

Intro

Nous avons fait notre TP sur l'indice boursier Intact Financial Corporation avec des données du site Yahoo Finance en utilisant le scraping

Processus

Le processus est simple, on exécute le fichier app.py à partir d'un IDE ou command prompt, il collecte les données entre 9h30 et 16h00 à des intervalles aléatoires et les place dans une table de base de données sqlite

Télécharger le repo ici :

https://github.com/Ludwigo13/ifc_collector

Dans le dossier du projet, ouvrir le fichier generate_schedules.py et modifier les dates (lignes 8 et 9) pour les journées ciblées. Prendre note que le programme génère une cédule entre 9h30 et 16h00 quand les marchés boursier sont ouverts.

```
generate_dates = (  
    "12-07-2024",  
    "13-07-2024"  
)
```

Pour rouler à partir de powershell, si vous avez une erreur Unauthorized parce que le script n'est pas digitally signed, exécuter cette commande en tant que admin :

Set-ExecutionPolicy -Scope Process -ExecutionPolicy Bypass

Naviguer jusqu'au dossier qui contient app.py et exécuter :

.\venv\Scripts\Activate

Exécuter avec le script avec la commande :

Python generate_schedules.py

Python app.py

Un insert se fait dans la table (juste une table) à chaque itération

Conclusion

Nous en avons appris beaucoup sur comment travailler avec github, pycharm, le scraping de données qui est un principe que je ne connaissais pas et même la finance.

Beautiful soup plus facile à utiliser, scrapy plus complet

Certains sites reconnaissent les principes de scraping et bloque

Une tâche planifiée pourrait être utiliser pour un peu plus d'automatisation

On pourrait avoir une table avec des dates pour une gestion différente de la cédule

Nous avons été chanceux dans un sens pour le choix de nos collègues. Je crois que j'aurais eu beaucoup de difficultés à travailler avec certains de mes camarades de classe dû à un écart entre nos connaissances. Ça aurait ajouté un poids au travail que je n'aurais pas pu gérer.

Je n'aime pas la redondance