**Section 17: Working with PDFs and Spreadsheet CSV Files**

**24.03.**

**130. Working with CSV Files in Python**

There is a csv built-in module for reading and writing csv files

Third party libraries:

Pandas

Openpyxl – designed for Excel files, supports formulas and other specific Excel functionalities

Google Sheets API – for many languages, make changes directly to spreadsheets hosted online

All these libraries can export to csv format

Reading using built-in csv library are needed the following steps:

Read csv file as a normal file with open

Use csv.reader(file) to parse the file

The output is then cast to list type

You end up with a list of lists, each inner list being a row in a csv file

!important: open the file with encoding=’utf-8’

Open a csv file with open and with newline=’’

Use csv.writer(output\_file, delimiter=’,’)

We can use .write\_row or write\_rows and pass a list or a list of lists with values

The last step is to close the file

Also we can append to a csv file if we open it in the append mode

**131. Working with PDF Files in Python**

There are multiple packages that allow us to read and write PDFs

Not all PDFs are readable with Python packages because of the format and included elements

Usually, the content of PDFs cannot be modified with the aid of Python packages

But there is the option to append pages to PDFs for example

An example of package used to manipulate PDFs is PyPDF2

Reading a PDF:

pdf = open(pdf\_name, ‘rb’)

pdf\_reader = PyPDF2.PdfFileReader(pdf)

On pdf\_reader object we can use: .numPages, .getPage(num), on the page object .extractText()

In the end pdf.close() must be called

Writing a PDF:

pdf\_output = open(pdf\_name, ‘wb’)

pdf\_writer = PyPDF2.PdfFileWriter()

pdf\_writer.addPage(page\_object\_extracted\_with\_pdf\_reader)

pdf\_writer.write(pdf\_output)

pdf\_output.close()

**25.03.**

**133. PDFs and Spreadsheets Python Puzzle Exercise - Solutions**

Regular expressions can be also used to remove / replace patterns

re.sub(pattern\_to\_search\_for, replacement\_pattern, text)