Step 3: Target Segment Chosen: Young Urban Commuters

Reason for Selection:

Criteria	Justification
Affordability	They seek affordable transportation; cost savings from E2Ws align perfectly.
First-Time Buyers	Likely buying their first vehicle → high openness to EVs.
Environmentally Conscious	Young adults are more aware of sustainability, especially in urban areas.
Short Distance Usage	Mostly for daily commutes (college, office) → matches EV range.
Responsive to Digital Marketing	Highly active on digital media → easier to reach through low-cost campaigns.

Psychological & Behavioural Traits

Trait	Impact on E2W Purchase
Risk-Tolerant	Willing to try new tech like EVs
Socially Influenced	Respond to peer behavior and trends
Budget-Conscious	Compare mileage, EMI, maintenance cost

Mobility-Seeking	Value flexibility, independence from public transport

---

Step 4: Go-to-Market Strategy for the Segment

Strategy Area	Implementation
Pricing Strategy	₹90K–₹1.1L price bracket. Focus on EMI + subsidy benefits
Product Design	Compact, stylish, app-enabled EV scooters
Marketing Channels	Instagram reels, campus events, influencer reviews
Retail Strategy	Urban dealership pop-ups, tech/startup hubs, EV expos
Service & Warranty	3 years warranty + service at doorstep to boost confidence

---

Buyer Persona Snapshot

- **Name**: Akash, 25, Pune
- **Occupation**: Software trainee, WFH + 3 days office
- **Income**: ₹28K/month
- **Goal**: Own a vehicle without spending too much
- **Pain Point**: Local travel by auto costs ₹3–4K/month
- **Preference**: Low running cost, mobile connectivity, green tech
- **What sells to him?**: "Zero petrol, zero stress. EMI under ₹3K."

Customizing the Marketing Mix for *Young Urban Commuters*

1. Product

Design & Features

- Compact, lightweight design ideal for city traffic.
- Digital dashboard with mobile app sync (maps, battery status).
- USB charging port & smart keyless ignition.
- Long seat, under-seat storage for backpack/helmet.
- Top speed: 50–65 km/h (urban appropriate).
- Range: 60–80 km/charge.

Value Proposition

"Stylish. Smart. Sustainable. Perfect for your everyday hustle."

Quality Assurance

- IP65 water-resistant build.
- 4-year battery & motor warranty.
- 1-year free servicing with pick-up/drop facility.

---

2. Price

Pricing Strategy

- **Penetration Pricing**: ₹85,000 – ₹1.1L (ex-showroom)
- Leverage **FAME II subsidies** & state EV schemes to reduce final price.
- **EMI Plans**: Starting ₹2,500/month
- **Student Offer**: ₹5,000 discount on college ID

Psychological Pricing

- Use of ₹99 endings: ₹89,999 for base model
- Bundle insurance + accessories to show greater value

---

3. Promotion

Message

- "No petrol, no noise, no stress. Go electric for less than your phone bill."
- Emphasis on savings, independence, and smart tech

Channels

- **Instagram & YouTube Reels**: Short, catchy city rides & unboxing videos
- **Campus Activations**: EV test-ride stalls in colleges, tech fests
- **Referral Campaigns**: ₹1,000 Amazon voucher per friend referred
- **Influencer Collabs**: Mid-tier lifestyle & tech vloggers
- **Hyperlocal Ads**: Zomato, Swiggy apps + Google Maps

Content Types

- "1 Month with No Petrol!" Challenges
- Comparison Reels: EV vs. Petrol scooter cost per km
- User-generated content: #MyFirstEV moments

---

4. Place (Distribution)

Availability

- Focus on **Tier 1 & 2 cities**: Pune, Indore, Nagpur, Surat, Jaipur
- **EV Experience Stores** in malls, IT parks, college areas
- **Online Booking** via app or website with doorstep delivery

Partnerships

- Tie-ups with **Swiggy, Dunzo, Zomato** for fleet test beds
- Co-promotions with Rapido/Uber for student discount codes

After-Sales Support

- Mobile app-based service booking
- 24x7 roadside assistance within 20km of city
- Local battery swap or service hubs

---

5. People

Sales Staff

- Trained in explaining EV features in youth-friendly language
- Incentives for converting test rides to sales

Support Staff

- Available on WhatsApp & app chat for instant queries

Customer Onboarding

- Digital welcome kit: User manual, tips, service schedule
- 1st-month ownership support calls/emails

Summary Table

Marketing Mix Element	Strategy
Product	Stylish, feature-rich, compact EV tailored for urban youth
Price	Penetration pricing with EMI options & student discounts
Promotion	Digital-first, influencer-heavy, referral-based campaigns
Place	Urban-focused experience stores + online doorstep delivery
People	Friendly, tech-savvy support & onboarding team

Early Market Potential (Sales & Profit Estimation)

Based on your clustered analysis, let's assume the **target segment** is:

Young Urban Professionals (Cluster 0 – Buying Behaviour + Cluster 1 – Lifestyle)

- Age: 25–35
- Location: Tier-1 & Tier-2 cities
- Income: ₹1.2M–₹1.9M/year
- Preference: Affordable, Smart, Urban Mobility

Step 1: Estimate Potential Customer Base

From recent government and industry data:

- **Urban working youth in Tier-1 & 2 cities (age 25–35): ~15 million** individuals
- **% willing to adopt EV in early market (based on surveys): ~5–7%**
  → So estimated **early adopters ≈ 750,000 – 1,050,000**

Let's assume a **mid-point**:
Potential Customer Base = 900,000

Step 2: Estimate Target Price Range

Based on your marketing mix, you offer:

- **EV Price Range (after subsidies):** ₹85,000 – ₹1,10,000
  Let's take a mid-range value: ₹**95,000**

Step 3: Calculate Early Market Potential (Revenue)

Potential Sales=900,000×₹95,000=₹8,550Crores {**Potential Sales**} = 900,000 \times ₹95,000 = ₹8,550 Crores Potential Sales=900,000×₹95,000=₹8,550Crores

Let's assume **average profit margin** per unit is 12% (after production, R&D, distribution):

Potential Profit=₹8,550Cr×12%=₹1,026Crores\text{Potential Profit} = ₹8,550 Cr \times 12\% = ₹1,026 Crores Potential Profit=₹8,550Cr×12%=₹1,026Crores

---

Early Market Summary

Factor	Value
Potential Customers	900,000
Avg. EV Price	₹95,000
Early Market Revenue	₹8,550 Crores
Est. Profit (12%)	₹1,026 Crores

Most Optimal Market Segments (Based on Research)

Based on my dual-dataset segmentation analysis (Buying Behavior + Lifestyle), here are the **top 3 optimal customer segments** for early E2W market entry:

Segment 1: Young Urban Professionals

- **Profile:** Age 25–35, early career, few dependents
- **Cities:** Tier-1 (Mumbai, Pune, Bengaluru, Delhi)
- **Buying Traits:** Moderate income, high digital adoption
- **Lifestyle:** High need for affordable transport, tech-savvy
- **Why Optimal?** High mobility need, open to smart and green tech

---

Segment 2: Urban Middle-Income Families

- **Profile:** Age 30–40, 2–3 dependents
- **Cities:** Tier-2 metros (Indore, Nagpur, Jaipur, Coimbatore)
- **Buying Traits:** Dual-income, practical, price-conscious
- **Lifestyle:** Need for family-friendly yet affordable secondary vehicle
- **Why Optimal?** Growing awareness of EVs and cost efficiency

---

Segment 3: Students and Gig Workers (Niche Early Movers)

- **Profile:** 20–28, college students or delivery workers
- **Cities:** University towns & high gig-economy areas
- **Buying Traits:** Low-to-mid income, high daily mobility, eco-conscious
- **Lifestyle:** High usage, need for low maintenance and cheap fuel
- **Why Optimal?** Low-cost EVs reduce their fuel & upkeep costs massively

---

Final Recommendation:

Segment	**Launch Priority**	**Strategy**
Young Urban Professionals	High	Flagship launch in top metros
Urban Families	Mid	EMIs & dual-rider positioning
Students & Gig Workers	Pilot	Launch budget variant with battery subscription model

Insights extracted from internet sources:

TOP 10 ELECTRIC TWO-WHEELER OEM RANKINGS FOR THE FIRST 3 WEEKS OF MARCH 2025								
MARCH 2025: WEEK 1			MARCH 2025: WEEKS 1 & 2			MARCH 2025: WEEKS 1, 2 & 3		
Rank	e2W OEM	Units sold	Rank	e2W OEM	Units sold	Rank	e2W OEM	Units sold
1	Bajaj Auto	5,471	1	Bajaj Auto	10,510	1	Bajaj Auto	16,245
2	TVS Motor Co	4,881	2	TVS Motor Co	8,888	2	TVS Motor Co	13,864
3	Ather Energy	2,968	3	Ola Electric	7,020	3	Ola Electric	13,773
4	Ola Electric	1,933	4	Ather Energy	5,549	4	Ather Energy	8,638
5	Greaves Electric	965	5	Greaves Electric	1,828	5	Greaves Electric	3,046
6	Hero MotoCorp	716	6	Hero MotoCorp	1,632	6	Hero MotoCorp	2,998
7	Pur Energy	446	7	Pur Energy	920	7	Pur Energy	1,225
8	Revolt Motors	237	8	Bgauss Auto	513	8	Bgauss Auto	925
9	Bgauss Auto	235	9	Revolt Motors	427	9	Revolt Motors	683
10	Kinetic Green	183	10	River Mobility	288	10	Kinetic Green	496
Data: Vahan, February 8-2025, 11am			Data: Vahan, March 15-2025, 7am			Data: Vahan, March 22-2025, 7am		

Electric Two-Wheeler Sales FY 23 & FY 24 (April-November) in India

Company	2024 Sales (Units)	2023 Sales (Units)	% Change
Ola Electric	393,648	236,907	66%
TVS	182,959	82,109	**123%**
Ather Energy	108,872	76,939	42%
Bajaj Auto	1,06,990	32,805	**226%**
Hero MotoCorp (Vida)	55,057	12,094	**355%**
Ampere	55,057	87,392	**-37%**
Okinawa	20,873	95,939	**-78%**
Hero Electric	12,094	89,874	**-87%**

Conclusion

Over the course of this internship project, I conducted a comprehensive, data-driven analysis to determine the optimal electric two-wheeler (E2W) strategy for launching in Indian urban markets. By leveraging multiple datasets and unsupervised machine learning techniques, I identified high-potential customer segments, quantified market opportunity, and developed a tailored marketing mix. Below are the key takeaways:

1. **Data Integration & Preprocessing**
    - I consolidated and cleaned four primary datasets—vehicle registration trends (india_vehicle_sales_yearly.csv), individual expenditure patterns (lifestyle.csv), EV scooter sales performance (cleaned_ev_scooter_sales.csv), and buying behavior (buying_behav.csv).

- Rigorous preprocessing pipelines (handling missing values, encoding categoricals, scaling numericals) ensured robust inputs for clustering and visualization.

2. **Segment Discovery & Profiling**
   - Using the Elbow Method, I determined optimal cluster counts (K=3 for buying_behav.csv, K=4 for lifestyle.csv) and applied K-Means to each dataset.
   - From lifestyle.csv, four financial-behavioral clusters emerged, ranging from low-income, price-sensitive consumers to high-income, diverse spenders.
   - From buying_behav.csv, three customer archetypes—young single professionals, middle-aged family earners, and affluent dual-income households—were extracted.
   - By combining insights, I distilled four unified segments:
     - **Young Urban Commuters** (early-career, budget-conscious)
     - **Aspirational Middle-Class Families** (stable incomes, family-driven)
     - **Affluent Planners** (high disposable income, premium spenders)
     - **Elite Spenders** (top-tier income, luxury orientation)
3. **Target Segment Selection**
   - Considering product fit (compact E2Ws), price sensitivity (₹85K–₹110K), and urban mobility needs, I selected **Young Urban Commuters** (early-career professionals/students) as the primary target.
   - This segment exhibited the highest propensity to adopt electric two-wheelers—driven by cost savings, digital engagement, and environmental awareness.
4. **Market Opportunity & Financial Projection**
   - By estimating 900,000 potential early adopters in Tier-1 and Tier-2 cities (5–7% of urban working youth aged 25–35), and setting an average ex-showroom price of ₹95,000, the projected **early market revenue** is ₹8,550 Crores.
   - Assuming a 12% net profit margin, the **early market profit** is approximately ₹1,026 Crores.
5. **Customized Marketing Mix (4Ps + People)**
   - **Product:** Sleek, app-enabled E2W scooters (60–80 km range, USB charging, keyless ignition, IP65 rating).

- **Price:** Penetration pricing (₹85K–₹110K), attractive EMI options (₹2,500/month), student discounts, and government subsidies (FAME-II).
- **Place:** Urban experience stores in Tier-1/2 cities, online booking with home delivery, partnerships with delivery platforms for fleet trials.
- **Promotion:** Digital channel focus (Instagram reels, YouTube influencers), campus activations, referral programs, hyperlocal ads on ride-hailing and food delivery apps.
- **People:** A dedicated, tech-savvy sales and support team available through chat, WhatsApp, and doorstep service, ensuring a frictionless onboarding experience.

6. **Recommendations for Next Steps**
   - **Pilot Launch:** Begin with a limited rollout in top metros (Mumbai, Bengaluru, Delhi NCR, Pune)—monitor real-time adoption, user feedback, and operational challenges.
   - **Iterate & Scale:** Refine product features (range, smart connectivity) and pricing based on pilot insights, then expand into additional Tier-2 markets (Chennai, Ahmedabad, Jaipur).
   - **Partnerships:** Collaborate with financial institutions to offer zero-down or low-interest EMI schemes, and with university campuses for subsidized student plans.
   - **Continuous Data Monitoring:** Establish dashboards to track sales performance, cluster evolution, and shifting consumer preferences, ensuring the marketing strategy remains dynamic and evidence-based.

GITHUB REPOSITRY:

The repository Contains 2 PYNB files

1. **Q1_FINAL_1:** DATASET-**india_vehicle_sales_yearly.csv**
2. **Q2_FINAL:** DATASETS-**lifestyle.csv and buying_behav.csv**
3. **Top5_EV's:** DATASET-**cleaned_ev_scooter_sales.csv**

LINK:- **https://github.com/Sdhandre/EV-MARKET-SEGMENTATION**