I’m currently using a h5 file to store a large amount of data I want to analyse this data is all stored at the lowest level and organised by the levels above. At these levels I want to be able to analyse the data underneath at the lowest level to establish trends and compare. At the lowest level there is two data frames stored as pd data frames with the suffix .txt\_metrics or .txt\_info

an example of the layout:

Zn-Cu-In-S(Zns)/D65-0.05mgml-ITO-PMMA(3%)-Gold-s5/G/9/1-fs-1v-0.05sv-100na.txt\_metrics

Zn-Cu-In-S(Zns)/D65-0.05mgml-ITO-PMMA(3%)-Gold-s5/G/9/1-fs-1v-0.05sv-100na.txt\_info

Where you can name the levels the following

Material / sample / section / device/ filename

With device being number 1-10 and section A-L

I have several samples to compare with one another and for each sample there is 10 sections containing 10 devices with each device having anywhere from 1- 100 filenames all of which I want to be able to compare easily.

I want you to help me with a few things:

I would like to for example take all the filenames that start with “1-“ storing the dataframe and the key in a dataframe for use later.

I would also like to take all the data in each section and compare it with another section for example I can find using the txt\_metrics which section had the largest resistance, as an example. This is so I can see if there is a trend that section G has the largest resistance across all samples. I would like to be able to do this for the device as well.

I have attempted this a few times but cant do it please give me some python code that would do this