

Practice #3: Data collection (1/2): Dataframe

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 companies and make them usable by your python script.

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

Steps:

1. Read your .csv with `pd.read_csv()`

The goal of this first step is to get your csv file into a dataframe to use it with Python.

- Use the pandas function `read_csv` to read your .csv file
- Add the option to automatically change the date to datetime object

2. Modify your dataframe

The goal of this second step is to add information on your dataframe for later use.

- Add three columns (like practice 2):
 - o "Name" with your name (same for every line)
 - o "Surname" with your surname (same for every line)
 - o "Date_of_download" with the date of practice 1 (same for every line)

3. Modify your dataframe

The goal of this third step is to get stock market variation instead of price. This will make all stocks to the same range. Later in this course, we will try to predict the stock variation and not the stock price.

- Create a new column called `percentage_change`. This column is the evolution of the open price of the stock from the previous opening price. The equation is:

$$C_i = \frac{P_i - P_{i-1}}{P_i} * 100$$

With:

- "C" : the percentage change
- "P" : the price of the stock
- "i" : the index

Index	Open	Percentage_change
0	P_0	0
1	P_1	$\frac{P_1 - P_0}{P_0} * 100$
2	P_2	$\frac{P_2 - P_1}{P_1} * 100$
3	P_3	...
4	P_4	
5	P_5	

Figure 1: Example

4. Save your dataframe to json

The goal of this fourth step is to save and share your data with other students.

- Save your dataframe in JSON format
- Reopen your dataframe to validate than the dataframe is the same
- Rename your JSON file with the symbol of the company
- Upload your file on the shared folder:
<https://drive.google.com/drive/folders/1I9PAswClzcpNeOM-N6nWXEf4MgqxWp-w?usp=sharing>

5. Retrieve common data

The goal of this fifth step is to recover data of other students.

- Download 3 .json files from the shared folder of step 4.
- Concatenate these 3 Dataframes together with a loop (your code should change as little as possible if we change the number and names of files)