**Convert multiple corrupt XLS files to XLSX | Python in Finance #5**

[[Edward Jones](https://medium.com/@pythonprogramming?source=post_page-----1da1dc1cf89d--------------------------------)](https://medium.com/@pythonprogramming?source=post_page-----1da1dc1cf89d--------------------------------)

[[Geek Culture](https://medium.com/geekculture?source=post_page-----1da1dc1cf89d--------------------------------)](https://medium.com/geekculture?source=post_page-----1da1dc1cf89d--------------------------------)

[Edward Jones](https://medium.com/@pythonprogramming?source=post_page-----1da1dc1cf89d--------------------------------)

·

[Follow](https://medium.com/m/signin?actionUrl=https%3A%2F%2Fmedium.com%2F_%2Fsubscribe%2Fuser%2Fd852141f761f&operation=register&redirect=https%3A%2F%2Fmedium.com%2Fgeekculture%2Fconvert-multiple-corrupt-xls-files-to-xlsx-python-in-finance-5-1da1dc1cf89d&user=Edward+Jones&userId=d852141f761f&source=post_page-d852141f761f----1da1dc1cf89d---------------------post_header-----------)

Published in

[Geek Culture](https://medium.com/geekculture?source=post_page-----1da1dc1cf89d--------------------------------)

·

2 min read

·

May 27, 2021

5

2

In this article you will learn how to convert a batch of corrupt XLS files to an XLSX file.

A screenshot of a computer

Description automatically generated with low confidence

Trade-off of automating this process

In this article, you will learn how to automate the conversion process from (corrupt) xls files to normal xlsx files.

Data can be found: [here](https://mega.nz/file/cJQBUChb#25rBjAheYG4sDsZAg19k5XsbhCv-_Jz-auVEn7n21uA)

**Packages**

* [pywin32](https://github.com/mhammond/pywin32): This package is basically vba for python. It allows us to interact and automate Windows applications with python.
* [os](https://docs.python.org/3/library/os.html): This package allows us to use the operating system.
* [glob](https://docs.python.org/3/library/glob.html): This package allows us to create a list with the different file locations

import win32com.client  
import os  
import glob

**Initialization**

* You initialize the win32com and let it run an excel application.
* Set the object “o.Visible = False” in order to hide the excel application you have created.
* Set the input directory: This is the directory where all your corrupt xls files are located.
* Set the output directory: This is the directory where you want to store the converted xlsx files.
* Create a list of file paths of the all the files in the input directory by using the glob.glob function

o = win32com.client.Dispatch("Excel.Application")  
o.Visible = Falseinput\_dir = r"C:\input\_directory"  
output\_dir = r"C:\output\_directory"  
files = glob.glob(input\_dir + "/\*.xls")

**Loop**

For each file, you perform the following manipulations:

* Assign the filename to the object “file”
* Set the ouput filename by concatenating the output directory and the output filename replacing the xls extension with the xlsx extension.
* Open the xls files by the excel application
* Save the xls file as an xlsx file using the output path as previously specified
* Close the file

for filename in files:  
 file = os.path.basename(filename)  
 output = output\_dir + '/' + file.replace('.XLS','.xlsx')  
 wb = o.Workbooks.Open(filename)  
 wb.ActiveSheet.SaveAs(output,51)  
 wb.Close(True)

Using these bits of code, you just quickly automated a task that would cost you normally 40 minutes.

Happy programming!

Cheers,

Ewoud