Electric Vehicle Sales by State in India

Internship Project Report

1. Introduction

This project focuses on analyzing Electric Vehicle (EV) sales across different states in India. The aim is to understand adoption trends, state-wise performance, and vehicle category preferences using data analysis and machine learning techniques.

2. Objective

• Analyze EV sales trends across India
• Perform exploratory data analysis
• Build a machine learning regression model to predict EV sales
• Evaluate model performance using Mean Squared Error

3. Dataset Description

The dataset 'Electric Vehicle Sales.csv' contains state-level EV sales data including year, month, vehicle category, vehicle type, and sales quantity.

4. Tools Used

Python, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, VS Code, GitHub

5. Data Preprocessing

Data preprocessing included datatype correction, datetime conversion, duplicate checks, and feature engineering such as month and day extraction.

6. Exploratory Data Analysis

EDA was performed to analyze EV sales trends year-wise, month-wise, state-wise, and by vehicle categories using visualizations.

7. Machine Learning Model

A Random Forest Regressor was used to predict EV sales quantity. Categorical features were encoded using one-hot encoding.

8. Results & Evaluation

The model achieved a Mean Squared Error (MSE) value of **130.58**, indicating acceptable predictive performance for an internship-level project.

9. Conclusion

This project successfully demonstrates the use of data analytics and machine learning to understand EV adoption trends in India. The insights can support policy planning and infrastructure development.

10. Code Appendix

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error

df = pd.read_csv("Electric Vehicle Sales.csv")

df['Year'] = df['Year'].astype(int)
df['Date'] = pd.to_datetime(df['Date'], errors='coerce')

df['Month'] = df['Date'].dt.month
df['Day'] = df['Date'].dt.day

df_encoded = pd.get_dummies(
    df,
    columns=['State','Vehicle_Class','Vehicle_Category','Vehicle_Type'],
    drop_first=True
)

df_encoded.drop(['Date','Month_Name'], axis=1, inplace=True)

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)
mse = mean_squared_error(y_test, y_pred)
print(mse)
```