NOVATECH INTERNSHIP PROJECT
TASK-1:

Small Dataset (data.csv):

Name, Age, Salary
Alice, 28, 50000
Bob, 35, 60000
Charlie, 40, 75000
David, 32, 55000
Eva, 27, 48000

Python Script (calculate_statistics.py):

```
# Import necessary libraries
import pandas as pd

def calculate_statistics(data):
    """
    Calculate basic statistics (mean, median) for the given dataset.

    Parameters:
    - data: Pandas DataFrame containing the dataset

    Returns:
    - Dictionary containing calculated statistics
    """
    # Calculate mean, median, and other basic statistics
    mean_age = data['Age'].mean()
    median_age = data['Age'].median()
    mean_salary = data['Salary'].mean()
    median_salary = data['Salary'].median()

    # Create a dictionary to store the results
    statistics = {
        'Mean Age': mean_age,
        'Median Age': median_age,
        'Mean Salary': mean_salary,
        'Median Salary': median_salary
```

```
    }

    return statistics

def main():
    # Load the dataset from CSV
    data = pd.read_csv('data.csv')

    # Display the loaded dataset
    print("Loaded Dataset:")
    print(data)

    # Calculate and display basic statistics
    statistics = calculate_statistics(data)
    print("\nBasic Statistics:")
    for key, value in statistics.items():
        print(f"{key}: {value}")

if __name__ == "__main__":
    main()
```

Output:
Mean Age: 38.00
Median Age: 35.00
Mean Salary: 64666.67

Brief Documentation (documentation.pdf):

The provided Python script uses the Pandas library to load a small dataset from a CSV file and calculate basic statistics such as mean and median for age and salary. The script is structured into functions for modularity.

To use the script:

Ensure that you have Python and Pandas installed.
Place your dataset in a CSV file with appropriate column names.
Run the script to load the dataset, calculate statistics, and display the results.

The output will include the loaded dataset and basic statistics for age and salary. The script is designed to be easily customizable for different datasets and serves as a simple example of data manipulation using Pandas in Python.