KADI-PROJECT/PYTHON-DATA-GYM

DAY 3: 29 April 2025; Activity Report

Authors;

(1) David Njiru
(2) Syprose Nyadida

MORNING SESSION; 9.00AM

- The day started with the Recap of the previous day's activities by Levis and the team
- This was followed by the discussion and continuation of the previous day's lecture on; **Reading and Concatenating Data**
- Participants were introduced to;

-The Use glob for pattern matching and working with hundreds of files

```
import glob
files = glob.glob("data/*.csv")
dfs = [pl.read_csv(f) for f in files]
```

-Concatenating DataFrames

```
combined = pl.concat(dfs)
```

-Reading Excel Files

```
import polars as pl
from polars import read_excel
# Requires: pip install openpyxl
read_excel("data.xlsx", sheet_name="Sheet1")
```

Note; Excel support in Polars improving (***If needed: fall back to pandas for .xls files)***

-Handling Time Columns

```
df = df.with_columns([
    pl.col("timestamp").str.strptime(pl.Datetime, "%Y-%m-%d %H:%M")
    ])
```

-Emphasis to make sure timestamps are parsed as datetime, Plotting & resampling need proper types

-Saving Your Work

```
combined.write_csv("combined_data.csv")
```

-Emphasis to Save intermediate steps which is good for debugging and sharing

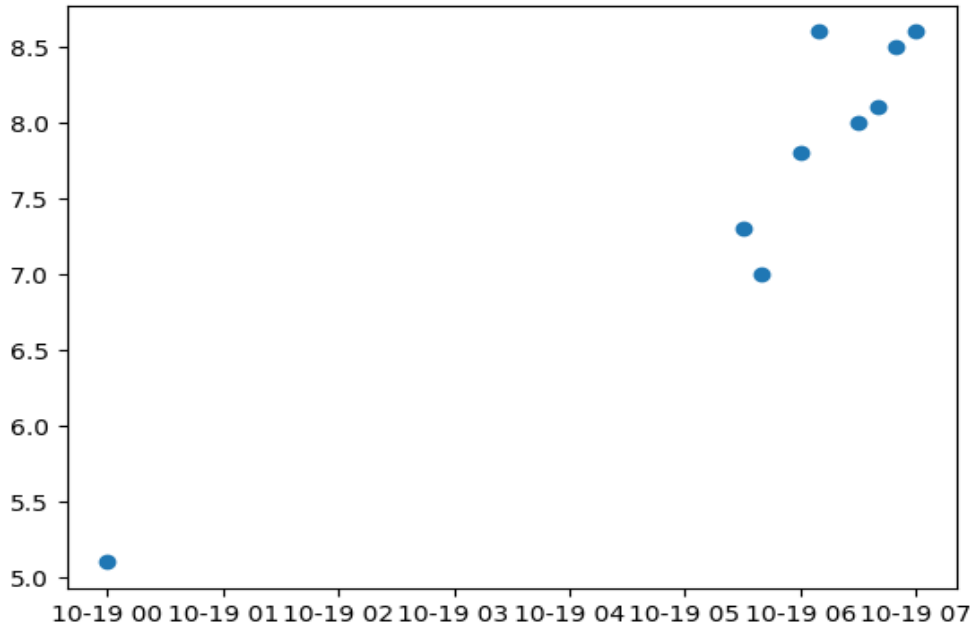
How to Read, lists into a Polars DataFrame for plotting (briefly)

```
import polars as pl
df = pl.DataFrame(zip(datetimes_utc, temperatures))
df.columns=["dtm", "T"]
df = df.sort("dtm")
df

df.plot.point(x="dtm", y="T")

%matplotlib widget
import matplotlib.pyplot as plt

fig = plt.figure()
plt.scatter(df["dtm"], df["T"])
```



Scatter plot for the temperatures

Emphasis on;

a) Use of the **datetime library**
b) Time zone class - which helps to know when the observation was taken.
   - Most stations use UTC, but one can convert to the local time zone.
   - The advantage of UTC is that it works well with large data

- Created a list of ***datetime (strptime).*** This helps the computer to know the format.

Example: (ds, "%Y%m%d%H%M"); Y=four-digit year, m=month, d=day, M=minute, H=hour.

List Comprehension – this is a way of creating a loop. Make use of the square brackets in the list.

Emphasis for Reading/Plotting data - Use of Matplotlib, Pyplot, and Pandas to plot data.

Use of built in functions to execute programs_ Operational functions

Example 1: Script for operational functions (int; float; etc)

```
def sum(a: float, b: float) -> str:
return a + b
c = sum(2.3, 5)
```

```
def sum(a: float, b: float) -> str:
"""Add 2 numbers
Args:
a (float): first number
b (float): seconf number
Returns:
str: the sum of two numbers
```

```
    """    return str(a + b)  ; c = sum(2.3, 5)


def multiply(a: int, b: float=3.14159) -> float:
"""Multiply 2 numbers

Args:
a (int): first number
b (float, optional): second number. Defaults to 3.14159.

Returns:
float: the product of two numbers
"""
return a * b

result = multiply(2)

print(result) 6.28318
```

Function calling another function and function in a function

```
def complex(a: int, b: int) -> int:
c = float(sum(a, b))
d = multiply(c)
return c,d

print(complex(2, 3))
print(type(complex(2, 3)[0]))
(5.0, 15.70795)
<class 'float'>
```

Getting user input in Jupyter notebook

```
a = input("Provide a number (NB: The input cell is at the top!):")
b = multiply(float(a))
   #Multiply two numbers; int*int = Int;  int*float = float; float*float=float
```

Wiki Page + Sarina

Discussion on how to initiate and deliver a project

Start with a general idea; aim of the project; data and methods; type of analysis - input/output (how to deal with the missing values); plotting maps/graphs, labels, units of data

Key points;

(i) Approach to opening different types of files; txt, .csv, netcdf etc
(ii) Use of pandas, polars ….to read, extract & open files; be sensitive to separators
(iii)Extract, slice, dice the subset files
(iv)Convert data to dataframes(df); extract bits and pieces to compute; averages, plot & visualize; save and export to folder

Introduction of the individual projects.

Each participant was given a chance to introduced their expected projects

Topic, aim of the project, data to use, data format, method of analysis, expected results

The afternoon session was mainly spent on project development, data acquisition and processing, consultation amongst the participants and working on data.

Participants were called to break for a group photo after which work on projects continued to dinner time.

- **Evening session.**

Participants were called to present progress of individual project work

Two participants presented;

The First participant was developing a script for formatting and packaging observed Meteorological data from the GAW stations for archiving in to a usable format

Then the second participant presented the analysis of daily rainfall totals observed from the rainfall stations for inclusion to the daily weather bulletin for dissemination to stakeholders.

Each of the two presentations indicated good progress on the projects work.

The lead Trainers in the workshop had very good comments to guide the progress and completion of the proposed works

The presentation of the projects progress reports marked the end of Day2 @ (9.00PM)

The workshop was therefore adjourned to allow participant to retire and reenergize ready for following days' program.

Any comments?

……...End……...