Project Proposal: Electric Vehicle Sales by State in India

Project Title:

Electric Vehicle Sales by State in India

Project Overview:

The aim of this project is to analyze and understand the adoption trends of electric vehicles (EV) across different states in India. The dataset provided includes comprehensive data on EV sales, vehicle types, and categories, which will be used to conduct various analyses, including market segmentation, trend analysis, and predictive modeling.

Tools Required:

- Python

- Machine Learning (ML)

- SQL

- Excel

Domain:

Data Analysis

Project Difficulty Level:

Intermediate

Dataset Description:

- The dataset includes state-level data on EV sales, vehicle types, and categories.

- The data is preprocessed, with all null values removed, and is provided in CSV format.

- Key features include state names, vehicle types, vehicle categories, and the quantity of EV sales.

Project Steps:

1. **Data Collection:**

   o Load the dataset using Python libraries.

   o Inspect the dataset to understand its structure and content.

2. **Data Preprocessing:**

   o Handle any remaining missing values.

   o Convert date columns to proper datetime formats.

   o Perform feature engineering, such as creating new columns from existing date columns.

3. **Exploratory Data Analysis (EDA):**

o   Visualize trends and relationships between variables using various charts and graphs.

o   Analyze EV sales trends over time, across states, and by vehicle categories.

4. **Feature Engineering:**

o   Create new features from the date column (e.g., extracting month and day).

o   Encode categorical variables using one-hot encoding.

5. **Modeling:**

o   Build a regression model (e.g., Random Forest Regressor) to predict EV sales.

o   Split the dataset into training and testing sets for model evaluation.

6. **Model Evaluation:**

o   Evaluate the model's performance using metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).

o   Interpret the results to understand the factors affecting EV sales.

7. **Visualization:**

o   Create visualizations to present the analysis results, trends, and feature importance.

Python Code (Example):

```
# Import necessary libraries
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


# Load the dataset
df = pd.read_csv('EV_Dataset.csv')
print(df.head())


# Data Preprocessing
```

```python
df['Date'] = pd.to_datetime(df['Date'])

print(df.isnull().sum())

df.fillna(df.median(), inplace=True)


# Exploratory Data Analysis (EDA)

plt.figure(figsize=(10, 6))

sns.lineplot(data=df, x='Year', y='EV_Sales_Quantity', hue='State')

plt.title('EV Sales by State over the Years')

plt.show()


# Feature Engineering

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)


# Modeling

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)


# Model Evaluation

mse = mean_squared_error(y_test, y_pred)

rmse = np.sqrt(mse)

print(f'Root Mean Squared Error: {rmse}')
```

Dataset Overview:

- **Rows:** 96,845 entries

- **Columns:** 8 columns

- **Data Types:**

  o 6 object (categorical/string) columns

  o 2 float (numerical) columns

Columns:

- Year: Year of sales (2014-2024)

- Month_Name: Month of sales

- Date: Full date in string format

- State: State in India where the sale occurred

- Vehicle_Class, Vehicle_Category, Vehicle_Type: Categorical data for vehicle classification

- EV_Sales_Quantity: Number of EV sales

Key Observations:

- **No missing values** in any column.

- **Sales data spans from 2014 to 2024.**

- **Highly skewed data in EV sales**: Median sales are 0, but the maximum sales for a single entry is 20,584.

- **Potential Issues:**

  o Date is stored as object; converting it to a date-time format would be beneficial.

  o Some categories like Vehicle_Type and Vehicle_Class may need cleaning or grouping for analysis.

- **Next Steps for Data Analysis:**

  o Convert Date to datetime format.

  o Analyze yearly and monthly sales trends.

  o Find top-performing states for EV sales.

  o Classify vehicle categories to find the most sold EV type.

  o Identify any seasonal patterns or trends over time.

- **Conclusion:**
- This project aims to provide insights into the factors affecting electric vehicle sales across different states in India. By analyzing and predicting EV sales, you can gain valuable insights for market planning and policy development.