

Offre de stage L3

Laboratoire L3I - Du 13 avril au 29 mai 2026.

Responsable : Armelle PRIGENT

Les données open data relatives à l'enseignement supérieur sont nombreuses et couvrent des thématiques variées : établissements, formations, indicateurs de réussite, d'attractivité ou encore données liées aux plateformes d'accès à l'enseignement supérieur ou aux diplômes de master. Des projets tels que supTracker existent et offre des visualisations sur la base des données issues des plateformes telles que Parcoursup et Monmaster.

https://beta.suptracker.org/?tab=statistiques-formation&type_bac=tous>acod=3083

Ils sont à ce jour insuffisants en termes de garantie de la qualité de la donnée et en termes d'extensibilité (ajout de nouvelles données).

Les informations disponibles sont réparties dans plusieurs jeux de données qui peuvent être exploités conjointement, sous réserve de résoudre des problèmes d'hétérogénéité de structure et d'identifiants.

L'objectif du stage est de développer un prototype d'application permettant d'extraire, préparer, structurer et visualiser un sous-ensemble pertinent de ces données.

Le travail consistera à développer des scripts en Python pour la récupération et la transformation des données, ainsi qu'à utiliser les bibliothèques appropriées pour leur nettoyage et leur préparation. Le stockage des données reposera sur une base de données orientée graphe, et le ou la stagiaire produira les requêtes nécessaires à l'extraction d'informations pertinentes. Enfin, le stage donnera lieu à la réalisation de tableaux de bord paramétrables.

Après un temps de formation aux technologies, le ou la stagiaire participera à :

- L'analyse des jeux de données existants pour la sélection des données pertinentes.
- L'analyse et la documentation de leur structure pour la production d'un dictionnaire des données ;
- La définition d'une méthode de rapprochement entre données issues de sources différentes ;
- La conception d'un modèle de données orienté graphe avec les données sélectionnées ;
- Le développement des scripts d'alimentation de la base ;
- L'écriture de requêtes d'analyse ;
- La réalisation d'un tableau de bord interactif par établissement et/ou par formation.

Technologies mobilisées : Python, pandas, Neo4j, Cypher, Streamlit

Compétences mobilisées : programmation, traitement de données, visualisation de données.

Pour candidater : vous adresserez un CV et lettre de motivation à aprigent@univ-lr.fr

La lettre de motivation devra justifier de vos expériences antérieures de programmation et/ou de traitement de données.