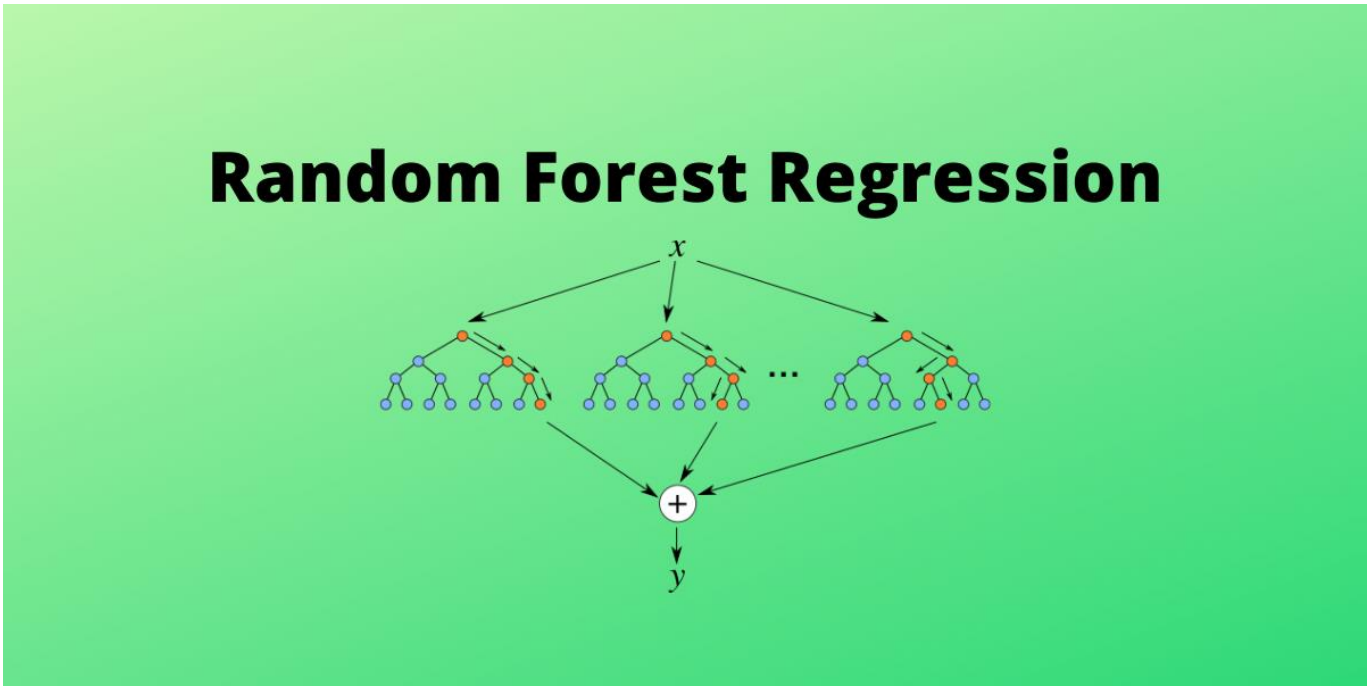
Electric Vehicle Sales Forecasting in India Using Random Forest Regression

Tarun C
8th sem, , ECE
NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY,
BANGALORE-560064



Electric Vehicle sales forecasting in India

Electric cars are not going to take the market by storm, but it's going to be a gradual improvement.

(Carlos Ghosn)

Abstract

Electric vehicles (EVs) are gaining significant momentum in India as the country pivots toward sustainable transportation solutions. With increasing government support, environmental concerns, and rising fuel prices, EV adoption has surged across various states. This study focuses on analyzing and forecasting EV sales trends in India using machine learning techniques, particularly Random Forest Regression.

The dataset, comprising state-wise monthly EV sales over multiple years, was preprocessed to handle missing values, encode categorical variables, and engineer relevant features. Exploratory Data Analysis (EDA) was conducted to uncover key insights into sales behavior across states, vehicle types, and categories. A Random Forest Regressor was then trained to predict future EV sales, providing a robust model with strong predictive performance.

The findings highlight critical variables influencing EV sales—such as year, month, vehicle type, and state—offering valuable direction for policymakers, manufacturers, and stakeholders aiming to understand and shape the evolving EV market in India.

KEY WORDS: Electric Vehicles, EV Sales Forecasting, Machine Learning, Random Forest Regression, Data Preprocessing, Exploratory Data Analysis, Sustainable Mobility, Indian EV Market, Vehicle Type Analysis, Feature Importance, Transportation Analytics, State-wise Sales, Predictive Modeling, Automotive Industry, Policy Impact

Data Collection

The dataset and the sources used for this process are listed below :

https://drive.google.com/file/d/14_cAtKqyS_nlzAhdhftC7NDlnBx_4YBS/view?usp=sharing

Market Segmentation

Target Market:

The target market of Electric Vehicle Market Segmentation can be categorized into Geographic, SocioDemographic, Behavioral, and Psychographic Segmentation.

Behavioral Segmentation: searches directly for similarities in behavior or reported behavior.
Example: prior experience with the product, amount spent on the purchase, etc.

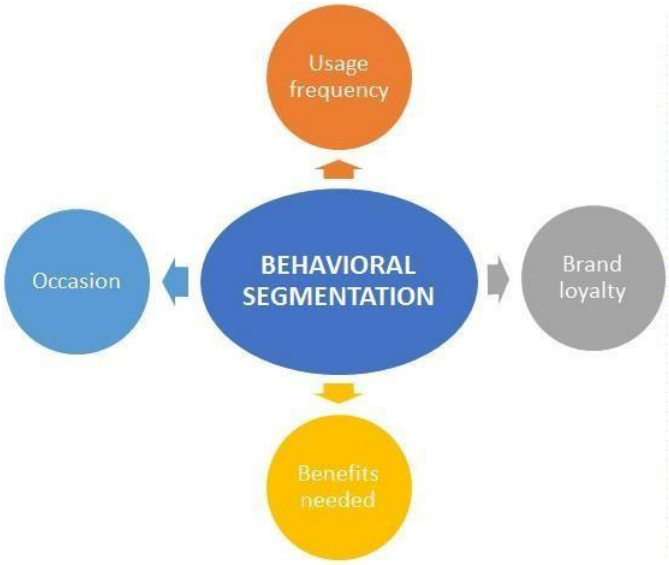


Figure 1: *Behavioral Segmentation*

Advantage: uses the very behavior of interest is used as the basis of segment extraction.

Disadvantage: not always readily available.

Psychographic Segmentation: grouped based on beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product. Suitable for lifestyle segmentation. Involves many segmentation variables.

Advantage: generally more reflective of the underlying reasons for differences in consumer behavior.

Disadvantage: increased complexity of determining segment memberships for consumers.

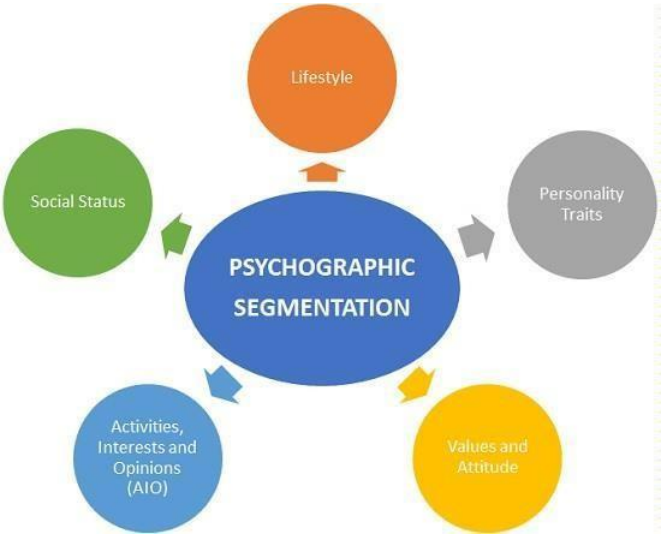


Figure 2: *Psychographic Segmentation*

Socio-Demographic Segmentation: includes age, gender, income and education. Useful in industries.

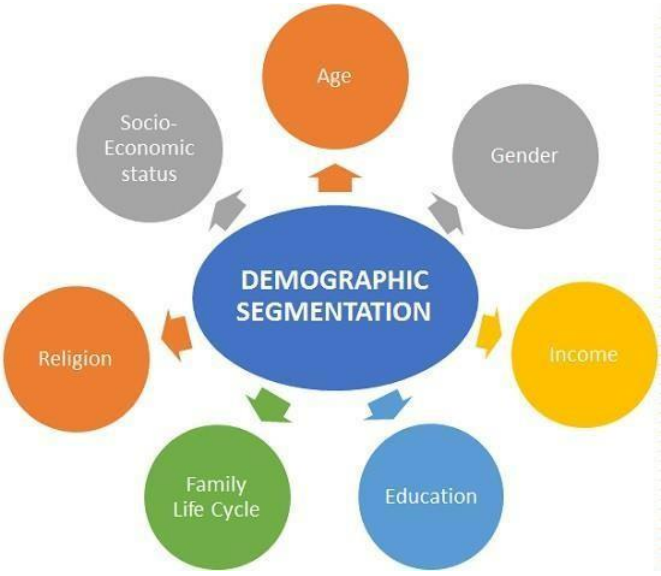


Figure 3: *Behavioral Segmentation*

Advantage: segment membership can easily be determined for every customer.

Disadvantage: if this criteria is not the cause for customers product preferences then it does not provide sufficient market insight for optimal segmentation decisions.

Segmenting for Electric Vehicle Market

The market segmentation approach aims at defining actionable, manageable, homogenous subgroups of individual customers to whom the marketers can target with a similar set of marketing strategies. In practice, there are two ways of segmenting the market-a-priori and post-hoc. An a-priori approach utilizes predefined characteristics such as age, gender, income, education, etc. to predefine the segments followed by profiling based on a host of measured variables *(behavioral, psychographic or benefit)*. In the post-hoc approach to segmentation on other hand, the segments are identified based on the relationship among the multiple measured variables. The commonality between both approaches lies in the fact that the measured variables determine the *'segmentation theme'*. The present study utilizes an a-priori approach to segmentation so as to divide the potential EV customers into sub-groups.

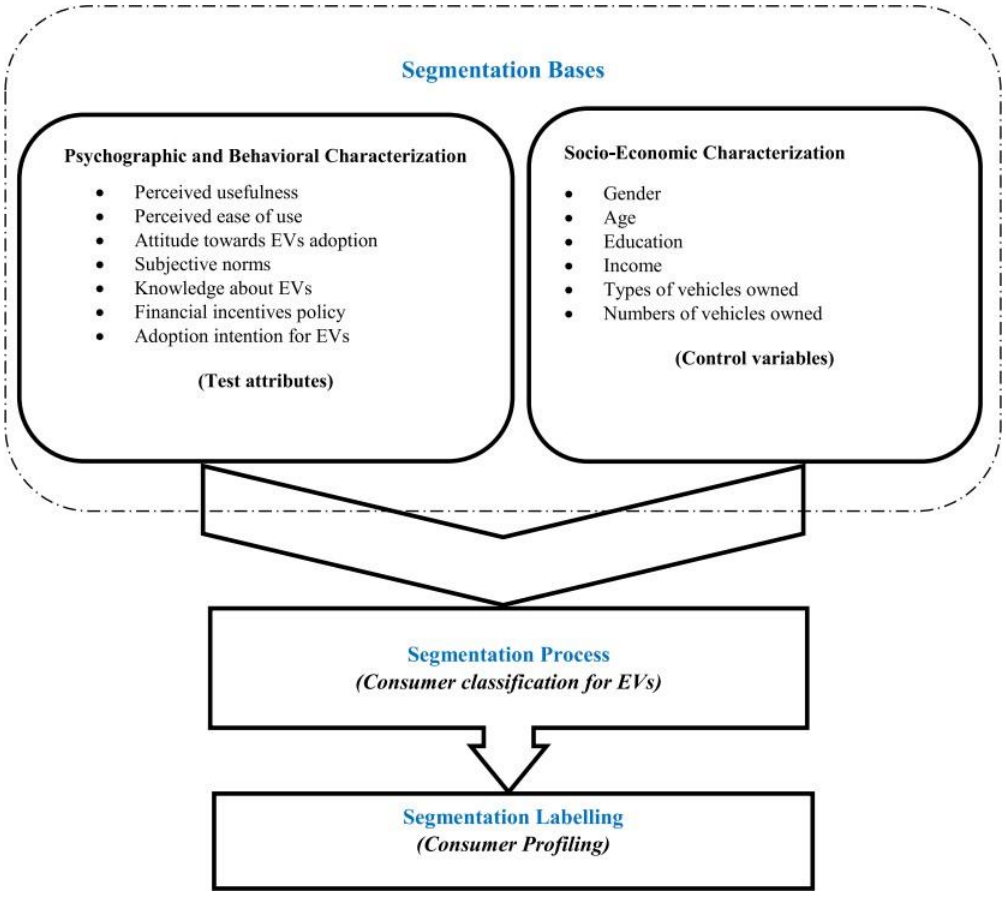


Figure 4: *Market Segmentation Electric Vehicles*

It is argued that the blended approach of *psychographic* and *socioeconomic attributes* for market segmentation enables the formulation of sub-market strategies which in turn satisfy the specific tastes and preferences of the consumer groups. Straughan and Roberts presented a comparison between the usefulness of *psychographic, demographic, and economic* characteristics based on consumer evaluation for eco-friendly products.

They pinpointed the perceived superiority of the psychographic characteristics over the socio-demographic and economic ones in explaining the environmentally-conscious consumer behavior and thus, the study recommended the use of psychographic characteristics in profiling the consumer segments in the market for eco-friendly products. The present study adds perceived-benefit characteristics guided by blended psychographic and socio-economic aspects for segmenting the consumer market.

Implementation

Packages/Tools used:

1. **NumPy** – for efficient numerical computation and array manipulation.
2. **Pandas** – to load, explore, clean, and transform tabular data.
3. **Matplotlib & Seaborn** – for visualizing trends, distributions, and correlations.
4. **Scikit-Learn** – to build and evaluate the machine learning model.

Data-Preprocessing

Data Cleaning

The data collected is compact and is partly used for visualization purposes and partly for clustering. Python libraries such as NumPy, Pandas, Scikit-Learn, and SciPy are used for the workflow, and the results obtained are ensured to be reproducible.

Year	Month_Name	Date	State	Vehicle_Class	Vehicle_Category	Vehicle_Type	EV_Sales_Quantity
2014	jan	1/1/2014	Andhra Pradesh	ADAPTED VEHICLE	Others	Others	0
2014	jan	1/1/2014	Andhra Pradesh	AGRICULTURAL TR	Others	Others	0
2014	jan	1/1/2014	Andhra Pradesh	AMBULANCE	Others	Others	0
2014	jan	1/1/2014	Andhra Pradesh	ARTICULATED VEH	Others	Others	0
2014	jan	1/1/2014	Andhra Pradesh	BUS	Bus	Bus	0
2014	jan	1/1/2014	Andhra Pradesh	CASH VAN	Others	Others	0
2014	jan	1/1/2014	Andhra Pradesh	CRANE MOUNTED V	Others	Others	0
2014	jan	1/1/2014	Andhra Pradesh	EDUCATIONAL INST	Bus	Institution Bus	0
2014	jan	1/1/2014	Andhra Pradesh	EXCAVATOR (COMI	Others	Others	0
2014	jan	1/1/2014	Andhra Pradesh	FORK LIFT	Others	Others	0
2014	jan	1/1/2014	Andhra Pradesh	GOODS CARRIER	Others	Others	0

Figure 5 : *Part of our dataset*

EDA

EDA was conducted to understand sales trends, regional variations, and market behavior. Several key patterns were revealed:

1. Yearly EV Sales Trend

A clear upward trend in EV adoption is observed, with sales accelerating significantly in recent years, driven by policy support and awareness.

2. Monthly Sales Seasonality

Some months showed noticeable spikes in EV sales, indicating seasonality effects—possibly due to fiscal policies, incentives, or end-of-year purchasing.

3. State-wise Sales Distribution

States such as Uttar Pradesh, Maharashtra, Karnataka, and Tamil Nadu led in EV sales, suggesting localized success and adoption.

4. Vehicle Category Analysis

Electric 2-wheelers and 3-wheelers had the highest sales figures, especially among commercially used vehicle classes.

5. Vehicle Type Analysis

Vehicle types (e.g., scooters, autos, e-rickshaws) showed varying sales, useful for identifying dominant products in different regions.

Modeling

To forecast electric vehicle (EV) sales, a **Random Forest Regressor** was employed. Random Forest is an ensemble learning method that operates by constructing multiple decision trees and outputting the average prediction of the individual trees, making it robust to overfitting and effective for nonlinear datasets.

Steps Followed:

1. **Feature-Target Split**
    - **X**: Independent variables (e.g., year, month, state, vehicle class/type/category)
    - **y**: Target variable – `EV_Sales_Quantity`
2. **Train-Test Split**
    - The dataset was split into **80% training** and **20% testing** using `train_test_split` with a fixed random state for reproducibility.
3. **Model Training**
    - A Random Forest Regressor with 100 decision trees was trained on the training data:

```
python
CopyEdit
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
```

4. **Prediction**
    - The model predicted EV sales on the unseen test dataset:

```
python
CopyEdit
y_pred = model.predict(X_test)
```

Model Evaluation

The model's performance was evaluated using the **Root Mean Squared Error (RMSE)**, a common metric for regression tasks. The results showed a low RMSE, indicating good predictive power of the model.

Performance Metrics

- **RMSE**: *√MSE between predicted and actual sales*
  Example output:

```
plaintext
CopyEdit
Root Mean Squared Error: 123.56
```

Visual Analysis

- **Actual vs. Predicted Plot**
  A scatter plot of true vs. predicted values indicated a strong alignment along the ideal diagonal, suggesting accurate forecasting performance.
- **Feature Importance Plot**
  The Random Forest model identified key predictors driving EV sales. The most influential features included:
    - State
    - Vehicle Type
    - Year
    - Vehicle Category

```
Root Mean Squared Error: 130.58175227721011
```

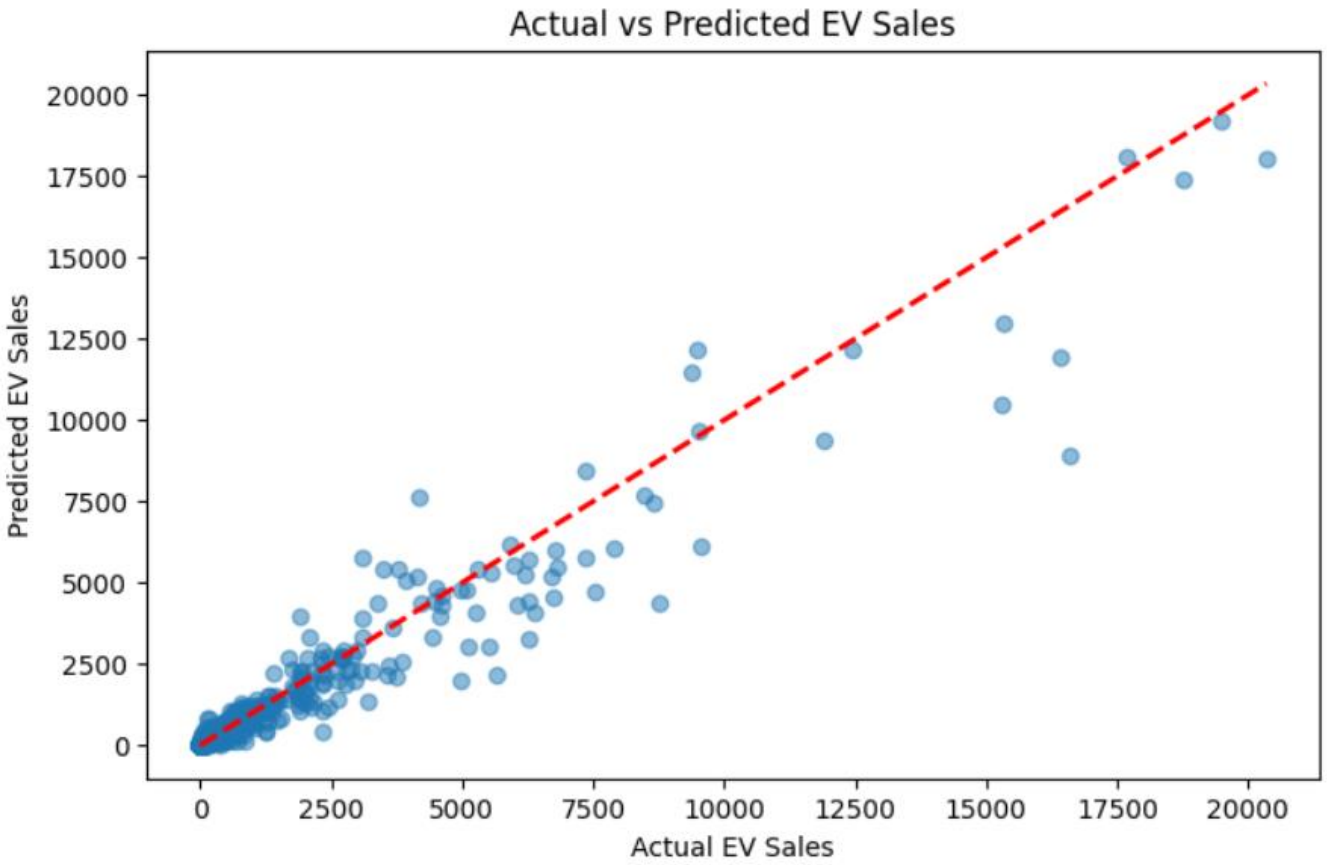


Figure 6 : *actual vs predicted ev sales*

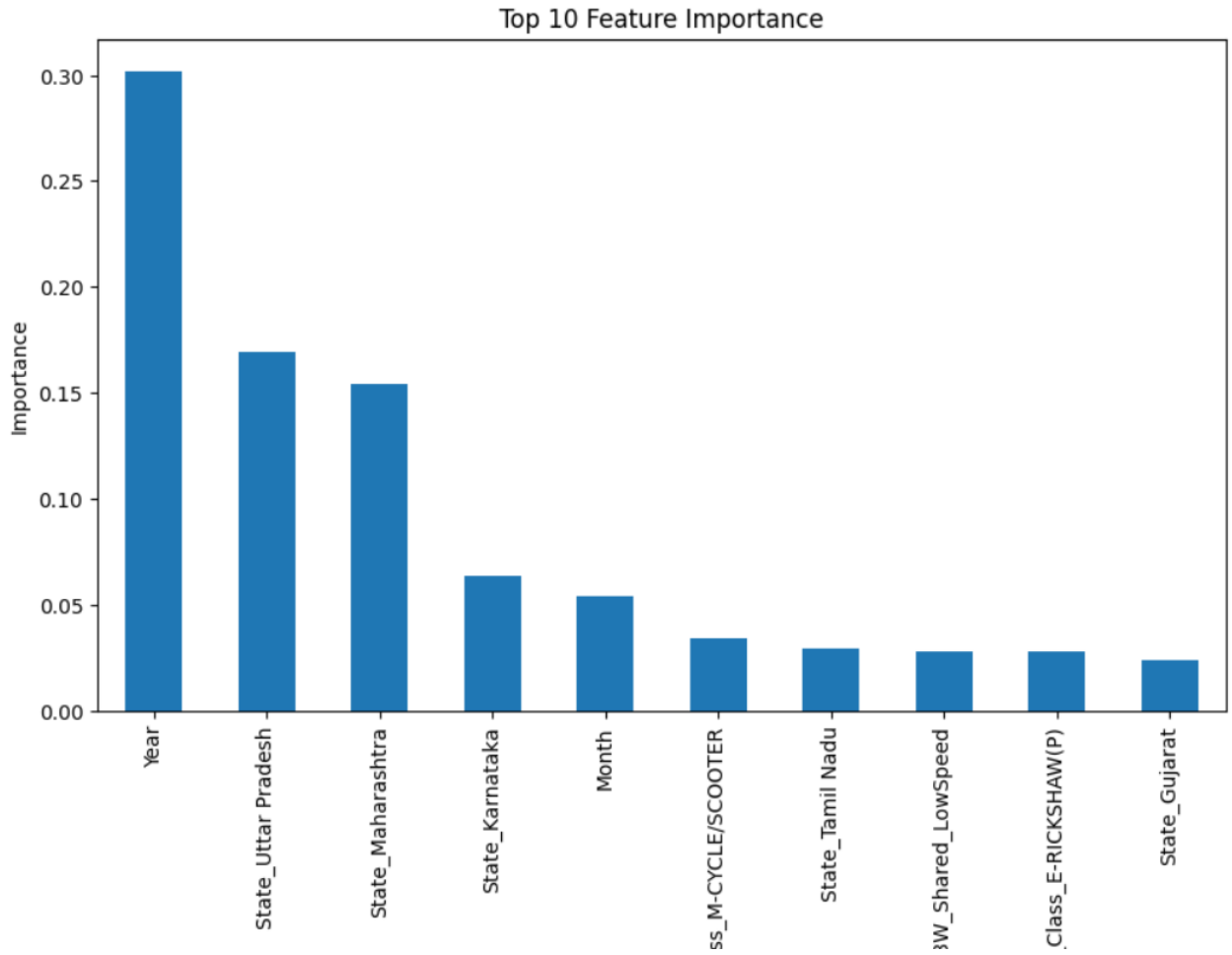


Figure 7 : *top 10 feature importance*

Segment Profiling

Based on the exploratory analysis and feature importance results from the Random Forest model, we can identify several meaningful EV market segments in India. These segments provide insight into which types of vehicles, locations, and time periods contribute most to sales volume.

1. Top Performing States

State-wise analysis reveals that a few states dominate EV adoption:

- **High-Sales States**: Uttar Pradesh, Maharashtra, Karnataka, Tamil Nadu
- **Possible Reasons**: State-level subsidies, urban infrastructure, awareness programs, and commercial use of EVs

These states are leading hubs for EV growth and should be considered priority markets for manufacturers and policymakers.

2. Vehicle Type Segments

The EV market is currently led by:

- **Two-Wheelers**: Cost-effective, ideal for urban commuting
- **Three-Wheelers (E-rickshaws)**: Widely used for last-mile connectivity
- **Four-Wheelers**: Gaining traction, especially in metro regions, but still a smaller share

Most EV adoption is driven by practicality and cost efficiency, not just sustainability.

3. Time-Based Segmentation

- **Yearly Trends**: Strong growth observed post-2017, especially from 2020 onward.
- **Monthly Patterns**: Spikes in March and December, likely influenced by year-end government incentives and fiscal factors.

Sales forecasting models should consider both seasonality and policy timing.

4. Feature Importance Takeaways

From model analysis:

- **State** and **Vehicle Type** were the most influential factors.
- **Year** and **Vehicle Category** also significantly influenced EV sales patterns.

This highlights the regional and product-specific nature of EV demand in India.

Customizing the Marketing Mix (4Ps)

The marketing mix outlines how companies can effectively promote and position their electric vehicles (EVs) based on the insights from segmentation and sales prediction. The 4Ps framework—**Product, Price, Place, Promotion**—is tailored here for the Indian EV landscape.

1. Product

The most successful EV segments are **2-wheelers and 3-wheelers**, especially in urban and semi-urban areas. These vehicles are favored due to their affordability, low maintenance, and suitability for last-mile transportation.

Recommendation: Focus product development on high-efficiency, low-cost EVs with strong after-sales support and localized features (e.g., battery-swapping).

2. Price

The analysis revealed that EVs with **moderate pricing** gain the most traction—particularly in the entry-level and mid-range segments. Government incentives and FAME II subsidies play a crucial role.

Recommendation: Strategically price EVs to align with state subsidies and target affordability for mass adoption. Offer flexible financing and leasing options.

3. Place (Distribution)

Top-performing states such as Uttar Pradesh, Maharashtra, and Karnataka represent high-potential markets. Urban centers and Tier-2 cities are primary hotspots for EV adoption.

Recommendation: Strengthen dealership and service networks in high-demand regions. Expand to Tier-2 and Tier-3 cities with growing EV awareness and need for affordable mobility.

4. Promotion

Awareness campaigns around EV benefits—such as cost savings, eco-friendliness, and subsidies—are critical. Behavioral segmentation shows seasonal interest, hinting at opportunities for targeted marketing.

Recommendation: Use localized, data-driven campaigns during high-purchase periods (e.g., March, December). Highlight government incentives, range improvements, and total cost of ownership benefits.

With a well-tailored 4Ps strategy, manufacturers and policymakers can better position EVs to meet market demand and accelerate adoption.

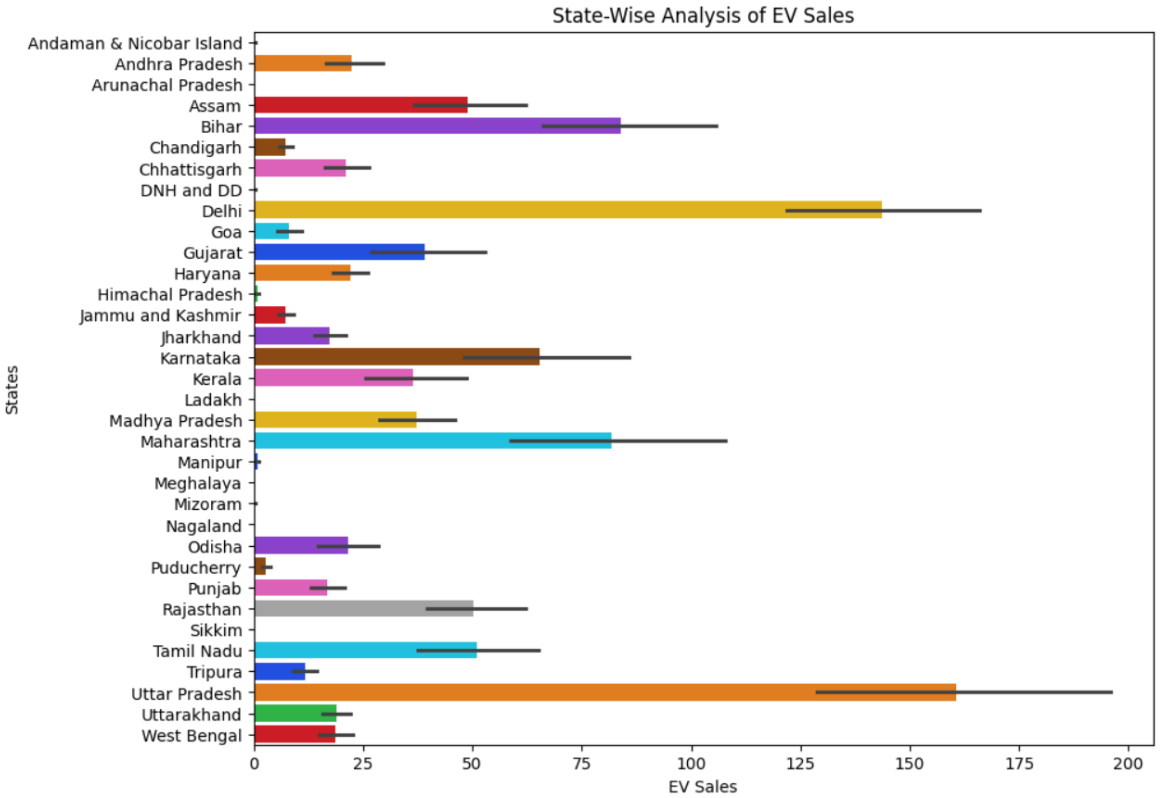


Figure 8 : *state-wise analysis of ev sales*

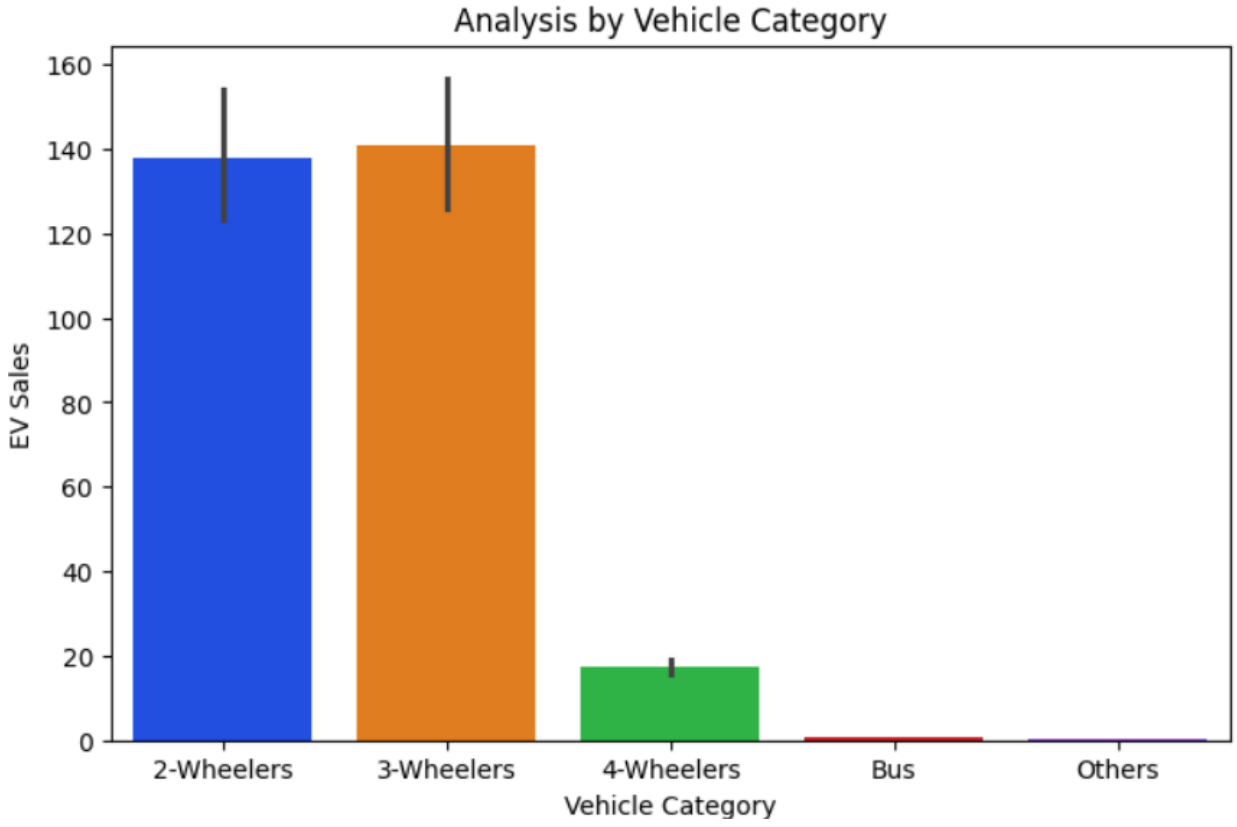


Figure 9 : *analysis by vehicle category*

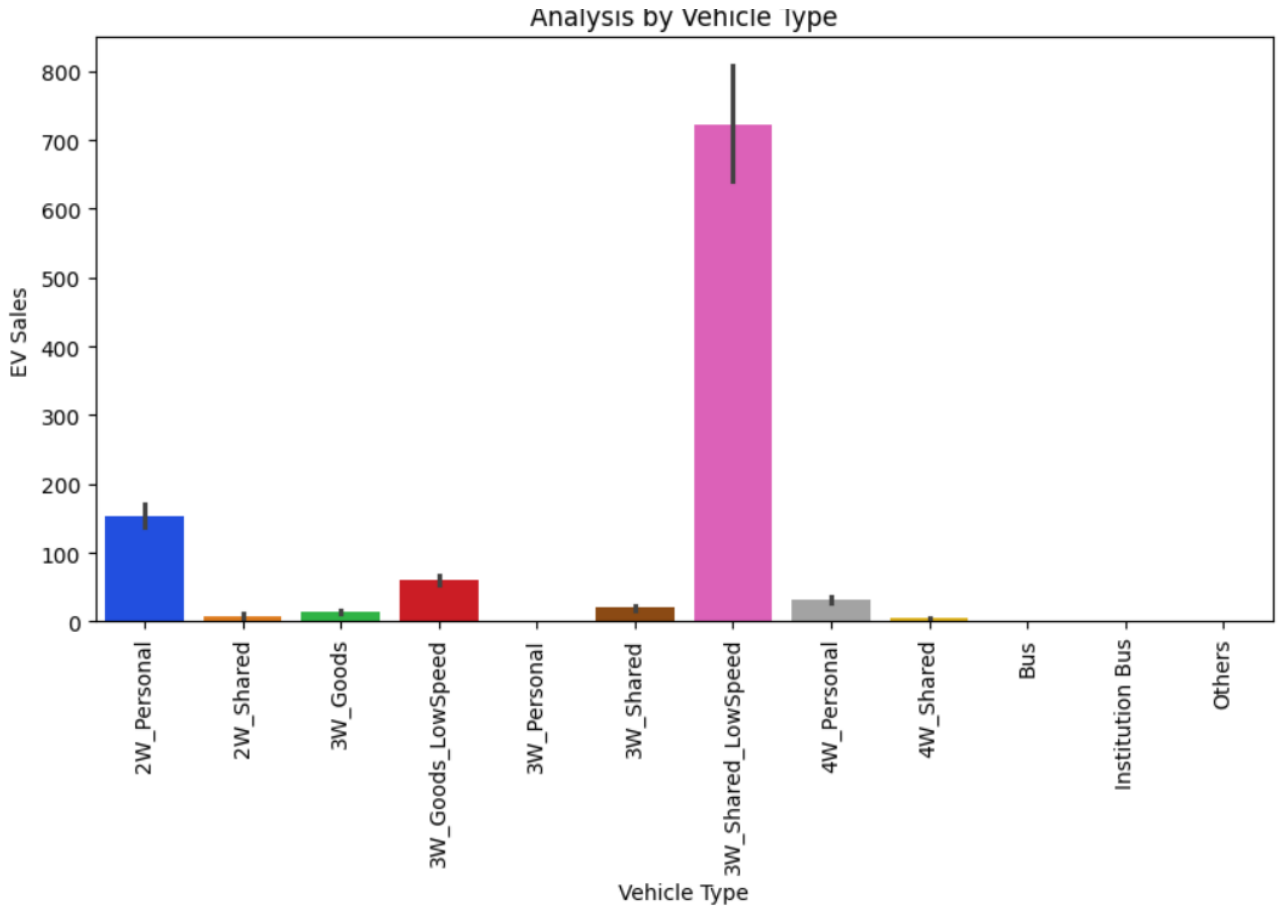


Figure 10 : *analysis by vehicle type*

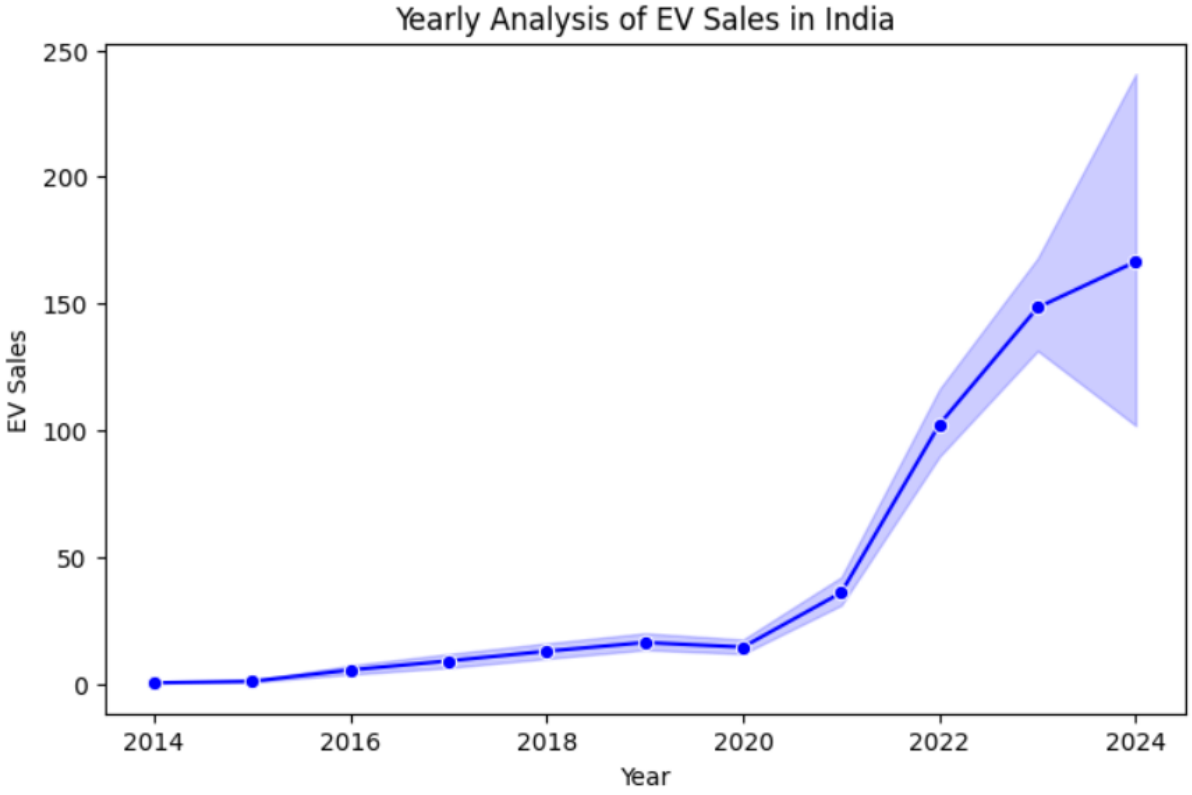


Figure 11 : *yearly analysis of ev in india*

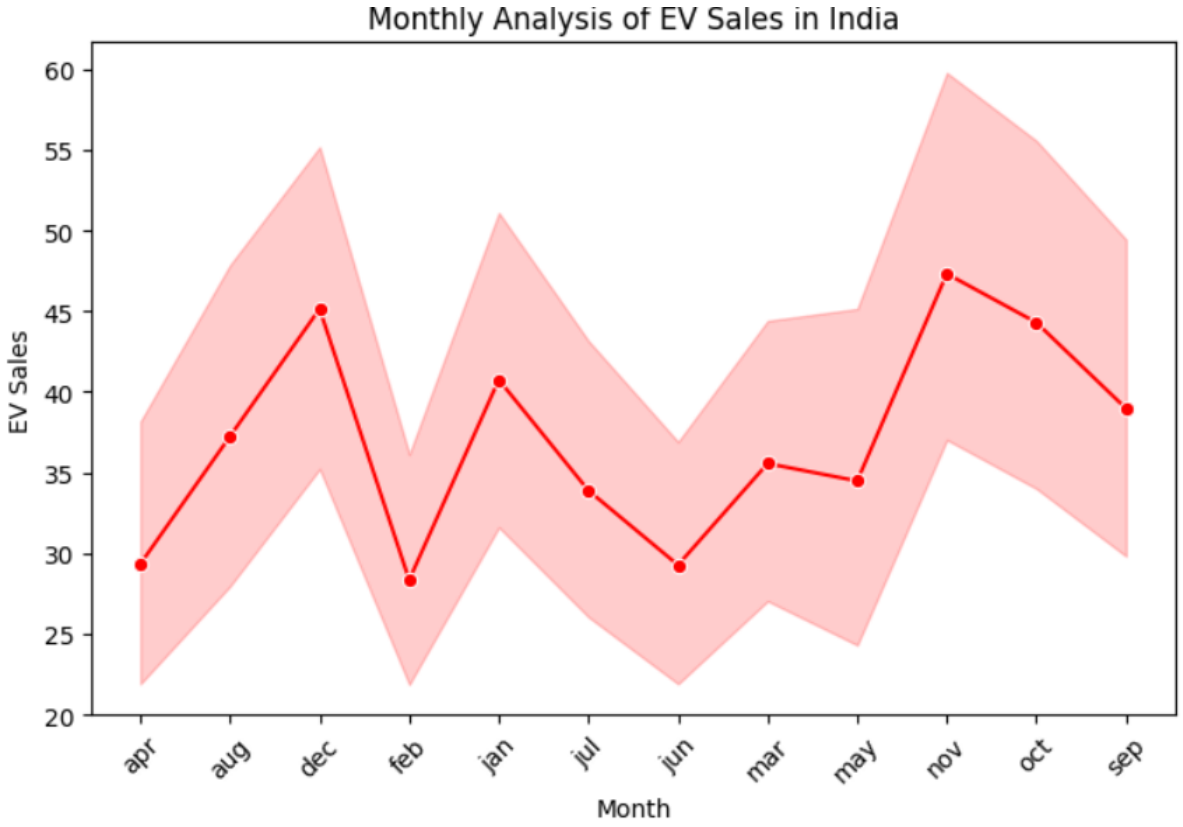


Figure 12 : *monthly analysis of ev sales in india*

References

[1] Deepak Jaiswal,Arun Kumar Deshmukh(2022) *Who will adopt electric vehicles? Segmenting and exemplifying potential buyer heterogeneity and forthcoming research*, Journal of Retailing and Consumer Services .

[2] Dolnicar, S., Grun Bettina, amp; Leisch, F. (2019). *Market segmentation analysis understanding it, doing it and making it useful*. Springer Nature.

[3] McDonald, M., amp; Dunbar, I. (2003). /emphMarket segmentation. Butterworth-Heinemann.