



To make Medium work, we log user data.
By using Medium, you agree to our
Privacy Policy, including cookie policy.

in app

Get started



Published in DataDrivenInvestor

You have **1** free member-only story left this month. [Sign up for Medium and get an extra one](#)



Daniel Ellis

Follow

Feb 24 · 3 min read ★ · Listen

Save



Writing Excel Files into CSVs (Using Python)

Converting the proprietary XLSX format into something more accessible.



Photo by [Markus Spiske](#) on [Unsplash](#)





To make Medium work, we log user data.
By using Medium, you agree to our
Privacy Policy, including cookie policy.

Sign in

Get started

workbook interface. In the process, the workbook is converted into a series of comma-separated tables.

Excel spreadsheet

The Dataset

We take a sample dataset representing the number of COVID related deaths from the Office of National Statistics website.

The file can be found [here](#).

Installing the Relevant Python Libraries

The next step is to set up our work environment. We start by installing the openpyxl and pandas libraries which we will then use within our processing script.

```
pip install openpyxl pandas
```

These are then imported within our python script:

```
from openpyxl import load_workbook  
import pandas as pd
```

Reading the Workbook

After initializing the libraries we want to load our workbook. This is done using:

```
src_file = './path/to/file.xlsx'  
  
wb = load_workbook(filename = src_file)
```

Getting Workbook Sheets





To make Medium work, we log user data.
By using Medium, you agree to our
Privacy Policy, including cookie policy.

[Sign in](#)[Get started](#)

This allows us to access it:

```
wb['my_sheetname']
```

or iterate through all given sheets using a for loop:

```
for sheet in sheets:  
    wb[sheet]
```

Extracting Tables from a Sheet

If the spreadsheet contains purely data, we can read its entirety and convert it into a pandas data frame with:

```
pd.DataFrame(wb[sheet])
```

If the data provider has been exceptionally unfriendly (which is often the case with excel users), they may have occluded the sheet using several tables within a single page. This means we have to perform some additional work to get our data.

Reading Multiple Tables

We start by looking at all the tables that may be contained on the sheet:

```
tablenames = wb[sheet].tables.keys()
```

Once we know what table we are interested in, we can use this to reference the sheet:

```
selection = wb[sheet].tables[table] # where table is in  
tablenames
```





To make Medium work, we log user data.
By using Medium, you agree to our
Privacy Policy, including cookie policy.

Sign in

Get started

```
wb[sheet][selection.ref]
```

Here the reference is in the form of an excel, cell selection query (e.g. `C4:H11`).

Converting the Table into a DataFrame

As we have a sheet selection we can read the cells into a data frame and consequently their values using:

```
df = pd.DataFrame(wb[sheet][selection.ref]).apply(lambda x:  
[y.internal_value for y in x])
```

We now need to correct the column headers and index:

```
df = df.set_index(0,inplace=False).rename(columns=df.iloc[0],  
inplace = False).iloc[1:]
```

and optionally save the table:

```
df.to_csv('outputfilename.csv')
```

Summary

In this tutorial, we have used `openpyxl` to extract the tables from an excel workbook and `pandas` to save them into a comma-separated-value file. The code used is available at the following GitHub repository:

GitHub - wolfix/excel2csv: Convert the inaccessible and proprietary office spreadsheets to...





I hope this tool proves u
organisations.

To make Medium work, we log user data.
By using Medium, you agree to our
Privacy Policy, including cookie policy.

in app

Get started

ormats are adopted by

Schedule a DDIChat Session in **Coding, Software, and Mobile Development**:

Experts - Coding, Software, and Mobile Development - DDIChat

DDIChat allows individuals and businesses to speak directly with
subject matter experts. It makes consultation fast...

app.ddichat.com

Apply to be a DDIChat Expert here.

Work with DDI: <https://datadriveninvestor.com/collaborate>

Subscribe to DDIntel here.



Get an email whenever Daniel Ellis publishes.

Subscribe

