

SPRINT DATA - PROJET FINAL

Objectif

L'objectif de ce projet est d'extraire des données d'un site e-commerce au choix, de les stocker dans une base de données MySQL, de les nettoyer avec Pandas, puis de les analyser et visualiser à l'aide de Matplotlib et Seaborn.

Vous devrez :

- **Scraper les données d'un site e-commerce** (nom, prix, catégorie, avis, lien du produit,...).
 - **Stocker ces données dans une base de données MySQL structurée.**
 - **Charger et nettoyer les données avec Pandas :**
 - Conversion des types de données
 - Gestion des valeurs manquantes
 - Ajout de colonnes calculées pertinentes
 - Normalisation des catégories
 - **Effectuer des analyses et visualisations pertinentes selon les données obtenues.**
-

Sites recommandés pour le scraping

Chaque groupe doit choisir son propre site e-commerce parmi les recommandations suivantes ou en proposer un autre sous validation :

- **Books to Scrape** – Site fictif de vente de livres avec prix et notation.
- **Trek Bikes** – Vente de vélos avec prix et descriptions.
- **Decathlon** – Articles de sport, avec prix, avis et descriptions bien structurées.
- **Boulanger** – Électronique et électroménager avec prix et avis clients accessibles en HTML.
- **Sephora** – Produits de beauté avec prix et avis clients.
- **Rakuten** – Marketplace avec divers produits (électronique, mode, livres, etc.).
- **Amazon, Cdiscount, Fnac, eBay** – (Scraping plus complexe nécessitant Selenium si besoin).

Vous pouvez proposer d'autres sites, mais le choix doit être validé avant de commencer.

Contraintes et attentes

- **Utilisation libre des outils de scraping** (Requests, BeautifulSoup, Selenium, Scrapy, etc.).
 - **Base de données MySQL** bien structurée avec des contraintes d'intégrité.
 - **Nettoyage intelligent des données**, avec une justification des choix.
 - **Analyses et visualisations laissées à la créativité des étudiants**, en fonction des données collectées.
-

Livrables attendus

- **Fichier ZIP** contenant tout le projet.
- **Présentation en PDF** résumant le travail réalisé.

Les modalités de rendu sont détaillées ici : [MODALITÉ DE RENDU - PROJET FINAL](#)

Formation des groupes

Les groupes seront composés de **5 à 6 personnes**. Vous devez créer vos groupes de projet en suivant les feuilles Excel dédiées à votre groupe Bachelors via ce lien : [Lien vers le fichier Excel](#)