

Assessment 2: Data problems

General Information

In this assessment you have to prove your data science skills. You can solve the problems in your chosen programming language and also work with existing packages.

We expect as a final product:

- on the one hand, a **report with the results** (PPT, Word, HTML, PDF);
- on the other hand, the **comprehensibly commented code**;
- We love plots, charts and graphics!

You have 12 hours to solve this assessment and send the report, your code to principaltrading@cryptofinance.ch. If the results are larger than 15MB, send us the download link (GITHUB, OneDrive, Dropbox etc.) to the project folder.

1.) Bitmex trading data: Data cleaning and manipulation

In the link you will find a ZIP containing the following datasets *data_nonfiltrata_a.csv* and *data_nonfiltrata_b.csv*. Actually, both data sets belong together. Together these sets form BITMEX trading data for Bitcoin (XBTUSD).

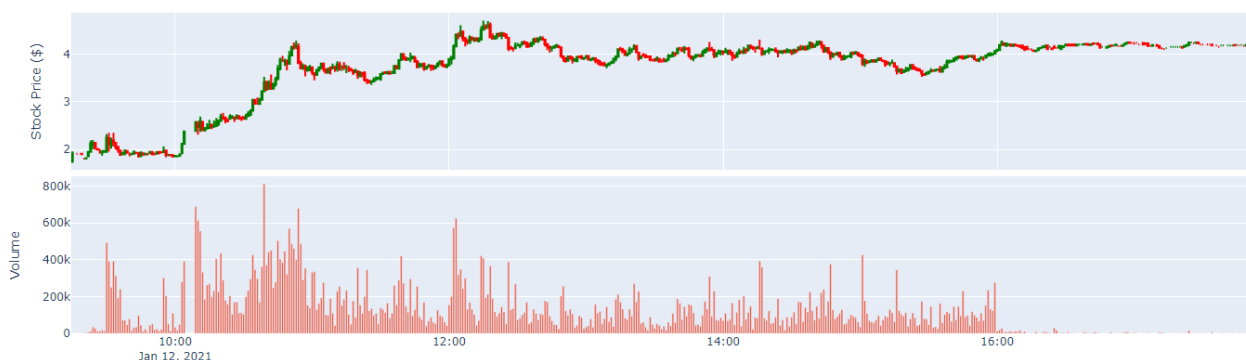
Structure of *data_nonfiltrata_a.csv*:

- **id** is the trading ID of the trade
- **timestamp** is the datetime timestamp in unix epoch time (<https://www.epochconverter.com/>)
- **symbol** is the traded trading pair
- **side** is the action taken "sell" or "buy".

Structure of *data_nonfiltrata_b.csv*

- **id** is the trading ID of the trade
- **amount** is the traded amount in USD
- **price** is the traded price in USD

- Merge both datasets using the common variable and sort the data by timestamp.
- Use the result of 1. a.) and omit/delete all data points (entire row) if they contain a NA or NaN. Do you also find potential data errors in the timestamp? If so, delete this data as well. How many rows does your new dataset have after these cleaning steps?
- Use the cleaned data set and create a new variable *timestamp_CET*, which puts the variable timestamp into a reader-friendly datetime format. The time zone should be Central European Time (CET) and don't forget the milliseconds.
(hint: use <https://www.epochconverter.com/> and play with the number of digits to get the right format)
- Now make a plot. Don't forget to label the plot in a comprehensible way (title, axes, etc.). The plot consists of two subplots. The upper subplot shows the price development over time (y-axis: price; x-axis: timestamp_CET). The lower subplot shows the volume over time (y-axis: amount; x-axis: timestamp_CET). Below you will find an example of such a plot.



- Clean the outliers in variable amount. Which procedure do you use and why? Show a boxplot and a descriptive summary of the amounts before and after the outlier cleaning

- f.) Calculate a 100 point simple moving average price of your cleaned dataset in 1.d) and compare the price and the moving average.
- g.) Calculate a 100 point amount weighted moving average price of your cleaned dataset in 1.d) and compare the price and the moving average of 1.f) and the amount weighted moving average price.
- h.) Derive your own 15min OHLC data from your cleaned dataset in 1.d). and plot the data as candlesticks or barcharts.
What is OHLC-DATA: <https://www.daytradetheworld.com/trading-blog/ohlc-charts/>
- i.) Derive your own 15min median and amount weighted price from your cleaned dataset in 1.d).
Compare the mediam, amount weighted price and the Close price of 1.f.) with each other.

2.) Analysis of multiple time series

For this exercise, you will work with market data from FTX, a crypto exchange. To access the data you will need the documentation of the Rest-API of FTX, which you can find [here](#) (see endpoints: *Get markets* and *get historical prices*).

- a.) Get the 60 min market data of the last 360 days for “ETH/USD” and “BTC/USD”. Merge the datasets to one dataset.

(hint: use pagination)

If you are unable to retrieve the data, please use the provided files *btc4hdata.csv* and *eth4hdata.csv* and merge these two.

- b.) Play around with the data (prices, returns, etc.) and show us plots and descriptive analyses of the two series. (We love labels and colours)

- c.) Are the two series cointegrated?

BONUS: Create and back test you own trading strategy

Let your creativity run free! Use any data you need from the FTX API (see question 2) . Create a simple trading strategy (momentum, reversal or even a pair trading strategy), optimise and backtest it on different pairs (small cap vs large cap tokens) and on different time frames (e.g. 15min, 1h, daily data). Present the results.

Keywords: brute force optimisation, in vs out of sample, risk adjusted performance