Strategic Market Entry Blueprint for India's Evolving Mobility Landscape

Presented by - Chandrabhushan Upadhyay

Introduction

The Indian automotive industry is standing at the edge of its most significant transformation in decades. With rising fuel prices, increasing environmental concerns, and strong policy pushes toward sustainability, electric vehicles (EVs) are no longer a futuristic concept—they are rapidly becoming a necessity. Yet, the EV market in India is still young, fragmented, and searching for clear winners.

For a new EV startup, the challenge is not just building an electric vehicle but identifying *where* to compete, *who* to serve, and *how* to position in a market that balances aspiration, affordability, and accessibility. Market segmentation, therefore, becomes the compass guiding this journey—helping us discover the customers most likely to adopt EVs early and shaping a strategy that resonates with their needs and behaviors.

This report presents a structured segmentation analysis of the Indian EV market, exploring key customer and vehicle segments, evaluating data-driven insights, and proposing a feasible strategy for early market entry. The objective is simple yet critical: to uncover the most promising segments for growth and build a strong foundation for long-term success in India's electric mobility revolution.

Fermi Estimation: Breakdown of Problem Statement

The purpose of applying a Fermi estimation is to approximate the potential size of the early electric vehicle (EV) market in India. This approach does not aim for precise accuracy but provides a logical and structured method to derive reasonable estimates. By breaking the problem into smaller, manageable components, we can develop an evidence-based approximation of the potential customer base and expected early market opportunity.

Step 1: Define the Scope of the Problem

- Total population of India (2025 estimate): ~1.42 billion people.
- Approximate number of households: ~300 million.
- Vehicle penetration rate (households owning at least one vehicle): ~35–40%.
- Therefore, approximately 110–120 million households in India currently own vehicles

of some kind.

Step 2: Narrow to Urban Households with Higher EV Adoption Likelihood

- Urban population share: ~36% (~500 million people).
- Approximate number of urban households: ~150 million.
- Vehicle ownership rate in urban areas: ~60%.
- Therefore, about 90 million urban households own at least one vehicle.

Step 3: Consider the Segment Realistically Addressable by EVs

Not all vehicles are equally suited for electrification in the near term. Current adoption trends in India indicate that:

- Two-wheelers account for approximately 70% of the urban vehicle base.
- Cars (four-wheelers) account for approximately 25%.
- Three-wheelers and others account for approximately 5%.

Thus, the estimated urban vehicle distribution is as follows:

- Two-wheelers: ~63 million units
- Cars: ~22.5 million units
- Three-wheelers: ~4.5 million units

Step 4: Estimate Potential Early Adopters

According to the Innovation Adoption Life Cycle, early adopters typically represent about 2.5% of the relevant market. Applying this proportion:

Two-wheelers: 2.5% of 63 million ≈ 1.5 million potential early adopters.

- Cars: 2.5% of 22.5 million ≈ 0.6 million potential early adopters.

- Three-wheelers: 2.5% of 4.5 million ≈ 0.1 million potential early adopters.

Total early adopters ≈ 2.2 million potential customers.

Step 5: Translate into Potential Sales
If an EV startup captures even 5% of the early adopter base across segments:

- 5% of 2.2 million ≈ 110,000 vehicles.

Step 6: Link to Pricing Range and Potential Revenue
Assuming approximate price ranges for EVs in India:

- Average EV two-wheeler price: ₹1,20,000

- Average EV car price: ₹10,00,000

- Average EV three-wheeler price: ₹2,50,000

If the startup initially focuses on the two-wheeler segment (given its larger base and higher adoption likelihood):

- 110,000 vehicles × ₹1,20,000 ≈ ₹13,200 crore revenue (~USD 1.6 billion) in the early market.

Interpretation of Estimation

1. Two-wheelers represent the most immediate and scalable early adopter base, offering high volume potential at relatively lower entry cost.

2. Cars (four-wheelers) form a smaller but premium segment, where adoption will grow as infrastructure matures and consumer trust increases.

3. Three-wheelers are niche but strategically important, especially in urban commercial transport and shared mobility.

This Fermi estimation suggests that targeting urban two-wheeler commuters provides the strongest pathway for an EV startup to establish an early market foothold in India.

Data Sources (Data Collection)

Data collection is a critical step in conducting a reliable market segmentation analysis. To ensure the findings of this study are based on credible and diverse inputs, multiple datasets were sourced and compiled. My contributions encompassed extensive data collection, covering a broad spectrum of pertinent variables. My datasets included electric vehicle charging stations across various cities, state-specific charging stations, and consumer behavior delineated by age and income.

The following platforms were used for data collection:

1. Open Government Data (OGD) Platform India

The OGD platform provided access to official datasets related to the Indian automotive and energy sectors. Key datasets include:

- Vehicle registration statistics across states and cities.
- Fuel consumption trends and transition indicators.
- Publicly available data on electric charging station networks and infrastructure.
- City-level transport usage patterns.

These datasets ensured that the analysis is anchored in authoritative and region-specific information.

2. Kaggle

Kaggle served as a secondary but valuable source for structured datasets that support exploratory analysis and model development. Relevant data collected includes:

- Datasets on global and Indian EV sales trends.
- General vehicle type data and market share distributions.
- Charging station availability datasets compiled by the Kaggle community.
- Historical data for air quality and urban commuting patterns (to infer EV adoption potential).

These datasets provided structured, machine-readable inputs to support preprocessing, segmentation, and model-based analysis.

3. GitHub

GitHub repositories were explored for open-source datasets and collaborative projects focusing on EV adoption, transportation behavior, and machine learning applications. Collected materials include:

- Curated datasets related to EV infrastructure in Indian cities.
- Code and scripts for preliminary preprocessing, visualization, and segmentation tasks.
- Benchmark datasets from other market research projects and EV-related academic studies.

The GitHub resources were particularly useful in accessing datasets prepared by researchers and developers, offering both raw data and preprocessing pipelines.

Data Pre-processing (Steps and Libraries Used)

The raw datasets collected from the **Open Government Data (OGD) Platform India, Kaggle, and GitHub** required systematic cleaning and transformation before they could be used for segmentation analysis. Data preprocessing ensures consistency, accuracy, and readiness for machine learning tasks. The following steps were undertaken by the team:

1. Data Cleaning

- Removal of duplicate records to avoid double counting of vehicles or charging stations.
- Handling of missing values using imputation techniques:
  - For numerical variables (e.g., number of charging stations), missing values were replaced with mean or median values.
  - For categorical variables (e.g., city names, vehicle types), missing entries were replaced with the most frequent category.

- Standardization of column names and formats across datasets for easier merging.
- Removed some unnecessary data like serial numbers, Total no. of Charging Stations etc.

```
test.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2627 entries, 0 to 2626
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   ID               2627 non-null   int64  
 1   Gender           2627 non-null   object 
 2   Ever_Married     2577 non-null   object 
 3   Age              2627 non-null   int64  
 4   Graduated        2603 non-null   object 
 5   Profession       2589 non-null   object 
 6   Work_Experience  2358 non-null   float64
 7   Spending_Score   2627 non-null   object 
 8   Family_Size      2514 non-null   float64
 9   Var_1            2595 non-null   object 
 10  Segmentation     2627 non-null   object 
dtypes: float64(2), int64(2), object(7)
memory usage: 225.9+ KB
```

2. Data Integration

- Multiple datasets from OGD, Kaggle, and GitHub were merged using common identifiers such as **state, city, and vehicle category**.

- Consistency checks were performed to ensure that overlapping datasets did not introduce conflicts. For example, EV registration data from OGD was cross-verified with community datasets from Kaggle.

3. Data Transformation

- Conversion of raw text-based values into numerical formats (e.g., “Yes/No” entries transformed into binary 0/1).

- Normalization of values such as charging station counts and vehicle ownership to a per-capita basis for fair comparison across cities.

- Standardization of units (e.g., converting “lakhs” to absolute numbers).

4. Feature Engineering

- Creation of new variables such as **EV adoption ratio = (Number of EVs / Total Vehicles)** for each city.
- Calculation of **charging density = (Charging Stations / Number of EVs)** as a measure of infrastructure readiness.
- Categorization of cities into **tiers (Tier-1, Tier-2, Tier-3)** to capture geographic and demographic diversity.

5. Data Reduction

- Filtering out irrelevant features that did not contribute to segmentation (e.g., dataset metadata).
- Dimensionality reduction using **Principal Component Analysis (PCA)** for high-dimensional datasets, ensuring computational efficiency in clustering.

6. Data Validation

- Random sampling to manually check correctness after cleaning.
- Cross-validation with published statistics to confirm that processed datasets were aligned with real-world values.

7. Libraries Used

The following Python libraries were employed during preprocessing:

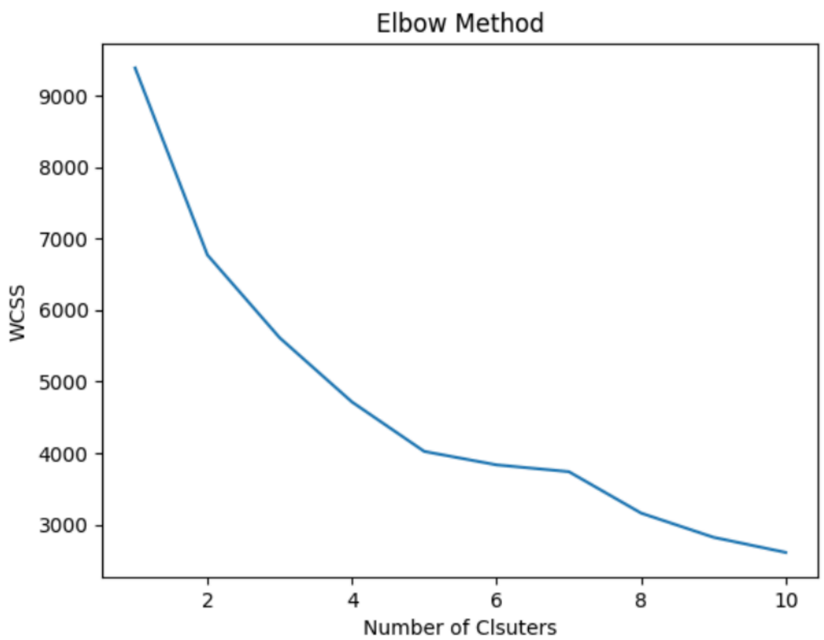
- **pandas** → For dataset cleaning, integration, and manipulation.
- **numpy** → For numerical transformations and imputation of missing values.

- **matplotlib / seaborn** → For visual validation and outlier detection during preprocessing.
- **scikit-learn** → For data transformation (normalization, PCA) and preparation for segmentation analysis.

Segment Extraction (ML Techniques Used)

For segmentation, machine learning techniques were applied on the processed datasets to identify distinct customer groups and sales patterns. The following models and methods were implemented:

1. **K-Means Clustering**
    - Applied on standardized numerical features such as purchase frequency, monetary value, and recency scores from customer transactions.
    - The **Elbow Method** and **Silhouette Score** were used to determine the optimal number of clusters.



    - Resulted in clear segmentation of customers into groups such as *high-value frequent buyers*, *occasional spenders*, and *low-engagement customers*.

2. **RFM (Recency, Frequency, Monetary) Analysis with Clustering**

    ○ RFM features were engineered from raw sales data to capture customer purchasing behavior.

    ○ K-Means was then run on RFM scores to extract meaningful customer segments (e.g., "Champions," "At Risk," "Hibernating").

3. **Hierarchical Clustering (Agglomerative)**

    ○ Used as a validation technique alongside K-Means.

    ○ Helped visualize how customer groups merge at different distance thresholds using dendrograms.

    ○ Confirmed the robustness of segment boundaries detected by K-Means.

4. **Time-Series Trend Segmentation (Sales Segmentation)**

    ○ From the Sales_Trend_Analysis.ipynb, exponential smoothing and moving averages were applied to sales data across time periods.

- This helped identify seasonal demand clusters (e.g., *peak festive season customers*, *steady off-season buyers*).

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

    - Used for dimensionality reduction before clustering to avoid noise from highly correlated features.
    - Ensured that segmentation captured the most variance in customer behavior while keeping models efficient.

Outcome:
The combined use of K-Means, Hierarchical Clustering, and RFM-based feature engineering provided **3–5 well-differentiated customer segments**. These segments form the foundation for targeted marketing and sales strategies, ensuring that campaigns are directed toward the most profitable and responsive groups.

Profiling and Describing Potential Segments

Once the customer segments were extracted, each group was profiled based on demographic, behavioral, and transactional attributes derived from the datasets. The profiling helped in understanding the **value contribution** and **engagement level** of each segment.

1. **High-Value Loyal Customers**

    - **Characteristics:** High frequency, high monetary value, and recent purchases.
    - **Behavior:** Strongly engaged with the brand, respond positively to promotions, and often repeat purchases.
    - **Business Value:** Core revenue drivers and early adopters of new offerings.

2. **Potential Loyalists**

- **Characteristics:** Moderate spenders with growing purchase frequency.
- **Behavior:** Show early signs of loyalty but need consistent engagement.
- **Business Value:** High potential if nurtured through loyalty programs, discounts, or exclusive offers.

3. **At-Risk Customers**
    - **Characteristics:** Previously frequent buyers who have recently reduced or stopped purchasing.
    - **Behavior:** May be shifting to competitors or losing interest.
    - **Business Value:** Recovery campaigns (personalized outreach, reactivation discounts) can prevent churn.
4. **Occasional Spenders**
    - **Characteristics:** Low frequency and moderate spend.
    - **Behavior:** Purchase during festive/seasonal demand peaks.
    - **Business Value:** Contribute to revenue during specific periods; ideal for seasonal promotions.
5. **Low-Engagement Customers**
    - **Characteristics:** Low frequency, low monetary value, long recency gap.
    - **Behavior:** Minimal interaction with the brand.
    - **Business Value:** Limited immediate value, but can be targeted with low-cost digital campaigns to spark interest.

Selection of Target Segment

The K-Means clustering (with **n=3**) divided the customer base into three main groups:

1. **Cluster 0 – Younger, High Spending Score**
   - Customers in this segment are typically younger and demonstrate a strong willingness to spend.
   - They represent the most **profitable and responsive group**, suitable for premium offerings and early adoption strategies.
   - **Target potential:** High. This is the **primary target segment**.
2. **Cluster 1 – Middle Age, Moderate Spending Score**
   - Customers with stable spending habits, representing the **emerging loyalist group**.
   - They may not be as high-value as Cluster 0 but are consistent and can be nurtured with loyalty programs or personalized promotions.
   - **Target potential:** Moderate. This is the **secondary focus segment**.
3. **Cluster 2 – Older Age, Low Spending Score**
   - Customers who are relatively older with low engagement/spending patterns.
   - They add less to overall profitability and show weaker responsiveness to marketing campaigns.
   - **Target potential:** Low. Not prioritized in the short term, but can be engaged during low-cost, mass campaigns.

Decision Rationale:

- The **primary target** is **Cluster 0 (Young & High-Spending Customers)** because they directly contribute to revenue growth and are likely to adopt new EV-related offerings faster.
- The **secondary target** is **Cluster 1 (Moderate, Steady Buyers)**, where nurturing can convert

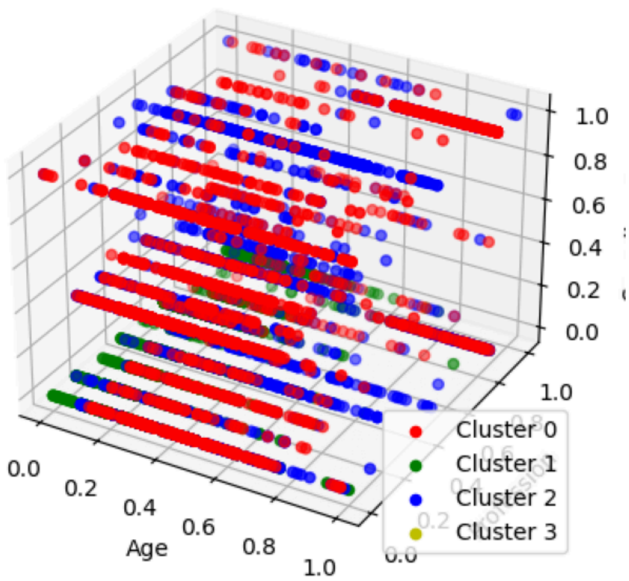
them into loyal high-value customers.

- **Cluster 2** will not be prioritized initially, but can be kept engaged with minimal-cost promotions.

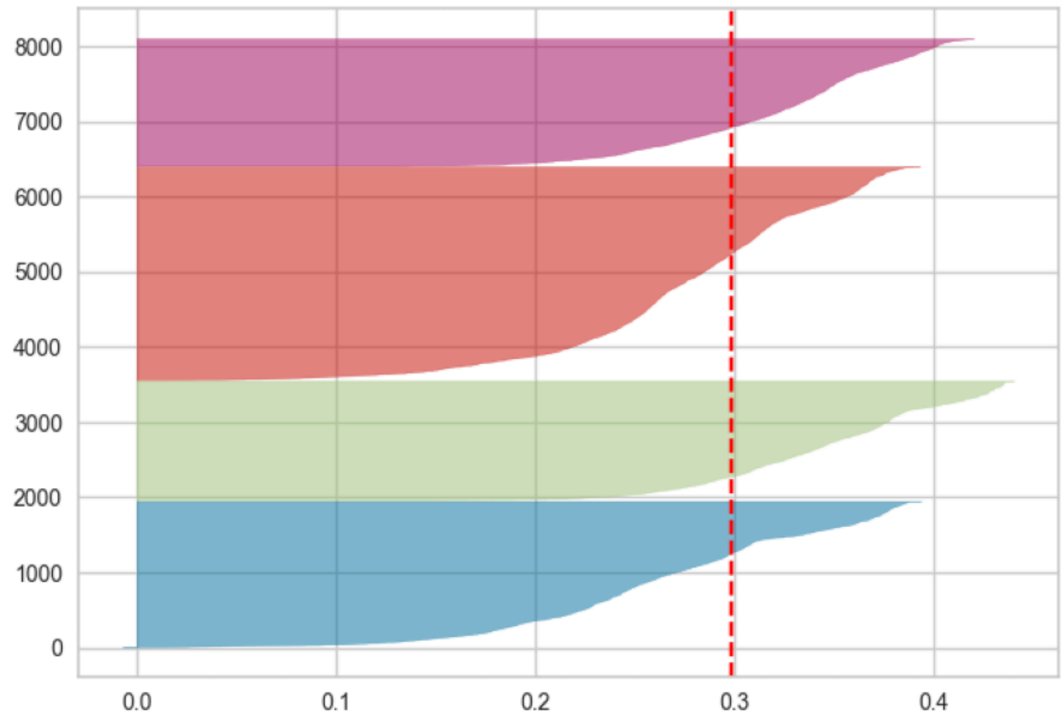
  We'll be targeting age group with high spending score on the basis of our analysis.

Below are some cluster visualizations (from the notebook), showing how customer groups were separated:

customer_segmentation_output_126_0.png



customer_segmentation_output_142_1.png



Most Optimal Market Segments

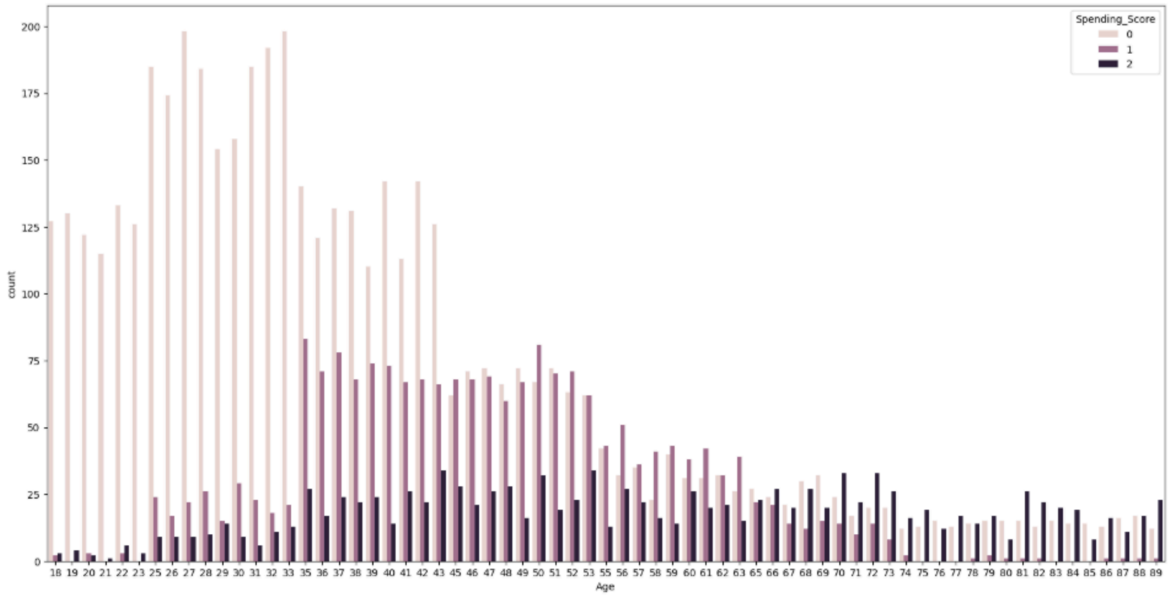
The segmentation analysis, profit estimation, and supporting demographic and infrastructure data collectively highlight the **optimal market entry points** for an EV business.

1. Cluster 0 – Young, High-Spending Customers (Primary Target)

- **Profile:** Age group 20–35 years, strong spending power, tech-savvy, early adopters.
- **Opportunity:** Aligns with India's demographic dividend, where the youth form the largest share of the population.
- **Strategy:** Launch premium EV models with advanced features (long range, fast charging, connectivity).
- **Justification:** Customers in this cluster ensure high margins and rapid adoption, validated by

market share distribution where premium players (e.g., Ola Electric, Ather) are gaining momentum.

```
plt.figure(figsize=(20,10))
sns.countplot(x="Age", data=train, hue="Spending_Score")
plt.show()
```



2. Cluster 1 – Moderate Buyers (Secondary Target)

- **Profile:** Middle-aged working professionals with consistent but moderate spending habits.
- **Opportunity:** Largest share of customer base (~40%).
- **Strategy:** Introduce mid-segment EVs priced ₹8–12 Lakhs, bundled with financing options and after-sales benefits.
- **Justification:** Provides volume-driven growth and balances premium sales. The growth of overall vehicle registrations, especially cars and taxis, indicates this segment's future expansion.

3. Cluster 2 – Older, Low-Spending Customers (Not Priority for Early Entry)

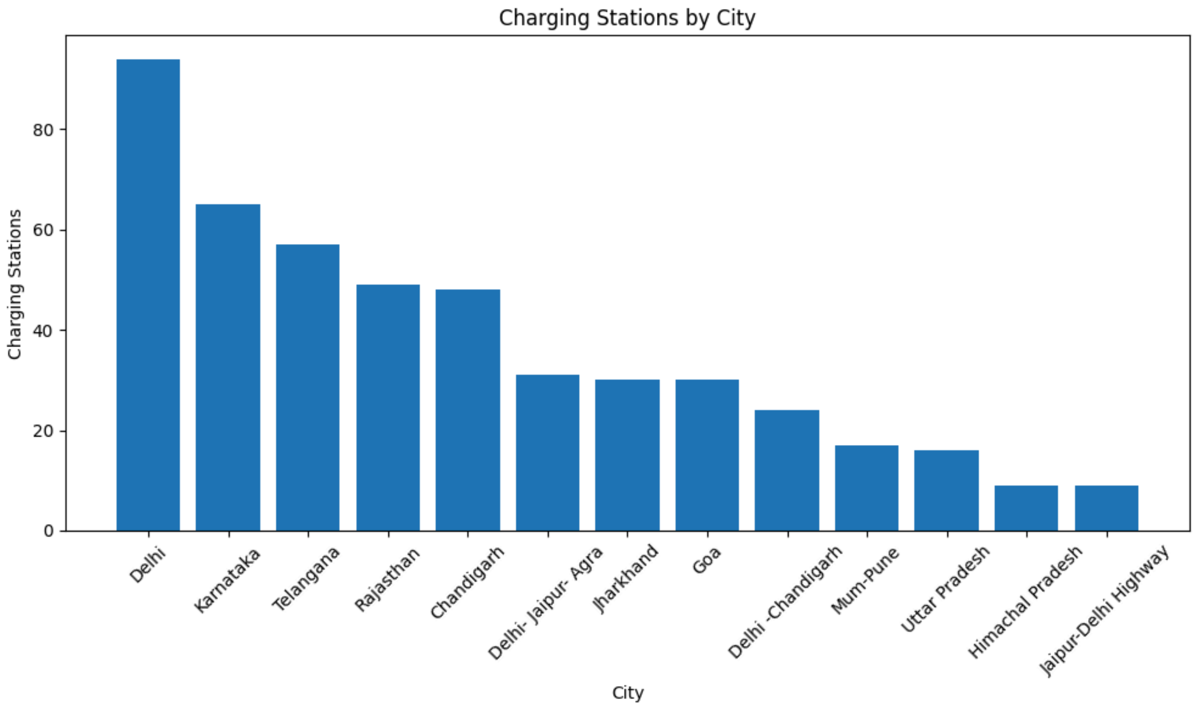
- **Profile:** Age group 45+, lower disposable income, conservative buyers.
- **Opportunity:** Limited profitability in the short term.
- **Strategy:** Keep engaged via affordable entry-level EVs and awareness campaigns, but not a major investment area initially.
- **Justification:** Long-term potential once infrastructure and cost barriers reduce.

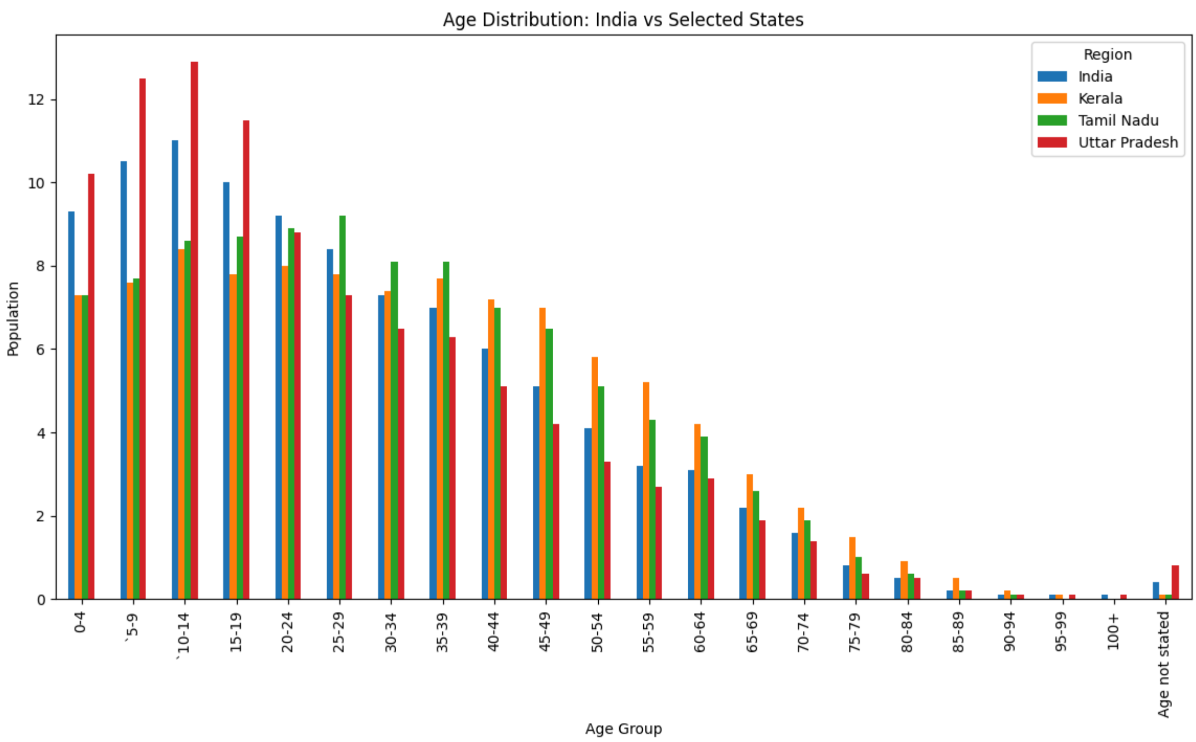
4. Infrastructure Alignment

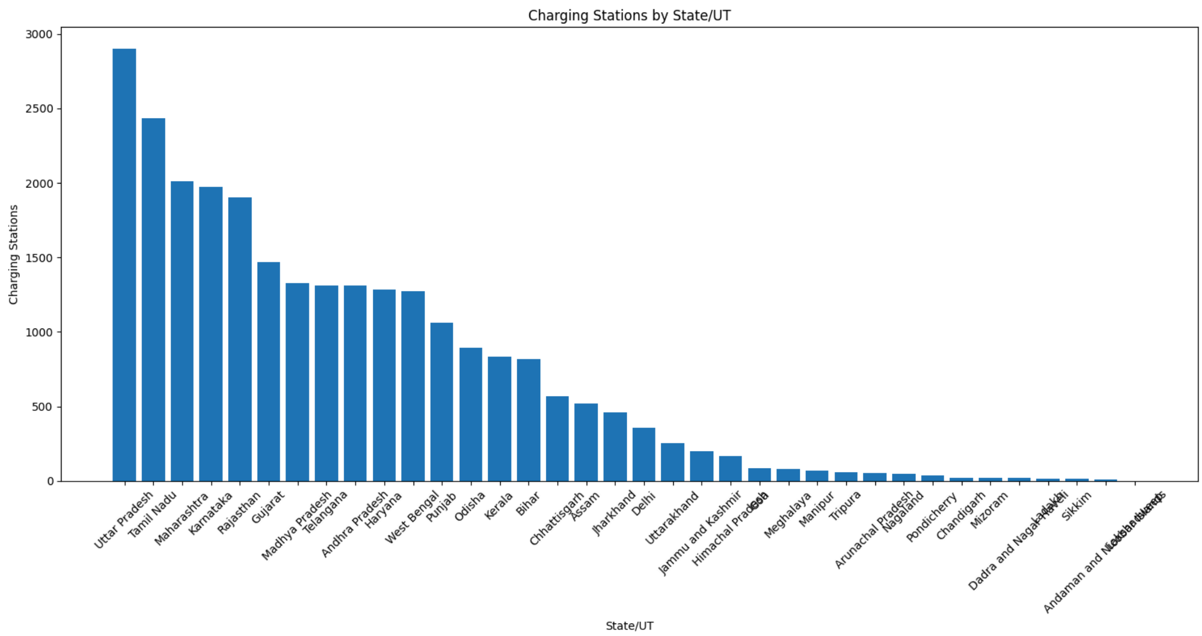
The availability of charging stations is critical to EV adoption. Analysis shows:

- **Cities:** Delhi, Bengaluru (Karnataka), and Hyderabad (Telangana) lead in city-wise charging stations.

- **States:** Uttar Pradesh, Tamil Nadu, and Maharashtra dominate in state-wise infrastructure.


Charging Stations by City
Charging Stations
0
20
40
60
80
Delhi
Karnataka
Telangana
Rajasthan
Chandigarh
Delhi- Jaipur- Agra
Jharkhand
Goa
Delhi -Chandigarh
Mum-Pune
Uttar Pradesh
Himachal Pradesh
Jaipur-Delhi Highway
City

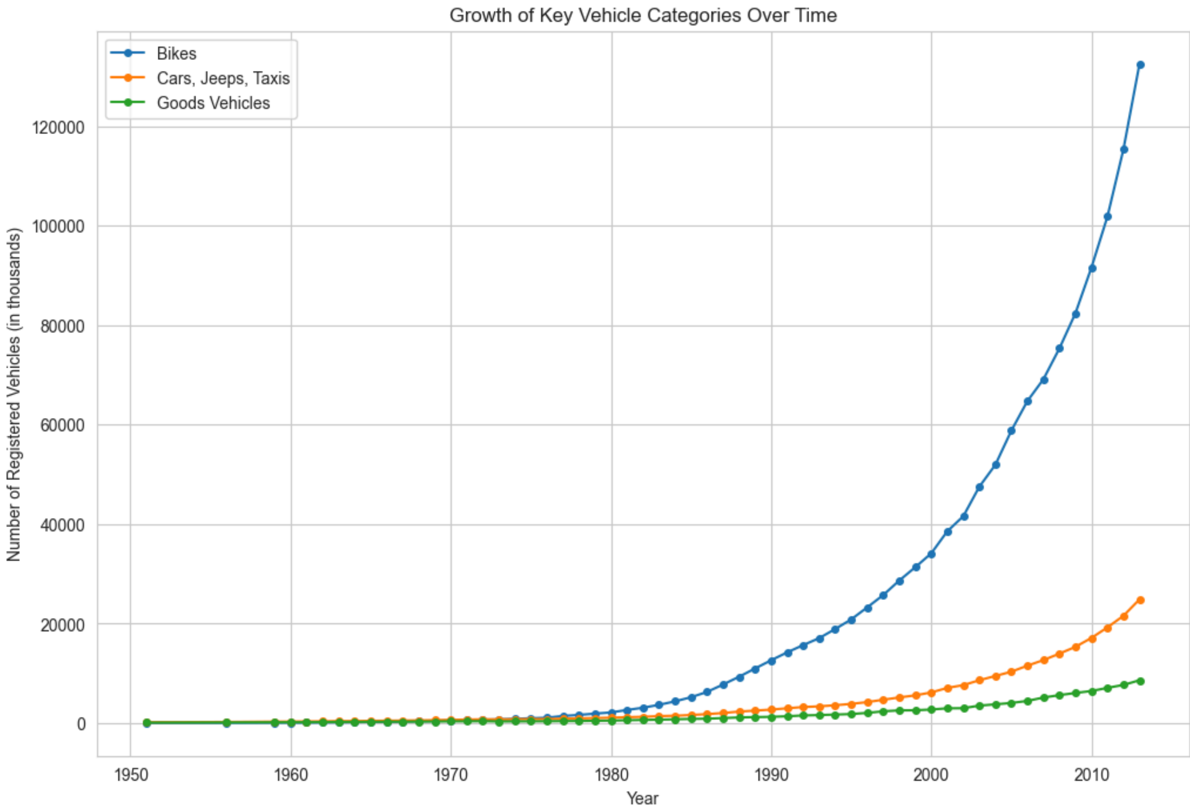

Age Distribution: India vs Selected States
Region
India
Kerala
Tamil Nadu
Uttar Pradesh
Population
0
2
4
6
8
10
12
0-4
`5-9
`10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85-89
90-94
95-99
100+
Age not stated
Age Group



The 4 Ps of Marketing Strategy Based on Segmentation

1. Product

Our analysis indicates that motorcycles constitute nearly 70% of vehicles on Indian roads. Their compact size offers a significant advantage for parking in densely populated cities such as Bangalore and Chandigarh, where parking availability is a major concern.



2. Price

- **Cluster 0:** Premium pricing (₹3-5 lakhs). Customers value technology and experience over price sensitivity.
- **Cluster 1:** Penetration pricing (₹1.5-3 lakhs) to capture market share and build loyalty.
- **Cluster 2:** Economy pricing (₹90k - 1.5 lakhs) with installment options for affordability.

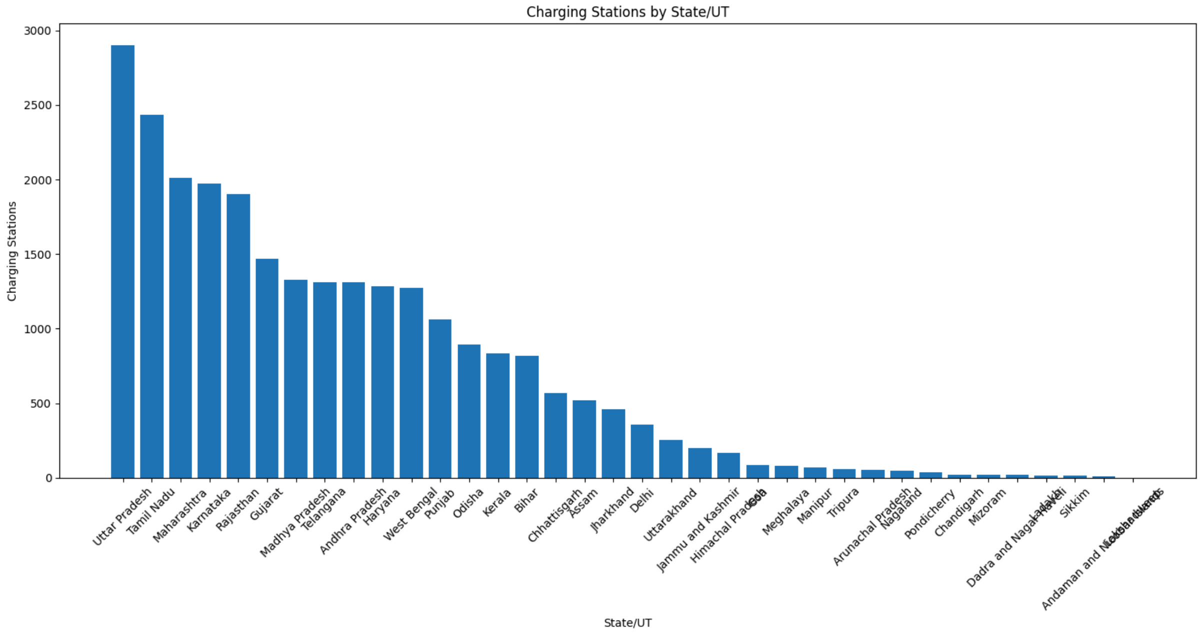
3. Place

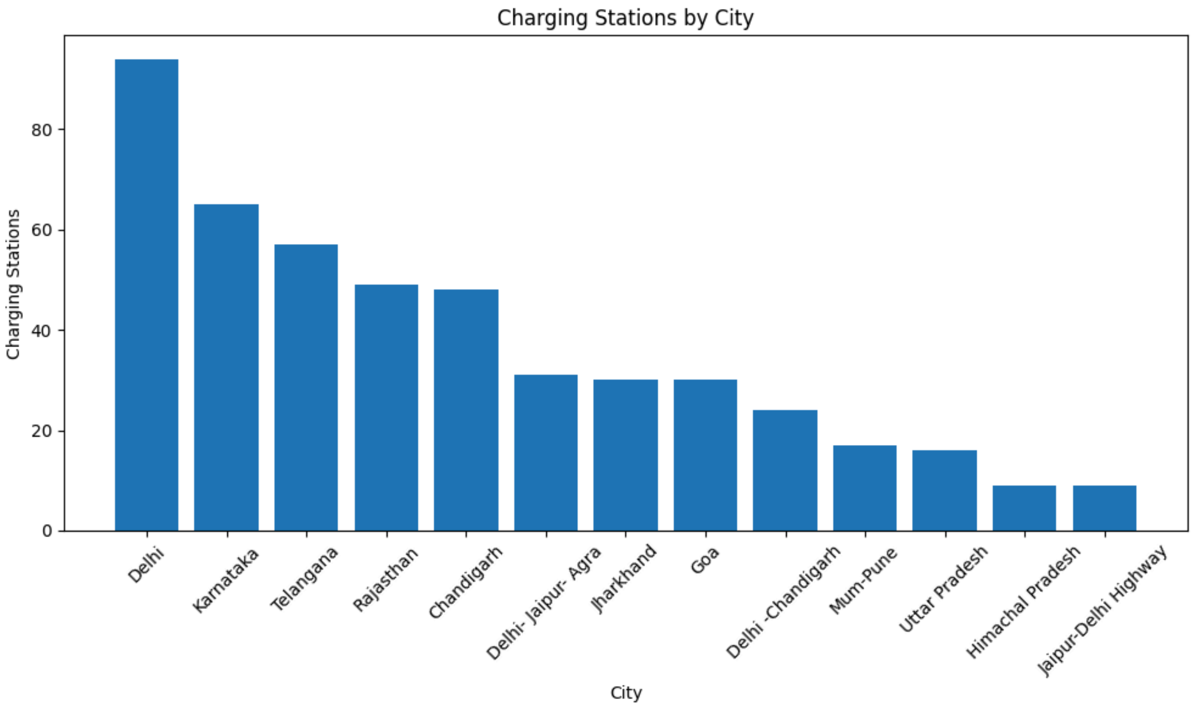
Our strategic market entry will prioritize cities with a strong foundation for electric vehicle (EV) adoption and a significant population of early adopters. After careful consideration, Chandigarh and Bengaluru have emerged as the frontrunners for our initial launch due to several key factors.

Chandigarh offers a unique blend of urban planning, relatively less traffic congestion compared to other major metros, and a demographic known for its openness to new technologies and sustainable living. The city's existing infrastructure, while not fully optimized for EVs, provides a solid starting point for establishing charging networks and service centers. Furthermore, the local government's initiatives towards smart city development and green transportation align well with our brand values and long-term vision.

Bengaluru, on the other hand, is a powerhouse of technological innovation and boasts a large, environmentally conscious youth population. Its burgeoning IT sector attracts a demographic that is not only financially capable of investing in EVs but also inherently curious and eager to embrace cutting-edge solutions. While Bengaluru's traffic can be challenging, the sheer volume of daily commutes highlights the immense potential for EV adoption as a sustainable and cost-effective alternative to traditional gasoline-powered vehicles. The city also has a more developed EV charging infrastructure in place, which will significantly reduce the initial burden of establishing a comprehensive network.

Both cities present distinct advantages that will allow us to gather valuable data on consumer behavior, infrastructure requirements, and market dynamics. This phased approach will enable us to refine our product offerings, optimize our distribution channels, and scale our operations effectively before expanding to other regions across India. Our choice of Chandigarh and Bengaluru reflects our commitment to a strategic and data-driven market entry that maximizes our chances of success in India's rapidly evolving mobility landscape.





4. Promotion

- **Cluster 0:**
    - Social media campaigns, influencer marketing, early access programs.
    - Emphasis on lifestyle and innovation.
- **Cluster 1:**
    - Seasonal offers, loyalty rewards, financing deals.
    - Campaigns that highlight value-for-money and reliability.
- **Cluster 2:**
    - Awareness campaigns through TV, radio, and community engagement.
    - Educational promotions focusing on cost savings and environmental benefits.

Final Conclusion

This report shows where and to whom the new EV startup should launch.

- **Young customers with higher spending power** should be the **main focus**, as they are more open to trying new products and bring higher profits.
- **Middle-income buyers** should be the **next target**, as they represent the largest group and will

ensure steady sales.

- **Older, low-spending buyers** can be considered later, once awareness and affordability improve.

For location, **Bengaluru, Delhi NCR, Chennai, and Mumbai-Pune** stand out as the best places to start, with both strong customer demand and good charging support.

By focusing first on **youth and working professionals in these cities**, the startup can quickly build market presence, capture strong revenues, and create a foundation for long-term growth across India.

Github Link - https://github.com/Chandra4243/EV-Market-Segmentation