# NLP Text Mining ATLAS

Analyse régionale des offres d’emploi


Université Lumière Lyon 2

Cyrille PECNIK - Mohamed Riad SAHRANE - Olivier BOROT

# Introduction à l'application

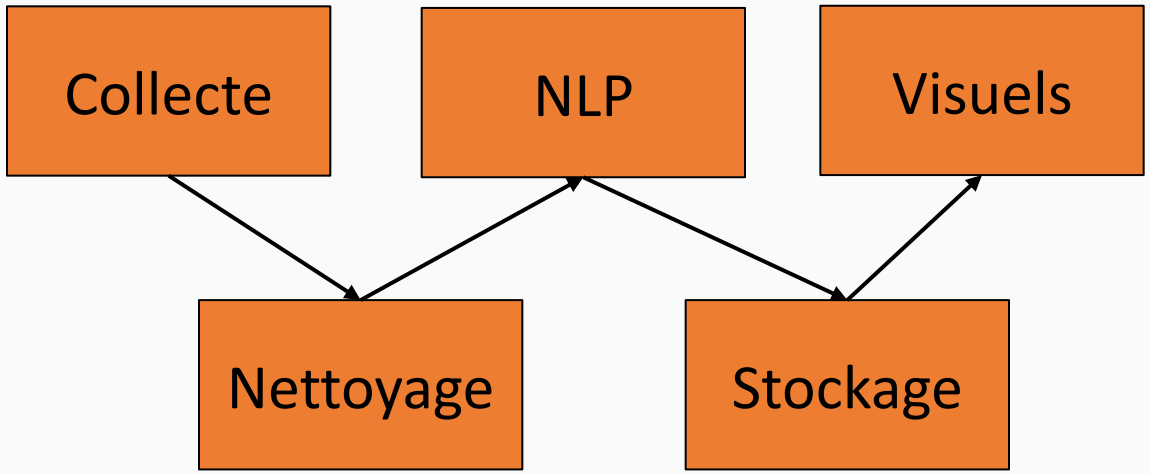
## Constat

Les offres d'emplois sont :

- Dispersées
- Non structurées
- Difficile à analyser globalement

## Solution

Pipeline automatisé complet




## Résultat

- 5000 offres collectées
- 35000 communes géolocalisées
- 8 thèmes
- 384 dimensions d'embeddings

# Sommaire

- Données utilisées
- Entrepôt de données
- Web application Streamlit
- Dockerisation
- Analyses
- Bilan critique

# Données utilisées

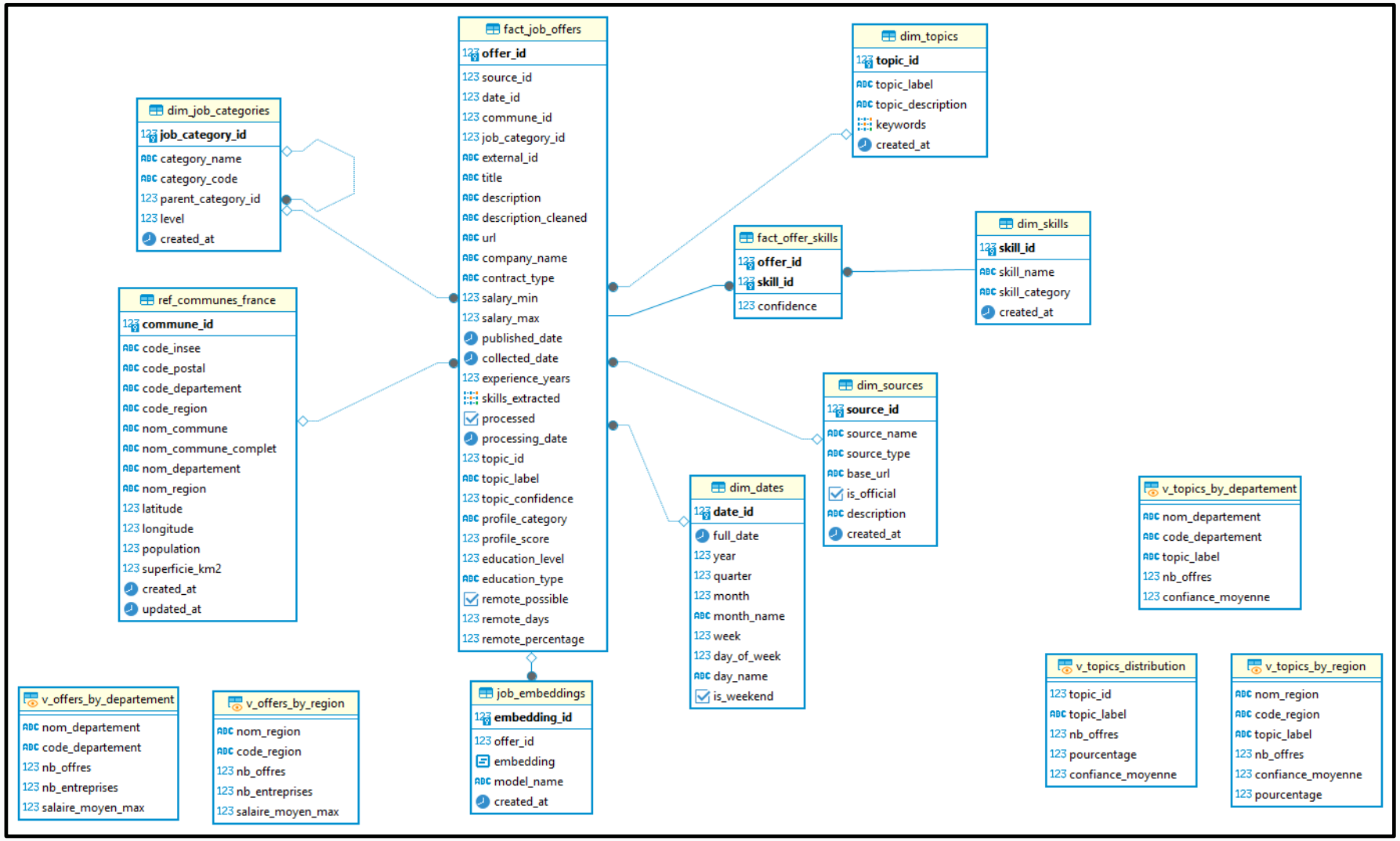
## France Travail

- Données structurées
- API OAuth2
- Limite de 3000 offres/requête
- Coordonnées fournies

## Welcome to the Jungle

- Web Scraping Selenium
- Aucune
- Dépend du scrolling infini (environ 500 offres)
- Géocodage nécessaire

# Entrepôt de données




- Base de données **Supabase** (PostgreSQL)
- Modèle en étoile
- dimensions :

Topics,

Skills,

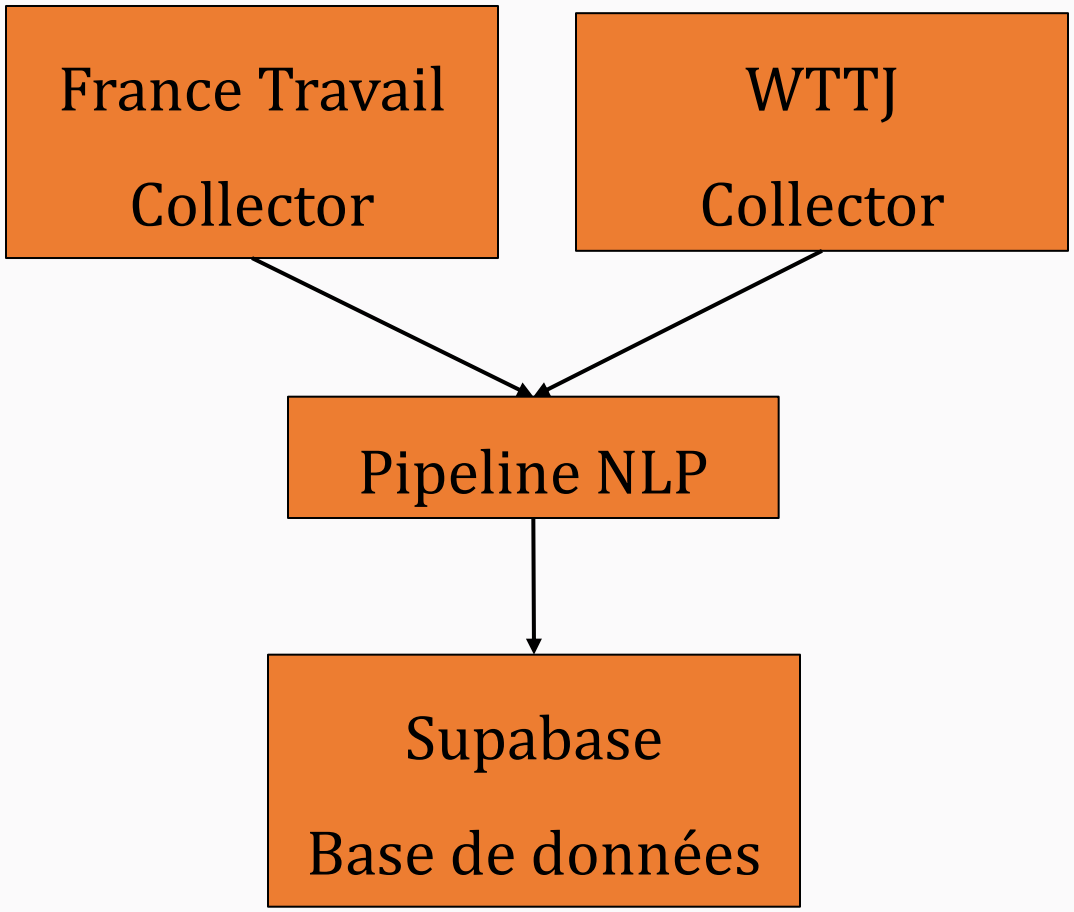
Sources,

Temps,

Localisation,

Catégorie métier

# Entrepôt de données




Pipeline d'alimentation de la base de données

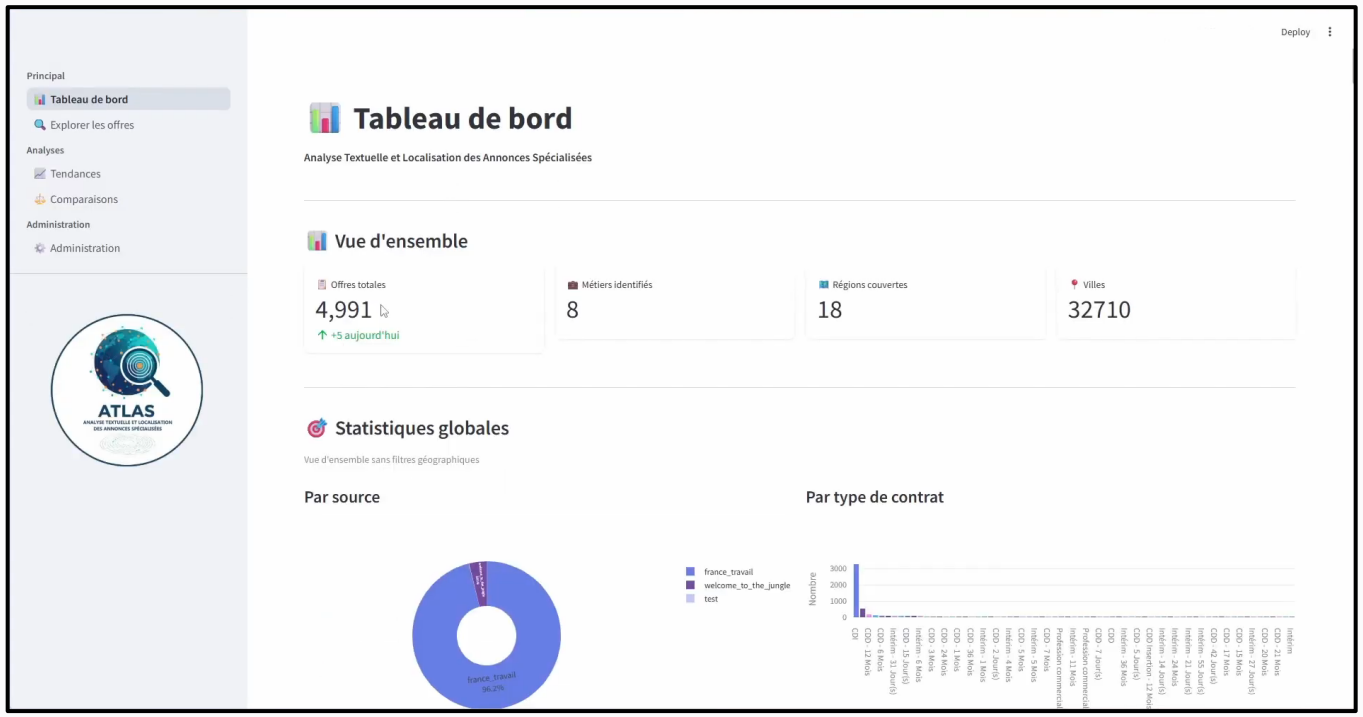
Atouts :

- Sauvegarde locale (dédoublonage)
- Enrichissement NLP
- Coordonnées GPS géocodés si non disponibles
- Détection des doublons
- Bonne gestion des vecteurs par PostgreSQL

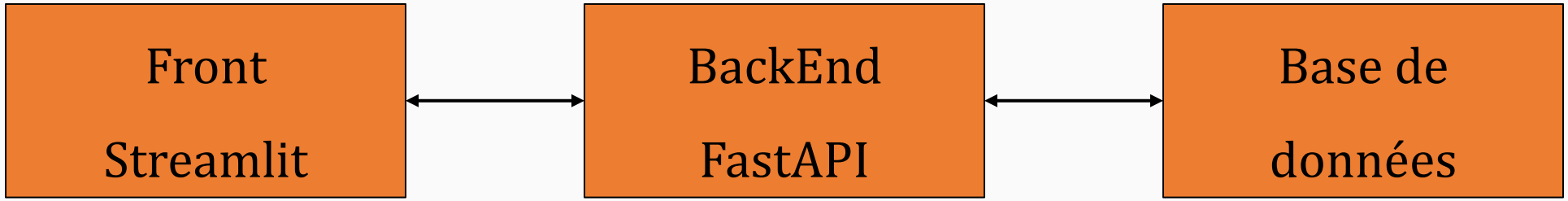
Risques pour la pérennité :

- Changement de l'API
- WTTJ peut bloquer le scraping
- Coûts pour Supabase pour scale

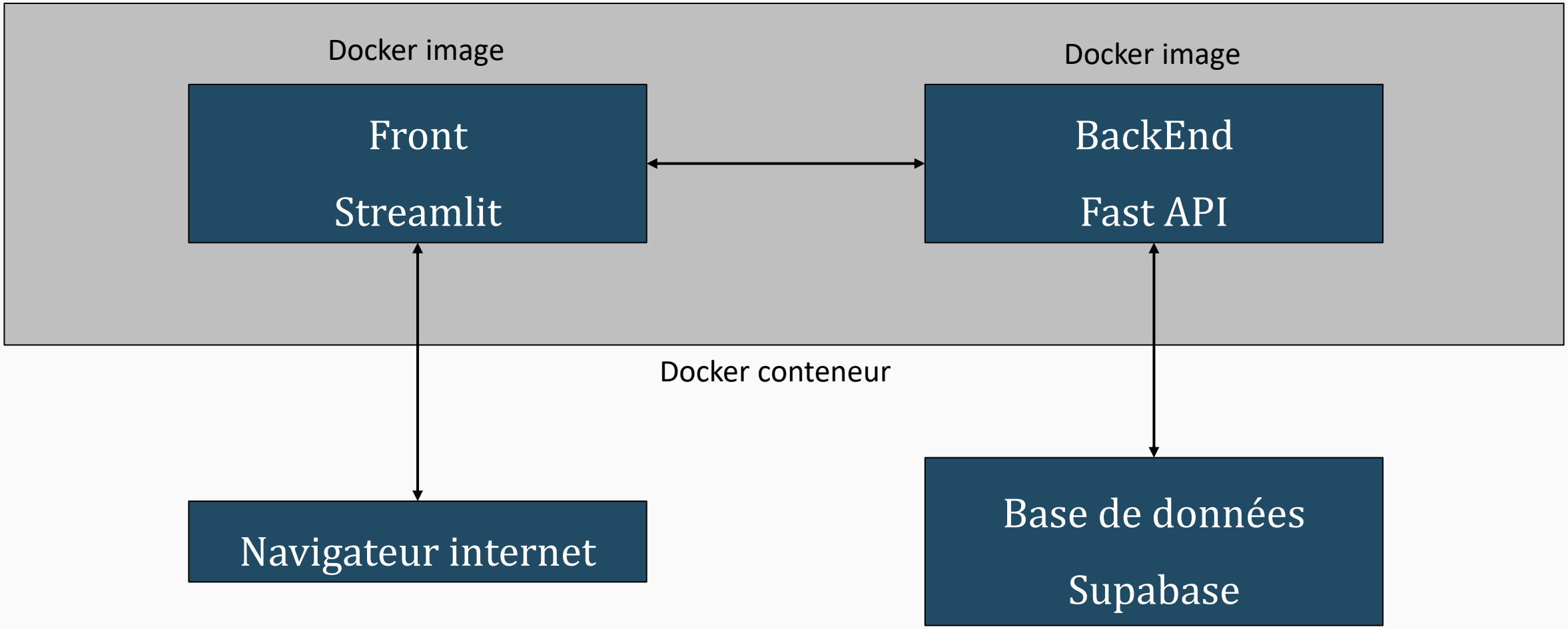
# Web app. Streamlit



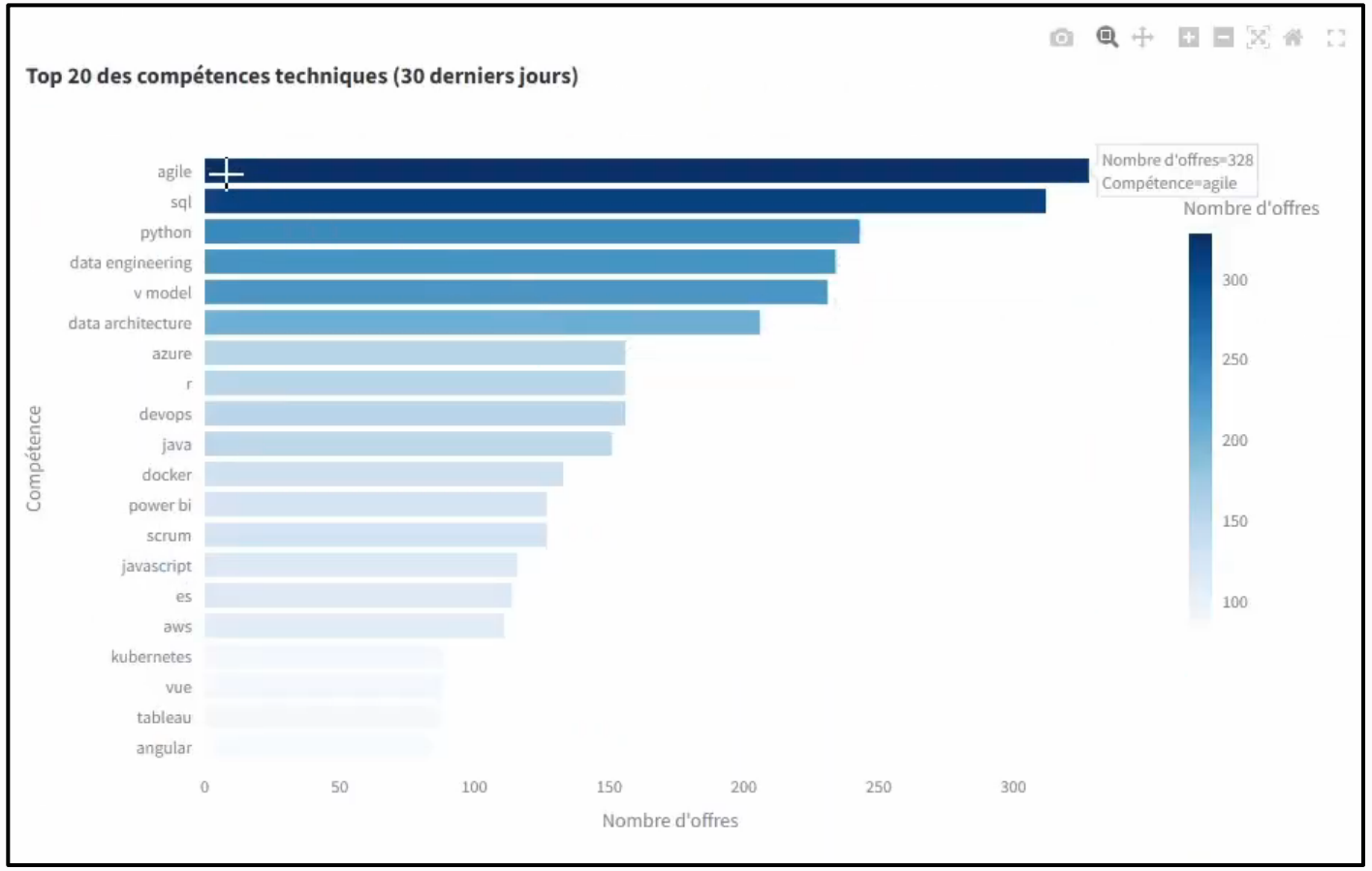

**Dashboard** : KPIs principaux, répartition par source/contrat
**Carte** : Visualisation géographique des offres (Folium)
**Comparaisons** : Comparaison Data Analyst vs Data Scientist vs Data Engineer
**Tendances** : Évolution temporelle des offres et salaires
**Administration** : Lancement manuel de la collecte, enrichissement NLP



# Dockerisation



# Analyses




**Analyses simples :**

- Nombre d’offres collectée
- Répartition par sources
- Répartition par types de contrat ...

**Analyses géographiques :**

- Par région / département / ville

**Analyses NLP :**

- Compétences les plus demandées
- Soft skills
- Évolution dans le temps

# Bilan critique

- Application efficace, Streamlit + API, enrichissements NLP
- Excellente base pour continuer à itérer et rajouter des améliorations

---

**Possibilités d'ajouter :**

- Scores de profil, score de qualité de l'offre …
- Analyses de biais, salariale avec benchmarks, de sentiment de l'offre …
- Enrichissement avec Glassdoor, nouvelles APIs, résumés d'offres …
- Detection de biais, de phrases problématiques, de contrats illégaux …
- Chatbot RAG …

# Merci de nous avoir écouté.


UNIVERSITÉ LUMIÈRE LYON 2

Cyrille PECNIK - Mohamed Riad SAHRANE - Olivier BOROT