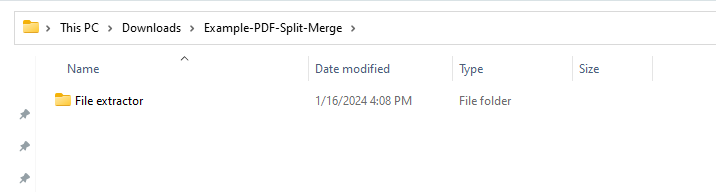
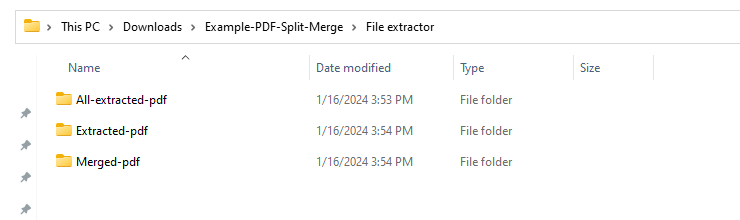
**PDF Attachments Split and Merge**

**Script Execution Instructions:**

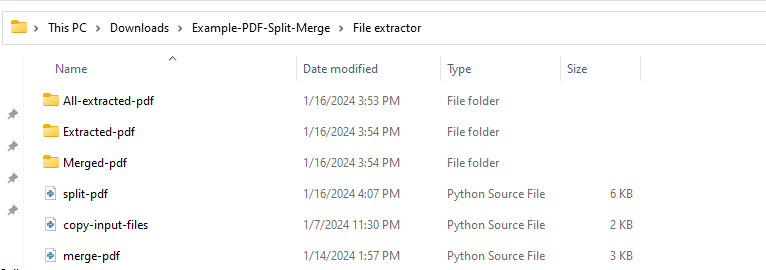
1. Create root folder: File extractor



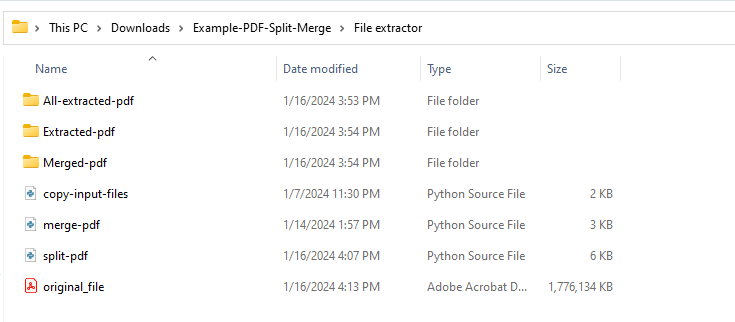
1. Create sub folders inside root folder: All-extracted-pdf, Extracted-pdf, Merged-pdf



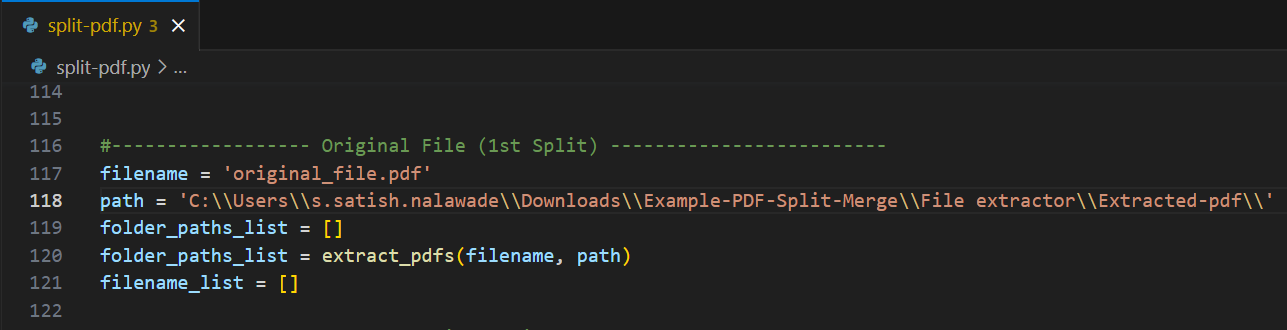
1. Create 3 python scripts with .py extention inside root folder: split-pdf.py, copy-input-files.py, merge-pdf.py

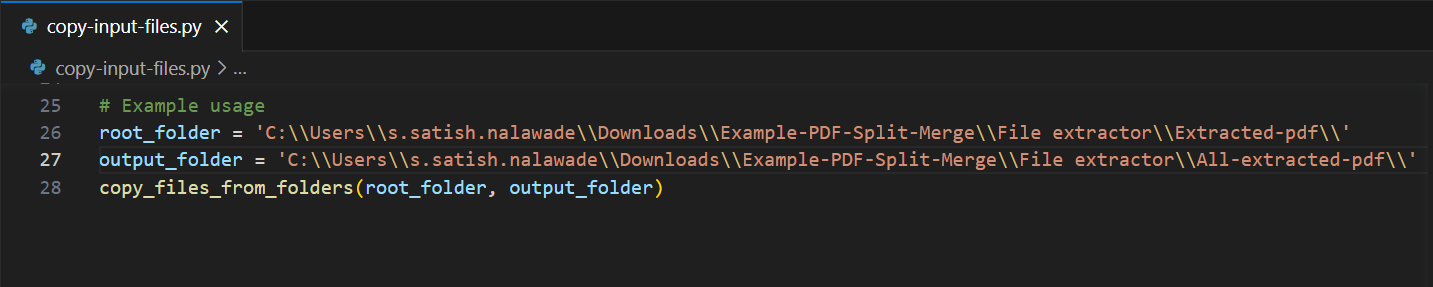


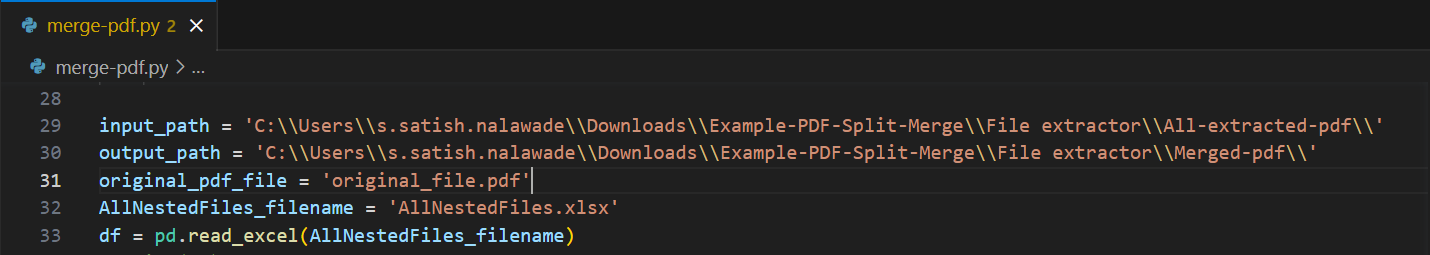
1. Place original\_file.pdf (original PDF file with attachments) inside root folder.



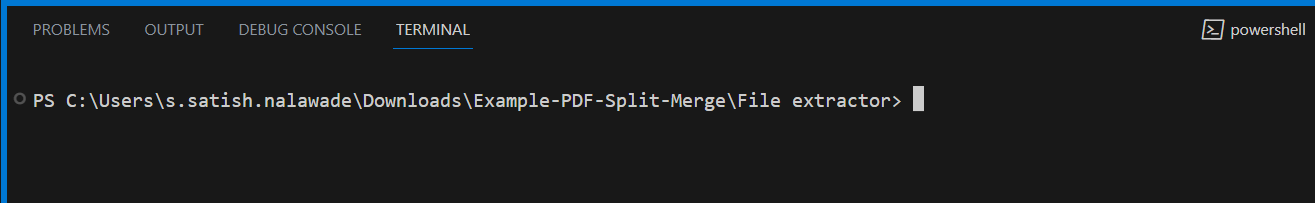
1. Set all the path and file name variables in all 3 scripts with correct values.



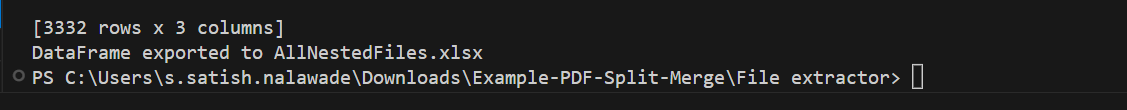
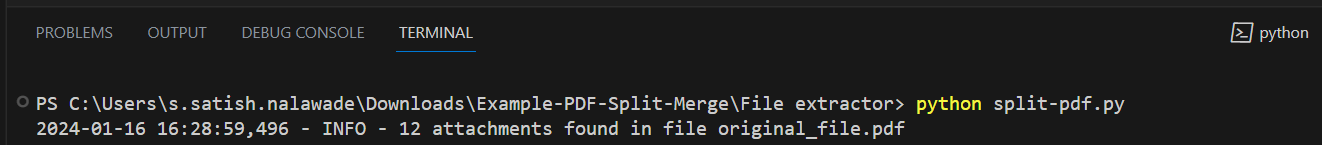


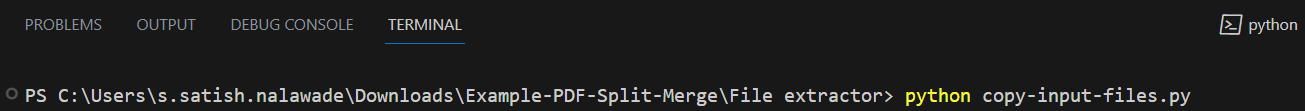


1. Open cmd or terminal and change current working directory to root folder (cd 'File extractor')

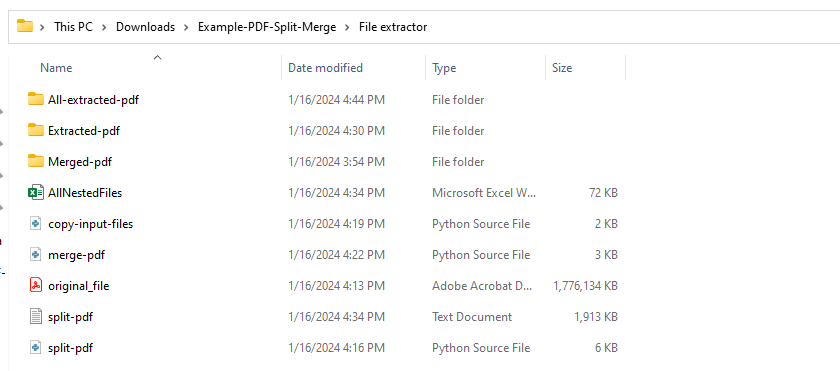


1. Execute python scripts in sequence by executing below commands in cmd or terminal: python split-pdf.py **->** python copy-input-files.py

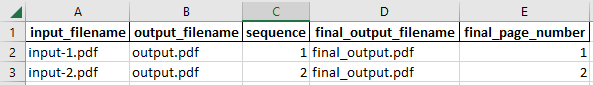




1. Update 'AllNestedFiles.xlsx' file with grouping logic.



Format of '*AllNestedFiles.xlsx*' file:



**input\_filename:** Name of the input files which needs to be merged.

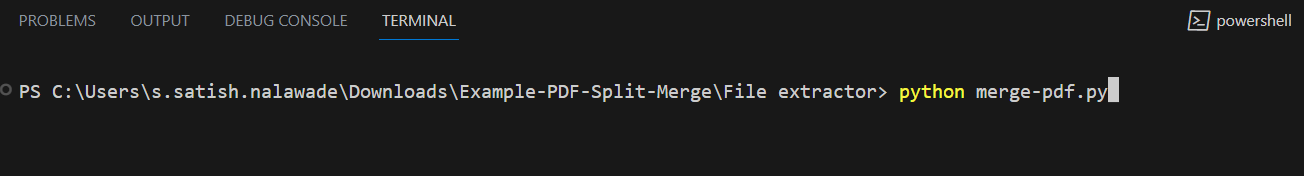
**output\_filename:** Name of the source file from which input files are extracted.

**sequence:** sequence number of the input files in original file tree hierarchy (sequence of attachments).

**final\_output\_filename:** Name of the merged file (output file) to be generated from input files.

**final\_page\_number:** Page number of the input files in the final merged file.

1. Execute python script by executing below command in cmd or terminal: python merge-pdf.py



1. Do these validation checks to ensure successful execution:
2. Check if all the attachments extracted from original PDF are present in 'Extracted-pdf' folder with the exact hierarchy with sub folders.
3. Check if 'AllNestedFiles.xlsx' file is present in 'File extractor' folder and contains all the expected information.
4. Check if all the attachments from 'Extracted-pdf' folder and its sub folders are copied in 'All-extracted-pdf' folder.
5. Check if all the merged pdf files are present in 'Merged-pdf' folder based on defined grouping logic in 'AllNestedFiles.xlsx'.

**Description:**

* **split-pdf.py:** extract attachments from original pdf with exact hierarchy in pdf tree represented by sub folders inside Extracted-pdf folder.
* **copy-input-files.py:** copy attachments from Extracted-pdf folder and its sub folders to specific destination folder All-extracted-pdf.
* **merge-pdf.py:** merge all the pdfs from All-extracted-pdf folder based on grouping logic defined in 'AllNestedFiles.xlsx' and place generated merged pdfs in Merged-pdf folder.