excel_analysis.py program setup

By: Spencer Varzandeh

Background

- Python program designed to export an excel sheet containing the desired function’s analysis of an .xlsx file
- Utilizes Python’s pandas module to create Dataframes that store an excel file’s data.

- Functionality:
  - newLayers - Gets the layers that are in the new excel, but not in the old.
  - missingLayers - Gets the layers that are in the old excel, but not in the new.
  - remNullLayers - Gets an excel that excludes the rows with null values in the selected column.

Setup

1. Install python
   a. Follow the link https://www.python.org/downloads/ (this is the main distribution for Python).
   b. Download the version 3.10.4.
   c. Allow the install client to run. Python should install within the Documents folder of your C: Drive.
2. Install Pandas
   a. From the windows Command Prompt, run the following command: ‘pip install pandas’.
   b. Allow the command to finish executing to install Pandas.
3. Download the python _master folder containing the excel_analysis.py program. Recommended to install the folder to your desktop.
4. Copy the excel files that you wish to analyze to the _master folder on your desktop. (This allows you to enter the relative file path (shorter), rather than the absolute path (the long file path from your C: drive).

Executing the program

1. Open a command prompt on your computer.
2. Navigate to the _master folder that you setup on your desktop in the previous slide.
    a. Use the command ‘cd folder_name’ to go inside folder_name, and use ‘cd..’ to go back one directory. The easiest way to do this is to use ‘cd..’ a few times until you reach your root directory, which is just your C: drive. Then from there, use ‘cd folder_name’, and enter folders until you reach the folder within your User named Desktop. From Desktop, cd into ‘_master’ to reach the program’s directory.
3. Into the command prompt, enter ‘python excel_analysis.py’.
    a. This runs the program named ‘excel_analysis.py’ within the current directory that you navigated through in the previous step.
4. Choose a function to run, and write that in the command prompt line. The function options should print if the program was run correctly.
5. The program will then ask for the directories of the two excels that you wish to analyze (or one excel for the remNullRows function). Input these, along with the desired sheet name and column name.
    a. For the directories of the two excel sheets, you can use relative path instead of absolute path by placing the two (or one) excels into the same folder as the program. Then, when the program asks for the file path, you need only to input the file’s name as the path.
6. The program should then export the Dataframe analysis as an excel file within the same directory which you can now view in Microsoft Excel, Google Sheets, or any other .xlsx viewer/editor.

QuickFix Notes

- If the program is exporting an empty excel file, the program may not be finding the excel files properly. If this is the case, retry the program using the absolute directory. This is best achieved by navigating to the file location through the File Explorer window and then copying the path all the way from the C: drive.
- If the program fails to run due to the error ‘pandas module could not be found’, your pandas install may have failed, or Python cannot find the path to Pandas in order to use the module. If you get to this error, attempt to uninstall and reinstall pandas. If this still fails, I recommend uninstalling both Python and Pandas, then reinstalling them both (following the instructions in Slide 3, Steps 1 & 2).

newLayers and missingLayers functions

- Both analyze the layers between two excels, and require that both excel sheets are compared from the same column name.
- newLayers finds the layers that are in the new excel, but not in the old excel.
- missingLayers finds the layers that are in the old excel, but not in the new excel.
- These are useful to find what new layers were added. The most current and widely used software for Excel analysis, AbleBits, costs money and requires that the excel documents analyzed contain the same number of rows and columns, making it incompatible for QGIS layer exports.
- THIS FUNCTION REQUIRES 2 EXCELS TO WORK PROPERLY.

remNullRows function

- Exports an excel that has removed null values from the specified column.
- Specifically for the Layers vs Plans analysis, this function can be run on the ‘NO_PLAN’ column to find which layers do not have an associated plan (or are template layers). The function can also be run on the ‘Plan_1’ column to display which layers do have (an) associated plan(s).

Future Build Notes

- As of now, the program works only on excel files.
- In the future, the program will work with excel, csv, or (Geo)JSON files.

Please feel free to email me with any questions, concerns, or suggestions:
svarzandeh@thinkbignets.com