

# Files

Sugarkhuu Radnaa

Py4Econ in Ulaanbaatar

*py4econ@gmail.com*

## Week 4: Learning objectives

Learn to read, edit and write various file types:

- Excel
- CSV
- PDF
- Word

+

glob module and text files.

Not to discuss, but useful file formats: json, html, xml, latex, pickle

# A short about glob module and text files

- `import glob`
- `glob.glob("dir/dir/dir/*.pdf")` – pdf files in the "dir/dir/dir/" directory

# Excel

- pandas
- openpyxl – read and write xls \xlsx
- xlrd – read xls \xlsx
- xlwt - write xls
- xlsxwriter
- xlwings

# CSV

- csv
- pandas

# PDF

- pyPDF2 – good for reading and editing
- fpdf – good for writing
- pdfkit – for writing
- tabula-py – working with tables
- textract
- reportlab

# Word

- python-docx

# Homework

- ① Task 1
  - ② Task 2
  - ③ Task 3
  - ④ Task 4
- Submit your result as a Github repository
  - Deadline: 1 week

**Note:** Create a github repo from the start and populate it with your results step by step.



# Task 1: Excel

Python excel module-уудаар өөрөө зохион жишээ дата үүсгэн (write a file by openpyxl or other packages), түүн дээр тооцсон томьёо, chart-тай эксел файл үүсгэнэ үү.

## Task 2: PDF

Энгийн хэд хэдэн "pdf" файл үүсгэн, тэднийг нэгтгэнэ үү.  
Нэгтгэхдээ glob модулийг ашиглаарай.

## Task 3: Word

task3.docx файлыг үүсгэх код зохионо уу.

## Task 4: CSV

- 1 Файлын нэр нь 3 латин үсгээс бүрдсэн байх тус бүр нь ердөө нэг үг бүхий 1000 ширхэг “csv” файлыг нэг фолдерт үүсгэ.
- 2 Эдгээр файлаас a-h хүртэлх үсгээр эхэлсэн файлууд дахь үгсийг нийлүүлсэн нэг файл үүсгэ.
- 3 Нийт хэдэн файлыг 2-р алхамд нэгтгэсэн бэ?

Thank you!