Strategic Market Analysis for EV Startup Manufacturing

By: Mahipal Kumar Singh
Date: 03/04/2025

Abstract

The electric vehicle (EV) market is growing fast, and new companies need smart strategies to succeed. This research uses sales data from 2016 to 2024 to segment the EV market into different vehicle categories. We analysed the data with tools like PCA and K-Means clustering to find patterns and group similar categories. Our findings show three main segments: high-growth categories like Two-Wheeler (NT) and Three-Wheeler (T), a medium-growth segment like Light Goods Vehicles (LGV), and a low-priority segment with smaller sales. For an EV startup, we recommend targeting Two-Wheeler (NT) and Three-Wheeler (T) because they have big sales, steady growth, and huge market potential. This paper gives clear insights to help startups decide which categories to focus on for manufacturing.

1. Introduction

Electric vehicles are the future of transportation. With climate change and rising fuel costs, more people and businesses are switching to EVs. For a new EV company, picking the right vehicle category to manufacture is a big decision. Should they make two-wheelers, three-wheelers, or maybe heavy goods vehicles? The wrong choice could waste time and money, while the right one could lead to success.

This research looks at EV sales data from 2016 to 2024 to figure out which categories are worth targeting. We used data science techniques like clustering to group vehicle categories based on sales, growth, and stability. Our goal is to help an EV startup make a smart choice by showing which categories are growing fast, which are steady, and which might not be worth the effort. The data comes from a file called "EV Categories Sales.csv," and we analysed it step-by-step to find clear answers.

In this paper, we'll explain the problem, walk through our analysis, show how we segmented the market, and suggest the best categories for a startup to target. Let's get started!

2. Problem Statement

Starting an EV company is exciting but tricky. The market has many vehicle types—two-wheelers, three-wheelers, light goods vehicles, heavy passenger vehicles, and more. Each category has different sales numbers, growth rates, and risks. For example, some might sell a lot but not grow much, while others might be small now but growing fast. An EV startup needs to know: Which category should we focus on to build our first product? Without clear data, they might pick a category that's too small or too risky.

The problem is that there's no simple way to decide. Sales data alone isn't enough—you need to look at trends, stability, and potential. This research solves that by analysing EV sales data from 2016 to 2024. We want to answer: Which vehicle categories have the best mix of high sales, strong growth, and steady demand? Our findings will guide a startup to pick the right target and avoid wasting resources.

3. Data Overview & Methodology

3.1 Data Collection

The dataset used in this study was collected from the Indian government's official transportation website, Parivahan. It contains sales data across different categories of vehicles from 2016 to 2024. The dataset provides insights into the market trends and adoption rates of EVs in India by tracking the annual sales of various vehicle types.

Here is overview of first 6 rows of Dataset.

	Vehicle Category	2024	2023	2022	2021	2020	2019	2018	2017	2016
0	FOUR WHEELER (INVALID CARRIAGE)	37	47	13	3	0	0	0	0	4
1	HEAVY GOODS VEHICLE	194	327	0	0	0	0	0	1	8
2	HEAVY MOTOR VEHICLE	0	3	2	6	0	0	0	1	8
3	HEAVY PASSENGER VEHICLE	3396	2397	1962	921	53	371	36	9	3
4	LIGHT GOODS VEHICLE	5737	2283	328	987	13	59	759	907	93
5	LIGHT MOTOR VEHICLE	67407	73298	33267	12313	3280	1034	1158	925	717

Fig 1: First 6 row of data

Data Source: https://vahan.parivahan.gov.in/vahan4dashboard/vahan/dashboardview.xhtml

3.2 Data Health Check

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15 entries, 0 to 14
Data columns (total 10 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Vehicle Category  15 non-null     object
 1   2024              15 non-null     int64
 2   2023              15 non-null     int64
 3   2022              15 non-null     int64
 4   2021              15 non-null     int64
 5   2020              15 non-null     int64
 6   2019              15 non-null     int64
 7   2018              15 non-null     int64
 8   2017              15 non-null     int64
 9   2016              15 non-null     int64
dtypes: int64(9), object(1)
memory usage: 1.3+ KB
```

Fig 2: Data information

The data has 15 rows (one for each category) and 10 columns (one for the category name and one for each year). For example, "TWO-WHEELER(NT)" sold 727,021 units in 2024, while "HEAVY MOTOR VEHICLE" sold 0. This shows some categories are huge, and others are tiny.

Next, we checked the data's health. There were no missing values, which is great—it means we can trust the numbers. We also noticed big differences in sales. "THREE-WHEELER(T)" sold 645,001 units in 2024, but "MEDIUM MOTOR VEHICLE" sold just 20. This range told us we'd need to scale the data later for fair comparisons.

3.3 Exploratory Data Analysis (EDA)
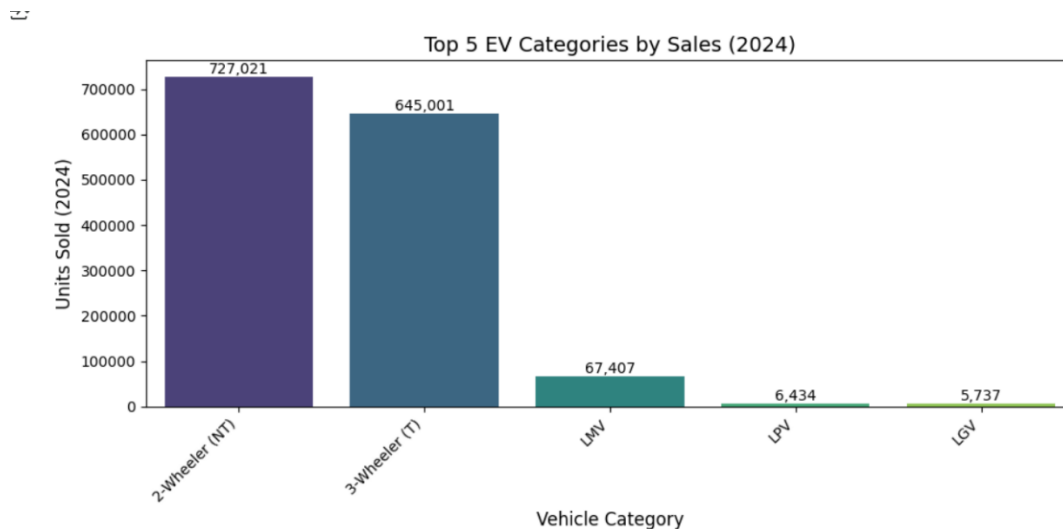
3.3.1 Top EV Categories by Sales (2024)

Fig 3: Top 5 EV Sales Categories in 2024

A bar plot visualizes the top five EV categories based on sales data for 2024. The analysis reveals that two-wheeler (NT) EVs have the highest sales, followed by three-wheeler (T) EVs, highlighting their dominance in the market. Other categories, such as Light Motor Vehicles (LMV), Light Passenger Vehicles (LPV), and Light Goods Vehicles (LGV), show significantly lower sales figures.

This indicates that two- and three-wheelers are the leading choices for EV adoption in India, likely due to affordability, urban mobility needs, and government incentives. The results suggest that startups should focus on these segments to maximize market reach and business growth.

3.3.2 EV Sales Trends (2016-2024)

The line plot illustrates the sales trends of the top five EV categories from 2016 to 2024. The analysis shows that two-wheeler (NT) EVs experienced the most significant growth, particularly from 2021 onwards, peaking in 2023 before a slight decline in 2024. Three-wheeler (T) EVs also showed a steady rise, maintaining their strong presence in the market.
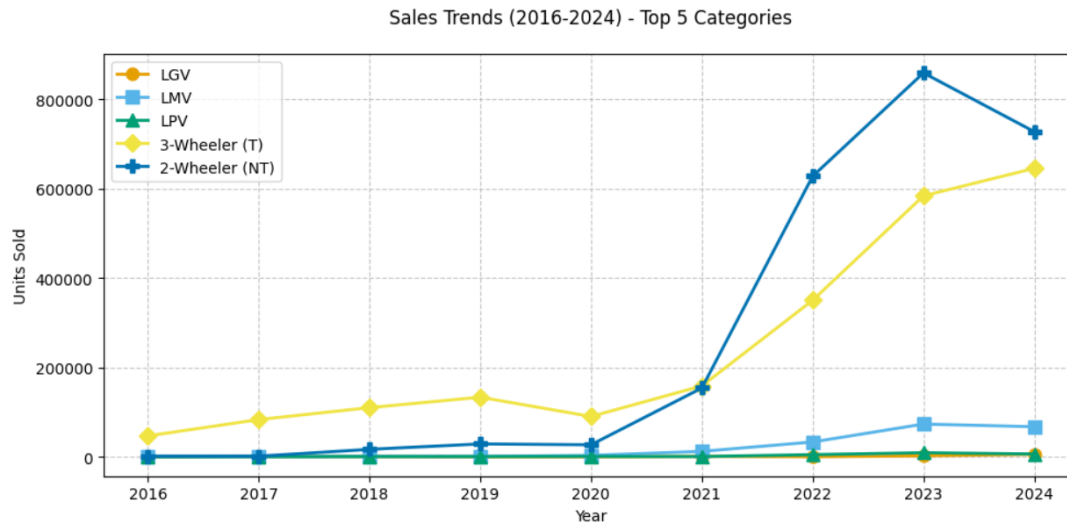
Fig 4: Top 5 EV Sales Categories (2016 -2024)

In contrast, Light Motor Vehicles (LMV), Low-Powered Vehicles (LPV), and Light Goods Vehicles (LGV) had relatively lower sales figures, with minimal growth over the years. The data suggests that the EV market in India is primarily driven by two- and three-wheelers, likely due to affordability, urban transportation demands, and government incentives.

3.3.3 Key Findings: EDA

3.3.3.1 Market Dominance: 2-Wheelers & 3-Wheelers

- 2-Wheeler (NT) leads with 727,021 units sold in 2024, followed by 3-Wheeler (T) at 645,001 units.

- Why This Matters for Startups:

  - Lower manufacturing complexity than 4-wheelers.

  - High demand in urban transport and last-mile delivery.

  - Government subsidies favor small EVs.

3.3.3.2 Steady Growth in Light Motor Vehicles (LMVs)

- LMVs rank third with 67,407 units in 2024, showing stable year-over-year growth.

- Strategic Insight:

  - Balanced demand across personal and commercial use.

  - Moderate competition, leaving room for innovation.

3.3.3.3 Niche Opportunities in Commercial EVs

- Light Goods Vehicles (LGV): 5,737 units in 2024.

- Heavy Passenger Vehicles (HPV): 3,396 units in 2024.

- Why Consider These Segments?

  - Less crowded than 2/3-wheeler markets.

- o Higher price points improve profit margins.

- o Rising demand in logistics and public transport.

Conclusion: The EV market is expanding rapidly, with 2- and 3-wheelers dominating adoption, while LMVs and commercial EVs present emerging opportunities.

---

4. Feature Engineering

To extract deeper insights from the EV sales data, we performed feature engineering to calculate key growth and stability metrics. Specifically, we computed:

1. **Compound Annual Growth Rates (CAGR)** for different timeframes:

   - o **1-Year Growth (2023–2024)** → Identifies short-term trends.

   - o **3-Year Growth (2021–2024)** → Measures mid-term momentum.

   - o **5-Year Growth (2019–2024)** → Reveals long-term adoption patterns.

   - o **Overall CAGR (2016–2024)** → Provides full-period performance.

2. **Sales Stability (Standard Deviation)** → Quantifies demand volatility for each category.
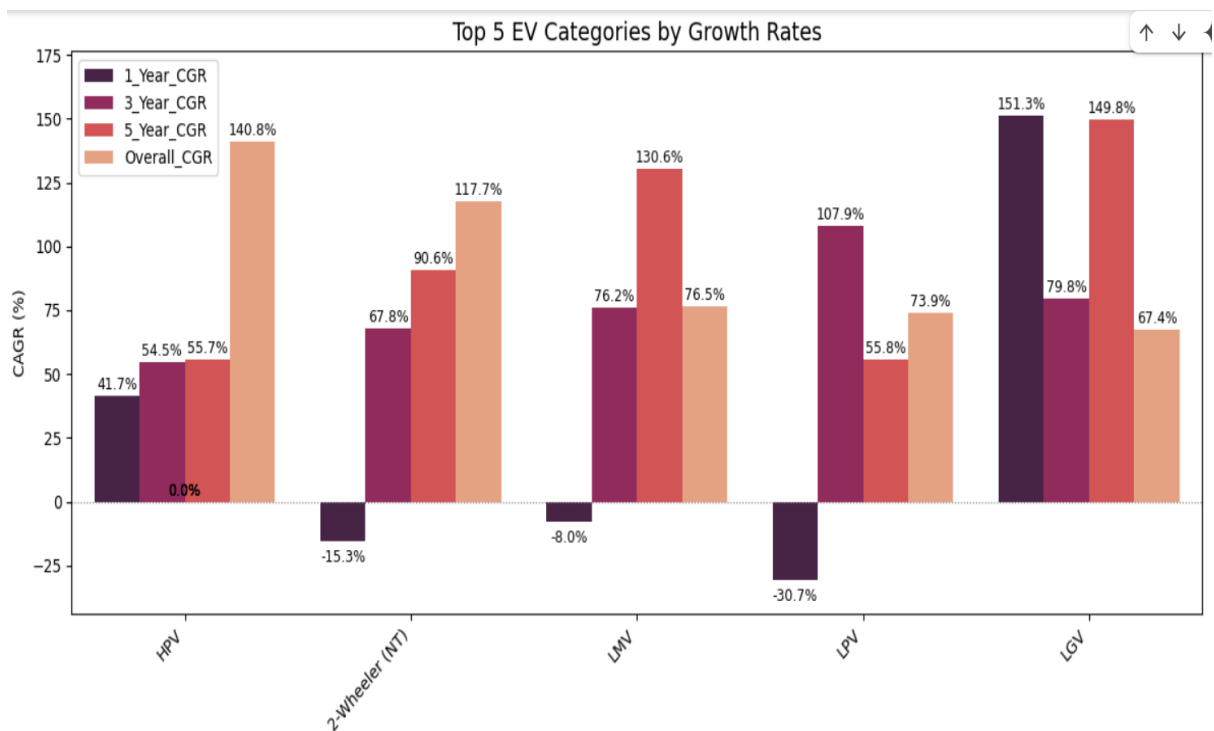
4.1 Top 5 EV Categories by Growth Rates

Fig 5: Top 5 EV Sales Categories Growth Rate (CGR)

The graph above highlights the top 5 EV segments with the highest Overall CGR:

Key Insights:

1. HPV (Heavy Passenger Vehicle) & 2-Wheeler (Non-Transport) show continuous growth across all periods. This indicates strong market expansion and increasing adoption.

2. LGV (Light Goods Vehicle) and LPV (Light Passenger Vehicle) have high recent CGR, indicating a rising demand in commercial and passenger EVs.

3. LMV (Light Motor Vehicle) saw a recent decline but maintains a strong long-term growth trend, suggesting a recovering market.

Strategic Takeaways: Startups should focus on fast-growing segments like HPV, 2-Wheeler (NT), and LGV for stable revenue.

---

5. Feature Selection

For clustering EV categories, we selected the following key features:

1. **Recent 2024 Sales** – Represents the latest market demand, ensuring that the clustering model captures current market trends.

2. **5-Year CGR (Compound Growth Rate)** – Provides a long-term perspective on how each EV category has grown, helping to identify sustained market expansion.

3. **Sales Stability (Sales Standard Deviation)** – Measures the consistency of sales over time. This helps differentiate between stable and volatile markets, which is crucial for understanding market reliability.

Why These Features?

- Capturing Market Trends: Recent 2024 Sales reflects the latest performance, ensuring that the clustering results are relevant for current decision-making.

- Identifying Growth Potential: 5-Year CGR highlights categories with consistent growth, helping to separate emerging vs. stagnant markets.

- Assessing Risk & Stability: Sales Stability ensures that we consider both high-growth and low-risk segments, crucial for strategic planning.

By using these features, clustering will help group EV categories based on demand patterns, long-term potential, and sales volatility, leading to better market segmentation and investment strategies.

---

6. Segment Extraction and Evaluation

To segment the EV market, we used K-Means Clustering with three key features: **2024 Sales, 5-Year CAGR (Compound Annual Growth Rate), and Sales Stability**. These features

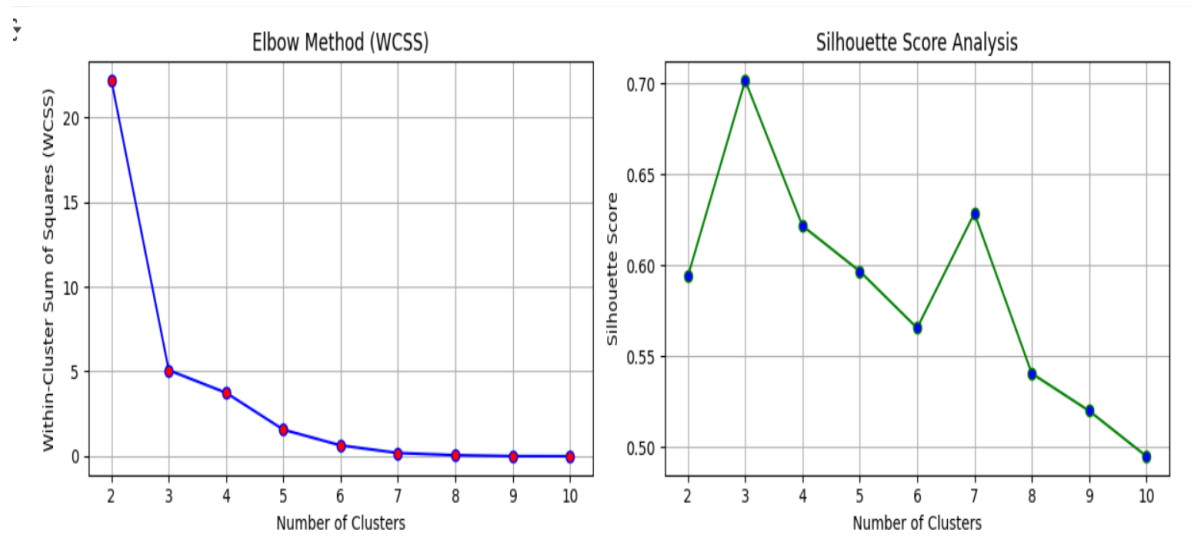ensure a balanced approach by considering recent sales performance, long-term growth trends, and market volatility.

Fig 6: WCSS Graph vs Silhouette Score

To determine the optimal number of clusters, we applied two methods: **WCSS (Within-Cluster Sum of Squares)** and **silhouette scores**. WCSS measures how compact the clusters are, while the silhouette score evaluates how well-separated, they are. Using the default parameters of K-Means, as you can see the **elbow plot showed a noticeable bend at 3 clusters**, suggesting an optimal grouping. The **silhouette score is also highest for 3 clusters (around 0.70)**, indicating well-defined clusters. While this result looks promising, we will conduct further evaluation to confirm its effectiveness.

6.1 Evaluating using K-Means Clustering

For further evaluation of clustering, this time we applied K-Means with 10 different random centroid initializations by setting n_init = 10. We tested for clusters ranging from 2 to 8 and calculated both the Silhouette Score and Davies-Bouldin Index for better assessment.

(Davies-Bouldin Index measures the compactness and separation of clusters, where lower values indicate better clustering.)

We observed a noticeable change between 2 and 3 clusters, as shown below:

	k	Silhouette	Davies-Bouldin	WCSS
0	2	0.713645	0.392348	15.160501
1	3	0.701443	0.277162	5.073549
2	4	0.675622	0.266211	2.897903
3	5	0.596620	0.180037	1.567707
4	6	0.603638	0.129526	0.443218
5	7	0.628510	0.160819	0.192610
6	8	0.540466	0.104656	0.062369

Fig 7: Silhouette, Davies Bouldin and WCSS Score

6.1.1 WCSS (Within-Cluster Sum of Squares) Analysis

- At k=2, WCSS is 15.16, meaning that the clusters still contain a lot of intra-cluster variance (less compact).

- At k=3, WCSS drops significantly to 5.07, indicating that clusters have become more compact and better separated.

- A steep drop in WCSS suggests that increasing the number of clusters improves intra-cluster similarity.

6.1.2 Silhouette Score Analysis

- k=2 gives the highest Silhouette Score (0.7136), but this does not necessarily mean better clustering.

- k=3 still has a high score (0.7014), indicating that clusters are well-separated.

- While the score slightly decreases at k=3, this is expected as the number of clusters increases.

- A small drop in Silhouette Score is acceptable if the WCSS improvement is significant.

6.1.3 Davies-Bouldin Index (Cluster Separation & Compactness)

- Lower values indicate better clustering.

- At k=3, the Davies-Bouldin Index is 0.277 (lower than 0.392 at k=2), meaning clusters are more distinct and well-separated.

- The drop in Davies-Bouldin Index at k=3 confirms that the additional cluster enhances cluster separability rather than adding unnecessary complexity.

6.1.4 Business & Interpretability Considerations

- With k=2, the segmentation might be too broad, merging dissimilar EV categories.

- With k=3, the segmentation allows for finer granularity, distinguishing categories more effectively.

- This improves actionable insights, helping to identify high-growth, stable, and fluctuating EV market segments.

6.1.5 Conclusion: k=3 is the Optimal Choice

- Significant WCSS reduction (from 15.16 to 5.07).

- Lower Davies-Bouldin Index (indicating better separation).

- Slightly lower but still strong Silhouette Score (0.7014 vs. 0.7136).

- More meaningful business segmentation for decision-making.

As we can see below cluster graph k=3 provides the best balance of compact, well-separated clusters with a meaningful business interpretation.
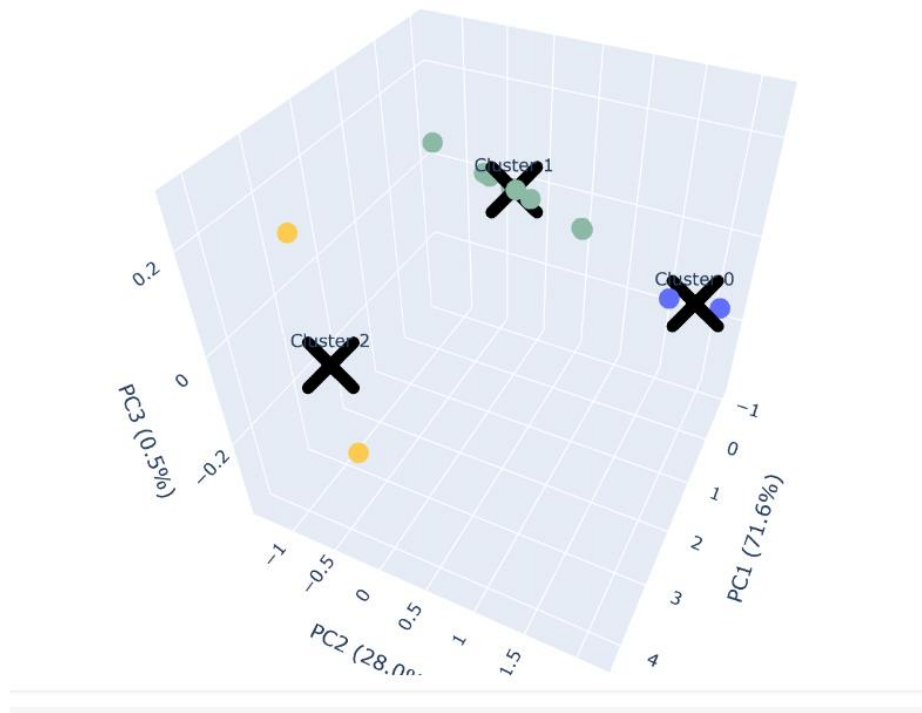
Fig 8: Top K-Means Cluster

6.2 Evaluating Using Gaussian Mixture Models (GMM)

For further evaluation of clustering, we applied Gaussian Mixture Models (GMM) instead of K-Means, using 10 different random centroid initializations by setting n_init = 10. We tested cluster sizes ranging from 2 to 8 and evaluated the clustering performance using AIC, BIC, Silhouette Score, and Davies-Bouldin Index for a more comprehensive assessment. Here is the below result:

	k	Silhouette	Davies-Bouldin	BIC	AIC
0	2	0.591793	0.796456	-132.160646	-145.613599
1	3	0.701443	0.277162	-163.747136	-184.280591
2	4	0.675622	0.266211	-178.723973	-206.337931
3	5	0.596620	0.180037	-177.686261	-212.380721
4	6	0.603638	0.129526	-175.295924	-217.070886
5	7	0.515594	0.093288	-169.610831	-218.466295
6	8	0.607962	0.127445	-173.841901	-229.777867

Fig 9: Matrix Evacuation of GM Model

The Silhouette score is highest at K=3 (0.7014), indicating good cluster separation, followed by K=4. The Davies-Bouldin Index is best at K=7 but remains reasonable for K=3. The Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) suggest K=4 as optimal, with diminishing improvements beyond that point. Comparing K-Means and Gaussian Mixture Model (GMM), K=3 or K=4 emerge as strong choices due to their balance across metrics.
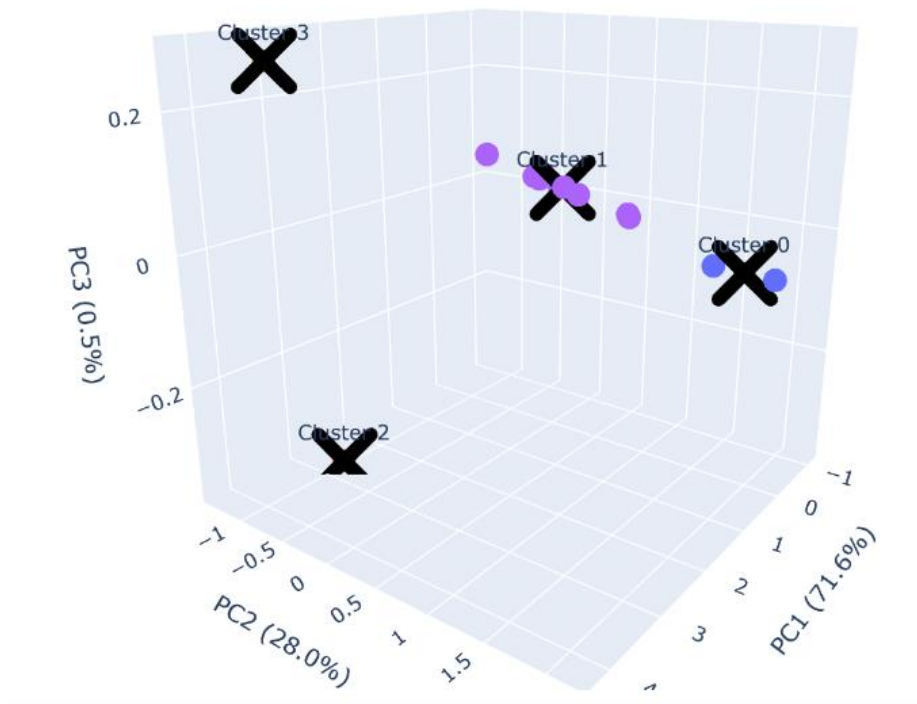
Fig 10: GM Model Cluster

6.3 Final Conclusion for Segment Evaluation

Based on the clustering analysis and vehicle category data, we should consider K=3 or K=4 as the optimal number of clusters. The Silhouette score is highest at K=3, ensuring well-separated clusters, while K=4 is supported by BIC and AIC for optimal model complexity. according to the sales trends and growth rates, clustering into 3 or 4 groups can effectively distinguish vehicle categories based on key growth metrics such as 1-year, 3-year, and 5-year CGR, overall CGR, and sales stability. K=3 may provide a simpler segmentation, differentiating major vehicle groups with distinct sales trajectories, whereas K=4 might offer a finer classification, capturing additional market variations.

For an EV startup, K=3 would be the better choice.

- K=3 provides a simpler and clearer segmentation, which is crucial for an early-stage startup to focus on key market categories.

- The Silhouette score is highest at K=3 (0.7014), indicating well-separated clusters.

- K=4 adds complexity but does not offer a significantly better separation.

- In an EV startup, the market can typically be divided into low-growth, moderate-growth, and high-growth segments, which aligns well with K=3. According to the sales data, 2-Wheeler (NT) and 3-Wheeler (T) fall under the high-growth segment, and they are grouped into the same cluster when using K=3. This confirms that K=3 effectively captures the key market trends, making it a suitable choice for strategic decision-making.

7. Profiling Cluster Segment

```
===================================================
Vehicle Categories per Cluster
===================================================

Cluster 0 (2 vehicles):
-----------------------------
['LGV', 'LMV']

Cluster 1 (11 vehicles):
-----------------------------
['4-Wheeler (IV)', 'HGV', 'HMV', 'HPV', 'LPV', 'MMV', 'MPV', 'Other', '3-Wheeler (NT)', '2-Wheeler (IV)', '2-Wheeler (T)']

Cluster 2 (2 vehicles):
-----------------------------
['3-Wheeler (T)', '2-Wheeler (NT)']
```

Fig 11: Vehicle Categories with cluster

Cluster 0: Light & Medium Passenger Vehicles

- **Includes: LGV (Light Goods Vehicle), LMV (Light Motor Vehicle)**

- Profile: This cluster represents compact, versatile vehicles commonly used for personal transport and small-scale goods movement. These vehicles are known for their moderate commercial demand and stable market presence.

Cluster 1: Diverse & Mixed Vehicle Segment

- **Includes: 4-Wheelers, Heavy & Medium Vehicles, Public Transport, and Other Vehicles**

- Profile: This is the most diverse cluster, including public transport (HPV, LPV), heavy vehicles (HGV, HMV), and specialized vehicles (MPV, MMV). These categories cater to various industrial, commercial, and passenger needs, making it a highly mixed and dynamic segment.

Cluster 2: High-Demand, Mass Market EVs

- **Includes: 3-Wheeler (T), 2-Wheeler (NT)**

- Profile: This cluster consists of popular and rapidly growing EV categories, primarily two- and three-wheelers used for transportation and delivery services. These vehicles have seen consistent adoption in urban areas, reflecting a strong market shift toward small, efficient electric mobility solutions.

---

8. Describing Segment

After profiling the cluster, we calculated stats for each cluster: average sales, average 5year CGR growth, sales stability (standard deviation). Here is below result:

```
            2024 Sales  5-Year CGR  Sales Stability
  Cluster
  0            36572.00      140.19         15739.82
  1             1107.73        2.26           651.22
  2           686011.00       63.81        292294.14
```

Fig 12: Cluster with mean value of sales, CGR, and SS

8.1 Cluster 0: Stable Performance

- This segment includes Light Motor Vehicles (LMV) and Light Goods Vehicles (LGV).

- It has moderate sales volume (36,572 units) but exhibits high sales stability (15,739.82) and the highest growth rate (140.19%) over the past five years.

- These vehicles maintain consistent demand, making them a steady market segment with reliable growth.

8.2 Cluster 1: Emerging Niche

- This segment includes a diverse set of vehicles such as Heavy Motor Vehicles (HMV), Heavy Goods Vehicles (HGV), Passenger Vehicles (HPV, MPV, MMV, LPV), and smaller 2-Wheelers & 3-Wheelers.

- It has low overall sales (1,107.73 units) and low growth (2.26%), with some categories experiencing negative or stagnant sales growth.

- This segment represents a niche market with limited but specialized demand, requiring further strategic development to boost growth.

8.3 Cluster 2: High-Growth Mass Market

- This segment consists mainly of 2-Wheelers and 3-Wheelers, including 2-Wheeler (Non-Transport) and 3-Wheeler (Transport) vehicles.

- It has the highest sales volume (686,011 units) and moderate growth (63.81%), with significant sales stability (292,294.14).

- This category is a key driver of the market, reflecting high consumer demand and strong expansion potential.

Based on the clustering analysis, we identify Cluster 0 as a stable and reliable market, Cluster 1 as an emerging niche with uncertain demand, and Cluster 2 as a high-growth segment driving overall industry expansion. These insights can help businesses and policymakers tailor their strategies to maximize market opportunities.

9. Selecting the Target Segment(s)

Selecting the right customer segments is a critical step in market segmentation. This process involves evaluating identified segments based on their attractiveness and the organization's ability to serve them effectively.

For an EV startup, we have categorized the vehicle market into three distinct clusters based on sales volume, growth rate, and sales stability. To determine the strategic importance of each segment, we introduce a Priority Score, calculated using weighted factors such as sales volume, growth rate, and stability. This scoring system helps identify which segments hold the greatest business potential, ensuring a data-driven approach to market selection.

9.1. High-Priority Segment

Cluster 2: High-Growth Mass Market

- Dominant Vehicle Category: 2-Wheeler (Non-Transport)

- Sales Volume: 686,011 units

- Growth Rate: 63.81%

- Sales Stability: 292,294.14

- Priority Score: 0.63 (Highest among all clusters)

◈ **Key Insights:**

- This segment represents the largest market share, with strong growth and demand stability.

- 2-Wheelers (non-transport) are widely used for personal mobility, making this a critical market for expansion and investment.

- **Strategic Focus:** Companies should prioritize investment in this segment due to its strong performance and potential for continued growth.

9.2. Medium-Priority Segment

Cluster 0: Stable Performance

- Dominant Vehicle Category: Light Goods Vehicles (LGV)

- Sales Volume: 36,572 units

- Growth Rate: 140.19% (Highest among all clusters)

- Sales Stability: 15,739.82

- Priority Score: 0.615

◈ **Key Insights:**

- Despite moderate sales volume, this segment demonstrates the highest growth rate, indicating strong market momentum.

- LGVs (Light Goods Vehicles) are essential for logistics and small-scale goods transportation, making them a vital part of commercial transport expansion.

- **Strategic Focus:** This segment is ideal for companies targeting commercial logistics markets with an emphasis on sustained growth and service optimization.

9.3. Low-Priority Segment

Cluster 1: Emerging Niche

- Dominant Vehicle Category: 2-Wheeler (Intermediate Variant - IV)

- Sales Volume: 1,107.73 units

- Growth Rate: 2.26% (Lowest among all clusters)

- Sales Stability: 651.22

- Priority Score: 0.11

◈ **Key Insights:**

- This segment has low sales volume and minimal growth, making it a niche market with limited business impact.

- The demand for 2-Wheelers (IV) appears stagnant, likely due to market saturation or shifting consumer preferences.

- Strategic Focus: This category requires further market analysis before any significant investment. Companies should monitor changes in demand but prioritize resources elsewhere.

9.4 Conclusion and Strategic Recommendations

9.4.1 Primary Focus on Cluster 2 (High-Growth Mass Market):

- This segment holds the largest market share, ensuring maximum returns on investment.

- Companies should expand their presence in the 2-Wheeler (NT) category through product innovation and competitive pricing strategies.

9.4.2 Targeted Investment in Cluster 0 (Stable Performance):

- The LGV segment shows strong growth potential, making it a strategic choice for commercial transport solutions.
- Companies should focus on expanding their logistics and fleet management offerings in this segment.

9.4.3 Cautious Approach to Cluster 1 (Emerging Niche):

- This segment currently lacks growth potential, so investments should be minimal unless market conditions shift.
- Further consumer research is needed to determine if demand will increase in the future.

10. Final Thought and Recommendation

While the sales data analysis has identified the most promising vehicle segments for EV manufacturing, additional critical analyses are required before moving forward with production. These analyses will ensure that the startup understands market demand, consumer preferences, geographical feasibility, and competitive landscape.

10,1. Sentiment Analysis: Understanding Customer Needs & Competition

- Sentiment analysis helps the company understand what customers like and dislike about existing EVs.

- By analysing social media, online reviews, and surveys, the startup can identify the weak points of competitors (e.g., battery issues, pricing concerns, lack of charging stations).

- This allows the company to offer better solutions, such as improved battery life, competitive pricing, or a stronger charging network.

10.2. Geographical Analysis: Choosing the Right Location

- Not every region has the same demand for EVs. Some areas have better charging infrastructure, more government incentives, or high customer interest.

- A proper geographical analysis will help the startup find the best place to start manufacturing—a location that ensures lower logistics costs and easy access to customers.

- Understanding which areas have the highest demand will help the company grow faster and capture the market efficiently.

10.3. Government Policies & Incentives

- Many states and countries offer subsidies and tax benefits for EV manufacturers.

- Studying these policies will help the startup reduce costs and make EVs more affordable for customers.

- Knowing registration, road tax, and emission rules will help in better planning.

10.4. Studying the Competition

- Understanding what existing EV brands are offering will help the startup position itself better.

- The company can identify market gaps, such as a lack of affordable EV options or weak after-sales support, and build a unique selling point (USP).

10.5. Supply Chain & Infrastructure Planning

- Finding reliable battery suppliers and charging station partners will be key to ensuring smooth operations.

- Partnering with logistics and fleet companies can help in faster delivery and better market penetration.

Before investing in factories and production, the startup should validate its market strategy with deep analysis. This will help in choosing the right product, the best location, and a competitive business model. By making data-driven decisions, the company can enter the EV market with confidence and a strong foundation for success.

Git Hub Link: https://github.com/mahipalds/EV-Startup-Market-Analysis