Assignment-1:

Working with the **Wine Quality Dataset** using Pandas, NumPy, and Matplotlib

1. Pandas: Data Manipulation and Exploration

Tasks:

- **Loading Data**:
  - Load the dataset from a URL or local CSV file.
  - Display the first few rows and check the structure of the dataset (.head(), .info(), .describe()).

- **Checking for Missing Values**:
  - Use .isnull() and .sum() to identify missing values in the dataset.

- **Data Filtering**:
  - Filter rows where alcohol content is greater than a specific value (e.g., 10%).
  - Filter rows for wines with a certain level of quality (e.g., wines rated as 7 or higher).

- **Sorting Data**:
  - Sort the dataset by columns like **quality** or **alcohol** in both ascending and descending order.

- **Group By and Aggregation**:
  - Group the dataset by **quality** and calculate the mean of other columns like **pH**, **alcohol**, or **residual sugar**.
  - Calculate the median **alcohol** for each **quality** level.

- **Adding New Columns**:
  - Create a new column that categorizes wines into "High Quality" (quality > 7) or "Low Quality" (quality <= 7).

2. NumPy: Statistical Analysis

Tasks:

- **Basic Statistics**:

  - Calculate the **mean**, **median**, **standard deviation**, **minimum**, and **maximum** for numeric columns such as **alcohol**, **pH**, and **citric acid**.

- **Correlations**:

  - Use NumPy to calculate the correlation between **alcohol** and **quality** or **residual sugar** and **quality**.

- **Working with Arrays**:

  - Convert columns like **alcohol** or **pH** into NumPy arrays and perform operations such as:

    - Normalizing the values.

    - Finding unique values in categorical columns like **quality**.

- **Statistical Analysis**:

  - Compare the mean **alcohol** content of "High Quality" wines (quality > 7) with that of "Low Quality" wines (quality ≤ 7) using statistical measures like the **t-test**.

Matplotlib: Data Visualization

Tasks:

- **Histograms**:

  - Plot a **histogram** to visualize the distribution of **alcohol** or **quality** across all wines.

- **Scatter Plots**:

  - Create a **scatter plot** to show the relationship between **alcohol content** and **wine quality**.

  - Create a scatter plot to visualize the relationship between **pH** and **citric acid**.

- **Box Plots**:

  o Generate a **box plot** to compare the distribution of **residual sugar** for different quality levels.

  o Create a box plot to show the spread of **pH** across different quality ratings.

- **Bar Plots**:

  o Create a **bar plot** to show the average **citric acid** or **alcohol content** for each wine **quality** level.

- **Line Plots**:

  o Plot a **line graph** to track the average **pH** level as a function of increasing **wine quality**.