📊 EDA & Regression Analysis App

A powerful Streamlit user interface for Exploratory Data Analysis (EDA) and building Regression Models without writing code.

🚀 Features

1. Exploratory Data Analysis (EDA)

- **Data Overview**: View dataset shape, columns, and data types.
- **Descriptive Statistics**: Summary stats (mean, std, min, max, etc.).
- **Missing Value Handling**:
  - Visualize missing data.
  - **Impute** values using Mean, Median, or Mode.
  - **Bulk Imputation** support for cleaning all columns at once.
- **Visualizations**:
  - **Correlation Heatmap**: Understand relationships between variables.
  - **Distribution Plots**: Analyze the spread of numeric data.
  - **Pair Plots**: Visualize scatter plots for multiple variables.

2. Regression Modeling

Build and compare multiple types of regression models:

- **Linear Regression**: Best for simple linear relationships.
- **Polynomial Regression**: Capture non-linear patterns (adjustable degree).
- **K-Nearest Neighbors (KNN)**: Distance-based regression (adjustable K).

- **Random Forest**: Robust ensemble method (adjustable trees).

Key Capabilities:

- **Dynamic Feature Selection**: Choose one or multiple independent variables (X) to predict your target (Y).
- **Smart Data Cleaning**: Automatically attempts to fix text-based numbers (e.g., "$1,200" → 1200) so you don't lose data.
- **Model Evaluation**: View **R² Score** and **Mean Squared Error (MSE)**.
- **Visualizations**:
  - **Regression Line** (for single feature).
  - **Actual vs Predicted Plot** (for multiple features).
- **Model Explainability**: Get text-based insights on how the model works and interpreting coefficients.

🛠️ Installation & Setup

1. **Prerequisites**: Ensure you have Python installed.
2. **Install Dependencies**:

   ```
   pip install streamlit pandas scikit-learn seaborn matplotlib
   ```

3. **Run the App**:

   ```
   streamlit run app.py
   ```

📂 Project Structure

- `app.py`: Main application entry point.
- `src/eda.py`: Functions for data loading, cleaning, and EDA visualizations.
- `src/model.py`: Logic for training models and generating predictions/plots.
- `archive/`: Folder containing default datasets.

💡 How to Use

1. **Upload Data**: Use the sidebar to upload your own CSV, or use the default Nuclear Energy dataset.
2. **Clean Data**: Go to the "EDA" tab to check for missing values and fill them if necessary.
3. **Train Model**: Go to "Regression Modeling", select your Target and Features, and click "Train Model".