Practice #2: Data collection (1/2): Dataframe Management

Example:

The goal is to predict the stock price using past stock prices and other information available online.

In this practice, you will retrieve stock market prices from 3 companies and make them usable by your python script.

The practice can be done with any interpreter (VSCode, Jupyter, Spyder, Pycharm, …).

Steps:

1. **With the correction of the practice done in lecture:**

- Open the .csv file (downloaded in the previous practice)
- Get data from the .csv file into lists
- Modify your lists to have datetime objects and float numbers instead of text
  Note: All these steps have been done during the correction of the practice in lecture. The code is available in Teams

2. **Prepare your data for the dataframe**

- Create dictionary with seven keys (one for each column of the .csv file) and the values from the previous lists

3. **Create a dataframe with pandas**

- Install pandas
- Create a dataframe from the previous dictionary

4. **Modify your dataframe**

- Change the names of the columns of your dataframe after creation
- Add three columns:
  - "Name" with your name (same for every line)
  - "Surname" with your surname (same for every line)
  - "Date_of_download" with the date of last practice (same for every line)
- Display the 10 first lines of your dataframe
- Display the 10 last lines of your dataframe
- Save your dataframe in a .json file

5. **Make selection on your dataframe**

- Display the following dataframe:
  - Data of 2021's summer
  - Data where volume is higher than the mean volume of 2021's winter
  - Only the open, close, and high prices
  - Ordered by opening price

OPTIONAL:

6. **Creation of the new dataframe**

- Create a new dataframe like:
- The final version of your data has to meet the following format:

Index	Date	Name of the company	Symbol of the company	Type of information	Value	Name of student	Surname of student	Date of download
1	2022-05-03	APPLE	AAPL	HIGH	206	MY_NAME	MY_SURNAME	2022-09-10
2	2022-05-03	APPLE	AAPL	LOW	200	MY_NAME	MY_SURNAME	2022-09-10
3	2022-05-03	APPLE	AAPL	OPEN	205	MY_NAME	MY_SURNAME	2022-09-10
4	2022-05-03	APPLE	AAPL	CLOSE	30	MY_NAME	MY_SURNAME	2022-09-10
5	2022-05-03	APPLE	AAPL	VOLUME	1427545	MY_NAME	MY_SURNAME	2022-09-10
6	2022-06-03	APPLE	AAPL	HIGH	202	MY_NAME	MY_SURNAME	2022-09-10
…								

Information on the final output:

- The final version is a dataframe
- The “date” and the “date of download” are under the datetime format
- The value is a float number
- The other columns are in a string format
- Name and surname are in latin letters
- All strings are in upper cases