

# Scraping Python

## Data Cohorte 6

### **Objectif du Projet :**

Sélectionner un sujet, identifier et analyser des sites Web pertinents, extraire les données et les stocker dans une base de données PostgreSQL.

#### **Livable 1**

##### 1. Identification de la Problématique :

- Définir le sujet et les objectifs spécifiques du projet de scraping.

##### 2. Sélection de Pages Web :

- Identifier au moins 3 pages Web à scraper :
- Une page accessible sans connexion.
- Une page nécessitant une connexion.
- Une page de réseau social.

##### 3. Sélection d'une API :

- Identifier une API traitant du même sujet que les pages sélectionnées.

#### **Livable 2**

##### 1. Identification des Données :

- Déterminer les données spécifiques à extraire de chaque page Web, en tenant compte des différences possibles entre les pages.

##### 2. Création de la Base de Données :

- Concevoir et créer la base de données PostgreSQL ainsi que les tables nécessaires en fonction des besoins identifiés.

#### **Livable 3**

##### 1. Extraction et Insertion des Données :

- Extraire les données des pages Web sélectionnées.
- Insérer les données extraites dans la base de données PostgreSQL.

#### **Livable 4 (A venir)**

##### 1. Visualisation des Données :

- Créer des visualisations JavaScript pour représenter les données extraites.

## **Outils et Organisation**

### 1. Gestion de Projet:

- Créer un tableau Trello pour suivre l'avancement des livrables.
- Créer un repository GitHub avec une branche par service.

### 2. Technologies Utilisées :

- Utiliser Python et JavaScript pour le scraping des pages Web.
- Utiliser PostgreSQL pour le stockage des données.
- Utiliser des bibliothèques et outils de visualisation JavaScript pour la représentation des données.