Practice 6

Manipulate stock exchange data (ctt shares in the last 10 years)

We have a CSV file “CTT10.CSV” that shows the values of CTT shares in the last 10 years.

Each line in the file has

Data fecho abertura max min vol var%

Date Last Open High Low Vol. Var. %

That is, the date, the closing value, the opening value, the maximum of the day, the minimum of the day, the volume of transactions and the variation.

We have to prepare a report from this file for an American customer.

When looking at our data, we noticed several problems already with the presentation of the data:

04.12.2021 3.50 3.57 3.64 3.46 1.57M -1.55%

a) The comma in the numbers must be replaced by a point

b) The sales volume has sometimes an M and sometimes a K. this will prohibit the use of numbers as numbers, so we must replace, for example, 1.57M by 1560000 or 4.56K by 456000

c) We want to add a column that contains positive or negative as a function of variation.

d) Change column names to English

e) The date is in European format we must change the date to appear 04-12-2021 that is day, month and year (NOTE dates and times are manipulated with formats. look for information about timestamp in pandas)

1- Write a function to correct these points and save the file with the name of cttcorrcted.csv

2- What was the date with the highest volatility (difference between min and max)

3- What was the best month (30 days) during the 10 years to win with this title This means if I had to keep my stock only 30 days when was the best 30 days?

4- What was the duration of the longest sequence of ascent of the title in the closings?

5- What was the date that saw the greatest turmoil in the market, ie large volumes with important variations? You can choose for example volume * (max-min) to get a measure of turbulence).