TP de groupe

Sujet :
Chaque groupe doit sélectionner un site web à scraper (par exemple : une boutique en ligne, un annuaire, un site d'actualités). Le site choisi doit offrir la possibilité d'extraire des informations structurées (titres, prix, descriptions, dates, catégories, etc.) sur un ensemble d'éléments (produits, articles, profils, etc.).

Objectifs pédagogiques :

- Mettre en œuvre trois technologies de scraping sur un même site : **BeautifulSoup**, **Selenium** et **Scrapy**.
- Comprendre les contextes d'utilisation de chaque outil et en évaluer l'efficacité, la rapidité, la simplicité et la pertinence.
- Comparer concrètement ces approches (performances, facilité d'extraction, maintenance du code, robustesse face aux modifications du site).
- Obtenir un jeu de données final propre, prêt à être utilisé ultérieurement (analyse, IA, etc.).

Livrables :

- Le code source complet pour les trois méthodes (BeautifulSoup, Selenium, Scrapy).
- Les CSV intermédiaires (données brutes et données nettoyées).
- Le fichier final `final_dataset.csv`. et `final_dataset.json`
- Un rapport (sous forme de `README.md` ou de document séparé) proposant une analyse comparative des trois outils.

Critères d'évaluation :

- **Maîtrise des 3 outils :** Code fonctionnel, opérationnel et adapté à chaque approche.
- **Qualité des données extraites :** Données complètes, structurées et propres.
- **Analyse comparative :** Mesures de performances concrètes, argumentation claire et pertinente sur les différences entre chaque outil.
- **Documentation et rapport :** Explications détaillées, démonstrations chiffrées, tableaux comparatifs, éventuels graphiques, et conclusions justifiées.