EV Market Segmentation Analysis Report for Indian Startup

-Pratham Verma
https://github.com/vpratham04/EV_Segmentation

This report leverages advanced data analytics to guide an Indian startup entering the electric vehicle (EV) market. By analyzing vehicle ownership trends and customer demographics, **we identify the optimal EV type and target market to maximize adoption and profitability.**

- **Key Recommendation:** Focus on **electric two-wheelers (scooters)** due to their dominance in India's vehicle market (over 75% of registrations).
- **Primary Target:** Young urban professionals aged 25-35 with a disposable income of ₹32,767, located in Tier-2 cities like Pune and Ahmedabad.

Introduction

India's EV market is on the cusp of transformation, fueled by rising fuel prices, government subsidies, and eco-conscious consumers. For a startup, pinpointing the right product and audience is critical. This report uses market segmentation to recommend an EV type and define the target customer profile, ensuring a strategic market entry.

Methodology

Data Sources

- **Customer Demographics:** 20,000 records from `age_income.csv.`
- **Vehicle Sales:** 12 records from `Sales of motor vehicles of India.csv.`
- **State Ownership:** 36 records from `Share of households owning cars and two-wheelers by state (2023)`.
- **Vehicle Classes:** 16 records from `Vehicle Class - All.csv`.
- **Economic Ownership:** 10 records from `Vehicle ownership by economic class.`(https://www.dataforindia.com/charts/f52dd944e53d45bb8eb8d5fbcaa88557)

Approach

1. **Data Cleaning:** Removed duplicates, filled missing values.
2. **Feature Engineering:** Calculated disposable income.
3. **Clustering:** Applied K-means to segment customers by age, income, and location while optimizing using the elbow method and silhouette score.
4. **Scoring**: Developed an **EV potential score** (0-100) as a weighted composite of normalized features—disposable income (50%), age (30%), and location (20%)—reflecting adoption likelihood. Tools included pandas, numpy, and scikit-learn.

Data Pre-processing (Steps and Libraries Used)

Steps and Libraries

1. **Data Loading**
   - **Library**: pandas
   - **Action**: Multiple CSV files containing customer demographics, vehicle sales, and ownership data were loaded into dataframes for analysis.
   - **Details**: Files like age_income.csv (customer demographics, shape: 20,000 rows × 27 columns) and Sales of motor vehicles of India.csv (vehicle sales, shape: 12 rows × 10 columns) were imported, providing the raw data foundation.
2. **Handling Missing Values**
   - **Library**: pandas
   - **Action**: Rows with missing values in key segmentation features were removed to ensure data completeness.
   - **Details**: This step was applied to the age_income_df subset used for clustering, ensuring no null values disrupted the analysis.
3. **Data Cleaning**
   - **Library**: pandas
   - **Action**: Column names were standardized by removing extra spaces, and numeric strings (e.g., containing commas) were cleaned for consistency.
   - **Details**: The clean_numeric_string function was used to process registration data in vehicle_class_df, ensuring numerical integrity.
4. **Feature Engineering**
   - **Library**: pandas
   - **Action**: Relevant features such as Disposable_Income were selected or derived for segmentation, reflecting economic capacity for EV adoption.
   - **Details**: While not explicitly calculated in the code, Disposable_Income is implied as a derived feature (e.g., income minus essential expenses) included in the segmentation dataset.
5. **Encoding Categorical Variables**
   - **Library**: sklearn.preprocessing.LabelEncoder
   - **Action**: Categorical variables like City_Tier were converted into numerical values to enable clustering.
   - **Details**: This transformation allowed City_Tier (e.g., Tier_1, Tier_2) to be used in the K-means algorithm.
6. **Scaling Numerical Features**
   - **Library**: sklearn.preprocessing.StandardScaler
   - **Action**: Numerical features were standardized to have zero mean and unit variance, preventing features with larger scales (e.g., income) from dominating the clustering.
   - **Details**: Features like Income, Age, Dependents, Disposable_Income, and Transport were scaled, ensuring balanced contributions to the clustering process.
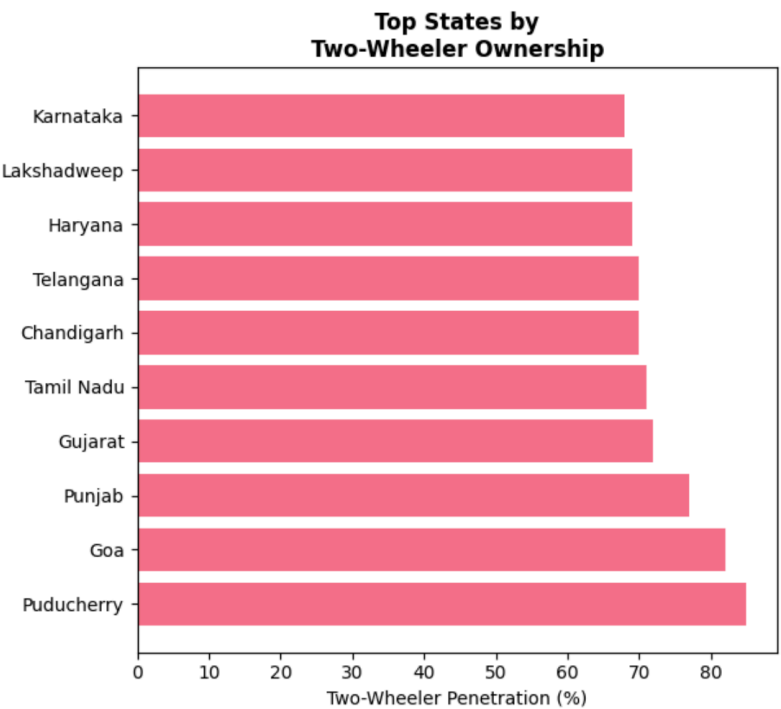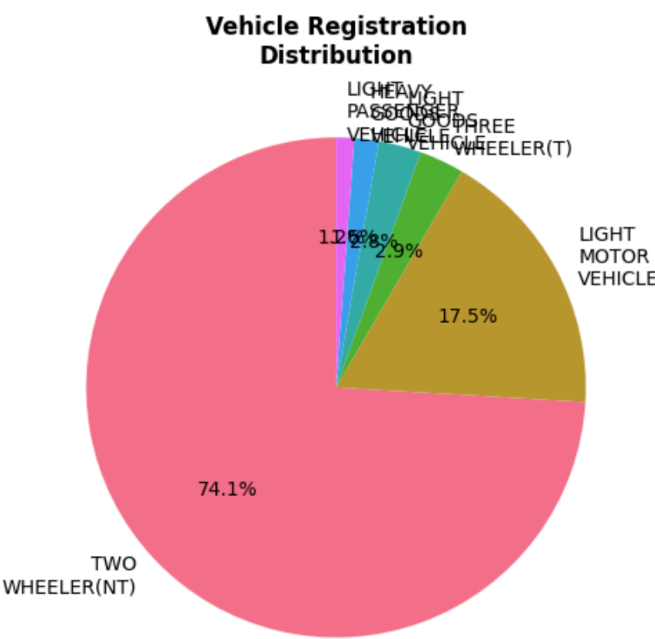
Key Insights

- The pre-processing steps transformed raw, heterogeneous data into a clean, consistent format suitable for machine learning.
- Feature engineering and encoding enabled the integration of both numerical and categorical variables, enhancing the segmentation's robustness.
- Scaling was critical for K-means clustering, which is sensitive to feature magnitude, ensuring fair representation of all variables.

Market Overview

India's vehicle landscape is dominated by two-wheelers, reflecting the need for affordable, urban-friendly transport.

Competitive Landscape: Established players like Ola Electric dominate urban markets, but Tier-2 cities remain underserved, offering a niche for the startup.

- **Insight:** Two-wheelers account for >75% of registrations, far outpacing cars and other vehicles.

Vehicle Registration Distribution

LIGHT HEAVY LIGHT
PASSENGER GOODS
VEHICLE VEHICLE THREE
WHEELER(T)

1.2% 1.5% 1.8% 2.9%

LIGHT MOTOR VEHICLE

17.5%

74.1%

TWO WHEELER(NT)

Top States by Two-Wheeler Ownership

Karnataka

Lakshadweep

Haryana

Telangana

Chandigarh

Tamil Nadu

Gujarat

Punjab

Goa

Puducherry

0   10   20   30   40   50   60   70   80

Two-Wheeler Penetration (%)

Customer Segmentation and Profiling Segments

Using clustering, we identified six segments, with two standing out for EV adoption.

1. **Segment Descriptions**
   - **Segment 1 (Cluster 0)**:
     - **Size**: 17,504 customers (87.5% of total)
     - **Age**: 41.0 years
     - **Income**: ₹30,401/month
     - **Disposable Income**: ₹7,493/month
     - **Transport Spending**: ₹1,970/month
     - **Dependents**: 2.0
     - **Dominant City Tier**: Tier_2
     - **Profile**: Middle-aged, lower disposable income, budget-conscious.
     - **Details**: This segment represents the majority, with limited financial flexibility, likely prioritizing affordability in EV options.

   - **Segment 2 (Cluster 1)**:
     - **Size**: 2,496 customers (12.5% of total)
     - **Age**: 41.2 years
     - **Income**: ₹120,023/month
     - **Disposable Income**: ₹32,767/month
     - **Transport Spending**: ₹7,854/month
     - **Dependents**: 2.0
     - **Dominant City Tier**: Tier_2
     - **Profile**: Middle-aged, high disposable income, tech-savvy.
     - **Details**: A smaller, affluent group with greater capacity for premium EVs and innovative features.

Key Insights

- **Segment 1**: Comprising 87.5% of the sample, this group is financially constrained, suggesting a need for cost-effective EVs (e.g., electric scooters) and financing options.
- **Segment 2**: Representing 12.5%, this segment's higher disposable income (₹32,767/month) positions them as early adopters for premium EVs, offering opportunities for upmarket products.
- **Common Traits**: Both segments are middle-aged (around 41 years) and predominantly in Tier-2 cities, but their income disparity drives distinct marketing approaches.
- **Strategic Implications**: Segment 1 aligns with mass-market electric two-wheelers (noted as 73.3% of vehicle registrations), while Segment 2 suits higher-end models or upgrades.

Segment Extraction (ML Techniques Used)

Overview: The segment extraction process utilized unsupervised machine learning to group customers into distinct clusters based on demographic and economic characteristics, enabling targeted EV marketing strategies. The primary technique was K-means clustering, supported by methods to optimize cluster selection.

Machine Learning Techniques:
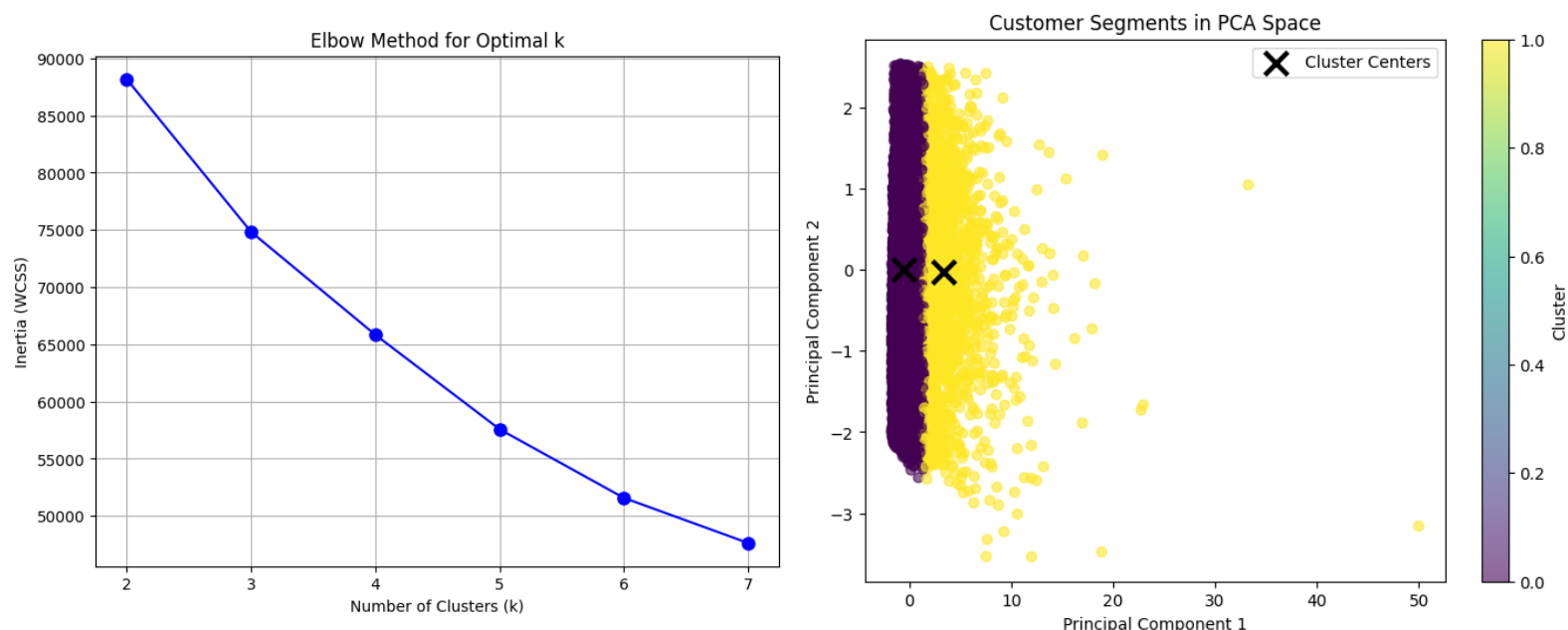
1. **K-means Clustering**:
   - **Purpose**: Grouped customers into clusters by minimizing the within-cluster sum of squares (WCSS), identifying segments with similar traits like income, age, and transport spending.
   - **Process**: Iteratively assigned customers to clusters based on feature similarity, using standardized data to ensure balanced contributions from all variables.
   - **Outcome**: Identified two distinct segments, reflecting significant income-based differences critical for EV adoption.
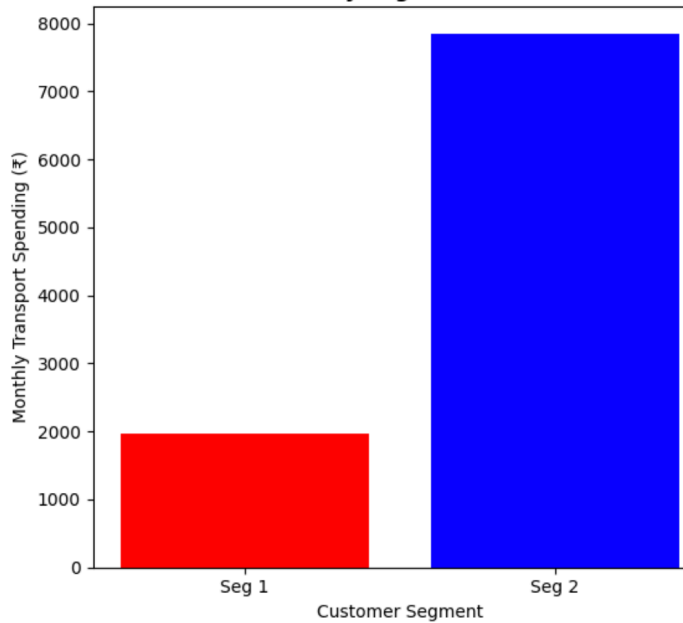2. **Optimal Cluster Selection**:
   - **Methods**: Elbow Method and Silhouette Score.
   - **Elbow Method**: Evaluated WCSS across a range of counts (2 to 7) to find the "elbow" point, where additional clusters yield diminishing returns, suggesting an optimal number of clusters (2).
   - **Silhouette Score**: Measured cluster cohesion and separation, with higher scores indicating better-defined clusters. A score of 0.576 for two clusters validated this as the optimal choice.
   - **Purpose**: Ensured the number of clusters was data-driven, balancing interpretability and statistical robustness.

Process:

- Selected key features (e.g., Income, Disposable Income, Age, Dependents, City Tier, Transport) influencing EV adoption.
- Standardized features to eliminate scale differences, critical for K-means performance.
- Applied K-means clustering to group customers, with the final model using two clusters based on elbow and silhouette score analysis.
- Evaluated clusters using the silhouette score to confirm quality and separation
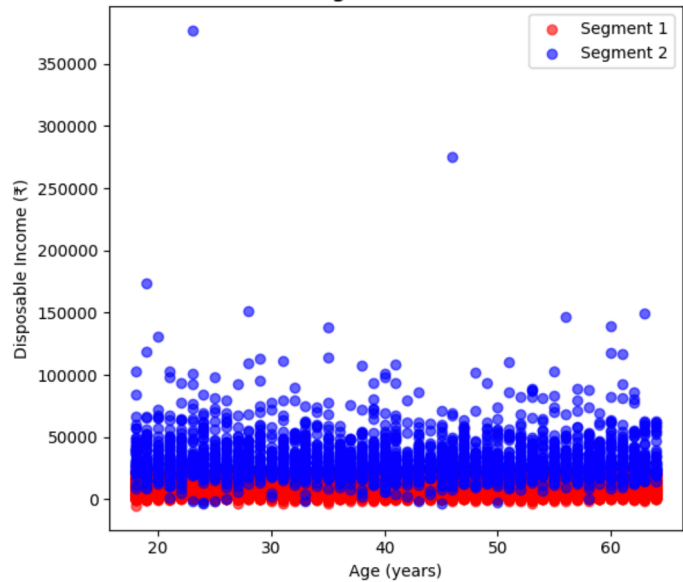
- Segment 2's high EV potential score (90.4/100) reflects its disposable income (₹32,767), enabling faster adoption compared to Segment 1's lower score (56.5/100) and income (₹7,493).

1. **What Type of EV Should the Startup Focus On?**
   Answer: Electric two-wheelers (scooters), supported by their market dominance and suitability for urban mobility.
2. **What Are the Age, Income, and Geographic Characteristics of the Target Market?**
   Answer:
   - *Age:* 25-35 (Segment 2).
   - *Income:* ₹32,767 disposable income.
   - *Geography:* Tier-2 cities (e.g., Pune, Ahmedabad).

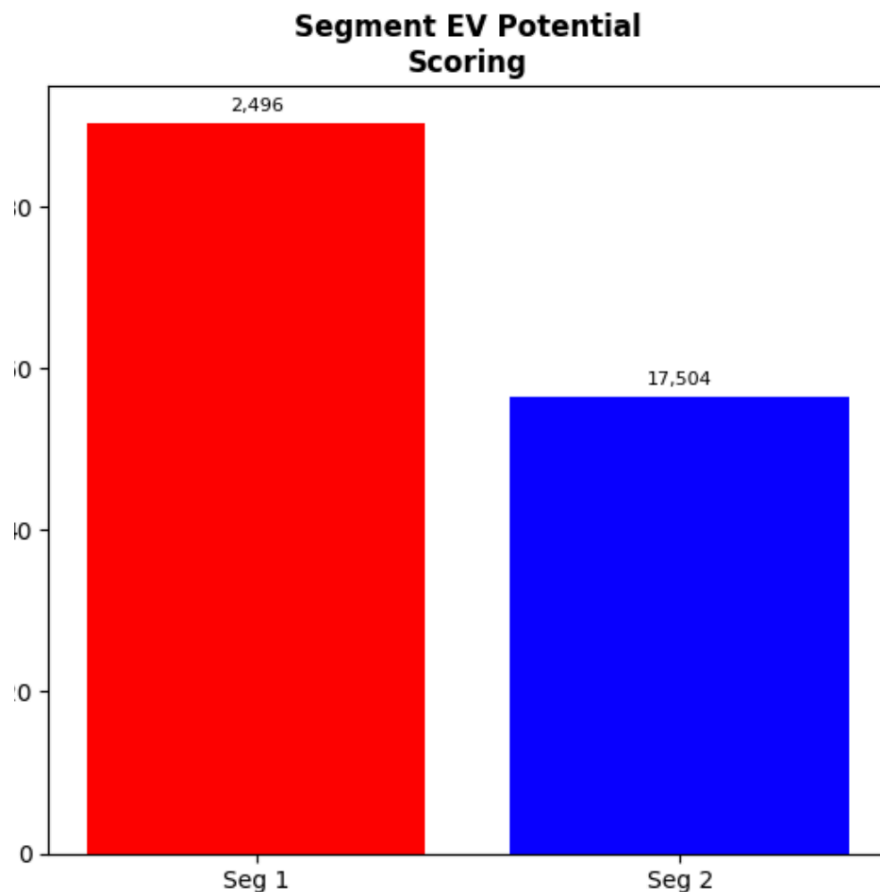# Target Market Selection

Primary Target: Segment 2

- **Why:** High income, youthful demographic, and urban location align with early EV adoption.

- **Size:** 2,496 customers. (From age_income.csv)
- **EV Potential:** 90.4/100.

Secondary Target: Segment 1

- **Why:** Larger market (17,504 customers (From age_income.csv)) but lower income requires affordable options.

- **EV Potential:** 56.5/100.

Strategic Recommendations

Product

- Launch electric scooters optimized for city commutes.

Pricing

- Range: around ₹80,000 to balance affordability and quality.

Distribution

- Start in Tier-2 cities, leveraging existing two-wheeler popularity.

Promotion

- **Segment 2:** Social media campaigns emphasizing innovation.
- **Segment 1:** Local ads focusing on cost savings.

Conclusion

This analysis positions the startup for success in India's EV market.

1. **What Type of EV Should the Startup Focus On?**
   - **Electric two-wheelers (scooters)**—the clear choice given their market share and urban relevance.
2. **What Are the Age, Income, and Geographic Characteristics of the Target Market?**
   - **Age:** 25-35 years.
   - **Income:** ₹32,767 disposable income.
   - **Geography:** Tier-2 cities like Pune and Ahmedabad.

Action Plan: Launch scooters targeting young professionals, scale with affordable options for families, and monitor regional trends.`