**Excel Files**

Although CSV files are a convenient and simple way to handle data; it is very common to need to be able to read or write Excel files directly. To this end there are several libraries available in Python for this purpose. One widely used library is the OpenPyXL library. This library was originally written to support access to Excel 2010 files. It is an open source project and is well documented. The OpenPyXL library provides facilities for

• reading and writing Excel workbooks,

• creating/accessing Excel worksheets,

• creating Excel formulas,

• creating graphs (with support from additional modules).

As OpenPyXL is not part of the standard Python distribution you will need to install the library yourself using a tool such as Anaconda or pip (e.g. pip install openpyxl). Alternatively, if you are using PyCharm you will be able to add the OpenPyXL library to your project.

The key element in the OpenPyXL library is the Workbook class. This can be imported from the module:

from openpyxl import Workbook

wb = Workbook()

A new instance of the (in memory) Workbook can be created using the Workbook class (note at this point it is purely a structure within the Python program and must be saved before an actual Excel file is created).

**The Openpyxl. WorkSheet Objects**

A workbook is always created with at least one worksheet. You can get hold of the currently active worksheet using the Workbook.active property:

ws = wb.active

You can create additional worksheets using the workbooks’ create\_sheet () method:

ws = wb.create\_sheet('Mysheet')

You can access or update the title of the worksheet using the title property:

ws.title = 'New Title'

The background colour of the tab holding this title is white by default. You can change this providing an RRGGBB colour code to the worksheet. sheet\_properties.tabColor attribute, for example:

ws.sheet\_properties.tabColor = "1072BA"

WorkBook Cells

It is possible to access the cells within a worksheet. A cell can be accessed directly as keys on the worksheet ,This returns a cell object:

ws['A1'] = 42 #to set the value

cell = ws['A1'] #to get the value

We can obtain the value of the cell using the value property:

print(cell.value)

Access to the cells using row and column notation:

d = ws.cell(row=4, column=2, value=10)

A row of values can also be added at the current position within the Excel file using append:

ws.append([1, 2, 3]) #This will add a row to the Excel file containing 1, 2, and 3

Ranges of cells can be accessed using slicing:

cell\_range = ws['A1':'C2']

Ranges of rows or columns can also be obtained:

col = ws['C']

col\_range = ws['C:D']

row10 = ws[10]

row\_range = ws[5:10]

The value of a cell can also be an Excel formula such as:

ws['A3'] = '=SUM(A1, A2)'

These workbooks can be saved using the save() method. This method takes a filename and writes the Workbook out in Excel format:

workbook.save('balances.xlsx')

*#example*

​

**from** openpyxl **import** Workbook

​

**def** main():

*#first create work book object*

workbook **=** Workbook()

*#Second get the current active worksheet in the workbook,*

ws1 **=** workbook.active

*#by default every workbook will come with one worksheet so create title for it*

ws1.title **=** "First WorkSheet"

*#Just to add bg color*

ws1.sheet\_properties.tabColor **=** '1072BA'

*#Set some values to the cell*

ws1['A1'] **=** 40

ws1['A2'] **=** 50

ws1['A3'] **=** '=SUM(A1,A2)' *#A3 cell should have formula for adding 2 cell values*

*#Create second work sheet in the same workbook*

ws2 **=**workbook.create\_sheet(title **=** "Second Worksheet") *#while creating the sheet we can pass the tile parameter also*

*#Set some values to the cell in the second worsheet*

ws2['A1'] **=** 3.42

*#we can append some values also in this worksheet*

ws2.append([20,30,40])

ws2.append([50,60,70])

ws2.cell(column **=** 2,row**=**1,value **=** 15) *#you can even add the values to the specified cell using this cell option and provide*

*#the values to it.*

*#Now save the Workbook with some name*

workbook.save('ExcelExample1.xlsx')

**if** \_\_name\_\_ **==**'\_\_main\_\_':

main()

​

​

​

**Loading a Workbook from an Excel File**

To import data from an existing Excel file can be done using the OpenPyXL load\_workbook() function. This function opens the specified Excel file (in read only mode by default) and returns a Workbook object.

from openpyxl import load\_workbook

workbook = load\_workbook(filename='sample.xlsx')

You can now access a list of sheets, their names, obtain the currently active sheet etc. using properties provided by the workbook object:

• workbook.active returns the active worksheet object.

• workbook.sheetnames returns the names (strings) of the worksheets in this

workbook.

• workbook.worksheets returns a list of worksheet objects.

*#Example*

​

**from** openpyxl **import** load\_workbook *#for creating the workbook import workbook , to read the existing workbook import load\_workbook*

​

workbook **=** load\_workbook(filename **=** "ExcelExample1.xlsx")

​

print("The current active worksheet is: ",workbook.active)

print("Number of availble worksheet are :",workbook.sheetnames)

print("To get the object of the worksheets :",workbook.worksheets)

​

*#get specific worksheet and access the values in that sheet*

​

ws1 **=** workbook["First WorkSheet"]

​

print(ws1['A1']) *#prints the object of the A1 cell in worksheet Firstworksheet*

​

print("-- " **\*** 10)

print(ws1['A1'].value) *#prints the value present in the cell*

print(ws1['A2'].value)

print(ws1['A3'].value)

​

print("--" **\*** 10)

​

​

ws2 **=** workbook['Second Worksheet']

print("--" **\***10)

print(ws2['A1'].value)

print(ws2['A2'].value)

print(ws2['B1'].value)

print(ws2['C1'].value)

print(ws2['C2'].value)

​

*#getting data through range*

​

*# cell\_range = ws2['A1': 'C3']*

*# print(cell\_range)*

*# for i,cell in enumerate(cell\_range):*

**for** cell **in** ws2.values: *#can use values property and access all the values present in that sheet*

print(cell)

​

**for** cell **in** ws2.iter\_rows(min\_row**=**1,max\_col**=**3,max\_row **=** 3,values\_only **=** **True**):

print(cell)

The current active worksheet is: <Worksheet "First WorkSheet">

Number of availble worksheet are : ['First WorkSheet', 'Second Worksheet']

To get the object of the worksheets : [<Worksheet "First WorkSheet">, <Worksheet "Second Worksheet">]

<Cell 'First WorkSheet'.A1>

-- -- -- -- -- -- -- -- -- --

40

50

=SUM(A1,A2)

--------------------

--------------------

3.42

20

15

None

40

(3.42, 15, None)

(20, 30, 40)

(50, 60, 70)

(3.42, 15, None)

(20, 30, 40)

(50, 60, 70)

For reference and more reading , please check this link

​

https://openpyxl.readthedocs.io/en/stable/