**Objet : Programme d'extraction des prix – Version bêta terminée**

Bonjour Sam,

J’ai terminé la version bêta du programme d’extraction des prix pour le site Books to Scrape, en suivant les étapes indiquées dans le fichier des exigences.

Vous trouverez :

* Mon dépôt GitHub ici : <https://github.com/JulienCoureau/OC_P2_BooksToScrape>
* Un fichier compressé avec les données extraites (CSV + images)

Le dépôt contient :

* Le script Python,
* Un fichier requirements.txt avec les bibliothèques nécessaires (requests, beautifulsoup4),
* Un README.md qui explique comment installer et utiliser le script.

Conformément à votre demande, les données extraites et les images ne sont pas dans le dépôt, mais bien organisées dans l’archive jointe.

**Utilisation possible dans un pipeline ETL**

Voici une idée simple pour utiliser ce script dans un pipeline ETL (Extract, Transform, Load) :

* **Extract** : Le script récupère les données depuis le site en ligne.
* **Transform** : Les données sont mises en forme (nettoyage, format CSV).
* **Load** : Le fichier CSV peut être importé dans un tableur, une base de données ou un autre outil d’analyse.

Ce script pourrait être lancé régulièrement (par exemple avec une tâche planifiée) pour surveiller les prix dans le temps.

**Évolution possible du ETL**

Ce projet pourrait évoluer naturellement pour automatiser entièrement l'ETL

**Automatisation planifiée** :

* Le script pourrait être exécuté automatiquement tous les jours à une heure précise.
  + Cela permettrait de récupérer de nouvelles données sans intervention humaine.

**Archivage automatique** :

* Chaque fichier CSV généré pourrait être sauvegardé dans un dossier "Archives" spécifique.
* Chaque fichier serait daté du jour de la collecte (YYYY-MM-DD) pour conserver l'historique.
  + Cela rendrait possible des **analyses d'évolution des prix des livres** sur plusieurs jours, semaines ou mois.

Bien cordialement,

Julien Coureau