

Projet 2 : Mise en Place et Optimisation d'une Infrastructure Data sur le Cloud

Projet Cloud – Collecte, stockage et visualisation des offres d'emploi Data sur AWS

Loïc RABEARIVONY AIA01



OBJECTIFS DU PROJET

- Collecter des offres d'emploi via scraping et APIs
- Stocker les données dans AWS RDS (PostgreSQL)
- Créer un pipeline ETL automatisé
- Visualiser dynamiquement les données dans un dashboard Streamlit

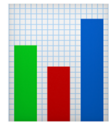


ARCHITECTURE DU PIPELINE

- 1. Scraping (AI-Jobs.net) et APIs (Adzuna, Pôle Emploi)
- 2. Nettoyage & normalisation des données
- 3. Stockage PostgreSQL (AWS RDS) + sauvegarde brute (AWS S3)
- 4. Analyse exploratoire (Notebook)
- 5. Dashboard interactif avec Streamlit

STACK

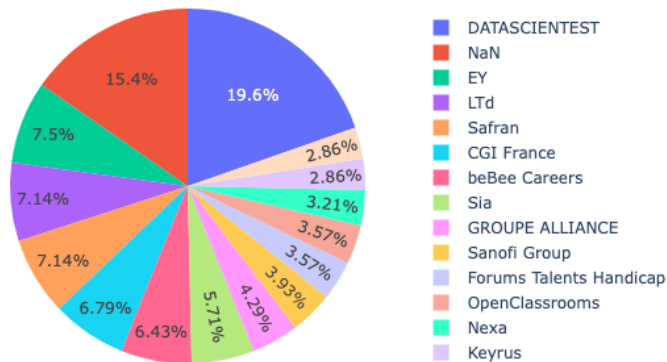
- - Python (Pandas, Seaborn, Matplotlib, Plotly, Streamlit)
- - PostgreSQL via SQLAlchemy
- - AWS : RDS (Base de données), S3 (Stockage fichiers)
- - GitHub (code source), dotenv (sécurité), Docker (A prévoir)



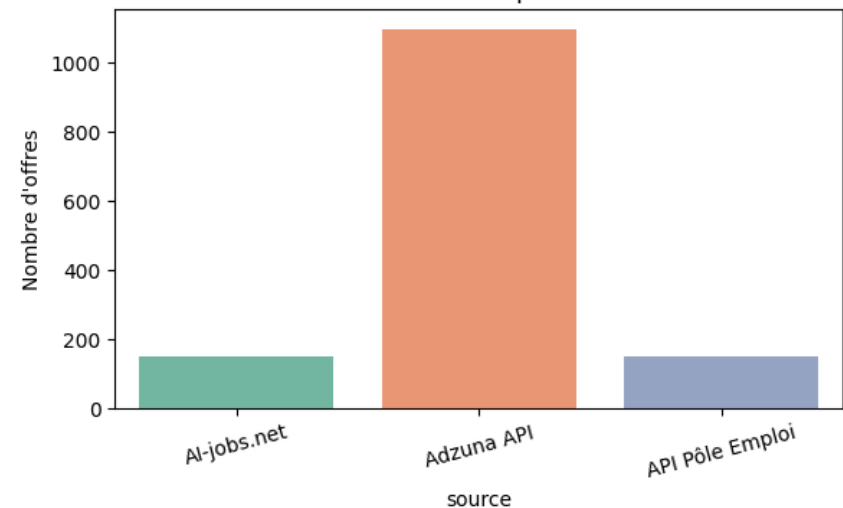
Visualisations Clés

- - Répartition des annonces par source / ville / entreprise
- - Histogrammes des salaires, contrats et catégories
- - Heatmap Postes x Villes, KPIs dynamiques

Top entreprises qui recrutent



Nombre d'offres par source





INSIGHTS

-  **Source dominante** : Adzuna Api
-  **Offres avec salaire** : 192
-  **Total d'offres** : 1399
-  **Entreprises uniques** : 631
-  **Villes uniques** : 366
-  **Poste le + fréquent** : formation data science
-  **Entreprise la + active** : DATASCIENTEST
-  **Salaire max moyen** : 41242 €
-  **Salaire min médian** : 38000 €
-  **% avec salaire** : 13.72%
-  **Postes uniques** : 968



DEMONSTRATION DASHBOARD

- Dashboard Streamlit déployé localement
- Filtres dynamiques et onglets analytiques
- Données connectées en live à AWS RDS



ROADMAP & PERSPECTIVES

- Mise en production (Docker + GitHub Actions)
- Enrichissement des analyses prédictives (salaire/poste)
- Amélioration UI/UX et accès utilisateur



CODE SOURCE & DOCS

- Code Streamlit + Notebook disponibles sur GitHub
- Documentation ETL, structure BDD et nettoyage intégrée
- ➡ https://github.com/space9cowboy/Projet_2_AIA01