IC 272: Lab3: Outlier detection, Standardization and Normalization of data

You are given with “landslide_data2_original.csv”. This dataset contains the readings from various sensors installed at 10 locations around Mandi district. These sensors give the details about the factors like temperature, humidity, pressure etc. Write a python program to do the following.

1. Read the data into a dataframe using pandas. Obtain the boxplot for the attributes “temperature”,”humidity” and “rain” . Observe the number of outliers in these attributes and their values. Outliers are the values that do not satisfy the condition: **(Q1 - 1.5 * IQR) < X < (Q3 + 1.5 * IQR)** where, **IQR** is the **Interquartile range (= Q3-Q1), where Q1** and **Q3** are the lower and upper quartiles. Replace these outliers with the median of the attribute. Plot the boxplot again and observe the difference. Do you still get outliers? Why?

2. Observe the range of the values in these three attributes (Use the data obtained after outlier correction). Find the minimum and maximum values in each attribute.
i) Perform the Min-Max normalization of this data
ii) Perform Min-Max normalization to have the range of values between 0-20.

3. Use the data obtained after outlier correction. Find the mean and standard deviation of the attributes. Standardize these three attributes using the relation $Xnew=(X-\mu)/\sigma$ where μ is mean and σ is standard deviation. Compare the mean and standard deviations before and after the standardization.