DATA LAKE CHALLENGE

Cartographie du marché de l'emploi Tech en Europe

eFrei

PARIS PANTHÉON-ASSAS UNIVERSITÉ

AMINE HAMZAOUI
BRAND TIKOUE TIKOUE
ANTHIME BAYANI
SALAH OUNI

SOMMAIRE

1 : CONTEXTE ET OBJECTIFS

Dans un contexte de pénurie de profils Tech en Europe, la Commission européenne a lancé le plan “TalentInsight”. Le but de ce projet est de cartographier en quasi-temps réel l’offre et la demande de compétences numériques à partir de 6 sources hétérogènes.

Notre mission : construire une plateforme analytique fiable, automatisée et exposable via une API RESTful.

OBJECTIFS DÉTAILLÉS :

- Scraper et miner des données d’emploi et de tendance tech dans plusieurs pays.
- Nettoyer, normaliser et centraliser dans un Data Warehouse.
- Exposer les données via une API Django REST.
- Fournir des analyses utiles pour les institutions, entreprises et chercheurs.

2 : ARCHITECTURE TECHNIQUE

2.1 INGESTION DES DONNÉES

- Outils utilisés : requests, BeautifulSoup, pytrends, API GitHub, CSV
- Format de sortie : JSON / CSV dans data/local/raw/[source]/

2.2 NETTOYAGE ET NORMALISATION

- Conversion des salaires en EUR
- Harmonisation des pays (codes ISO2)
- Formatage des dates, nettoyage des doublons

2.3 STOCKAGE

- Données brutes : dans le dossier raw/
- Données nettoyées : dans datasets_clean/ puis chargées dans le DWH
- DWH : SQL (SQLite ou PostgreSQL ou MySQL), avec tables en étoile :
- d_date, d_country, d_company, d_skill, d_source
- f_salary_stats, f_trends

3 : SOURCES COLLECTÉES

3.1 ADZUNA API

- Accès : API avec clé gratuite
- Champs : salaire médian, secteur, skills
- Défis : authentification, pagination

3.2 GITHUB

- Accès : GitHub REST API
- Champs : langage, stars, localisation
- Défis : pagination, auth token, limites

3.3 GOOGLE TRENDS

- Accès : PyTrends
- Champs : popularité mots-clés tech
- Défis : parsing JSON

3.3 STACK OVERFLOW SURVEY

- Accès : CSV annuel
- Champs : stack préférée, salaire, techno
- Défis : taille du CSV, mapping pays/langue

4 : API REST

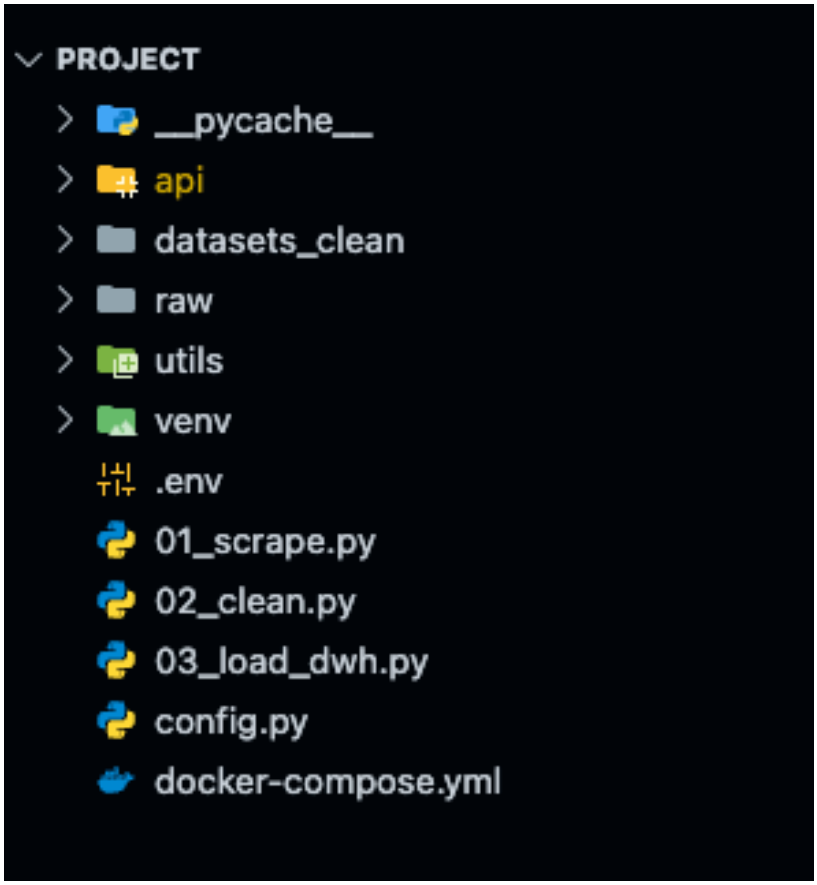
TECHNOLOGIES

- Framework : Django 5 + Django REST Framework
- Auth : Token via en-tête
- Pagination : 50 résultats par défaut

ENDPOINTS

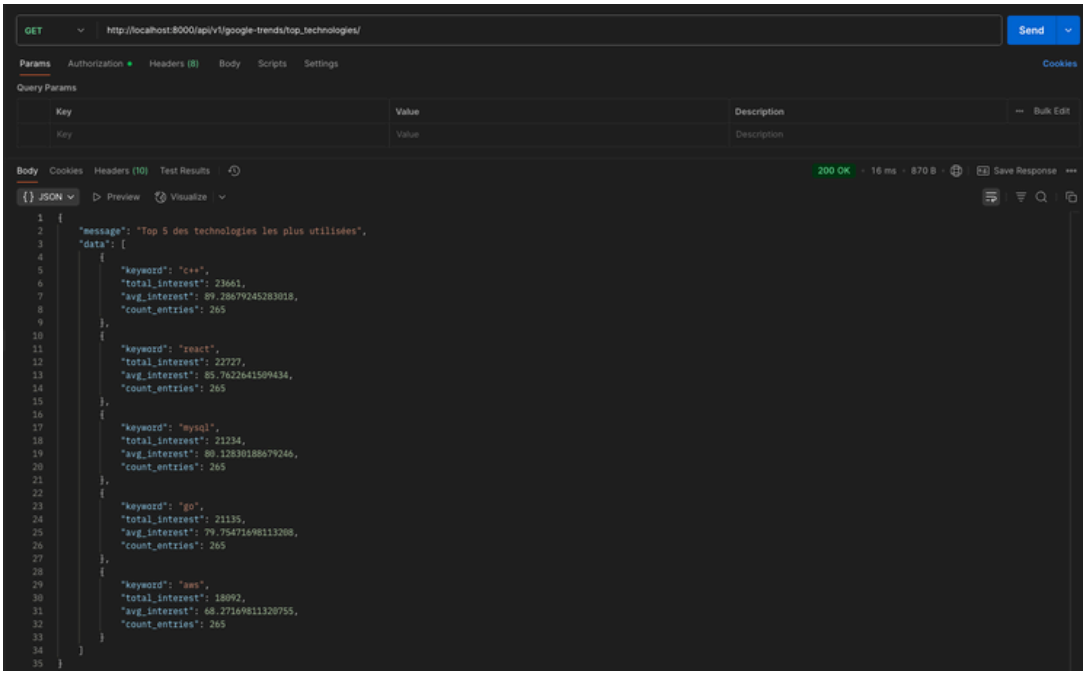
- /api/v1/salary-daily/
- Ex : ?country=FR&skill=Python
- /api/v1/skill-trends/
- Popularité par date et pays

5 : STACK TECHNIQUE & ORGANISATION

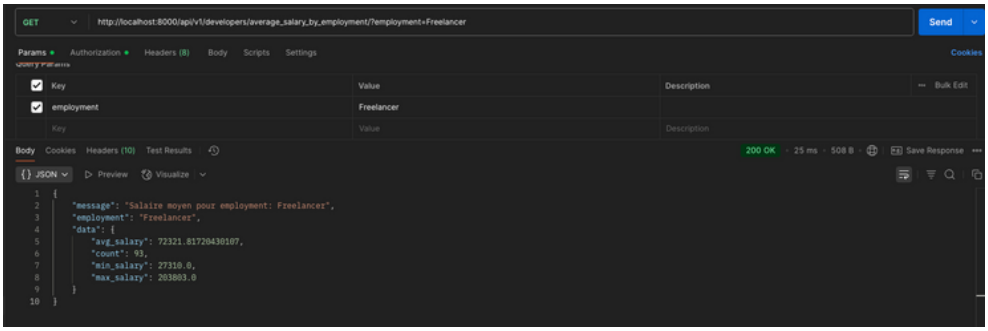


6 : CAPTURES D'ÉCRAN ET TESTS

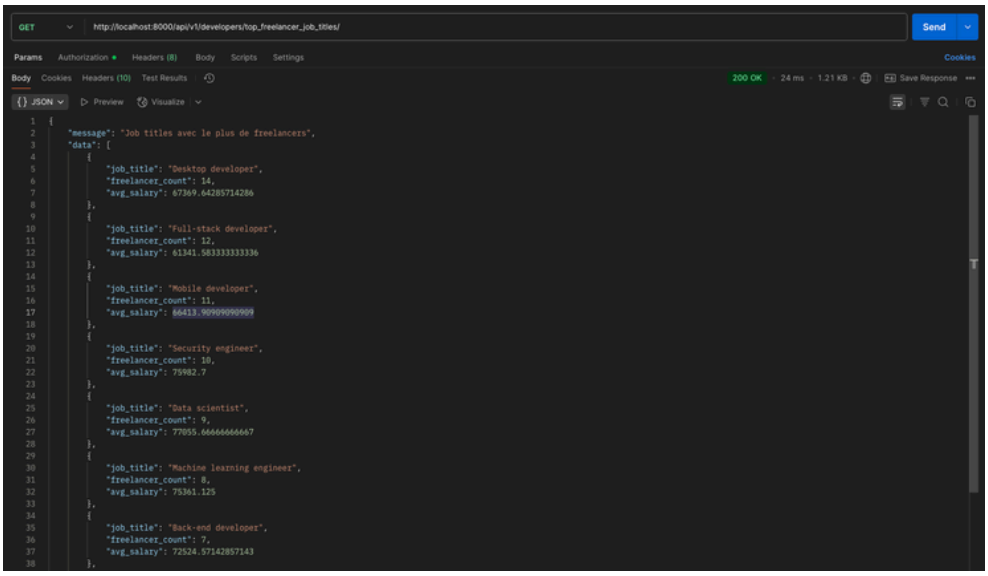
TOP 5 TECHNOS LES PLUS UTILISÉES



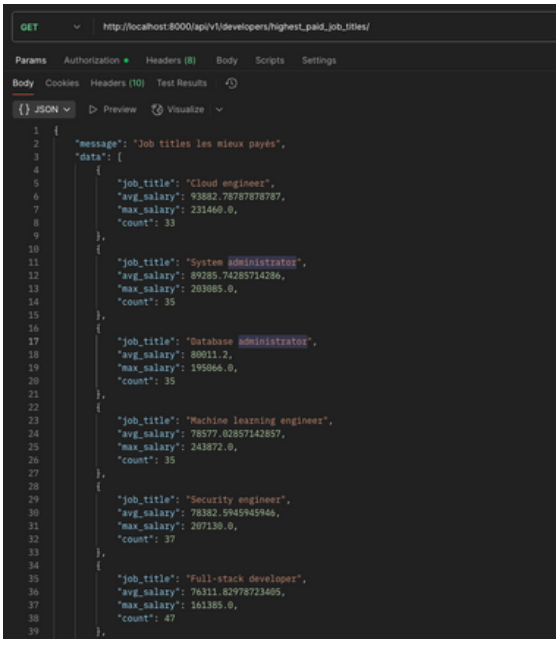
SALAIRE MOYEN DES FREELANCES



LISTE DES JOBS AVEC LE PLUS DE FREELANCES



LISTE DES JOBS LES MIEUX PAYÉS



7 : CONCLUSION

Ce projet nous a permis de maîtriser la construction d'un Data Lake et d'un Data Warehouse à partir de données hétérogènes, ainsi que la conception d'une API RESTful exploitable par des acteurs publics, privés ou académiques.

En croisant tendances de recherche, salaires et compétences recherchées, cette plateforme offre un socle solide pour orienter les politiques publiques et stratégies de formation.