Home Assignment

You have a table called 'factory_test.csv' that contains test data for 51 devices from the population of devices tested in the factory- the “sample group”.

The dataset contains 5 dependent variables (features) Y1 ... Y5 and 4 independent variables X2 ... X5, such that the variable Yi corresponds to Xi, except for Y1 which has no X variable. To analyze this data, write a python script using the specific libraries numpy, pandas, and matplotlib. **No other libraries are allowed for this analysis.**

To submit your result, send us a link to a **git repository** with **one** of the following options:

1. **.py script and README.md file** displaying your work + explanation
2. Pre-evaluated **.ipynb** notebook displaying your work + explanation

According to the requirements below.

The script shall contain the following parts:

1. Data Loading and Description

   Load the data from 'factory_test.csv' and provide a description of its contents.

2. Data Visualization and Explanation

   Visualize the data and explain the observed patterns or trends.

3. Data Preprocessing

   Perform any necessary preprocessing on the data.

4. Data Exploration

   You have also received another file called ‘new_devices.csv’ containing data for 3 additional devices. Please rank these new devices based on their probability of belonging to the sample group.

5. Bonus

   Suggest and implement a method for testing new devices based on your knowledge of the sample group and specify which devices (sample group and new devices) fail your tests and why.