Analyzing Stock Market Values

Data Processing and Analytics (DPA) Course - International Master DISS UCBL 25/26. Upload the Jupyter notebook and the report on Tomuss (TP2 column) by the 2nd December.

The objective of this project is to use Spark Streaming to analyze the value of different stocks through time.

The Dataset

The data is contained in a CSV file named **stock.csv.** For this project, we will consider only two columns, the last one that is the name of the stock, and the second one that is its value. You don't have to worry about parsing the file, the notebook "Kafka_Producer_for_Project" will take care of that.

Development environment

For this project, I suggest you use the same environment you used for the TP related to Spark and Spark Streaming.
Use the notebooks you find on Moodle, in particular

● The notebook **Kafka_Producer_for_Project** reads the file and ingests the data into Kafka with the schema < name, price, timestamp >. Consider that for the timestamp it uses the current one.

● Put the csv with the data in the same folder.
● The notebook **Project Template** registers Spark to the stream and puts it in a

manageable form. Be careful not to edit the cell already there unless you know what you are doing ;).

Tasks

Imagine you want to create a dashboard that shows statistics in real time about what's at the stock market. For instance - given a time window of your choice (it can vary between tasks, and motivate it in the report) - compute:

1. The stocks that had the smallest daily variance
2. Stocks that gained the most every hour and every day
3. Top-10 most traded symbols daily
4. Top 5 stocks per volume daily
5. Design a stop-loss and take profit system. Choose a (or some) stock (s) and test it on it (them).
6. Create a wallet with 5 stocks and track the evolution of the wallet.

Expected Deliverable

- The Jupyter Notebook. The code should be working.
- A document (pdf) containing
    - An explanation of what you did
    - The results of the analysis for each task
- Everything compressed in a ZIP file (rar = -2 points)

- Bonus point if you manage to use some visualizations (OPTIONAL) - A user guide on how to run the notebook (e.g., which cells do what, or any

other information that makes it easy for us to understand how to make your code work)

Evaluation

- 2pts for task 1,2,3 and 6.
- 5pts for task 4 and 5
- 2 pts for the report
- 1pt (bonus) for visualizations
- 21 points total