

Project

Deadline March 26th

Rules

- This is an individual project
- This project should be done with the least help of anyone (but internet)
- Since it applies on different websites, there is no reason for your codes to be similar
- Similar codes are severely downgraded

Objective

Having up-to-date info from continuous scraping of a website, displayed on your online dashboard.

1. Find a website that have some dynamic information which changes each minute or so (select one with a stable html structure though, e.g. <https://www.coingecko.com/en/coins/bitcoin>)
2. That specific information should be retrieved (scrapped) using **bash**
3. That specific information should then be displayed on a dashboard, made with Python using the package **Dash**
4. A graph should also display the time series (values scrapped along time)
5. The dashboard should be updated every 5 minutes using **cron**
6. A daily report should also be displayed in the dashboard, updated at 8pm each day, and should provide several coherent metrics (e.g. for a financial series, the report could contain daily volatility, open and close price, evolution, etc.)
7. Additionnal features are welcome if they add value to the dashboard.

Specifications

- The code should be versioned using git and hosted on Github
- The code should run h/24 on your online hosted Linux virtual machine (be careful to prevent any cost from your cloud provider)
- The scrapper should be written in **bash** only without any advanced language or package (e.g. ~~no Python~~)
- You should use **regex** in your scrapper.

- The code and architecture should be clean, easily readable and easy to debug
- You will provide the address of your virtual machine to your teacher (to access the live dashboard)
- You will provide the address of your git repository hosting this project to your teacher (to access your code)
- You should enjoy the process since it might be very useful for your career!