

# **EV MARKET SEGMENTATION** **ANALYSIS**

**-BY ANIKET**

**GITHUB REPOSITORY- [LINK](#)**



## **Summary of the Project**

### **Problem Statement:**

Identify high-potential regions and strategies for launching an electric vehicle (EV) business in India using data-driven market segmentation and sales forecasting.

### **Background:**

India's electric vehicle (EV) market is expanding rapidly due to rising fuel costs, environmental concerns, and government incentives. Yet, EV penetration is extremely divergent by state because of infrastructure, regulation, and consumer differences. Startups struggle to realize which markets will work and what drives sales. With a data-driven platform, India's EV growth potential and strategic investment can be discovered.

### **Problems**

The Indian EV industry is expanding fast, but EV start-ups face some major challenges:

- Which are the Indian states that have the highest potential for EV adoption?
- How would the market be segmented to serve various customer categories and vehicle types?
- What are the forces that propel EV sales, and how can future sales be forecast?
- Can policy effects on EV growth be simulated before product launches?

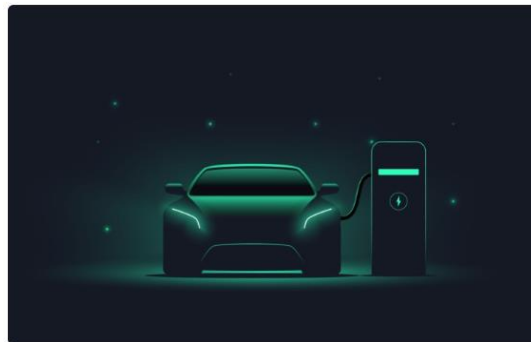
### **Solution Approach**

This project offers end-to-end, data-driven solution by:

- Clustering algorithms to segment the EV market
- Time-series modelling to forecast EV sales trends
- Regression to analyse leading drivers of sales
- Policy impact modelling to predict demand uplift
- Deploying the solution as an interactive Streamlit web app

# Streamlit Integration – Interactive Dashboard

## Electric Vehicle Sales Data Analysis



Upload your EV Sales CSV file

Drag and drop file here  
Limit 200MB per file • CSV

Browse files

Please upload a CSV file to begin analysis.

Throughout the analysis, there is a complementing Streamlit web application, which is able to:

- CSV Upload Interface – users to upload any EV sales dataset
- Real-time Visuals – Live plots, heat maps, line graphs, and predictions
- Sliders and Widgets – For testing EV subsidies, fuel prices, and infra growth
- On-click Data Download – Download cleaned dataset
- Interactive Forecast Viewer – Designed plots with Plotly

Advantages of Streamlit:

- No backend setup required
- Easy UI for end-users
- Easy deployment and sharing

## Overview and Pre-processing of Data

- Dataset Columns: State, Vehicle\_Class, Vehicle\_Category, Vehicle\_Type, EV\_Sales\_Quantity, Date

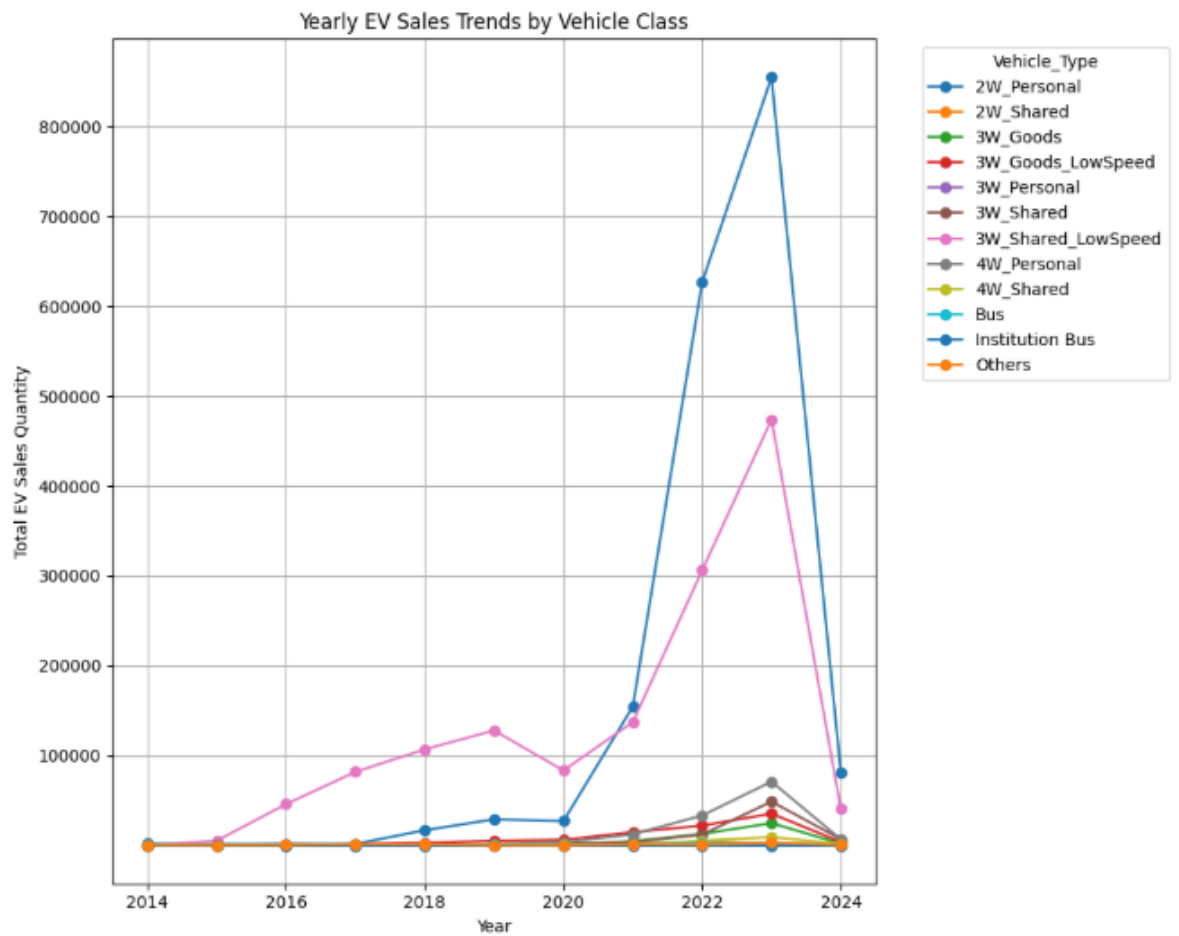
	Year	Month_Name	Date	State	Vehicle_Class	Vehicle_Category	Vehicle_Type	EV_Sales_Quantity
0	2014.0	jan	1/1/2014	Andhra Pradesh	ADAPTED VEHICLE	Others	Others	0.0
1	2014.0	jan	1/1/2014	Andhra Pradesh	AGRICULTURAL TRACTOR	Others	Others	0.0
2	2014.0	jan	1/1/2014	Andhra Pradesh	AMBULANCE	Others	Others	0.0
3	2014.0	jan	1/1/2014	Andhra Pradesh	ARTICULATED VEHICLE	Others	Others	0.0
4	2014.0	jan	1/1/2014	Andhra Pradesh	BUS	Bus	Bus	0.0
5	2014.0	jan	1/1/2014	Andhra Pradesh	CASH VAN	Others	Others	0.0
6	2014.0	jan	1/1/2014	Andhra Pradesh	CRANE MOUNTED VEHICLE	Others	Others	0.0
7	2014.0	jan	1/1/2014	Andhra Pradesh	EDUCATIONAL INSTITUTION BUS	Bus	Institution Bus	0.0

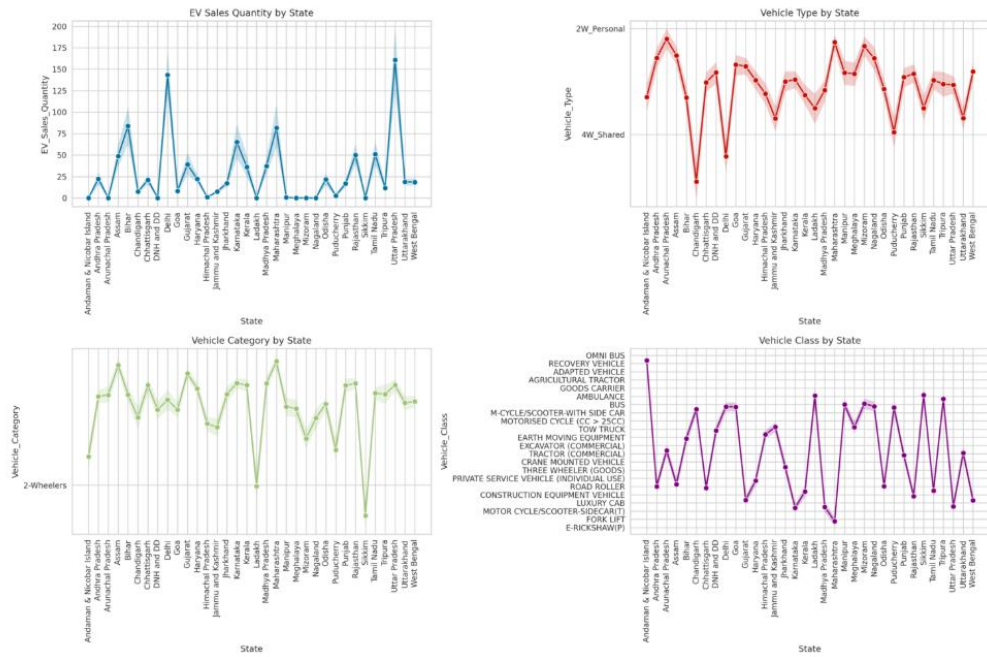
- Pre-processing Steps:
  - Extracted Year from Date

- Treated types, missing values
- One-hot encoded categorical data
- Normalized numerical columns

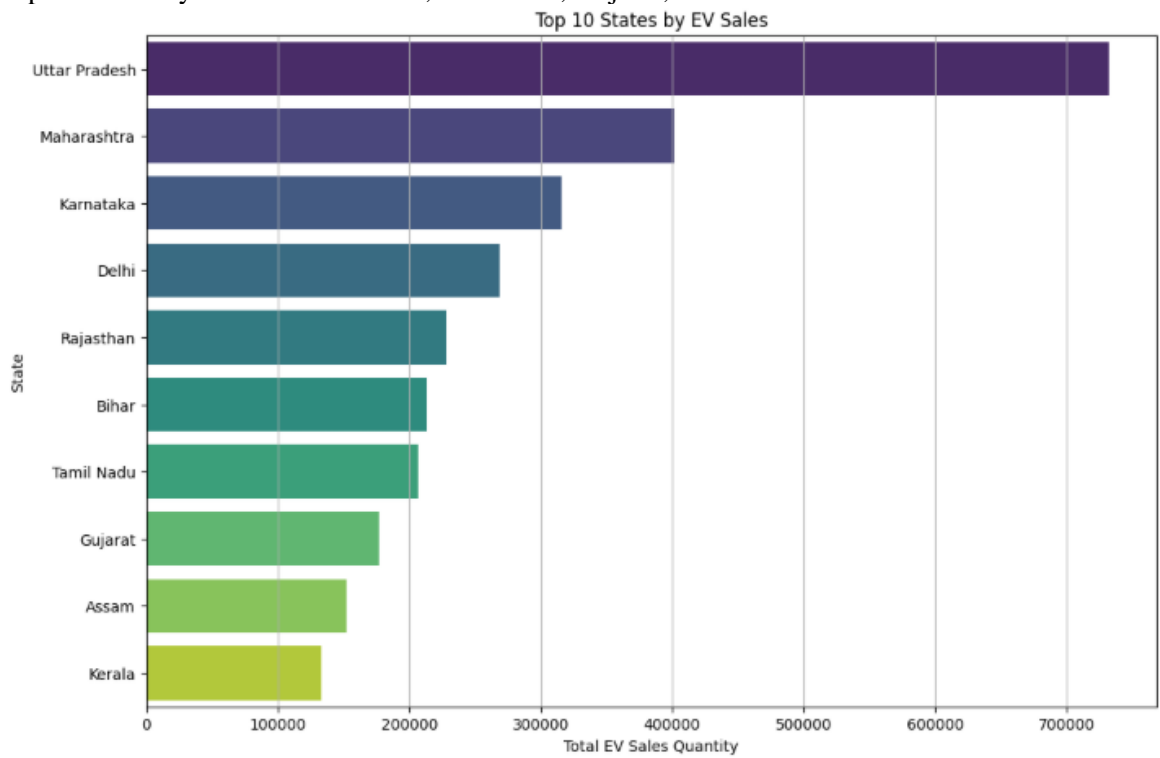
## **Exploratory Data Analysis (EDA)**

- Dataset Summary and inspection of type
- Year-wise EV Sales Trend: consistent growth throughout the years

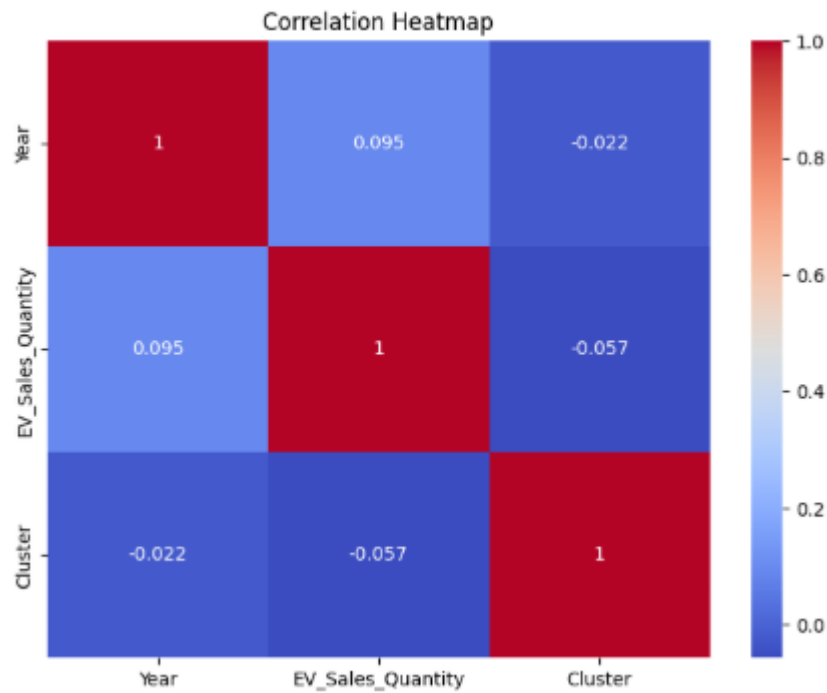




- Top 10 States by Sales: Maharashtra, Karnataka, Gujarat, etc.

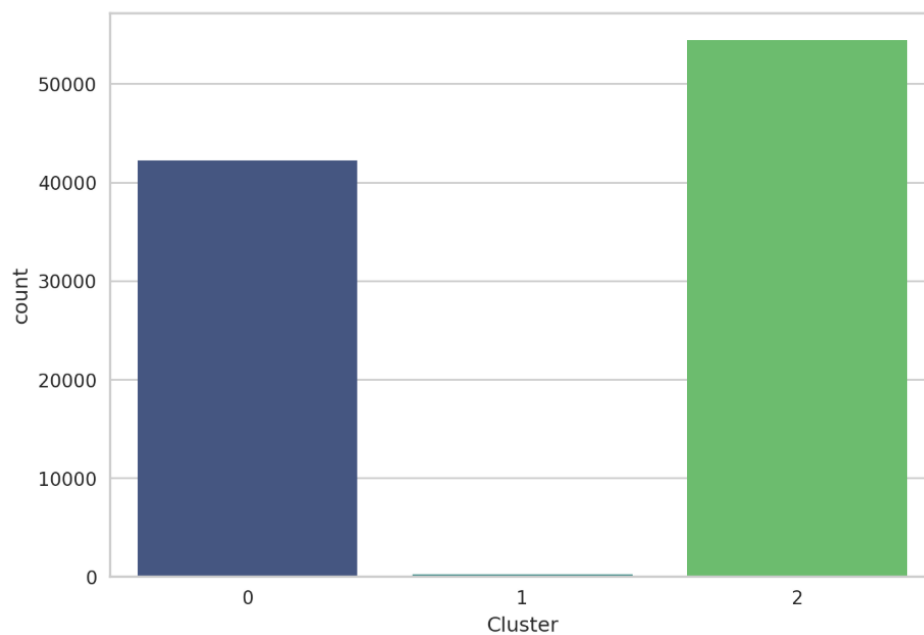


- A correlation matrix is represented visually by a correlation heatmap.

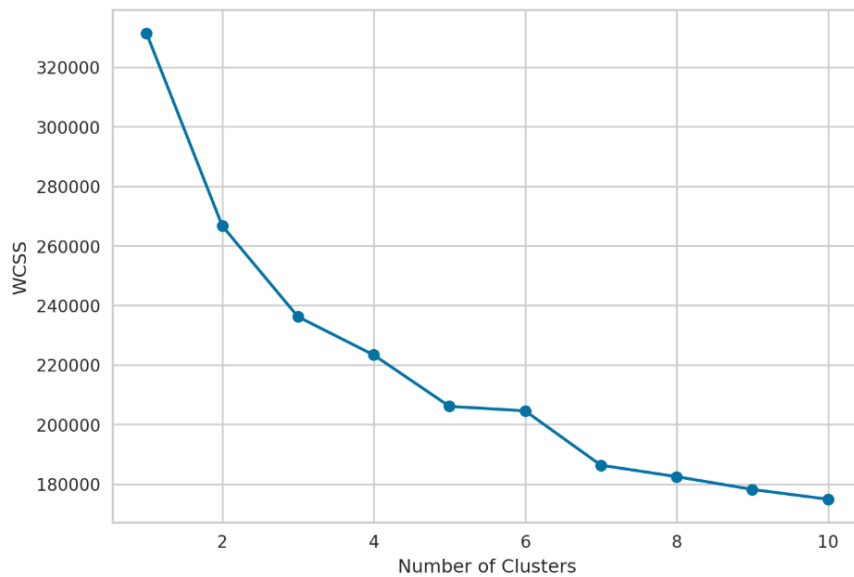


## Market Segmentation – K-Means Clustering

- Features Used: EV\_Sales\_Quantity, State, Vehicle\_Class, Vehicle\_Category
- Clusters Found:
  - Cluster 0: Strong infrastructure states with high adoption
  - Cluster 1: Growth opportunities in emerging markets
  - Cluster 2: Long-term potential in low-penetration regions

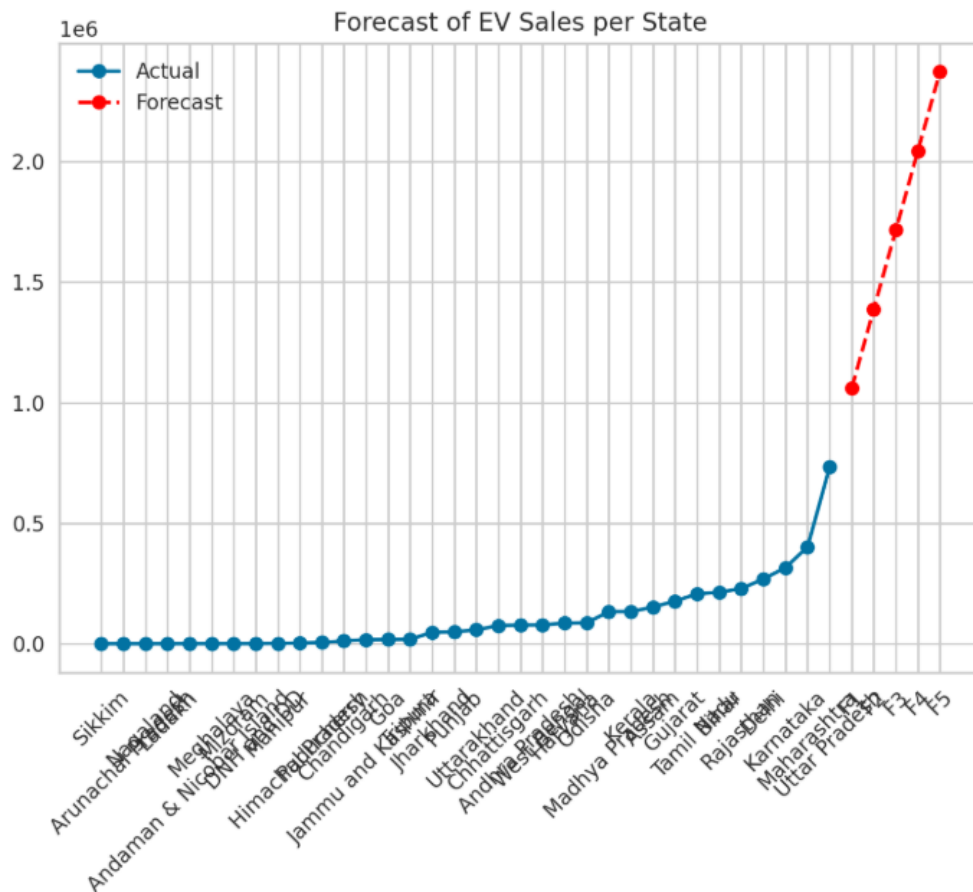


- Elbow Method utilized for best K selection (k=3)



## Exponential Smoothing Time-Series Forecasting

- Cumulative Sales by State
- The Holt-Winters model is used to forecast the following five periods.
- The outcomes are:
  - Consistent expansion in states with strong performance
  - Mid-range states with high potential

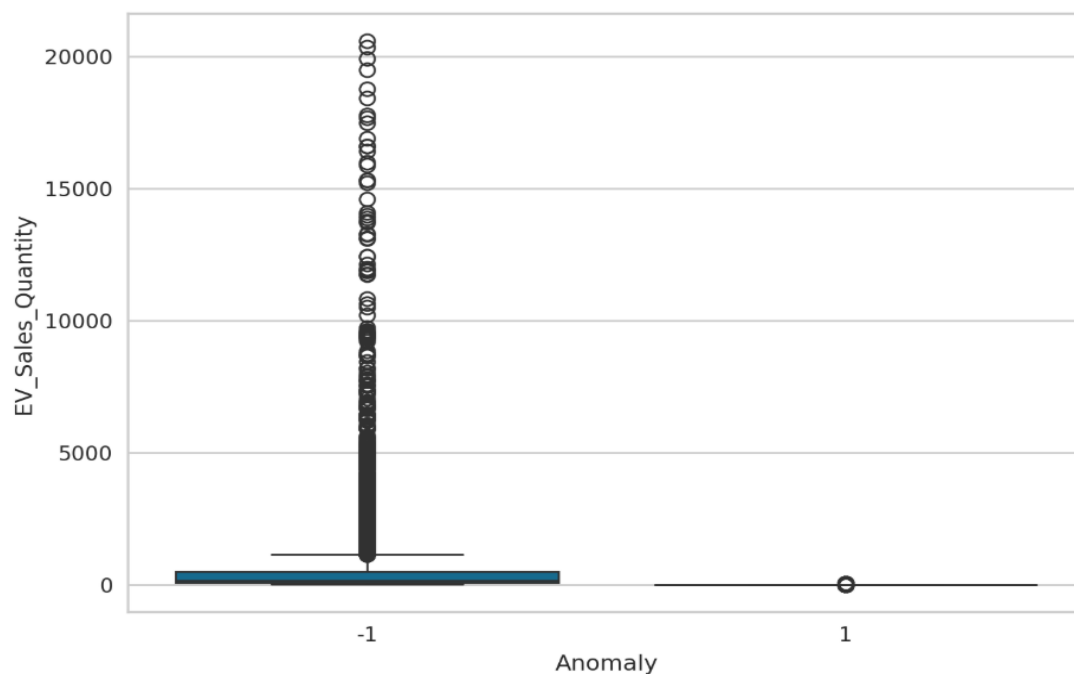


## **Random Forest Regressor**

- Features for Predictive Modelling: Year, Vehicle\_Type, Vehicle\_Class, State encoded, etc.
- The target is EV Sales Quantity.
- Mean Squared Error (MSE) is the metric for performance.
- Observation: Geographies and auto models are ordered according to feature importance by their impact on sales.

## **Anomaly Detection – Isolation Forest**

- State-by-state outliers in EV sales were identified by isolation forest anomaly detection.
- Unforeseen increases or decreases revealed by boxplot visualization



## **Interactive Analysis And Policy Impact Simulator**

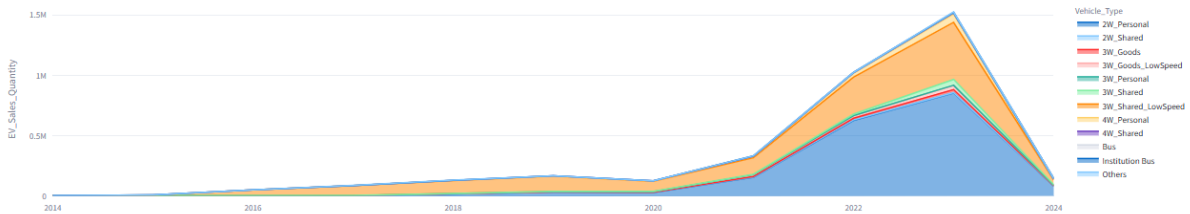
- Sliders with a clean design for:
  - The EV Subsidy (₹)
  - The cost of gasoline (₹/liter)
  - Growth of Charging Infra (%)
- The dynamically calculated and displayed projected increase in sales allows users to model:
  - Impact of government policy
  - Fuel price variations



- Infrastructure investments

## Interactive Market Share

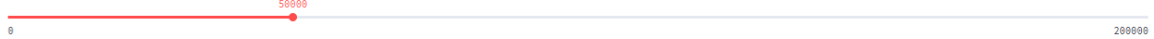
Market Share Over Time



## EV Policy Impact Simulator

Adjust parameters to simulate EV adoption trends:

Annual EV Subsidy per Vehicle (₹)



Petrol Price per Litre (₹)



Annual Growth in Charging Infrastructure (%)



Projected Annual EV Sales Increase

22,500 units

## Key Insights & Solution Summary

### PROBLEM

WHERE TO LOCATE AN EV STARTUP?

HOW TO SEGMENT THE MARKET?

WHAT ARE THE MOST IMPORTANT DRIVERS OF EV SALES?

IS IT POSSIBLE TO PREDICT FUTURE SALES?

IS IT POSSIBLE TO MODEL POLICY IMPACTS?

### SOLUTION

Target Cluster 1 states (e.g., Telangana, Gujarat) with nascent demand

Used K-Means to discover 3 strategic clusters

Region and vehicle type were found to be the primary drivers by the Random Forest model.

Yes, employing exponential smoothing with Holt-Winters

Yes, demand estimates using Streamlit sliders

## Final Recommendation

- Go-to-Market Strategy:
  - Soft launch with scalable demand in mid-range states
  - Watch out for small commercial EVs and two-wheelers.
- Invest in:
  - Charging infrastructure (most impactful sales driver)
  - R&D for low-cost, rapid-charging variants
- Make a Difference By:
  - The price of low-cost EVs
  - Scaling from mid-range adoption in use cases of delivery and logistics

## Output Delivery

- Downloadable cleaned dataset as CSV
- Whole project deployable in a single command using Streamlit
- Integrated native real-time performance measures and visual exploration

 **Download Cleaned Data**

Download CSV

---

## Outcomes

**1. Segmented the Indian EV market** into three distinct clusters using K-Means based on sales volume, vehicle class, and region:

- High-performing states
- Emerging markets
- Low-adoption regions

**2. Forecasted future EV sales trends** using Exponential Smoothing, showing a consistent growth pattern in mid-tier and top-tier states.

**3. Identified key factors influencing EV sales** using a Random Forest Regressor:

- Region/state and vehicle type were top predictors of EV sales volume.

**4. Detected anomalies in EV sales** across states using Isolation Forest, helping identify unexpected market surges or underperformance.

**5. Built an interactive Policy Impact Simulator** that projects EV sales growth based on changes in:

- Government subsidies
- Fuel prices
- Charging infrastructure expansion

**6. Developed a fully interactive Streamlit dashboard** for:

- Real-time data upload and visualization
- Model insights and forecasting
- Downloadable output and policy simulation

**7. Recommended strategic regions** like Telangana and Gujarat for new EV startups, based on sales potential and competition levels.

**8. Outlined future enhancements** such as LSTM forecasting, NLP sentiment analysis, and recommendation systems for EV selection.

## **Roadmap (Preview) for Upcoming Advanced Features**

- LSTM Time-Series Forecasting: deep-learning-based time-series forecasting
- NLP Integration: sentiment extraction from news headlines and user reviews on EV policy
- Suggestion Engine: car suggestions according to user needs

## **Conclusion**

- This end-to-end EV sales analysis solution leverages:
- Machine Learning, Forecasting, and Visualization
- Streaming into an open Streamlit web application
- Enabling stakeholders to make strategic choices in India's fast-expanding EV market

**GITHUB REPOSITORY- [LINK](#)**