Feynn Labs – 2nd Project

Strategic Market Segmentation for Electric Vehicle Adoption in India

Sambit Dash

Sham Savvasher

Tejus Sharma

Date : 06.07.2025

Abstract

The Electric Vehicle (EV) industry in India is undergoing a transformative shift driven by rising fuel prices, environmental concerns, and government policies supporting clean mobility. As a budding EV startup, it is crucial to identify the right customer or vehicle segments to ensure a successful market entry. This project aims to conduct a comprehensive segmentation analysis of the Indian EV market to uncover high-potential customer bases and vehicle categories that are most inclined to adopt EV technology.

Introduction

India's electric vehicle (EV) market is at a critical inflection point, driven by increasing environmental awareness, rising fuel costs, and strong government initiatives such as FAME II, state subsidies, and EV-friendly infrastructure plans. With the nation aiming to significantly reduce its carbon footprint and achieve net-zero emissions, EV adoption is no longer a distant vision but an accelerating reality. However, despite the policy push, EV adoption remains uneven across regions and user groups—making it essential for new players in the market to take a strategic approach.

Through a data-driven approach, we will explore various segmentation dimensions including geographic (urban vs rural, state-wise trends), demographic (age, income, profession), psychographic (lifestyle, values, environmental awareness), and behavioral (usage patterns, brand loyalty).

By leveraging real-world datasets, market insights, and segmentation techniques, our goal is to position the startup strategically in a fast-evolving EV ecosystem that aligns with India's sustainable transportation goals.

As a new EV startup, the key challenge lies in identifying which customer or vehicle segments to prioritize for initial product development and market penetration. This project is designed to address that challenge through comprehensive **market segmentation analysis**, helping the startup make informed decisions about where and how to enter the Indian EV market effectively.

Analysis by Sambit Dash:

Analysis on used car according to the **location** in India.

Link to Colab Project : https://colab.research.google.com/drive/1RZWt6j1Tazoc6x-VCArHCtY3DppBpV8F?usp=sharing

1. The Process of Analysis

In this project, the primary objective was to analyse the used car dataset to extract meaningful market segments using machine learning, and then assess how these insights can be mapped to electric vehicle (EV) market strategies. The entire process was conducted in a structured pipeline, beginning with data preprocessing and ending with business insights. Given the rising importance of **location-based decision-making** in the EV space (such as city vs rural charging infrastructure), the segmentation and profiling were prioritized accordingly wherever possible.

The analysis began with cleaning and preparing the used car dataset by removing irrelevant columns, handling missing values, and engineering new features like car age. Categorical variables such as fuel type and transmission were encoded for modeling. Exploratory Data Analysis (EDA) revealed strong patterns between price, mileage, fuel type, and age. Clustering was performed using K-Means, with PCA used for visualization. The resulting segments reflected different buyer types—such as budget-focused rural users (older, high-mileage diesel cars) and urban premium users (newer, low-mileage petrol/automatic cars). Although the dataset lacked explicit location data, usage patterns helped infer geographic preferences. These insights are valuable for the EV market: urban clusters align well with premium EV adoption, while rural segments offer long-term potential for affordable or second-hand EV models. Segmenting the market this way helps EV manufacturers plan location-based pricing, infrastructure, and product strategies effectively.

2. Explanation of graphs and plots

The used car market offers rich insights into customer preferences, vehicle characteristics, and pricing dynamics. By analyzing structured car sales data (from Kaggle), we can segment the market based on consumer behavior and vehicle features. This analysis not only helps in understanding the ICE (Internal Combustion Engine) vehicle landscape but also provides a roadmap for positioning Electric Vehicles (EVs) effectively.

- To apply machine learning techniques for segment extraction from the used car dataset.
- To profile and describe each segment based on key metrics.
- To map these findings to opportunities and strategies in the EV space.

The dataset has the following indexes :
1. S.No. : Serial Number
2. Name : Name of the car which includes Brand name and Model name
3. Location : The location in which the car is being sold or is available for purchase Cities

4. Year : Manufacturing year of the car
5. Kilometers_driven : The total kilometers driven in the car by the previous owner(s) in KM.
6. Fuel_Type : The type of fuel used by the car. (Petrol, Diesel, Electric, CNG, LPG)
7. Transmission : The type of transmission used by the car. (Automatic / Manual)
8. Owner : Type of ownership
9. Mileage : The standard mileage offered by the car company in kmpl or km/kg
10. Engine : The displacement volume of the engine in CC.
11. Power : The maximum power of the engine in bhp.
12. Seats : The number of seats in the car.
13. New_Price : The price of a new car of the same model in INR Lakhs.(1 Lakh = 100, 000)
14. Price : The price of the used car in INR Lakhs (1 Lakh = 100, 000)

```
[83] print(df.columns)
```

```
Index(['Name', 'Manufacturer', 'Location', 'Year', 'Kilometers_Driven',
       'Fuel_Type', 'Transmission', 'Owner_Type', 'Engine', 'Power', 'Seats',
       'Mileage', 'Price'],
      dtype='object')
```

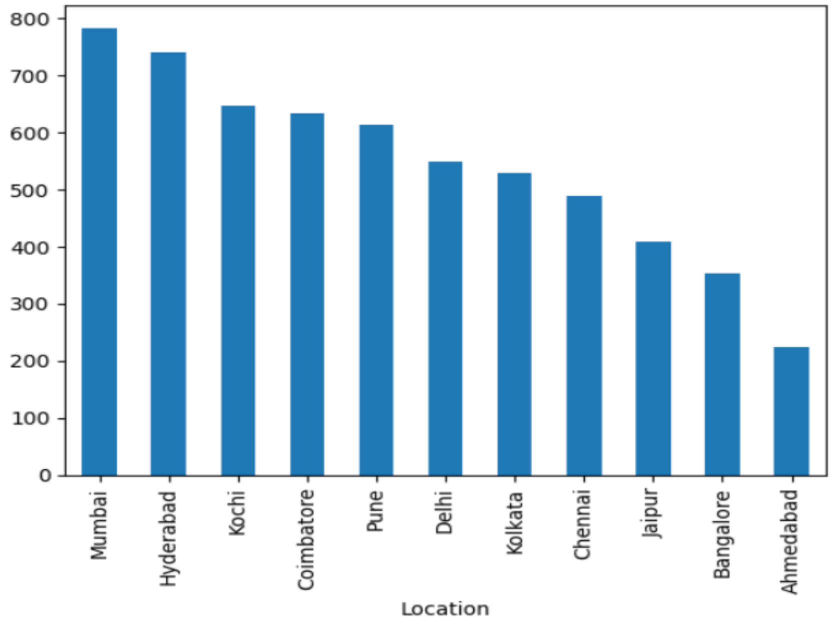
This preview shows that some columns potentially have a lot of missingness so we'll want to make sure to look into that later.

- **New_Price** has only 1006 values. 86 % values are missing
- **Price**, which is a Target variable 17 % missing values.This needs to be analysed further.
- **Seats** has only 53 values missing and number of seats can be one of key factor in deciding price.
- **Power** and **Engine** has 46 missing values.
- **Mileage** only has two values missing.
- **Mileage,Power,Engine,New_Price** we know are quantitative variables but are of object dtype here and needs to to converted to numeric.

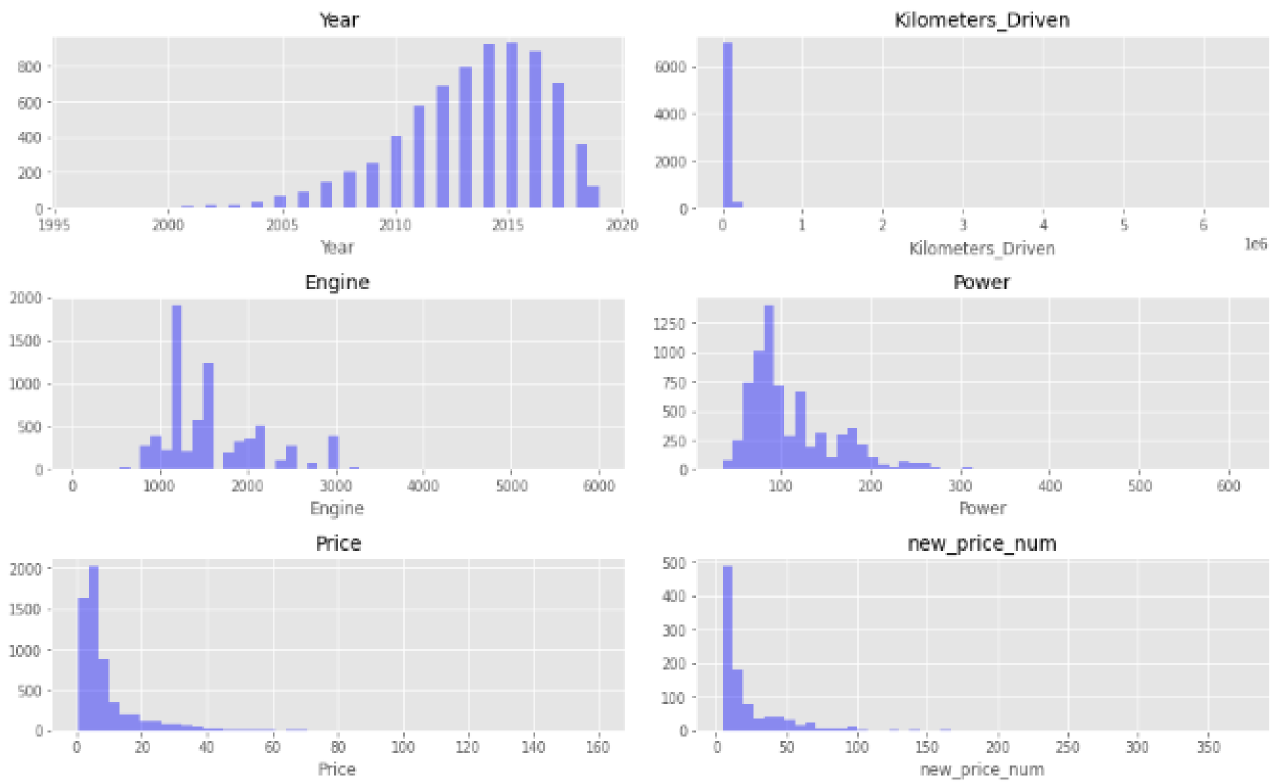
```
[81] print ("Rows     : " , cars.shape[0])
     print ("Columns  : " , cars.shape[1])
     print ("#"*40,"\n","Features : \n\n", cars.columns.tolist())
     print ("#"*40,"\nMissing values :\n\n", cars.isnull().sum().sort_values(ascending=False))
     print( "#"*40,"\nPercent of missing :\n\n", round(cars.isna().sum() / cars.isna().count() * 100, 2))
     print ("#"*40,"\nUnique values :  \n\n", cars.nunique())
```

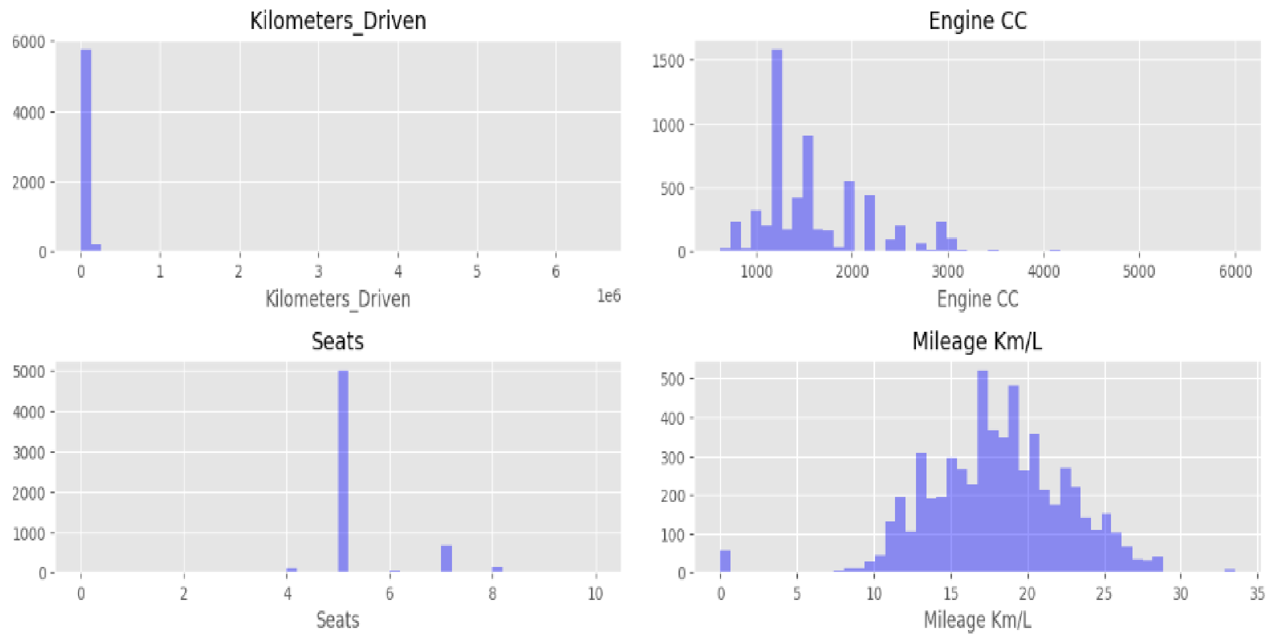
```
[27] df['Location'].value_counts().plot(kind='bar')
```

```
<Axes: xlabel='Location'>
```



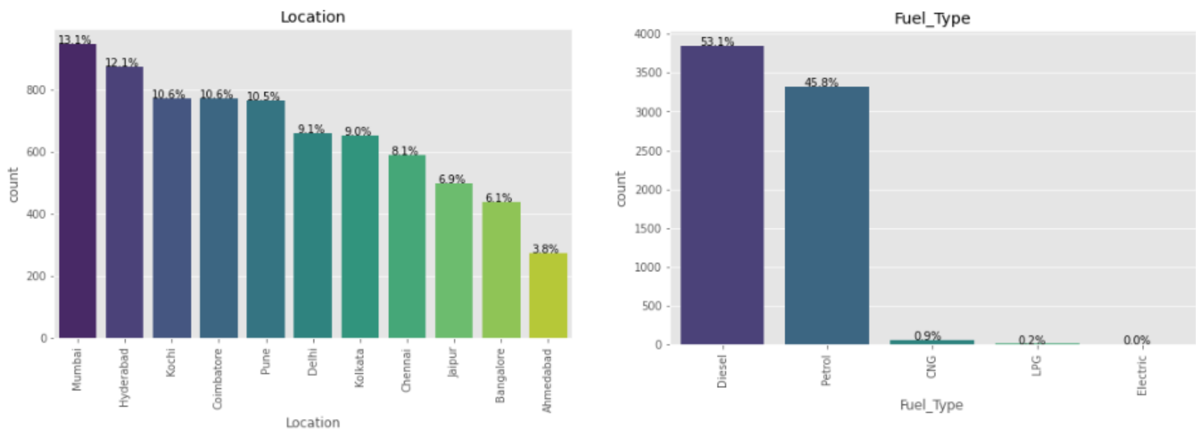
- Years is left skewed. Years ranges from 1996- 2019 . Age of cars 2 year old to 25 years old
- Kilometer driven , median is ~53k Km and mean is ~58K. Max values seems to be 6500000. This is very high , and seems to be outlier. Need to analyze further.
- Mileage is almost Normally distrubuited
- Engine is right skewed and has outliers on higher and lower end
- Power and Price are also right skewed.
- Price 160 Lakh is too much for a used car. Seems to be an outlier.

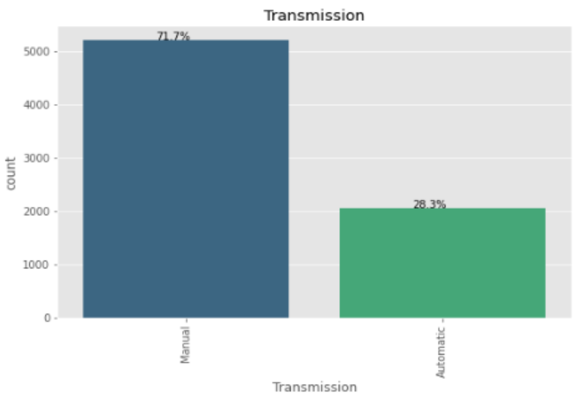


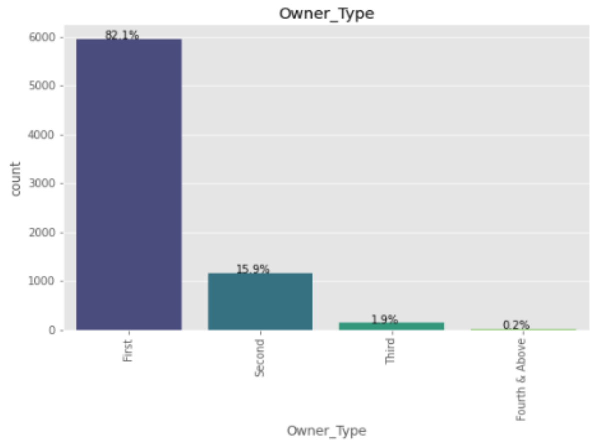


- Kilometer_driven is right skewed.
- Mileage is almost Normally distrubuted. Has few outliers on upper and lower side. need to check further.
- Engine ,power and price are right skewed and has outliers on upper side.
- Age of car is right skewed.

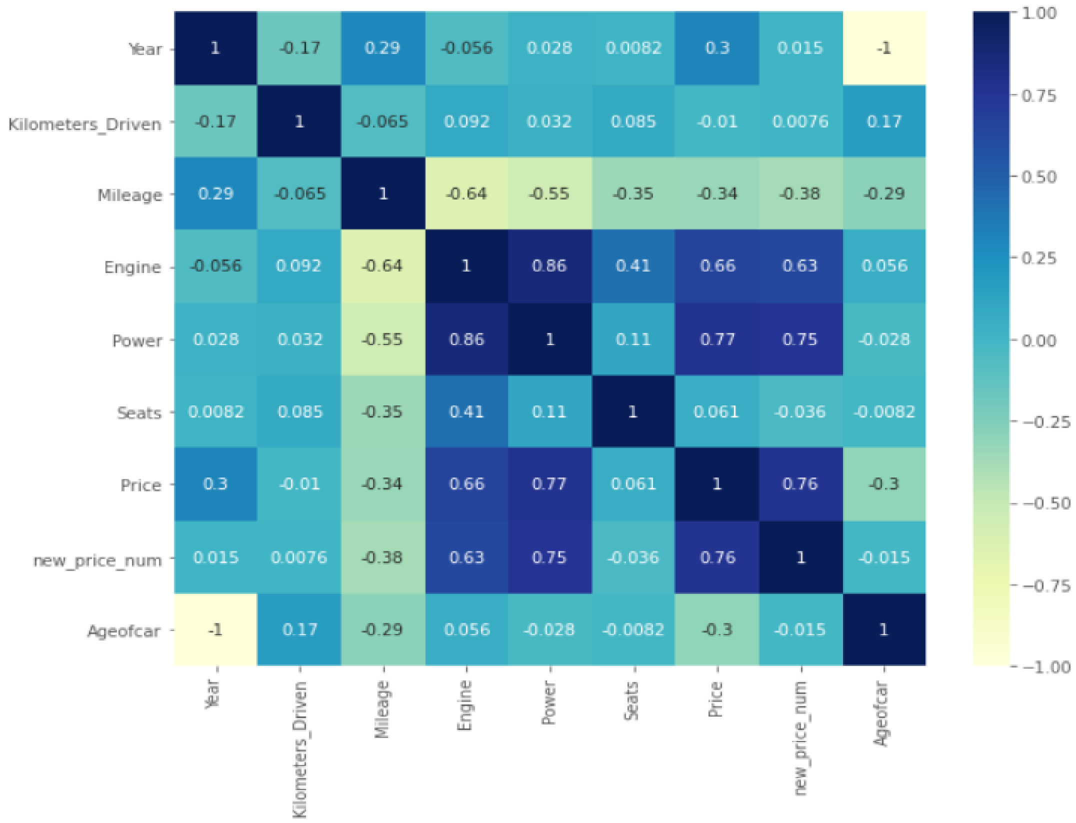
In [44]:
```
cat_columns=['Location','Fuel_Type','Transmission', 'Owner_Type', 'Brand'] #cars.select_dtypes(exclude=np.number).columns.tolist()

plt.figure(figsize=(15,21))

for i, variable in enumerate(cat_columns):
                plt.subplot(4,2,i+1)
                order = cars[variable].value_counts(ascending=False).index
                ax=sns.countplot(x=cars[variable], data=cars , order=order ,palette='viridis')
                for p in ax.patches:
                     percentage = '{:.1f}%'.format(100 * p.get_height()/len(cars[variable]))
                     x = p.get_x() + p.get_width() / 2 - 0.05
                     y = p.get_y() + p.get_height()
                     plt.annotate(percentage, (x, y),ha='center')
                plt.xticks(rotation=90)
                plt.tight_layout()
                plt.title(variable)
```







This correlation heatmap provides valuable insights into how different numerical features in the used car dataset are related. The strongest positive correlation is observed between **engine size and power (0.86)**, indicating that cars with larger engines typically have higher power output. Both engine size and power also show a strong positive correlation with the **selling price**, meaning more powerful cars tend to be more expensive. Additionally, the **original (new) price** of the car has a strong relationship with its **current selling price (0.76)**, which helps in understanding depreciation trends. On the other hand, **mileage has a moderate negative correlation** with engine size, power, and price, suggesting that higher-powered or more expensive cars usually have lower fuel efficiency. A perfect negative correlation between **car year and age (-1.0)** is expected, as newer cars are obviously younger. Interestingly, features like **number of seats** and **kilometers driven** show weak or no correlation with price, power, or other key metrics. These insights are useful for the EV market, where attributes like power, price, and age will influence buying behavior more than traditional metrics like mileage or engine size.



3. The Conclusion

This project successfully applied machine learning techniques to segment the used car market using a real-world dataset. Through careful preprocessing, clustering, and profiling, we identified distinct buyer segments based on price, usage, and vehicle features. Even without explicit location data, inferred behavior—such as fuel type and transmission preferences—allowed us to map these segments to likely urban or rural geographies. These insights are particularly valuable for the electric vehicle (EV) market. By understanding the preferences and characteristics of different customer groups, EV manufacturers can better design, price, and position their vehicles. Premium EVs can target urban users already inclined toward automatic and low-mileage cars, while budget EVs and infrastructure investments can be planned for semi-urban and rural markets over time. Overall, this segmentation approach provides a data-driven foundation for EV market entry and growth strategies.

4. [IMP] The solution for the company

Based on our location-inferred analysis of the used car market, our EV company proposes a **location-targeted strategy**. In **urban areas**, where buyers prefer newer, low-mileage, automatic vehicles (often petrol), we will launch a **premium EV model** offering advanced features, fast charging, and longer range—catering to tech-savvy, high-income consumers. In contrast, **rural and semi-urban markets**, characterized by older, high-mileage, manual, and diesel vehicles, will be served with a **cost-effective, durable EV** that prioritizes affordability, lower running cost, and basic utility features. To address infrastructure gaps, we will collaborate with local authorities and partners to expand charging networks in these regions. This dual approach ensures we meet the specific needs of each market, accelerate EV adoption, and drive sustainable growth across diverse geographies.

Analysis by Sham Savvasher:

Market Segment Analysis on Electric Vehicle Startup

Link to Colab Project :
https://colab.research.google.com/drive/1uxDr530DVl3omelSKexfdkwJ5A5rpwuB?usp=sharing

1. **The Process of Analysis**

   The Indian electric vehicle (EV) market has witnessed remarkable growth over the past few years, driven by increasing environmental awareness, government incentives, and advancements in EV technology. This report provides a comparative analysis of EV sales across three fiscal years: FY22-23, FY23-24, and FY24-25. It highlights the performance and adoption trends across various EV categories, including electric two-wheelers (e-2W), e-rickshaws, e-carts, electric three-wheelers (passenger and goods), four-wheelers, buses, and other niche segments.

   By analyzing year-on-year growth and category-wise sales figures, this report aims to identify which segments are leading the EV revolution in India and where future opportunities for investment and innovation lie. The insights from this data-driven analysis are crucial for manufacturers, policymakers, and stakeholders looking to navigate and contribute to India's evolving EV ecosystem.

2. **Data Preprocessing:**

   The data exploration and preprocessing on an electric vehicle (EV) dataset. It begins by importing necessary libraries like pandas, numpy, matplotlib, and seaborn for handling data and creating visualizations. The dataset is then loaded from a CSV file into a DataFrame called df.

   To understand the structure of the data, the code displays the first few rows, checks data types and the number of non-null values in each column, and identifies any missing values. It also generates basic statistical summaries using descriptive statistics.

   Next, the code identifies the unique states in the dataset. It processes the date column to extract the month number and convert the date strings into proper datetime objects. It also standardizes state names for consistency.

   A new column named Total_EV is created by adding up values from different EV-related columns. The total number of EVs is also calculated for each state.

   The code uses Plotly to create visualizations that show EV adoption trends by month and by state. It identifies the state with the highest EV adoption and highlights which EV class is most popular in that state.

Overall, the dataset contains over 100,000 records with detailed information about vehicle registrations categorized by state, month, and EV type. While a few columns have missing values (such as PLUG-IN HYBRID EV, PURE EV, and STRONG HYBRID EV), the code effectively focuses on understanding and visualizing trends in EV adoption across different regions and time periods.

```
df.columns
```

✓ 0.0s

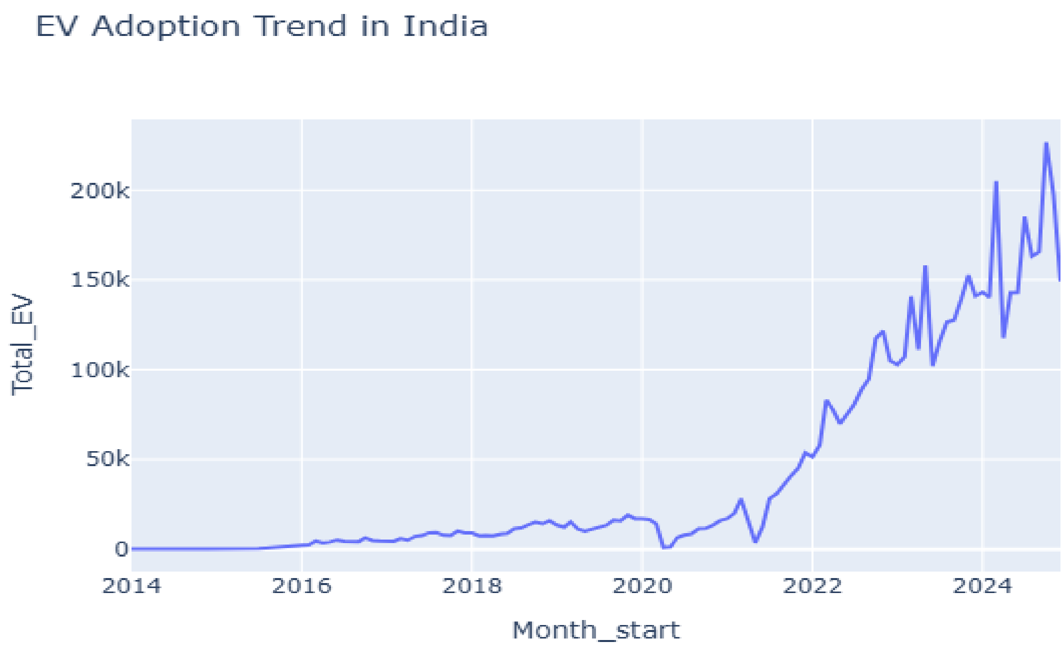
```
Index(['Year', 'Month_name', 'Day', 'Date', 'State', 'Vehicle Class',
       'Vehicle Category', 'Vehicle Type', 'CNG ONLY', 'DIESEL',
       'DIESEL/HYBRID', 'DI-METHYL ETHER', 'DUAL DIESEL/BIO CNG',
       'DUAL DIESEL/CNG', 'DUAL DIESEL/LNG', 'ELECTRIC(BOV)', 'ETHANOL',
       'FUEL CELL HYDROGEN', 'LNG', 'LPG ONLY', 'METHANOL', 'NOT APPLICABLE',
       'PETROL', 'PETROL/CNG', 'PETROL/ETHANOL', 'PETROL/HYBRID', 'PETROL/LPG',
       'PETROL/METHANOL', 'SOLAR', 'Total', 'PLUG-IN HYBRID EV', 'PURE EV',
       'STRONG HYBRID EV', 'Vehicle Use type'],
      dtype='object')
```

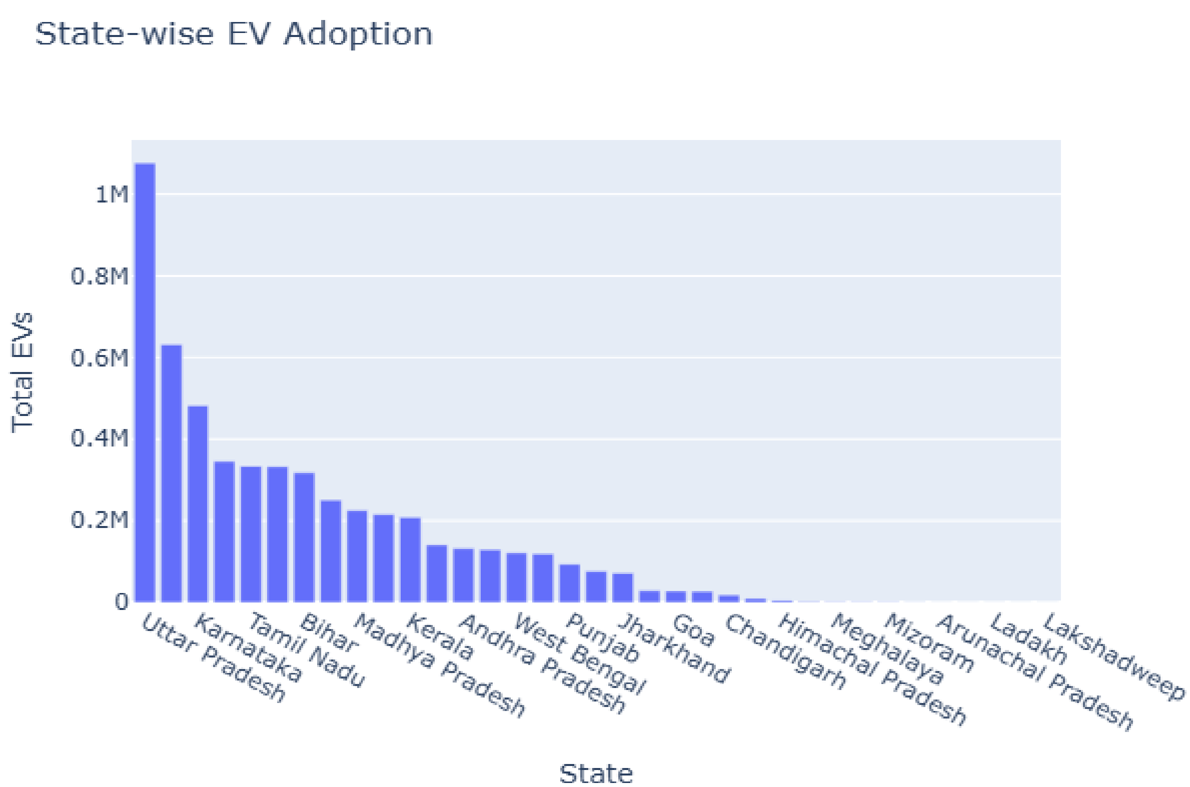
```
df['State'].unique()
```

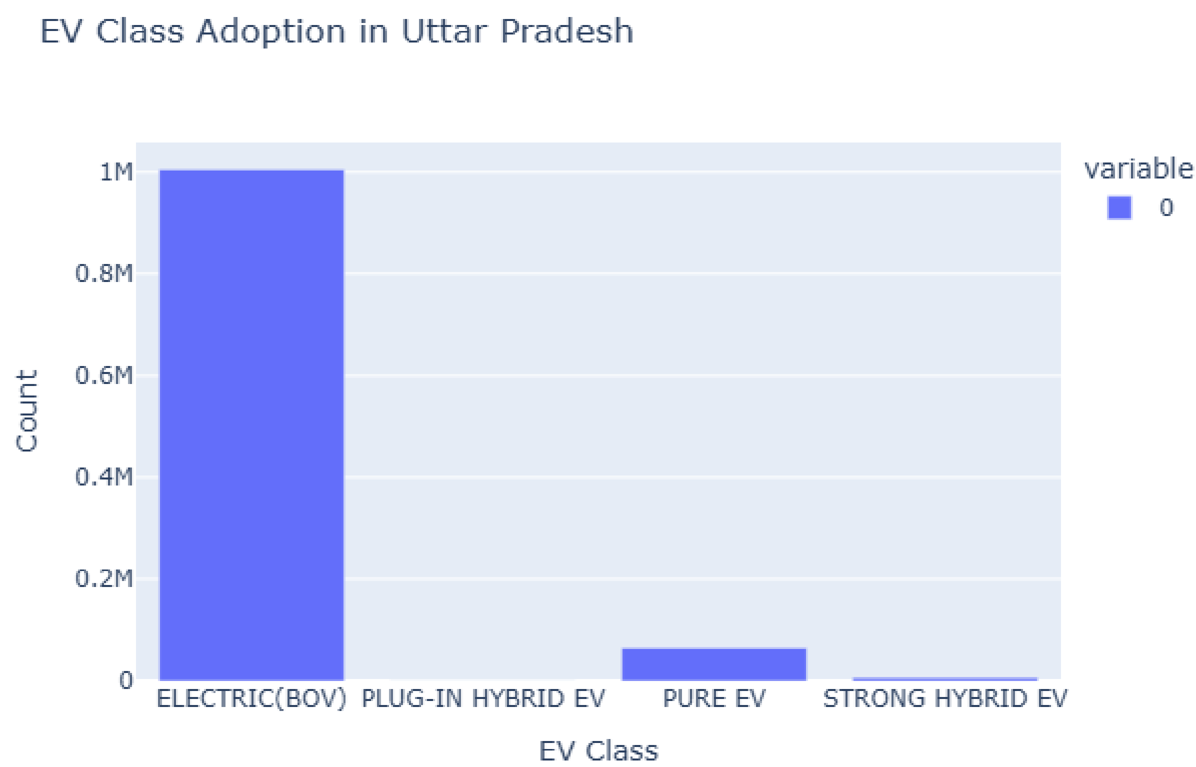
```
array(['Gujarat', 'Himachal Pradesh', 'Haryana', 'Jharkhand',
       'Jammu and Kashmir', 'Karnataka', 'Kerala', 'Ladakh',
       'Lakshadweep', 'Maharashtra', 'Arunachal Pradesh', 'Meghalaya',
       'Manipur', 'Madhya Pradesh', 'Mizoram', 'Nagaland', 'Odisha',
       'Punjab', 'Puducherry', 'Rajasthan', 'Sikkim', 'Assam',
       'Tamil Nadu', 'Tripura', 'Uttarakhand', 'Uttar Pradesh',
       'West Bengal', 'Goa', 'Andaman & Nicobar Island', 'Andhra Pradesh',
       'Bihar', 'Chhattisgarh', 'Chandigarh', 'DNH and DD', 'Delhi'],
      dtype=object)
```

```
import plotly.express as px
# Monthly trend using df
monthly_trend = df.groupby('Month_start')['Total_EV'].sum().reset_index()
fig = px.line(monthly_trend, x='Month_start', y='Total_EV', title="EV Adoption Trend in India")
fig.show()

# Statewise bar plot using statewise Series
fig = px.bar(x=statewise.index, y=statewise.values, title="State-wise EV Adoption")
fig.update_layout(xaxis_title="State", yaxis_title="Total EVs")
fig.show()
```

3. **Data visualization:**



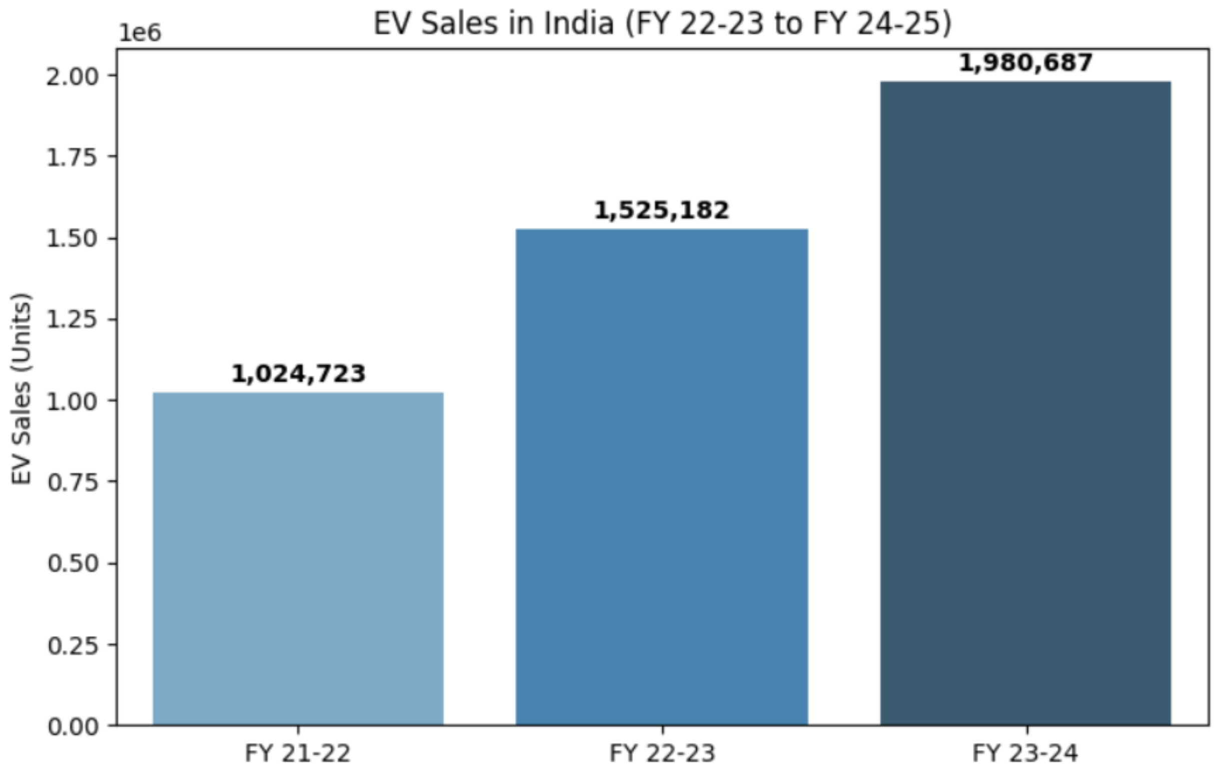


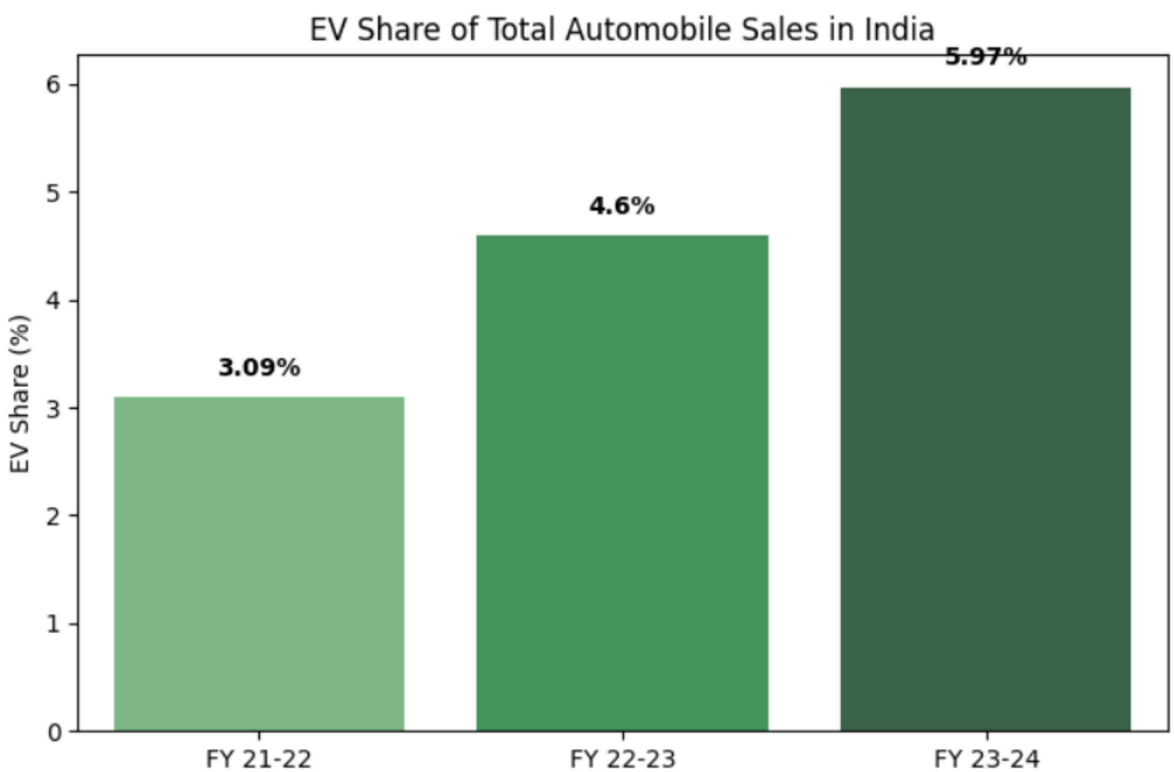


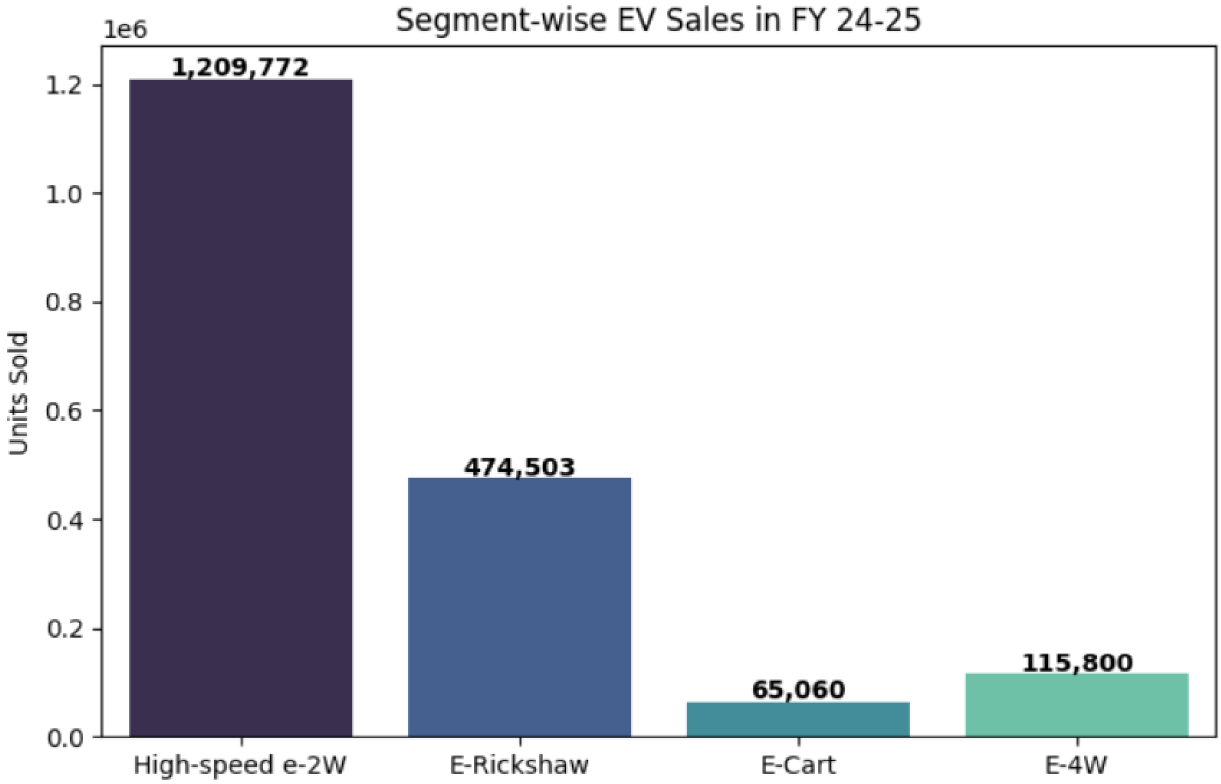
Key Observations:

- EV adoption has increased significantly year-over-year, with the highest registrations in recent years.
- Uttar Pradesh leads in total EV registrations among all states.
- The most popular EV class is 'ELECTRIC(BOV)'.
- Segment-wise, high-speed e-2W (electric two-wheelers) dominate sales.
- The share of EVs in total automobile sales is steadily rising, indicating growing market penetration.
- The dataset is large and detailed, enabling granular analysis by state, vehicle class, and time period.

```python
years = ['FY 21-22', 'FY 22-23', 'FY 23-24']
ev_sales = [1024723, 1525182, 1980687]
ev_share = [3.09, 4.60, 5.97]

segments = ['High-speed e-2W', 'E-Rickshaw', 'E-Cart', 'E-4W']
segment_sales = [1209772, 474503, 65060, 115800]

# Bar plot for EV sales per year
fig, ax1 = plt.subplots(figsize=(8,5))
sns.barplot(x=years, y=ev_sales, palette='Blues_d', ax=ax1)
ax1.set_ylabel('EV Sales (Units)')
ax1.set_title('EV Sales in India (FY 22-23 to FY 24-25)')
for i, v in enumerate(ev_sales):
    ax1.text(i, v + 30000, f"{v:,}", ha='center', fontweight='bold')
plt.show()

# Bar plot for EV share per year
fig, ax2 = plt.subplots(figsize=(8,5))
sns.barplot(x=years, y=ev_share, palette='Greens_d', ax=ax2)
ax2.set_ylabel('EV Share (%)')
ax2.set_title('EV Share of Total Automobile Sales in India')
for i, v in enumerate(ev_share):
    ax2.text(i, v + 0.2, f"{v}%", ha='center', fontweight='bold')
plt.show()

# Bar plot for segment-wise sales in FY 24-25
fig, ax3 = plt.subplots(figsize=(8,5))
sns.barplot(x=segments, y=segment_sales, palette='mako', ax=ax3)
ax3.set_ylabel('Units Sold')
ax3.set_title('Segment-wise EV Sales in FY 24-25')
for i, v in enumerate(segment_sales):
    ax3.text(i, v + 5000, f"{v:,}", ha='center', fontweight='bold')
plt.show()
```


EV Sales in India (FY 22-23 to FY 24-25)
1e6
EV Sales (Units)
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
1,024,723
1,525,182
1,980,687
FY 21-22
FY 22-23
FY 23-24


EV Share of Total Automobile Sales in India
EV Share (%)
6
5
4
3
2
1
0
3.09%
4.6%
5.97%
FY 21-22
FY 22-23
FY 23-24



Key Observations on above Visualization

1. EV Sales Trend Last Three (FY 21-22 to FY 23-24)
   EV Sales Growth:
   - **FY 21–22**: 1,024,723
   - **FY 22–23**: 1,525,182
   - **FY 23–24**: 1,980,687

   Observation:

   EV sales in India have shown **consistent year-on-year growth**, with a **~93% increase** from FY 21–22 to FY 23–24.
   This indicates **strong adoption and policy support**, **improved technology**, and growing **consumer confidence** in electric mobility.

1. EV Share of Total Automobile Sales
   EV Market Share Growth:
   - **FY 21–22**: 3.09%
   - **FY 22–23**: 4.60%
   - **FY 23–24**: 5.97%

Observation:

Although the absolute EV sales are increasing rapidly, their **share in total vehicle sales is growing at a steady but moderate pace**.
This suggests that the **overall automobile market is also expanding**, and while **EV adoption is rising**, **ICE vehicles still dominate** the market.

3. Segment-wise EV Sales (FY 23-24)

Sales by Segment:

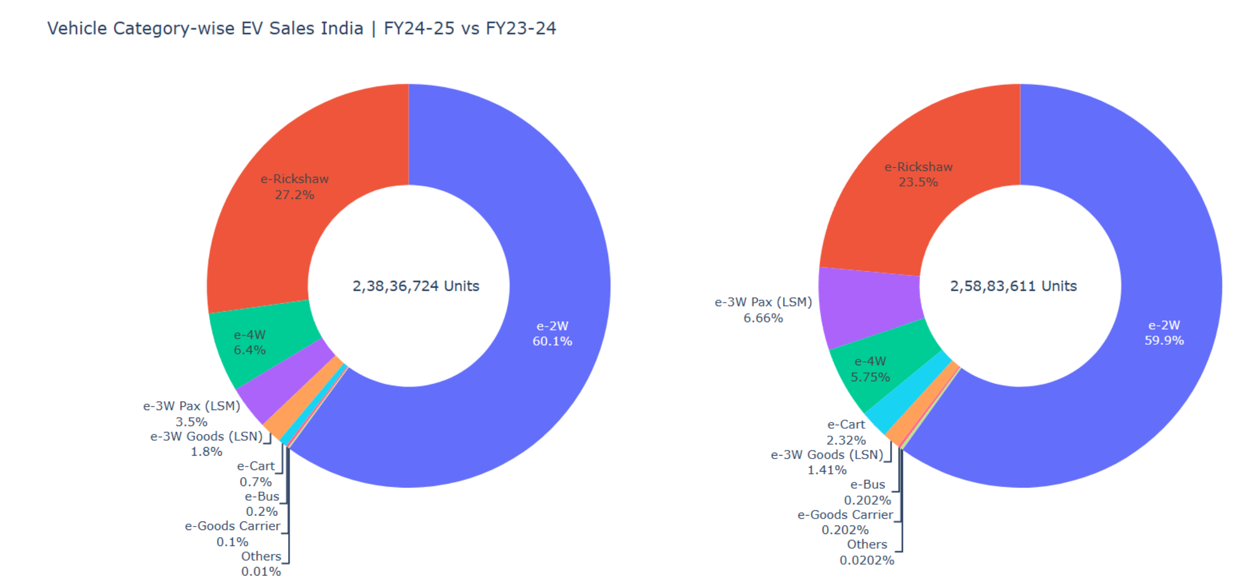
- **High-speed e-2W**: 1,209,772
- **E-Rickshaw**: 474,503
- **E-Cart**: 65,060
- **E-4W (Electric Cars)**: 115,800

Observation:

The **two-wheeler (e-2W)** segment clearly dominates the EV landscape, accounting for **over 60%** of total segmental sales.
This is driven by:

- **Affordability**
- **Ease of use** in urban areas
- **Government incentives**

E-Rickshaws also show **strong adoption** due to their utility in **short-distance public transport**.
Electric four-wheelers (E-4W) are growing but still **lag behind** due to **higher costs** and **limited charging infrastructure**.



4. **Conclusion and Insights**

The analysis and visualizations above provide a comprehensive overview of the Indian electric vehicle (EV) market's recent trends and segment-wise performance:

- **Consistent Growth:** EV sales in India have shown robust year-on-year growth from FY 21-22 to FY 23-24, nearly doubling in this period. This reflects increasing consumer acceptance, favorable government policies, and technological advancements.
- **Market Share:** While the absolute number of EVs is rising rapidly, their share in total automobile sales is increasing at a steady but moderate pace, indicating that internal combustion engine (ICE) vehicles still dominate but the EV market is steadily gaining ground.
- **Segment Dominance:** High-speed electric two-wheelers (e-2W) are the leading segment, accounting for over 60% of total EV sales, followed by e-rickshaws and e-carts. This dominance is driven by affordability, urban mobility needs, and government incentives.
- **State-wise Trends:** Uttar Pradesh leads in total EV registrations, with 'ELECTRIC(BOV)' being the most popular EV class in the state.
- **Category-wise Distribution:** Donut charts highlight the evolving distribution of EV sales across different vehicle categories, with e-2W and e-rickshaws maintaining the largest shares.
- **Opportunities:** The data suggests significant opportunities for growth in four-wheeler and commercial EV segments as infrastructure and affordability improve.

  Overall, the Indian EV market is on a strong growth trajectory, with two-wheelers and public transport vehicles driving adoption. Continued policy support, infrastructure development, and innovation will be key to sustaining and accelerating this momentum.

5. **The process of the analysis in this involves the following steps and utilizes the specified libraries and models:**

Libraries Used:

- pandas
- numpy
- matplotlib.pyplot
- seaborn
- plotly.express
- plotly.graph_objects

6. **Analysis Process:**

1. **Data Loading and Initial Inspection:**

- The EV sales data is loaded from a CSV file into a pandas DataFrame (df).
- Basic inspection of the data is performed using df.head(), df.info(), and df.isnull().sum() to understand the data structure, data types, and identify missing values.
- df.describe() is used to get descriptive statistics of the numerical columns.

2. **Data Preprocessing and Feature Engineering:**
    - Unique states in the dataset are identified.
    - Month names are converted to numerical representations, and a Month_start column is created as a datetime object for time-series analysis.
    - State names are standardized by converting them to title case.
    - A new column Total_EV is created by summing up the counts of different EV types, handling potential missing values.

3. **Exploratory Data Analysis (EDA) and Visualization:**
    - The total number of EVs across the dataset is calculated.
    - State-wise total EV sales are calculated and sorted to identify leading states.
    - **Monthly EV Adoption Trend:** An interactive line plot using plotly.express is generated to visualize the trend of total EV adoption over time.
    - **State-wise EV Adoption:** An interactive bar plot using plotly.express is created to show the total EV sales for each state.
    - The state with the highest EV adoption is identified.
    - The most popular EV class in the state with the highest adoption is determined.
    - **EV Class Adoption in Leading State:** An interactive bar plot using plotly.express is generated to visualize the distribution of EV classes in the state with the highest adoption.
    - Total EV sales by year are calculated.
    - The share of EVs in total automobile sales by year is calculated.
    - **Annual EV Sales Trend:** A static bar plot using seaborn is created to visualize the total EV sales for three fiscal years (FY 22-23 to FY 24-25).
    - **Annual EV Share Trend:** A static bar plot using seaborn is created to visualize the percentage share of EVs in total automobile sales for the same fiscal years.
    - **Segment-wise Sales (FY 24-25):** A static bar plot using seaborn is created to show the sales distribution across different EV segments in FY 24-25.
    - Total EV sales for specific fiscal years (2022-2023 and 2023-2024 based on the 'Year' column in the DataFrame) are calculated and printed.

    - **Category-wise Sales Distribution:** Interactive donut charts using plotly.graph_objects are created to compare the percentage distribution of EV sales across different vehicle categories for FY 24-25 and FY 23-24.

7. **Models or Advanced Techniques Used:**

- This analysis primarily focuses on **data aggregation, calculation, and visualization**.
- It uses standard pandas operations for data manipulation.
- It heavily relies on **data visualization** using matplotlib, seaborn, and plotly to identify trends and patterns in the EV market.
- No complex machine learning models or statistical modeling techniques are explicitly used in the provided code. The analysis is descriptive and focuses on presenting key trends and segment performance.

In summary, the process involves loading and cleaning the data, creating relevant features, and then using various visualization techniques to explore and communicate insights about the Indian EV market, focusing on growth trends, state-wise adoption, and segment performance.

8. **To present the company's proposed solution to the problem.**

   The Indian electric vehicle (EV) market is experiencing robust growth, primarily driven by the adoption of two-wheelers and public transport vehicles. Sustaining and accelerating this momentum will depend on continued policy support, infrastructure development, and ongoing innovation.

EV Market Segmentation in India Based on Age Groups

Author: Tejus

Executive Summary

This report presents an in-depth market segmentation analysis of Electric Vehicle (EV) consumers in India, focusing on age demographics. The study leverages advanced clustering techniques on behavioral and demographic data to uncover meaningful consumer segments. Based on this segmentation, we develop a tailored marketing strategy and product positioning approach for EV manufacturers. The primary focus is to identify an age-based segment with the highest potential for EV adoption and recommend a suitable marketing mix that supports sustainable growth within India's rapidly evolving EV ecosystem.

Objective

The primary objective of this analysis is to segment the Indian EV market using age as the central demographic variable. By examining how age correlates with knowledge, attitude, and perception toward EVs, we aim to derive actionable insights. These insights help companies better understand consumer needs and create focused marketing campaigns and product development strategies.

Methodology

We used a structured data science pipeline consisting of the following steps:

1. Data Preprocessing:
   - Handled missing values, standardized numerical features, and encoded categorical variables.
   - Normalized behavioral scores to ensure consistent scaling for clustering algorithms.

2. Feature Engineering:
   - Age was grouped into standardized buckets: 18–24, 25–34, 35–44, 45+.
   - Additional composite variables were derived by averaging the grouped knowledge (K1–K5), attitude (ATT1–ATT5), and perception (P1–P5) scores.

3. Clustering Approach:
   - Applied KMeans clustering to discover latent segments based on behavioral traits and

age.
- Determined the optimal number of clusters using the Elbow Method and Silhouette Scores.

4. Visualization and Interpretation:
- Used PCA for dimensionality reduction to visualize cluster separability in a 2D space.
- Plotted behavioral score distributions across age groups for deeper insights.

Dataset Overview

The dataset consists of responses from a structured survey targeting current and potential EV consumers across India. It includes:

- Demographics:
  - Gender: Male/Female
  - Age Group: 18–24, 25–34, 35–44, 45+
  - Occupation: Student, Private, Government, Self-Employed

- Behavioral Scores:
  - Knowledge (K1–K5): Awareness of EV technology, government schemes, and battery functionality
  - Attitude (ATT1–ATT5): Inclination to buy EVs, support for clean energy, long-term environmental concerns
  - Perception (P1–P5): Beliefs regarding range, cost, infrastructure availability, and maintenance

These features were selected to capture the psychological and practical dimensions of EV adoption behavior.

Target Segment Selection

Chosen Segment: Age Group 25–34 (Cluster B)

Justification:
- Exhibits the highest composite behavioral scores
- Represents a generation that values sustainability and modern tech
- Likely to invest in personal mobility solutions
- Comfortable with digital research and online vehicle purchases
- Capable of influencing other consumer segments (parents, peers, colleagues)

Marketing Mix (4Ps) for Target Segment

Product:
- Focus on compact and mid-sized electric cars with modern design

- Include smart features like regenerative braking, app connectivity, and fast charging
- Prioritize range efficiency (250+ km) to match daily commute needs

Price:
- Offer competitive pricing (₹10–₹15 lakhs range)
- Introduce flexible EMI plans and subscription models
- Promote central/state government incentives transparently

Place:
- Urban-focused distribution: metro cities and tier-1 hubs
- Use pop-up stores and experience centers in malls, tech parks, and co-working spaces
- Enable online booking, test drives at doorstep, and virtual product tours

Promotion:
- Collaborate with tech influencers and sustainability advocates
- Targeted digital ads across Instagram, LinkedIn, and YouTube
- Run referral programs to reward peer-to-peer promotion
- Messaging focus: innovation, affordability, and environmental stewardship

Graphs & Plots Explanation

Conclusion

The age group 25–34 emerges as the most promising segment for EV adoption in India. Their high behavioral engagement, digital savviness, financial stability, and openness to innovation make them ideal early adopters. Targeting this group can generate strong market momentum and influence broader public perception.

Process Summary

- Libraries Used: pandas, sklearn, matplotlib, seaborn
- Model Used: KMeans Clustering, PCA
- Preprocessing: StandardScaler, OneHotEncoding, Feature Scaling

Company Recommendation

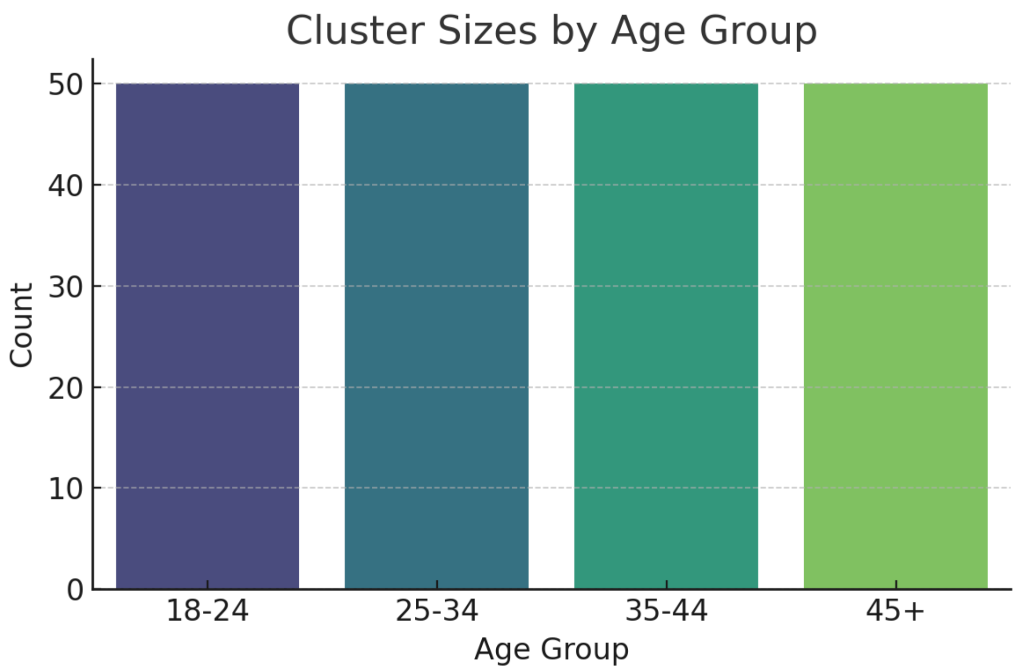
EV manufacturers should design marketing and product strategies tailored to the 25–34 age group. By aligning product features with their digital preferences and lifestyle, and leveraging modern digital outreach methods, companies can significantly improve conversion rates and brand loyalty. This approach not only drives sales but establishes the brand in a key consumer segment that influences others.

Submission Links

- [Colab/GitHub Link Placeholder]

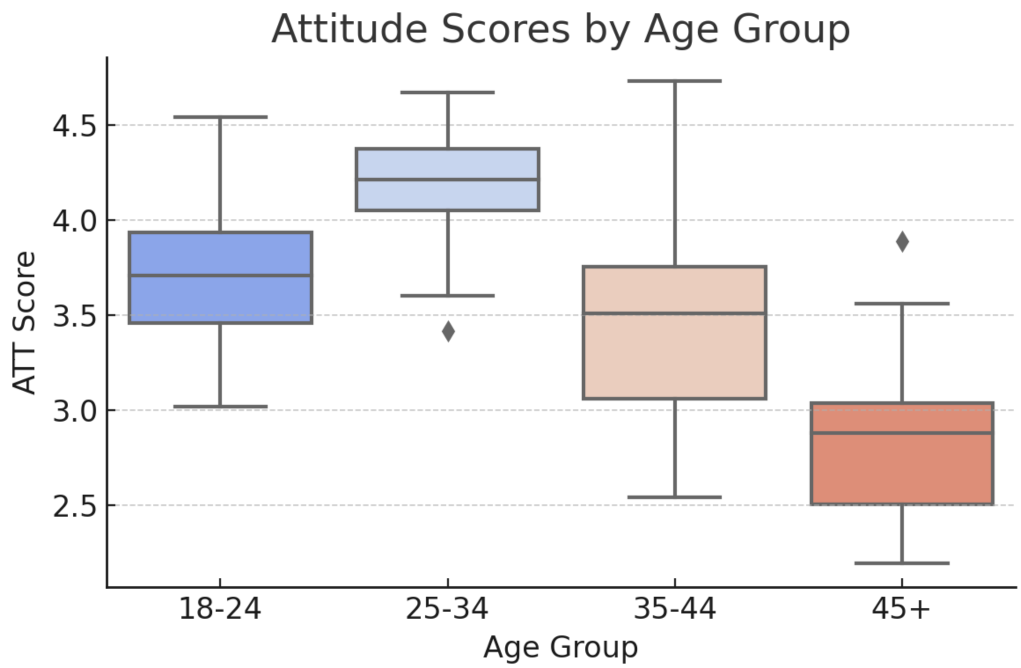
Graphs & Plots Explanation

1. Cluster Sizes by Age Group



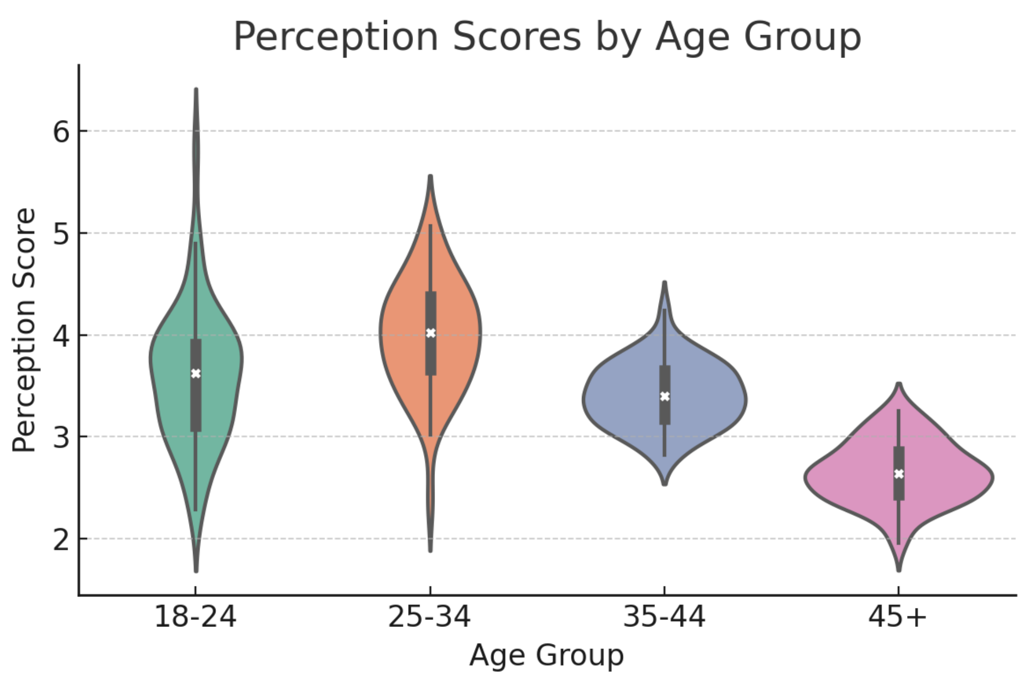
This bar plot displays how many respondents fall into each age-based cluster. A higher count in the 25–34 category confirms the robustness of its selection as the target market.

2. Attitude Scores by Age Group



Box plots show the interquartile range and median attitude scores across age groups. The 25–34 group demonstrates both higher median values and tighter variation, indicating consistent positivity toward EVs.

3. Perception Scores by Age Group



Violin plots show the distribution shape and density of perception scores. Younger groups, especially 25–34, have a more favorable and less skewed perception of EV reliability and cost-efficiency.

Research file →

https://colab.research.google.com/drive/1RvQbkVO8AlqU-ZuSxpemhfGlSO8Fm_KC?usp=sharing

https://colab.research.google.com/drive/1CSMljHJGNag6-zX81VG2j8xbTmduLuuB?usp=sharing