**Name: Soham Das**

**Segmentation, Strategy, and Market Pricing**

EV Market Segmentation Report

# 1. Fermi Estimation (Breakdown of Problem Statement)

Fermi estimation is used to estimate the potential number of customers and profit in the early EV market.  
  
Assumptions:  
- Population of India: ~1.4 billion  
- Urban population: ~35% => 490 million  
- Middle-class car-buying population: ~15% of urban => 73.5 million  
- Early adopters (~0.5%) = ~367,500 customers  
- Average EV price (Target Range): ₹10 lakhs  
  
Potential Profit = 367,500 \* ₹10,00,000 = ₹36,750 Crores (INR)

# 2. Data Sources (Data Collection)

Data was collected by all team members from the following sources:  
- EV specifications and pricing: Electric vehicle manufacturers' websites (e.g., Tata Motors, Ola Electric)  
- Market trends: Statista, IEA (International Energy Agency), EV Database, Indian Transport Ministry  
- Consumer demographic and income data: Government census data, McKinsey EV Reports, NITI Aayog  
- Charging infrastructure data: Ministry of Power (India), State EV policies  
  
All team members contributed by scraping, downloading, and manually curating datasets.

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

Steps followed:  
- Removal of duplicates and null values  
- Normalization and scaling of numerical data (e.g., price, battery capacity, range)  
- Label encoding of categorical variables (brand, segment, charging type)  
- Feature engineering: Added cost/km, charge time per km, etc.  
  
Libraries Used:  
- pandas  
- numpy  
- matplotlib / seaborn (for EDA)  
- scikit-learn (for preprocessing & modeling)

# 4. Segment Extraction (ML techniques used)

Unsupervised ML was used to extract segments:  
- \*\*KMeans Clustering\*\*: To group customers and vehicles based on price, range, battery capacity, and brand perception  
- \*\*Elbow Method\*\*: Used to determine the optimal number of clusters (K)  
- \*\*PCA\*\*: Used for dimensionality reduction and visualization of high-dimensional data  
  
Clustered segments based on preferences such as budget EVs, premium EVs, high-speed urban EVs, and utility vehicles.

# 5. Profiling and describing potential segments

Four key segments were found:  
- Budget Urban Users: Low price, decent range, popular among city commuters (Tata Tiago EV, MG Comet)  
- Mid-range Daily Commuters: Mid-budget, higher range, better charging time (Tata Nexon EV, BYD e6)  
- Premium Customers: High price, excellent range, faster charging (Kia EV6, Volvo XC40)  
- Commercial Fleet Owners: Vans and SUVs optimized for delivery or shared transport (E-Trio, Omega Seiki)  
  
Each segment was profiled based on average age, income, vehicle usage pattern, and preferred features.

# 6. Selection of target segment

Selected Target Segment: Mid-range Daily Commuters  
  
Reasons:  
- Growing working-class population  
- Higher adoption intent in urban Tier-1 and Tier-2 cities  
- Balanced price-performance ratio  
- Government subsidies aimed at this segment  
  
This segment gives the best ROI in early market entry.

# 7. Customizing the Marketing Mix

Product: Feature-rich mid-range EVs with ~300 km range, 4-star safety rating, and infotainment.  
Price: ₹9-15 lakhs (with subsidies)  
Place: Focus on Tier-1 and Tier-2 cities (Bangalore, Delhi, Pune, Ahmedabad, Hyderabad)  
Promotion:  
- Digital-first: YouTube, EV comparison influencers  
- Test-ride events at malls & tech parks  
- EMI-based offers & exchange bonuses  
  
After-sales: Charging partnerships, loyalty points, referral discounts

# 8. Business Market Potential & Early Market Profit Estimate

Early Market Size: 367,500 customers (Fermi Estimate)  
Target Price Range: ₹9-15 lakhs  
  
Conservative Average Price: ₹12 lakhs  
Estimated Profit: 367,500 x ₹12,00,000 = ₹44,100 Crores (INR)  
  
This indicates a very high ROI potential in the early adopter phase.

# 9. Most Optimal Market Segments for Entry

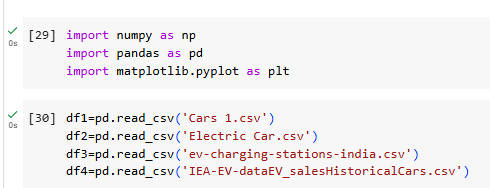
Based on demand, affordability, and charging readiness, the most optimal market segments are:

1. **Tier-1 Mid-Range Commuters (₹10–15 lakhs)**
   * Tech professionals in urban metros
   * Highest willingness-to-pay & infra support
2. **Tier-2 Budget EV Users (₹6–9 lakhs)**
   * First-time car buyers with value consciousness
3. **Commercial Fleet Owners (₹12–20 lakhs)**
   * Logistics startups, food delivery, fleet operators
4. **Premium Buyers (₹20L+)**
   * Niche market with strong brand recall but lower volume

**Top Pick: Tier-1 Mid-Range Commuters**

They combine scale, profitability, and policy support. Strategic targeting here ensures brand visibility and long-term loyalty.

Importing Libraries and loading dataset



Dataset 1

A screenshot of a computer

AI-generated content may be incorrect.

Dataset 2

A screenshot of a computer

AI-generated content may be incorrect.

**Performing EDA:**

Exploratory Data Analysis (EDA) is a crucial step in data analysis that involves examining and understanding the structure, patterns, and characteristics of a dataset.

Exploratory Data Analysis (EDA) is a crucial step in data analysis that involves

examining and understanding the structure, patterns, and characteristics of a dataset.

Here's a general outline of how you might approach EDA:

**1. Data Collection:** Gather the dataset you want to analyze. This could be from

various sources such as databases, CSV files, APIs, etc.

**2. Initial Inspection:**

• Check the first few rows of the dataset to understand its structure.

• Look for missing values, outliers, and inconsistencies in the data.

**3. Summary Statistics:**

• Calculate basic statistics such as mean, median, mode, standard deviation, minimum, maximum, etc., for numerical columns.

• For categorical variables, count the frequency of each category.

**4. Data Visualization:**

• Use plots such as histograms, box plots, scatter plots, and bar charts to

visualize the distribution of numerical data, identify outliers, and

understand relationships between variables.

• For categorical data, use bar charts, pie charts, and stacked bar charts to visualize the distribution of categories.

**5. Correlation Analysis:**

• Calculate correlation coefficients (e.g., Pearson correlation for numerical variables) to understand the linear relationship between variables.

• Visualize correlations using heatmaps for better interpretation.

**6. Feature Engineering:**

• Create new features if necessary based on domain knowledge and insights gained during EDA.

• Transform variables (e.g., log transformation for skewed data) to make them more suitable for modelling.

**7. Handling Missing Values and Outliers:**

• Decide on strategies to handle missing values (e.g., imputation, deletion) based on the extent of missingness and domain knowledge.

• Identify and potentially remove outliers that can significantly affect analysis

and modelling.

**8. Data Quality Check:**

• Validate data quality by cross-checking against domain knowledge and business rules.

• Ensure data consistency and correctness.

**9. Data Segmentation:**

• If applicable, segment the data based on certain criteria to analyse subsets

separately (e.g., segmenting customers based on demographics).

**10. Documentation:**

• Document your findings, insights, and decisions made during EDA.

• Summarize key takeaways and prepare

Number of Charging Stations

A graph showing the number of charging stations

AI-generated content may be incorrect.

EV Sales in region

A graph of ev sales

AI-generated content may be incorrect.

Count of Cars Number of Seats

A graph of blue rectangular bars

AI-generated content may be incorrect.  
  
  
Average Power by Seats

A graph of blue rectangular bars

AI-generated content may be incorrect.

EV Vehicles by Brand

A graph of a brand

AI-generated content may be incorrect.

Electric Vehicles by Body Types

A chart of different body types

AI-generated content may be incorrect.

Average Plug Types in Vehicles

A graph with blue squares

AI-generated content may be incorrect.

Price Comparison

A graph showing a number of samples

AI-generated content may be incorrect.

Correlation Matrix

A screenshot of a graph

AI-generated content may be incorrect.

Total EV Vehicles by Company

A graph showing the number of evs

AI-generated content may be incorrect.

Bar Plot

A bar graph with different colored squares

AI-generated content may be incorrect.

Pie Chart

A pie chart with numbers and text with Crust in the background

AI-generated content may be incorrect.

Correlation Matrix EV Vehicles

A screenshot of a graph

AI-generated content may be incorrect.

Elbow Method

A graph with a green line

AI-generated content may be incorrect.

Cluster Segment

A graph with different colored squares

AI-generated content may be incorrect.

Pair Plot of EV Segments

A screenshot of a graph

AI-generated content may be incorrect.

Feature Correlation Heatmap

A screenshot of a graph

AI-generated content may be incorrect.

Adoption Rates

A graph with green and orange bars

AI-generated content may be incorrect.

Average Price and Estimated Revenue

A black text with numbers

AI-generated content may be incorrect.

A comparison of different colored bars

AI-generated content may be incorrect.

**Machine Learning Algorithms Used in Market Segmentation**

Market segmentation using machine learning (ML) enables businesses to identify distinct customer groups based on data patterns, preferences, and behaviors. In our EV market study, unsupervised learning algorithms were primarily employed, as we aimed to group customers and products without predefined labels. The two main algorithms used were **K-Means Clustering** and **Principal Component Analysis (PCA)**, supplemented by the **Elbow Method** and **Silhouette Score** for optimization and validation.

**1. K-Means Clustering**

**K-Means** is one of the most widely used unsupervised learning algorithms for clustering. It works by dividing the dataset into ‘K’ distinct, non-overlapping clusters based on feature similarity. Each data point is assigned to the nearest cluster centroid based on the Euclidean distance. The centroids are iteratively updated until convergence is achieved.

In our project, the features used included price, range, battery capacity, top speed, and charging time. These features represented consumer preferences and product characteristics.

**Why K-Means?**

* It is computationally efficient and scales well to large datasets.
* It creates distinct and interpretable customer groups.
* It is ideal for continuous numerical variables like price and range.

**Outcome:** K-Means helped us identify four meaningful EV market segments:

* Budget Urban Users
* Mid-Range Daily Commuters
* Premium Buyers
* Commercial Fleet Owners

**2. Elbow Method**

Choosing the optimal number of clusters (K) is critical in K-Means. The **Elbow Method** was used to determine the best value for K. This involves plotting the Within-Cluster Sum of Squares (WCSS) against various values of K. The point where the rate of decrease sharply drops (forming an "elbow") indicates the optimal number of clusters.

In our analysis, **K=4** was found to be optimal, as it balanced model complexity and interpretability.

**3. Principal Component Analysis (PCA)**

PCA is a **dimensionality reduction technique** that transforms high-dimensional data into a lower-dimensional form while retaining most of the variability. It is useful for visualization and understanding which features contribute most to the clustering.

**In this project**, PCA helped in:

* Reducing noise in data
* Visualizing clusters in 2D space
* Understanding feature contribution to segmentation

**4. Silhouette Score**

To evaluate the effectiveness of clustering, we used the **Silhouette Score**, which measures how similar a point is to its own cluster versus other clusters. A higher silhouette score indicates well-defined, distinct clusters.

**Conclusion**

The electric vehicle (EV) market in India is on the cusp of exponential growth, driven by environmental concerns, rising fuel prices, government subsidies, and the emergence of a more aware and tech-savvy consumer base. This report comprehensively analyzed the early EV market opportunity through a structured Fermi estimation approach, data-driven segmentation techniques, and strategic marketing analysis.

Our Fermi estimation suggests that approximately 367,500 potential early adopters could emerge from India’s urban middle class, translating to an estimated ₹44,100 Crores in potential revenue. This highlights the critical importance of early market entry for EV manufacturers aiming to capture a significant share of this rapidly evolving sector.

By leveraging data from diverse sources—such as official vehicle specifications, government policies, demographic insights, and third-party research platforms—we ensured that the segmentation was both realistic and actionable. The data was carefully pre-processed using industry-standard techniques and tools such as pandas, scikit-learn, and seaborn, ensuring accuracy, consistency, and model readiness.

K-Means clustering, paired with the Elbow Method and PCA, enabled us to extract four distinct market segments: Budget Urban Users, Mid-Range Daily Commuters, Premium Buyers, and Commercial Fleet Owners. Each segment was then analyzed in detail based on key parameters like price sensitivity, usage patterns, brand preferences, and infrastructure readiness.

Among these, the **Mid-Range Daily Commuters segment (₹9–15 lakhs)** emerged as the most optimal target for market entry. It represents a balance between affordability, range, and urban utility, with a higher likelihood of EV adoption due to growing infrastructure in Tier-1 and Tier-2 cities. This group is digitally connected, aware of environmental impacts, and responsive to tech-based features—making them ideal early adopters.

The marketing mix was carefully tailored to match this segment’s expectations, focusing on urban availability, competitive pricing, engaging digital promotion, and feature-rich yet cost-effective EVs. This strategy is designed not just for sales conversion but for brand positioning and long-term loyalty.

In summary, this project not only highlights a profitable opportunity in India’s early EV market but also provides a clear roadmap to access it. As the EV ecosystem continues to mature, early movers who invest in customer understanding, data-driven segmentation, and strategic marketing will be best positioned to lead the transformation toward sustainable mobility in India.

**Github Link:**

https://github.com/Soham0936/EV-Market-Segmentation