WMO Normals Formatting Plan

Am I trying to format it all? The normal data set

1. Open the WMO Normals Template (under sheet template)
2. Open the WMO Normals Data Set (atm only one sheet, so open first sheet)
3. Open StationList (Query1 file)
   1. Extract the name from the WMO Normals Data Set
      1. Search through until find it 554
   2. Find the name in the station list under the station name header (col B)
   3. Station Name – B7 in the template
   4. Country Name – col J -> B6 in the template
   5. WMO Number - col F -> A10 in the template
   6. Latitude – col L -> B10 in the template
      1. Have to put in right format
   7. Longitude – col M -> C10 in the template
      1. Have to put in right format
   8. Station Height – col N -> D10 in the template
   9. WIGOS ????
   10. Close the StationList
4. Open NormalID to WMO Parameter and parse through this list (Not all station id will be used?)
   1. Want to populate a sheet for each of the stations
   2. Create an array of the normal IDs
      1. Most will be size 1, except for the quintiles
   3. Store the parameter code in a int value
   4. Use the parameter name to find the row on the template (search col B of the template), and go three down
      1. We use parameter name because sometimes the code has multiple things
   5. Iterate through the calculation name section of the normal ID until hit NOY (use for loop so we have an iterator variable). For each:
      1. Put in the WMO number we stored from above into col A of the template
      2. Put the parameter code in col B of the template
      3. Put this given calculation into col C of the template
      4. Put the calculation code in col D of the template (Ask how often the calculation codes change)
         1. Consider parameters and calculation methods could change
      5. Find the normal ID from the array based on the iterator variable (will have to adjust)
         1. If the value is NOY, find the normal ID by using iterator variable minus 1
      6. Find the corresponding place in the data set (col C)
      7. Iterate until the normal ID is not the same
         1. Take col F values
         2. For NOY take year count (col K) values
5. Close all workbooks

Question: how adaptable are the documents? Will they all follow the same format or will they change in the future? Have you seen the standards change drastically

Notes:

1. 13 means annual.
2. Assumption: the file given is the right station, so do not need to check the name is correct
3. Idea: keep a String name of the element
   1. If the element in question is the same, just continue.
   2. If the element in question is different, rename the string and find the row index it needs to be at.
4. Original data file does not contain column B and D
5. Station ID is virtual station ID
6. Year count col K
7. Extreme values need col G date of occurrence
8. Might want to add an element or two extra elements, so think of a way to add that
9. Put notes as modification (station data parameters)
10. Csv might be easier
11. Station height has to be rounded to whole number
12. Curtual station ID
13. Only do the ones with virtual station ids
14. Python might be better

Have done my mid feb

WMO Normals Formatting Plan Revised

1. Import pandas as pd
2. Specify the path to the WMO Normals Data Set csv
   1. Read the csv file into a data frame
3. Specify the path to the template csv (WMO Normals Template csv)
   1. Read the csv file into a data frame
4. Specify the path to the StationList csv
   1. Read the csv file into a data frame (make sure to specify the header is different)
5. Note: maybe reformat the stationList so that we know past a certain index are values that we want to input into the list (this can increase reusability if there are different parameters we want to input in the future)
   1. Therefore, we can find the index of each of the values once instead of have to find for each station (\*)
6. Specify the path to the NormalID to WMO Parameter ID
   1. Read the CSV file
7. Or loop through the Normals Data Set
   1. Use group by Station ID and then iterate through the different groups
   2. make a copy of the template data frame (.copy())
   3. get the value of the station if and find it from the virtual station ID in the station List
   4. Input all the general info
      1. Using the parameters specified above and the column headers, input the information in the right place (\*)
   5. Parse through the NORMAL\_ID (could use group by again)
      1. For each, find the corresponding WMO parameter and name (SPLIT UP QUINTILES)
      2. Find this WMO parameter in the template (create a finder function)
      3. Loop through the calculation name
         1. Usually col value (F)
         2. For NOY, loop through YEAR\_COUNT (K)
         3. For extremes, need to have date (FIRST\_OCCURENCE (G)
      4. Insert line??

Alternate for 8:

1. Loop through the StationList
   1. Check if there is a value for the virtual station ID
   2. If there is, make a copy of the template data frame (.copy())
   3. Input all the general info
      1. Using the parameters specified below and the column headers, input the information in the right place
   4. Create a variable that holds the given virtual station ID
   5. Use df[condition] on the data set to select the rows for that specific station