EV Indian Market Data Analysis

## Prepared By:

Paarth Makkar

## Project Link:

[GitHub Link](https://github.com/Xenocide-paarth/Market-Segmentation-Analysis)



Background

The Indian automotive landscape is on the cusp of a significant transformation, driven by rising fuel prices, increasing environmental concerns, and advancements in electric vehicle (EV) technology. While the potential for EV adoption in India is immense, the market is still nascent and requires careful navigation. Our client faces the crucial challenge of identifying the most promising customer segments and geographic regions to focus our initial product development and market entry efforts. A poorly targeted approach could lead to wasted resources and a missed opportunity in this burgeoning market. Therefore, a robust market segmentation analysis is essential to inform our strategic decisions.

# Key Challenges & Considerations:

1. **Diverse Customer Needs:** Indian consumers exhibit a wide range of needs and preferences regarding vehicle type, price point, range, and usage patterns. A "one-size-fits-all" approach is unlikely to succeed.
2. **Infrastructure Limitations:** The charging infrastructure for EVs is still developing in India, posing a challenge for widespread adoption, particularly in certain regions.
3. **Price Sensitivity:** Indian consumers are highly price-conscious, making the cost of EVs a critical factor in adoption. Balancing affordability with desired features is crucial.
4. **Range Anxiety:** Concerns about the limited range of EVs and the availability of charging stations remain a barrier for many potential buyers.
5. **Varying Awareness & Acceptance:** Awareness and acceptance of EVs vary significantly across different demographics and geographic locations.
6. **Competition:** The EV market is becoming increasingly competitive, with both established automakers and new entrants vying for market share.

# Aim of the Report:

This report aims to conduct a comprehensive market segmentation analysis of the Indian EV market. By analysing demographic, psychographic, geographic, and behavioural data, we will identify distinct customer segments and their specific needs. This analysis will enable us to:

* **Prioritize target customer segments:** Determine which segments are most likely to adopt EVs in the near to medium term.
* **Identify optimal geographic regions:** Pinpoint areas with high potential for EV adoption based on factors like infrastructure development, consumer affluence, and geography.
* **Inform product development:** Tailor our EV offerings to meet the specific needs and preferences of our target segments.
* **Develop targeted marketing strategies:** Craft effective communication and outreach campaigns to reach and engage our target customers.
* **Formulate a feasible market entry strategy:** Define a phased approach to entering the Indian EV market, focusing on the most promising segments and regions.

# Expected Outcome:

This segmentation analysis will provide a data-driven foundation for our startup's market entry strategy, maximizing our chances of success in the dynamic Indian EV market. We will be able to confidently decide which customer segment and geographic region to focus on initially, allowing us to allocate resources effectively and build a strong foothold in the market.

Data

The data used in this report was obtained from: [Indian EV Market Data](https://www.kaggle.com/datasets/srinrealyf/india-ev-market-data)

This dataset provides a detailed overview of the electric vehicle (EV) market in India from 2001 to 2024. It includes monthly sales data, sales data categorized by manufacturer, and vehicle class-wise sales data for different manufacturers. This rich dataset is ideal for market analysis, trend forecasting, and research on the adoption and growth of electric vehicles in India.

# EV Maker by Place

List of popular EV Makers and their location of Manufacturing Plant.

# Operational PC

Total Operational Public Charging Station for EV available in each state.

# Vehicle Class

Total vehicles (includes electric and all other fuels) registered (manufactured) by category from 2001 - Aug 2024

# ev\_cat\_01-24

Total electric vehicles manufactured from 2001 - Aug 2024 and vehicle category

# ev\_sales\_by\_makers\_and\_cat\_15-24

Total electric vehicles manufactured by makers from 2015 - Aug 2024 with the vehicle class

# Acknowledgements

This dataset was compiled and web – scrapped from [Vahaan4 Dashboard](https://vahan.parivahan.gov.in/vahan4dashboard.xhtml)

Navigating the EV Market as a Startup: Key Questions for Success

Entering the electric vehicle (EV) market as a startup presents both immense opportunity and significant challenges. Two fundamental questions must be addressed early on to pave the way for a successful market entry:

# What type of EV will the company produce?

For a startup, resources are limited, and market experience is often nascent. Attempting to cater to every segment is not only impractical but also strategically unwise. A focused approach is essential. The core challenge lies in **resource allocation**. A startup cannot afford to spread its limited funding across a broad spectrum of EV types. Furthermore, a lack of deep market understanding can lead to missteps in product development and market positioning.

Therefore, **specialization is key**. Instead of trying to be everything to everyone, the startup should identify a specific niche within the EV market. This could be defined by vehicle type (e.g., focusing solely on electric scooters, commercial vehicles, or a specific class of cars), price point (entry-level, mid-range, or luxury), or target application (e.g., urban commuting, last-mile delivery). By concentrating efforts, the startup can develop deep expertise in its chosen niche, build a strong brand identity, and optimize resource utilization.

# To whom will it be Sell?

Understanding the target customer is paramount. A successful market entry hinges on identifying and catering to the specific needs and preferences of the intended buyer. This requires developing a detailed **customer profile** encompassing:

1. **Demographics:** Age, Gender, Income Level, Education, Occupation, and family status.
2. **Geographics:** Location – States, Area – Urban; Suburban; Rural, and access to charging infrastructure.

By meticulously analysing these factors, the startup can gain valuable insights into its target customer. This understanding will inform crucial decisions related to product design, pricing strategy, marketing messages, and distribution channels. For instance, a startup targeting young urban professionals might prioritize sleek design, advanced technology features, and convenient charging solutions, while a startup focusing on cost-conscious commuters might emphasize affordability, efficiency, and practicality.

In conclusion, a focused product strategy and a deep understanding of the target customer are critical for an EV startup to thrive in a competitive market. Answering these two fundamental questions with clarity and precision will lay the foundation for a successful and sustainable business.

Data Set

While analysing, Python libraries such as Numpy, Pandas, Scikit – learn, matplotlib, seaborn, bioinfokit and SciPy are used for workflow and the results are ensured to be reproducible.

A screenshot of a computer program

AI-generated content may be incorrect.

Our Data contains information about the types of vehicles manufactured from year 2001 to 2024. Higher manufacturing indicates higher demand for that specific vehicle. This information is spread over 16 vehicle types and over 283 rows worth of data. The entries are spread over months.

Data Quality is good as there were neither any missing records nor duplicate data. But the data was not standardised. Standardisation is necessary for Principal Component Analysis (PCA).

PCA

In today's data-rich environment numerous variables that may be correlated, make it challenging to identify the underlying patterns and extract meaningful insights. Principal Component Analysis (PCA) is a powerful statistical technique that addresses this challenge by reducing the dimensionality of data while preserving its essential information.

# How PCA Works

PCA transforms the original variables into a new set of uncorrelated variables called principal components. These components are linear combinations of the original variables, and they are ordered in such a way that the first principal component captures the maximum variance in the data, the second principal component captures the second-highest variance, and so on. By focusing on the first few principal components, we can effectively reduce the dimensionality of the data while retaining most of its valuable information.

# Benefits of Using PCA

* **Dimensionality reduction:** PCA simplifies complex datasets by reducing the number of variables, making them easier to analyse and interpret.
* **Feature extraction:** PCA helps identify features that have the most variance in the data.
* **Noise reduction:** PCA can filter out noise and irrelevant information, leading to more accurate and reliable insights.

# PCA for Market Segmentation in the Indian EV Market

The Indian EV market presents a complex landscape with diverse customer segments, varying needs, and multiple factors influencing adoption. PCA can be a valuable tool for market segmentation in this context. Here's how:

1. **Identifying Key Factors:** PCA can analyse a wide range of variables related to customer demographics, psychographics, vehicle preferences, and charging infrastructure availability. By identifying the principal components, we can determine the key factors that drive EV adoption in India.
2. **Segmenting Customers:** Based on the principal components, we can cluster customers into distinct segments with similar characteristics and needs. This allows for targeted marketing and product development strategies.
3. **Understanding Segment Preferences:** PCA can reveal the specific preferences of each segment regarding vehicle type, price range, features, and charging solutions. This information can guide the development of tailored EV offerings.
4. **Predicting Adoption:** By analysing the relationship between principal components and EV adoption, we can predict the likelihood of different segments adopting EVs in the future. This helps prioritize target markets and allocate resources effectively.

# Conclusion

PCA is a valuable technique for simplifying complex data and extracting meaningful insights. In the context of the Indian EV market, PCA can help identify key factors driving adoption, segment customers into distinct groups, understand their preferences, and predict future adoption patterns. By leveraging PCA, businesses can make informed decisions regarding product development, marketing strategies, and market entry, ultimately increasing their chances of success in this dynamic market.

EDA

Converting data into standard format is easy task. For being any column of our data,

Represents the standardised column of our dataset. Where:

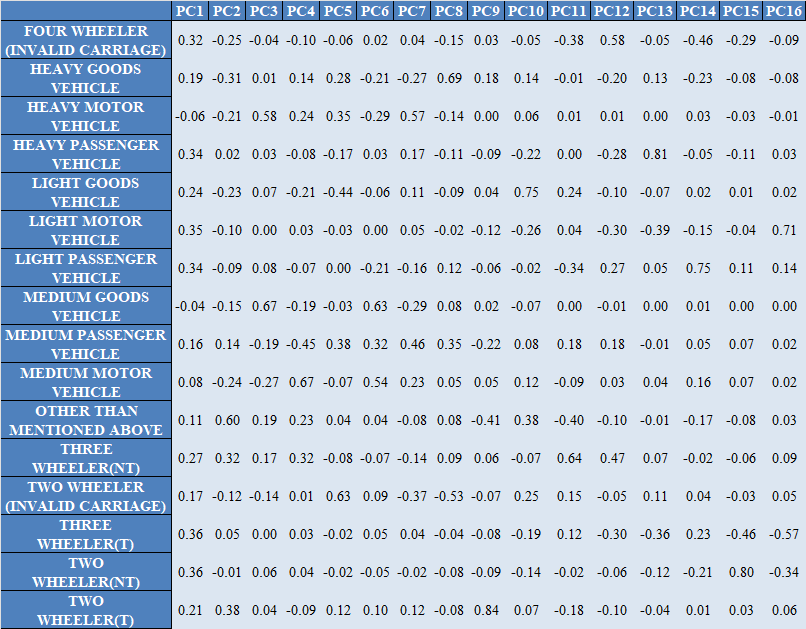
1. Standardised column of our dataset
2. Mean of .

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Standard Deviation** | **Proportion of Variance** | **Cumulative Proportion** |
| **PC1** | 2.676 | 0.447 | 0.447 |
| **PC2** | 1.283 | 0.103 | 0.550 |
| **PC3** | 1.024 | 0.066 | 0.616 |
| **PC4** | 0.997 | 0.062 | 0.678 |
| **PC5** | 0.983 | 0.060 | 0.738 |
| **PC6** | 0.959 | 0.057 | 0.796 |
| **PC7** | 0.925 | 0.053 | 0.849 |
| **PC8** | 0.856 | 0.046 | 0.895 |
| **PC9** | 0.743 | 0.034 | 0.929 |
| **PC10** | 0.666 | 0.028 | 0.957 |
| **PC11** | 0.525 | 0.017 | 0.974 |
| **PC12** | 0.383 | 0.009 | 0.984 |
| **PC13** | 0.324 | 0.007 | 0.990 |
| **PC14** | 0.304 | 0.006 | 0.996 |
| **PC15** | 0.217 | 0.003 | 0.999 |
| **PC16** | 0.133 | 0.001 | 1.000 |

1. Standard Deviation of .

This gives the **mean** and **var/std** of our dataset equal to 0 and 1 respectively. PCA provides a useful way of exploring data and creating a perceptual map. A perceptual map offers insights into how car manufacturing evolves over time more importantly, which car models tend to be similar. Principal components analysis is not computed to reduce the number of variables. This approach– also referred to as factor-cluster analysis– is inferior to clustering raw data in most instances. Here, we calculate principal components because we use the resulting components to rotate and project the data for the perceptual map.

Observe that the table lists all **16** principal components row wise. The **Standard deviation** of each component tell the amount of variability explained by each component independently. **Proportion of Variance** tell the proportion of total variance present in the data set explained by each component. **Cumulative Proportion** is the running sum of **Proportion of Variance.** Ideally, we would want the **Cumulative Proportion** to be over 80% for first 2 segments. What we achieved from this dataset is quite decent. Observe how over 50% of cumulative proportion of variance is explained by first 2 principal component. Extracting factor loadings of each Vehicle types that contribute to these 2 principal components can help up guide the interpretation of principal components. Loadings indicate how the original variables are combined to form principal components.



A few segmentation variables with highest loadings for principal component 1 are FOUR-WHEELER (INVALID CARRIAGE), HEAVY PASSENGER VEHICLE, LIGHT MOTOR VEHICLE, LIGHT PASSENGER VEHICLE, THREE-WHEELER(T) AND TWO-WHEELER(NT). Observe how these loadings differ widely for principal component 2. Loadings having the highest loading in PC1 have decreased loadings for PC2 and vica – versa, indicating how both components capture different segments of vehicles.

These initial exploratory insights represent valuable information for segment extraction. Results indicate that some attributes are strongly related to one another. Creating a perceptual map to gain more insights to what data suggests.

A graph with text and numbers

AI-generated content may be incorrect.

A few distinct vehicles classes are: MEDIUM MOTOR VEHICLE, MEDIUM GOODS VEHICLE, TWO-WHEELER(NT). The remaining attributes tend to be overlapping, indicating high heterogeneity among each segment. In contrast, the distinct vehicle classes tend to be unique, separate from others possibly due to demand and their key differences.

Following this analysis, we were able to identify a few distinct vehicle classes that might be a good candidate for target customers that we would analyze further. This was just exploratory data analysis. Moving on to extracting segments from data.

Extracting Segments

K-means clustering is a widely used unsupervised machine learning algorithm that addresses this challenge by partitioning data points into distinct clusters based on their similarity.

# How K-means Works

The K-means algorithm aims to find *k* clusters in the data, where *k* is a user-defined parameter. It works iteratively, starting with an initial set of *k* cluster centers (or "centroids"). The algorithm then assigns each data point to the nearest cluster centroid based on distance. Next, the centroids are recalculated as the mean of the data points assigned to each cluster. This process of assigning data points and updating centroids continues until the cluster assignments stabilize, meaning the centroids no longer change significantly.

# Benefits of Using K-means

* **Simplicity and Efficiency:** K-means is relatively easy to understand and implement, making it a popular choice for many applications. It is also computationally efficient, especially for large datasets.
* **Scalability:** K-means can handle large datasets efficiently, making it suitable for real-world business scenarios.
* **Interpretability:** The results of K-means clustering are relatively easy to interpret, as the cluster centroids represent the characteristics of each group.
* **Versatility:** K-means can be applied to a wide range of data types and business problems, including customer segmentation, market research, and anomaly detection.

# K-means for Market Segmentation

In a business context, K-means clustering is particularly useful for market segmentation. By analyzing Vehicle data, such as make and model, businesses can use K-means to identify distinct vehicle types having similarities in demand and features. This information can then be used to develop targeted vehicles that the market needs and improve customer satisfaction.

# Analysis

We calculated solutions for two to eight market segments using standard k – means analysis with ten random restarts. This was because we do not know in advance what the best number of market segments is. If we calculate a range of solutions, we can compare them and choose the one which extracts segments containing similar consumers which are different from members of other segments. We compare different solutions using Scree Plot.

A graph of blue rectangular bars

AI-generated content may be incorrect.

There seems to be a distinct elbow for 2 segments, however, the elbow is not wide enough to draw strong conclusions. We can use **mixture of distributions** approach to calculate latent class analysis using finite mixture of binary distribution.

# Using Mixture of Distributions

The mixture model maximizes the likelihood to extract segments (as opposed to minimizing squared Euclidean distances, as in case for k – means).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **iter** | **converged** | **k** | **k0** | **logLik** | **AIC** | **BIC** | **ICL** |
| **0** | 4 | TRUE | 2 | 2 | -2867.46 | 5738.92 | 5746.21 | 5745.86 |
| **1** | 4 | TRUE | 3 | 3 | -2422.97 | 4851.94 | 4862.88 | 4862.29 |
| **2** | 4 | TRUE | 4 | 4 | -2172.44 | 4352.87 | 4367.45 | 4366.85 |
| **3** | 9 | TRUE | 5 | 5 | -2003.72 | 4017.44 | 4035.66 | 4034.92 |
| **4** | 9 | TRUE | 6 | 6 | -1783.86 | 3579.72 | 3601.59 | 3600.84 |
| **5** | 8 | TRUE | 7 | 7 | -1605.79 | 3225.57 | 3251.09 | 3250.13 |
| **6** | 6 | TRUE | 8 | 8 | -1540.29 | 3096.58 | 3125.75 | 3124.97 |
| **7** | 7 | TRUE | 9 | 9 | -1410.59 | 2839.19 | 2872.00 | 2871.20 |
| **8** | 8 | TRUE | 10 | 10 | -1232.25 | 2484.50 | 2520.95 | 2519.95 |
| **9** | 8 | TRUE | 11 | 11 | -1060.97 | 2143.95 | 2184.05 | 2182.82 |
| **10** | 5 | TRUE | 12 | 12 | -993.49 | 2010.97 | 2054.72 | 2053.42 |
| **11** | 4 | TRUE | 13 | 13 | -975.98 | 1977.96 | 2025.35 | 2024.04 |
| **12** | 4 | TRUE | 14 | 14 | -928.08 | 1884.15 | 1935.19 | 1933.70 |

We plot the information criterion with a customized label for the y – axis to choose a suitable number of segments.

A graph with blue and orange lines

AI-generated content may be incorrect.

This figure plots information criteria values AIC and BIC on the y-axis for the different number of components (segments) on the x-axis. As can be seen, the values of all information criteria decrease quite dramatically until 7 components (market segments) are reached. If the information criteria are strictly applied based on statistical inference theory, even 14 segments are not sufficient to segment the market as AIC and BIC values keep on decreasing.

Visual inspection of the figures suggests that 7 market segments might be a good solution if a more pragmatic approach is taken. This is the point at which the decrease in the information criterion flattens.

Profiling Segments

|  |  |
| --- | --- |
| **Segments** | **Proportion** |
| 1 | 0.35% |
| 2 | 74.56% |
| 3 | 4.95% |
| 4 | 10.60% |
| 5 | 0.35% |
| 6 | 8.83% |
| 7 | 0.35% |

Core of Segmentation analysis is complete: market segments have been extracted. Now we need to understand what the seven-segment k-means solution means. The first step in this direction is to create a segment profile plot. The segment profile plot makes it easy to see key characteristics of each market segment. It also highlights differences between segments. The plot is kept easy to interpret, by keeping similar attributes close to one another. We achieved this by calculating a hierarchical cluster analysis. Hierarchical cluster analysis used on vehicle types (rather than dates) identifies– attribute by attribute– the most similar ones. The table on the right depicts the proportion of vehicle types present in each segment.

A screenshot of a computer

AI-generated content may be incorrect.

Observe how the names of the segmentation variables are written on the left side of the plot. The horizontal lines indicate median vehicle types in the entire sample present. The bars plot the median vehicle types *within each segment.* To understand market segments, Managers need to do two things:

1. Compare the bars for each segment to see what makes each segment distinct from all consumers in the market.
2. Compare bars across segments to identify differences between segments

Observe how the largest segment (2) comprises plenty of 2-wheeler vehicles (NT) and some light motor vehicles. Observe how the second largest segment (4) comprises mainly of 2-wheeler vehicles (NT) and 3-wheeler (T) vehicles.

Describing & Targeting Segments

Since TWO-WHEELER (NT) is a numerical column, we used a parallel box and whisker plot to assess the association of this vehicle type with segment membership.

A graph of a box diagram

AI-generated content may be incorrect.

Overall, observe how segments 2 and 6 (simply add 1 to each label to get correct segments) are quite similar in terms of TWO-WHEELER (NT) manufacturing, each consolidating to within 10000 units.

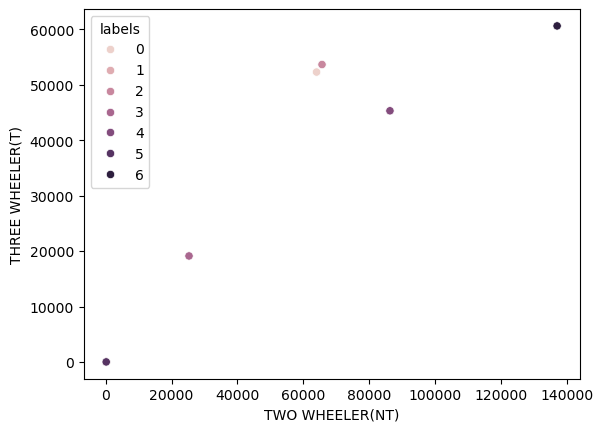
Also, Segment 3 tends to be manufactured more than segment 4 overall. These two segments are quite differentiated in terms of TWO-WHEELER (NT) manufacturing. Segments 5 and 7 generate mostly around 100000 and 1400000 units respectively. Each represents their unique features of being generated more than other segments.

Observe how TWO-WHEELER (NT) and THREE-WHEELER (T) were quite similar in terms of features from the last plot. Although, THREE-WHEELER (T) seemed to be manufactured less than the former. We should expect the same behavior in the boxplot. Generating the plot to draw some more inference.

A graph of a box diagram

AI-generated content may be incorrect.

Our assumption was right. THREE-WHEELER (T) how similar characteristics to TWO-WHEELER (NT). Aside from the salient features of the latter, it seems that THREE-WHEELER (T) tends to be manufactured less than TWO-WHEELER (NT) suggesting both are quite like each other and depicts a similar market demand and segment.



The Scatter plot for these 2 variables suggests that segment 7 seeks high manufacturing for these 2 specific vehicle types. However, given the nature of vehicle types in segment 7 (Refer to boxplot for TWO-WHEELER (NT) & THREE-WHEELER (T)), It seems to be quite concentrated, implying lack of heterogeneity among segments. This might lead to a situation of low demand and high supply. Catering to this segment might be disastrous.

On the other hand, segments 2 and 3 seem to be quite like each other. Observing segment profile plot suggests that segment 2 manufactures High TWO-WHEELERS (NT) but Low THREE-WHEELER (T). But, segment 3 manufactures High THREE-WHEELER (T). Implying potential demand for THREE-WHEELER (T) among segment 3 but not in 2. The startup can consider making TWO-WHEELER (NT) for segment 2 and THREE-WHEELER (T) for segment 3.

Also, notice how segments 1 and 3 are quite close to each other. It would be beneficial to target these two segments together for higher market exposure, provided company funds allow for targeting both.

Best location to set up manufacturing plant

To determine the best location to set up a manufacturing plant, we would need to maximize areas with the highest government penetration in EV space, more favorable regulation and specifically the highest number of PCS located. Immediately, problem arises what should be the scale of location? Should we target a specific area, town, city, state or country as a whole? Considering the availability of data and State should be an obvious pick. Below is listed the top 5 states having the highest numbers of PCS installed.

A graph of blue rectangular bars

AI-generated content may be incorrect.

Blindly accepting Maharashtra is set to be a doomed option. This is because, due to the higher number of PCS, more companies developing EVs are likely to be present in Maharashtra. To reduce the competition, a good balance between most PCS installed and least companies present in specific area is needed.

To achieve this, we would need to observe top 5 state with highest number of Manufacturers present. Doing so, there might be at least one state having decently installed PCS and low to nil manufacturers present. The following graph presents the top 5 states having the highest number of manufacturers present.

A graph of blue rectangular bars with names

AI-generated content may be incorrect.

Immediately, “Delhi” seems to be the inconsistent state out of the bunch, Delhi having second highest PCS installed but only 4 manufacturers present as compared to 15 in Maharashtra.

Diving deep into the analysis, below is Count plot of PCS v/s Companies present in each state. This plot is helpful because we are looking for states having low Companies present but high number of PCS installed. The plot is meticulously segmented into 4 clusters each representing high-low combination between No. of Companies and No. of Operational PCS.

A graph with numbers and dots

AI-generated content may be incorrect.

We have chosen 4 clusters to depict different types of states present in our data:

1. Cluster 0: State having High companies and low PCS installed
2. Cluster 1: State having High companies and High PCS installed
3. Cluster 2: State having Low companies and High PCS installed
4. Cluster 3: State having Low companies and Low PCS installed

We are looking for candidates in demand and newly emerged candidates to filter out states to set up operations. Ideally, we would want to have low competition and higher facilities available. Cluster 2 corresponds to Low companies present and high PCS installed. Perfect for our need. However, observing the lack of options present in cluster 2, we can consider cluster 3 as well, i.e. states having low PCS installed and low Company penetration, but not all states under 3 qualify for being a potential candidate. We need state with decently large PCS installed but close to nil companies present, i.e. upper segment of cluster 3 states. These include Delhi, Rajasthan and Telangana in decreasing order of importance.

Hence, the startup should consider setting up operation in Delhi, Rajasthan or Telangana to have good exposure to developing market while also enjoying least competition.

Pricing

We started by importing sales dataset for different vehicles. Our data permits us to analyse 2W, 3W, LMV and MMV vehicle types. Perfect for our chosen segments!

Our task is to determine how much vehicles of 2W and 3W have been sold for past 10 years and what does the trend suggest for next 3 years. A simple approach is to model regression of sales over years. That is, Linear Regression. Here a plot of Sales for 2W.

A graph with blue dots

AI-generated content may be incorrect.

The reason we chose logarithmic scale was because the sales seem to rise exponentially after 2018. Such data will not be a good fit for our linear regression, plotting logarithm of sales v/s years removes the exponential trend. Later, using some clever algebra, we can reintroduce the exponential trend to get exact sales figures.

Fitting a “Linear Model” refers to finding a “Line” of best fit that lies close to all datapoints. Thus, it is a line that “Minimizes” the error or distance between line and the datapoints. Ideally, we would want the line to fit all datapoints with 0 error, implying datapoint lying on the line itself. But, this is not possible practically.

The equation for Slope and Intercept of the Line:

Is given by:

|  |  |
| --- | --- |
|  |  |

Where:

1. Mean of Logarithm of Sales
2. Mean Year

Observing the scatter plot, we did not include year 2015 for both 2W and 3W because of 2 reasons:

1. Sales figure of 10 years ago holds much less value than sales of 2024
2. The datapoint for 2W for year 2015 is far off than other sales figures. This will cause poor model fitting.

Hence, it was better to not include 2015 sales figure for model building.

A graph with blue dots

AI-generated content may be incorrect.

Finding out a line of best fit is as simple as putting all the values in the formula. Below is the required value for Intercept and Slope of model fitted for 2W and 3W Sales in past 9 years.

**----------2W-----------**

**Intercept = 7.3519**

**Slope = 0.8606**

**----------3W-----------**

**Intercept = 7.3519**

**Slope = 0.2768**

Here’s the plot for fitted lines:

A graph of sales and vehicles

AI-generated content may be incorrect.

A graph of sales

AI-generated content may be incorrect.

Customizing Marketing Mix

Currently, our model predicts sales on a logarithmic scale by equation:

**For 2W**

Where year is inputted in absolute terms.

**For 3W**

Inputting values for years 2025, 2026, and 2027 yields results:

|  |  |  |
| --- | --- | --- |
| **Years** | **Sales 2W** | **Sales 3W** |
| 2025 | 3604027 | 631925 |
| 2026 | 8522244 | 833437 |
| 2027 | 20152078 | 1099209 |

In comparison, the actual sales for the years 2024 were 550351 and 372016 for 2W and 3W respectively. Looking at the numbers, that approximate growth of around **5x** and **2x** for 2W and 3W respectively.

Assuming the startup set ups its production in Delhi, where 4 companies would be rivaling post market exposure for the startup and each company sales equally likely, that correlates to 2W sold and 3W sold. Assuming 1 Lakh price for 2-Wheeler and around 2 Lakh Price of 3W, the startup would be earning gross income of around

9,010 Cr and 3,160 Cr for 2-Wheeler and 3 – Wheeler respectively.

**References:**

1. [3-Wheeler price](https://trucks.cardekho.com/en/news/detail/top-5-budget-friendly-electric-auto-rickshaws-in-india-power-range-price-included-2758.html)
2. [2-Wheeler price](https://www.bikedekho.com/electric-scooters)

Conclusion

The Indian EV market presents a rapidly evolving landscape with immense potential and significant challenges. Through our analysis, we have identified key customer segments, preferred vehicle types, and strategic locations for market entry. The use of **Principal Component Analysis (PCA)** and **K-means clustering** has enabled a data-driven segmentation of the market, revealing that **Two-Wheelers (NT) and Three-Wheelers (T)** are the most promising vehicle categories for an EV startup.

The insights from clustering suggest that **Segments 1 and 3** could provide early expansion opportunities as they seem to inculcate the right blend of uniqueness among other segments and some sense of heterogeneity among the segment as well. Both of these segments suggest development of Two-Wheeler (NT) and Three-Wheeler (T) vehicles. Furthermore, our location analysis indicates that **Delhi, Rajasthan, and Telangana** offer the best balance of infrastructure support and competitive advantage for setting up a manufacturing plant.

Additionally, our predictive sales modelling suggests a **5x growth in Two-Wheeler EV sales** and **2x growth in Three-Wheeler EV sales** over the next three years, highlighting the industry's rapid expansion. Based on projected sales volumes and competitive dynamics, a strategic pricing model can ensure profitability while maintaining market competitiveness.

In summary, this study provides a comprehensive, data-driven foundation for a successful EV market entry strategy in India. By focusing on the right customer segments, choosing optimal vehicle categories, and establishing manufacturing in key regions, the startup can effectively position itself for long-term success in the growing Indian EV industry.