

Scraping Python

Data Cohorte 5

Sujet

L'objectif du projet est de sélectionner un sujet, puis de trouver des sites Web proposant du contenu sur ce sujet afin de les analyser et d'extraire les données.

Une fois les données extraites, elles doivent être conservées dans une base de données PostgreSQL.

➤ **Livrable 1** (Jour 1)

- ✓ Identifiez votre problématique
- ✓ Identifiez au moins 2 pages que vous devez scraper
- ✓ Identifiez-en plus un API traitant le même sujet

➤ **Livrable 2** (Jour 1)

- ✓ Identifiez les données que vous devez récupérer au niveau des différentes pages, en sachant que vous pouvez avoir des données différentes d'une page à une autre.
- ✓ Créez votre base de données et vos tables en fonction de votre problématique.

➤ **Livrable 3** (Jour 2)

- ✓ Récupérez les données pour les insérer dans la BD

NB

Créez un Trello ou vous mettez les différents livrables

Créez un Github avec une branche par service

Le scraping de certaines pages est fait en Python et d'autre en JavaScript

Les groupes de travail du projet Django sont maintenus. Chaque membre du groupe prend le sujet correspondant au service qu'il a réalisé.