**Data Cleaning**

The initial data provided by Colin from Linus Tech Tips was in a Microsoft Excel spreadsheet with all the tests put together. To make each test more friendly to load into Pandas Dataframes, each of the tests were copied into individual csv files. This allows for each test to be loaded into a separate Dataframe without having to split them apart using Pandas operations. Column names in the tests were renamed to be more consistent with each other by taking the version that was the most concise. These first steps were done directly to the csv files, outside of Dataframe manipulation in Python. Now the data is ready to be loaded into Pandas Dataframes.

The data itself is very complete with only a total of 11-25 missing values for each of the sensors out of around 4900 rows for each of the tests. Removing NA values in the Dataframe was enough to get usable data. Duplicate columns were assigned a “.X” for each duplicate of the same column. A cursory look at each of the Dataframes shows no extreme outliers in the low temperature range or in the high temperature range when looking at the min and the max values.