Data Analysis and Visualization Dashboard (Python Mini Project)

Project Overview:

You are required to develop an **interactive Data Analysis and Visualization Dashboard** using **Python** and **Streamlit**. The system allows users to upload any dataset (in .csv format), explore its structure, perform descriptive statistics, visualize data patterns, and interactively handle missing data.

The dashboard demonstrates how **data analytics** works in real-world scenarios — from loading and exploring datasets to drawing insights using visualizations.

Key Python Concepts Used:

- **NumPy** → Numerical computations (mean, median, standard deviation)
- **Pandas** → Data handling, filtering, cleaning, and summarization
- **Matplotlib** → Base plotting for custom visualizations
- **Seaborn** → Advanced, aesthetic statistical plots
- **Streamlit** → Creating interactive dashboards and visual analytics tools

Project Requirements:

1. Create a **Streamlit app** named data_analysis_dashboard.py.
2. Allow users to upload **CSV datasets** for analysis.
3. Provide options for:
    - Data preview and structure
    - Summary statistics
    - Visual exploration through different plots
    - Missing data analysis and cleaning
    - Downloading descriptive reports
4. Include an **interactive sidebar** for user navigation and filters.
5. Use **input validation** to ensure uploaded data is valid and readable.

Core Features:

1. Dataset Upload

- Upload .csv files via st.file_uploader().
- Automatically read the dataset using Pandas.
- Display dataset preview, number of rows/columns, and data types.

2. Data Summary

- Generate descriptive statistics.
- Compute **mean, median, and standard deviation** using NumPy.
- Display data types and column summaries.

3. Data Visualization

Interactive visualization panel using Streamlit sidebar options:

- **Histogram** → View numeric column distribution
- **Bar Chart** → Frequency of categorical variables
- **Box Plot** → Detect outliers
- **Heatmap** → View correlation between numeric columns
- **Scatter Plot** → Relationship between two variables
- **Pairplot** → Multi-variable relationships

4. Missing Data Handling

- Identify missing values per column.
- Display missing data percentage visually using a heatmap.
- Provide options to:
    - Drop missing rows
    - Fill missing values with mean

5. Download Summary

- Export descriptive statistics as a downloadable .csv report.

6. Interactive Navigation

- Sidebar menu for switching between:
    - Data Summary
    - Visualization
    - Missing Data
    - Download Report
    - About Section

Input Validation Rules:

- Only .csv files should be accepted.
- Empty or corrupted files should display an error message.
- Warnings should appear if there are no numeric or categorical columns.
- User actions (like filling missing values) should be confirmed before applying.