**Practice #8: Clean the dataframe**

**Objectives:**

* Open the final dataframe
* Clean it from:
  + NaN values
  + 0 values

**Steps:**

1. **Open the dataframe and select**

*You have a file named “df\_final\_US\_EUR.parquet” open it and create a dataframe from it.*

* Use the “pd.read\_parquet()” function in pandas to create a dataframe directly from the file (you have to install “pyarrow” and “fastparquet”)
* Select one company of the big dataframe. You will do the next step on this selected dataframe.

1. **Calculation on the dataframe**

*You can now create columns for later use in class: the percentage change.*

* Create 5 columns with the percentage change of the columns. For the open price for example, this column is the evolution of the open price of the stock from the previous opening price. The equation is:

With:

- “C” : the percentage change

- “P” : the price of the stock

- “i” : the index

Table

Description automatically generated

* Remember to use NaN for the first value and if the daily value is 0 (division by zero is impossible).

1. **Clean the dataframe (the complete one)**

*You should clean the big dataframe before using it to predict with machine learning.*

* Make a list of all kinds of errors existing in the dataframe (NaN values, 0 in the price columns and (0 and negative values) in the volume column). How many zero and nan in each column?
* Clean them.