MENTOX PROJECT-1

- Dataset : Titanic Kaggle Dataset

- **Objective**: Practice data cleaning, normalization, and basic data visualization using only **NumPy**, **Pandas**, **Matplotlib**, and **Seaborn**.

- **Introduction**
  The Titanic dataset is one of the most famous datasets in data science. It contains information about passengers aboard the RMS Titanic, including whether they survived the ship's sinking on April 15, 1912. This assignment will guide you through basic data analysis techniques focusing on data cleaning, visualization, and normalization.

- **Libraries Required:**
  - NumPy
  - pandas
  - matplotlib
  - seaborn

- **Task 1: Understand the Dataset**
  Objective: Load the dataset and perform basic inspection.
  Instructions:
  - Load the dataset using **pandas**.
  - Display the first 5 rows.
  - Use .info() and .describe() to understand the structure and summary statistics.
  - Identify which columns have missing values.

- **Task 2: Handle Missing Data**
  Objective: Perform data cleaning by handling missing values.
  Instructions:
  - Drop columns with too many missing values (e.g., "Cabin").
  - Fill missing values in the "Age" column with the **mean** or **median**.
  - Fill missing values in "Embarked" with the **most frequent value (mode)**.

- **Task 3: Normalize Numerical Data**
  Objective: Apply normalization on numerical columns.
  Instructions:
  - Use **Min-Max Scaling** or **Standardization** on columns like "Age" and "Fare".
  - Show a before-and-after comparison using .head() or simple **bar charts**.

- **Task 4: Encode Categorical Variables**
  Objective: Convert categorical data into numerical form.
  Instructions:
  - Use **label encoding** or **one-hot encoding** on columns like "Sex" and "Embarked".
  - Drop the original columns after encoding if needed.

- **Task 5: Visualize the Data**

  Objective: Generate simple and meaningful visualizations.

  Instructions:

  - Create **bar charts** showing counts of passengers by:
    - "Pclass" (Passenger Class)
    - "Sex"
    - "Survived"
  - Create **pie charts** to show distribution of:
    - "Sex"
    - "Embarked"
