Programming and Scripting

Lab Topic 07-Files extra excel

## Introduction.

A ex student of this course asked me to look at an issue he has with an excel file.

I would not expect any of you to tackle this, but I though it might be interesting for you to see me fumble through this.

## Requirements:

He gets data in an excel spread sheet that is in this format there are two sets on the same sheet

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Date | 00:30 | 01:00 | 01:30 | 02:00 | 02:30 | 03:00 | 03:30 | 04:00 |
| 10120 | 10 | 1 | 1 | 2 | 0 | 0 | 0 | 0 |
| 20120 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 30120 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 40120 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 |
| 50120 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

He wants to produce another sheet in this format

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date | Time | set1 | set 2 | difference |
| 01/01/20 | 00:30 | 10 | 100 | 90 |
| 01/01/20 | 01:00 | 1 | 11 | 10 |
| 01/01/20 | 01:30 | 1 | 21 | 20 |
| 01/01/20 | 02:00 | 2 | 22 | 20 |

…….

A copy of a real file can be found here.

<https://github.com/andrewbeattycourseware/programmingandscripting/blob/master/topic07-X-extra-reading-excel/MPRN123456789Jan2020.xlsx>

## Tasks that need to be done

1. Read from an excel spreadsheet
2. Write to an excel spreadsheet
3. Read all the values from column 1 identifying when the first set ends
   1. Convert that data into a proper time format
4. Read the data from col2 to colN store in list. Of Dict objects
5. Do this for all rows in set1, putting data into one set
6. Write that set into new workbook (or new sheet in this workbook)
7. Identify where set 2 starts
8. Do the same for set 2
9. Write out results for set 2
10. Go through both lists subtracting one from the other (do we need to check for corrupt data?)
11. Write out the difference

Ok there is a lot to do with this.

But I have broken the project into a number of different tasks we can try solving these one by one.

## Read from an excel sheet

After doing a quick search on reading excel files in python, I came across

<https://openpyxl.readthedocs.io/en/stable/>

openpyxl

to use this we need to import the package

from the command line

pip install openpyxl

and after reading the cookbook link here

<https://openpyxl.readthedocs.io/en/stable/usage.html#read-an-existing-workbook>