📘 Internship Project Report

Title: Electric Vehicle Sales by State in India

Internship Organization: Unified Mentor Pvt. Ltd.
Intern Name: Himanshu Kaushik
Domain: Data Analyst Intern
Duration: 3 Months
Tools Used: Python, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Excel, SQL
Dataset Source: Unified Mentor

🔍 1. Objective

The aim of this project is to analyze the **Electric Vehicle (EV)** sales trends across different Indian states and predict future EV sales using machine learning. This analysis supports strategic decisions in market planning and infrastructure development for EV adoption.

📊 2. Dataset Overview

- **Records:** 96,845

- **Columns:** 8

- **Time Range:** 2014 – 2023

- **No missing or duplicate values**

🧹 3. Data Preprocessing

- Converted Date column to datetime

- Converted Year to integer

- Encoded categorical features using astype('category')

- No missing or duplicate records were found

📈 4. Exploratory Data Analysis (EDA)

🔶 Yearly EV Sales Trend

EV sales increased drastically from 2019 onward, peaking in 2023.

🔶 State-wise Sales

Top 5 States by EV Sales:

- Maharashtra

- Karnataka

- Uttar Pradesh

- Rajasthan

- Gujarat

◆ **Vehicle Class Insights**

Top vehicle classes sold:

- Motor Car

- Motor Cycle/Scooter

- Goods Carrier

- Bus

◆ **Vehicle Category & Type Insights**

- Most EVs are classified as "Others", followed by "2-Wheelers"

- Top vehicle types include: 2W_Personal, 4W_Shared, Bus, 3W_Goods

◆ **Monthly Trends**

Sales are fairly consistent but higher in **December and March**, possibly due to year-end push and fiscal year closing.

🧠 5. Feature Engineering

- Extracted Month and Day from Date

- One-hot encoded:

  o State

  o Vehicle_Class

  o Vehicle_Category

  o Vehicle_Type

🤖 6. Predictive Modeling

Model Used: Random Forest Regressor

- **Features:** All categorical & time features (encoded)

- **Target Variable:** EV_Sales_Quantity

Model Training:

python

CopyEdit

from sklearn.ensemble import RandomForestRegressor

from sklearn.model_selection import train_test_split

from sklearn.metrics import mean_squared_error

```
X = df_encoded.drop('EV_Sales_Quantity', axis=1)

y = df_encoded['EV_Sales_Quantity']


X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)


model = RandomForestRegressor(n_estimators=100, random_state=42)

model.fit(X_train, y_train)

y_pred = model.predict(X_test)
```

Evaluation:

python

CopyEdit

```
mse = mean_squared_error(y_test, y_pred)

rmse = np.sqrt(mse)
```

📊 7. Visualizations

- **Line Plot:** Yearly EV sales trends

- **Bar Plots:**

  o EV sales by state

  o EV sales by vehicle category and type

  o Vehicle class and type contributions

- **Scatter Plot:** Actual vs Predicted EV Sales

- **Feature Importance Bar Plot:** Top predictors of EV sales

---

✅ 8. Conclusion

- **Maharashtra, Karnataka, and UP** are leaders in EV adoption.

- **2W & 3W Personal** segments dominate the market.

- **Policy focus** should consider these regions and categories for infrastructure investment.

- **Machine learning models** like Random Forest offer a reliable way to predict EV sales for market planning.

📝 9. Key Learnings

- Hands-on with **real-world data cleaning** and **feature engineering**

- Understood **EV adoption trends** in India

- Learned how to **train, evaluate, and interpret** a regression model

- Used **data visualization** for better storytelling

Snapshots :

EV Sales by State

Actual vs Predicted EV Sales

Year-wise EV Sales in India