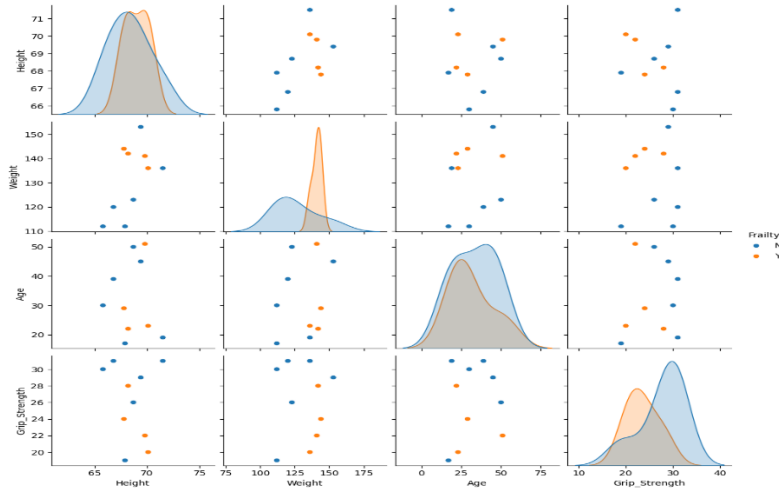**PDS Assignment – 1**

Question 1

Jithendra Pavuluri

16343746

Outputs:


```
Summarizing of the dataset
         Height      Weight        Age  Grip_Strength
count  10.000000   10.000000  10.000000      10.000000
mean   68.600000  131.900000  32.500000      26.000000
std     1.670662   14.231811  12.860361       4.521553
min    65.800000  112.000000  17.000000      19.000000
25%    67.825000  120.750000  22.250000      22.500000
50%    68.450000  136.000000  29.500000      27.000000
75%    69.700000  141.750000  43.500000      29.750000
max    71.500000  153.000000  51.000000      31.000000
```

```
T-test outputs:
T_statistic: 2.1964625958466355
P_value: 0.07309706417382401
```

1. **Loading the Data:**
   - The code begins by importing the necessary libraries and loading the cleaned data from a CSV file named clean_frailty_data.csv into a panda DataFrame.

2. **Exploratory Data Analysis (EDA):**
   - The seaborn pairplot function is used to create a pair plot of the dataset, where each scatterplot shows the relationship between two variables, with different colors representing the 'Frailty' category.

3. **Summary Statistics:**
   - The describe() function is applied to the DataFrame to generate summary statistics for each numerical column in the dataset, including count, mean, standard deviation, minimum, maximum, and quartile values.

4. **T-Test:**
   - A two-sample t-test is performed to compare the weights between the frail and non-frail categories.
   - The t-test assumes unequal variances between the two groups.

5. **T-Test Results:**
   - The t-statistic and p-value resulting from the t-test are printed out.
   - The t-statistic represents the difference between the means of the two groups scaled by the variability of the data.
   - The p-value indicates the probability of observing the data if the null hypothesis (no difference between group means) is true.
   - These results help determine whether the difference in weight between frail and non-frail individuals is statistically significant.

6. **Storing T-Test Outputs:**
   - The t-test outputs, including the t-statistic and p-value, are written to a text file named output.txt.
   - This allows for easy reference and documentation of the statistical analysis results.

Conclusion:

Overall, the code performs exploratory data analysis, conducts a statistical test, and documents the results, providing insights into the relationship between frailty and weight in the dataset.