



Ce projet est cofinancé
par l'Union européenne



La Région
Auvergne-Rhône-Alpes

OBJECTIFS

Objectifs pédagogiques

- Récupération de données structurées et non-structurées
- Intégration de données dans une base existante
- Introduction à l'industrialisation de dataflow

Démarche pédagogique (projet, ressources ...)

Ce module se décompose en différentes parties introduisant des méthodes pour aller récupérer des données et/ou les intégrer dans des jeux de données existants.

Les deux premières parties sont consacrées au scraping et à l'utilisation d'apis, la troisième à la structuration des données à l'aide d'ETL python ainsi qu'à l'industrialisation des dataflow et la quatrième à l'utilisation d'ETLs graphiques.

Pour finir un projet sera effectué afin de remettre toutes les compétences vues en pratique.



Compétences

Itération 1

- Scraping :
 - Récupération de données avec scrapy
 - Récupération de données tabulaire
 - Pagination avec selenium
 - Récupération d'info d'un fichier csv sur internet

Itération 2

- API
 - Utiliser une API avec des paramètres

Itération 3

- Modèle de données
 - Modéliser des données extraites
 - Normaliser un modèle de données

Itération 4

- ETL
 - Introduction Extract Transform Load et Extract Load Transform
 - Opération ETL: passer d'un fichier à une DB
 - Opération ELT: ajout de population par commune
 - Intégrer des données contenant des erreurs



MODALITÉS

Durée

16 jours soit 112 heures au total.

Lancement le 22/10/2025 et clotûre le 14/11/2025.

Formateur(s)

David NADJAR

Jules Tison

TRAME

		Planning		Jour	Sujet	Activités
22/10	Extraction de données	David	David	1	Scraping	Scrapy
23/10	Extraction de données	David	Autonomie	2	Scraping	Selenium
24/10	Extraction de données	David	Autonomie	3	Scraping	Selenium
27/10	Extraction de données	David	David	4	Scraping/API	Fin du scraping, API
28/10	Extraction de données	David	David	5	API	API
29/10	Extraction de données	David	Autonomie	6	API	API
30/10	Extraction de données	David	Autonomie	7	API	API
31/10	Extraction de données	David	Autonomie	8	API	Fin de l'API, début modèle de données
03/11	Extraction de données	Jules	Jules	9	ETL/1	ETL + fuzzywuzzy
04/11	Extraction de données	Jules	Jules	10	ETL/1	ETL et Production
05/11	Extraction de données	Autonomie	Autonomie	11	ETL/2	ETL et Production
06/11	Extraction de données	Jules	Jules	12	ETL/2	Knime
07/11	Extraction de données	Jules	TRE	13	ETL/2	Knime + visu
12/11	Extraction de données	Jules	Jules	14	projet	avec cas d'étude BI
13/11	Extraction de données	Jules	Jules	15	projet	avec cas d'étude BI
14/11	Extraction de données	Jules	Jules	16	projet	avec cas d'étude BI