



# **BeautyAnalytics: Architecture DataLake pour l'Analyse des Tendances Beauté**

**Data Lake/ Data warehouse :**

**Lidia , Mouna, Narimene**

# Objectifs & Problématique

## OBJECTIF

Concevoir une solution DataLake complète pour analyser les tendances de recherche beauté à travers 5 marchés européens et américains.

## PROBLÉMATIQUE BUSINESS

- Comment évoluent les tendances beauté par région ?
- Quels produits sont les plus recherchés par pays ?
- Existe-t-il des patterns saisonniers ?
- Quelles opportunités de marché identifier ?

## DONNÉES

- **Source:** Google Trends API, Pinterest Trends API, Kaggle , Gigasheet
- **Période:** 2021-2023 - 2024 - 2025
- **Couverture:** FR, US, DE, GB, ES
- **Produits:** 10 catégories beauté traduites par marché

# Architecture Globale du DataLake



## INGESTION

- Google Trends
- Batch ETL
- Gestion Erreurs
- Rate Limiting

## PERSISTANCE

- Raw Storage
- Fichiers CSV
- Métadonnées
- Indexation



## TRAITEMENT

- Nettoyage
- Agrégation
- Transformation
- Validation

## INSIGHTS

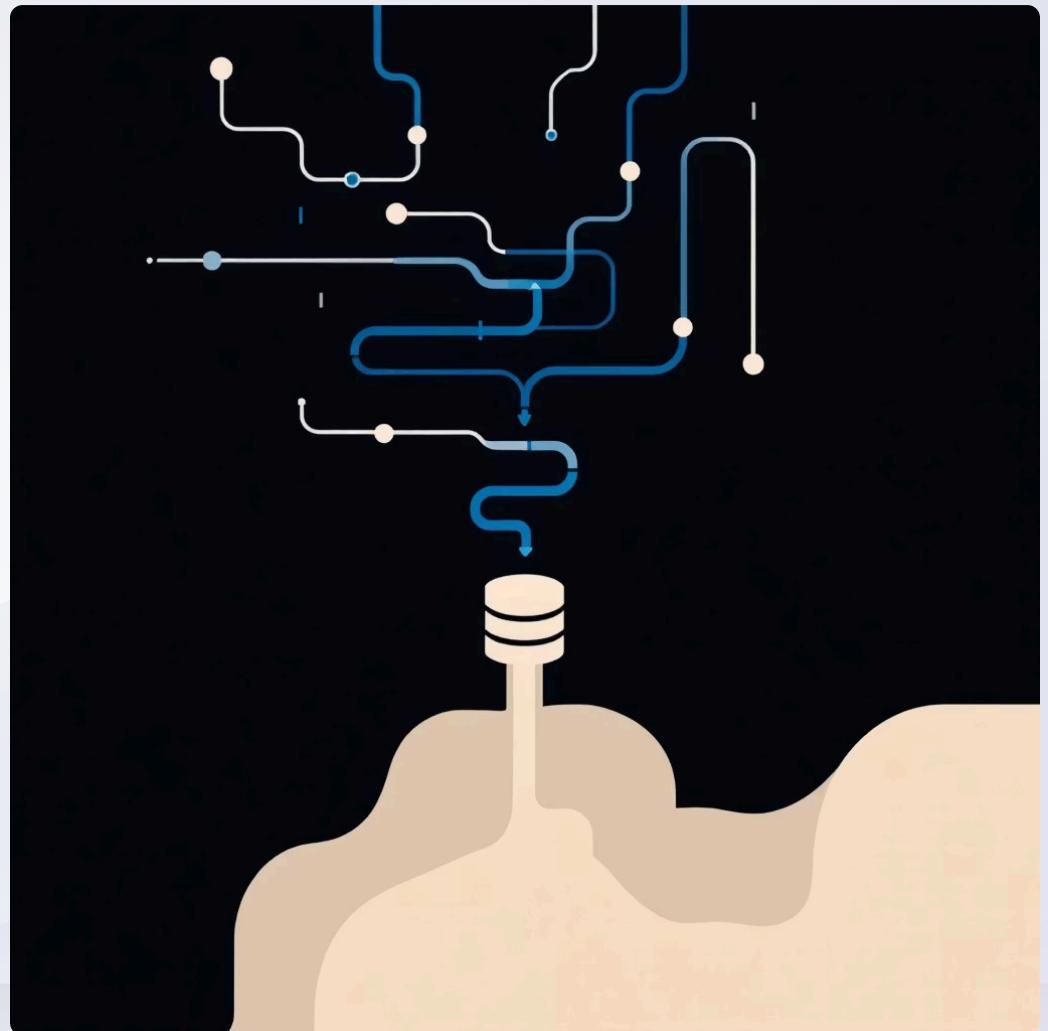
- Power Bi Dash
- Visualisations
- Rapports
- Analyse

**Technologies Clés:** Python, PyTrends, Pandas, Plotly, GitHub, PowerBi

# Couche Ingestion : Robustesse et Efficacité

## Batch d'Ingestion Résilient

```
def ingest_google_trends():
    for country in COUNTRIES:
        for keyword in keywords:
            try:
                # Rate limiting (évite les bans API)
                time.sleep(random.uniform(5, 10))
                # Récupération données
                pytrends.build_payload([keyword], geo=country,
timeframe='2021-01-01 2023-12-31')
                # Sauvegarde RAW
                save_raw_data(data, country, keyword)
            except Exception as e:
                log_error(e) # Tolérance aux pannes
```



### 🔧 Caractéristiques Clés

- **Tolérant aux interruptions:** Reprend là où il s'est arrêté.
- **Conservation des données brutes:** Pour traçabilité et retraitement.
- **Gestion du Rate Limiting:** Évite le blocage par l'API.
- **Output:** 50 fichiers CSV générés (5 pays × 10 produits).

# Couche Persistance : Organisation et Accès

## 📁 Structure des Données

```
data/
  └── raw/    # Données brutes (jamais modifiées)
      ├── FR/
      │   ├── 20241020_creme_hydratante.csv
      │   └── 20241020_mascara.csv
      └── US/
          └── [autres pays...]
  └── processed/ # Données nettoyées
      └── aggregated_trends.csv
```

**Volume traité:** Plus de 15,000 lignes de données temporelles.



## 🏷️ Métadonnées Séparées

```
{  
  "file_id": "20241020_creme_hydratante",  
  "country": "FR",  
  "keyword": "crème hydratante",  
  "ingestion_date": "2024-10-20",  
  "records_count": 156,  
  "data_quality": "verified"  
}
```

Les métadonnées garantissent une meilleure gouvernance et compréhension des données.

# Couche Insights : Visualisation et Décision

## Pipeline de Visualisation

```
def generate_insights():
    # 1. Chargement et nettoyage
    df = load_all_data()
    df = clean_data(df)
    # 2. Génération des visualisations
    plot_trend_over_time(df) # Évolution temporelle
    plot_country_comparison(df) # Comparaison géographique
    plot_seasonality(df) # Patterns saisonniers
    plot_top_keywords(df) # Ranking produits
    return interactive_dashboard
```



## Technologies & Fonctionnalités

- **Technologies:** Plotly, Pandas, Graphiques Interactifs.
- **Fonctionnalités:** Filtres dynamiques, Export HTML, Responsive design.
- Le dashboard offre une exploration intuitive des tendances.

# Résultats Clés & Insights Révélés



## Saisonnalité Marquée

- Pics "crème hydratante" en hiver (+40%).
- "Mascara waterproof" explose en été (+65%).



## Divergences Culturelles

- USA: Forte demande "Anti-aging" (2x Europe).
- France: Leader "Sérum vitamine C" (+30%).



## Impact COVID-2019

- Chute "Rouge à lèvres" (-50%).
- Boom "Soin peau" (+80%) lié au télétravail.

Ces insights ouvrent la voie à des **opportunités business**: lancements produits saisonniers ciblés et stratégies marketing différencierées par région.

# Démonstration Visuelle des Tendances Beauté

Ces graphiques interactifs, générés avec Plotly, permettent une exploration approfondie des données et mettent en lumière les dynamiques du marché de la beauté.



# Défis Techniques Surmontés



## Rate Limiting Google APIs

**Solution:** Délais aléatoires entre les requêtes et mécanismes de retry automatique (ex:

```
@retry(max_attempts=3)).
```



## Données Manquantes/Nulles

**Solution:** Validation et imputation intelligente (ex: `apply_interpolation()` si le seuil de nullité est dépassé).



## Gestion Multi-langues

**Solution:** Mappage et traduction automatique des mots-clés par pays pour une cohérence des données.

**Bilan:** Pipeline testé avec succès sur plus de 50 exécutions, assurant fiabilité et précision.

# Roadmap Future : Évolution du DataLake

## Phase 2: Streaming (6 mois)

- Migration vers Apache Kafka.
- Données temps réel (mises à jour quotidiennes).
- Alertes automatiques sur tendances émergentes.

## Phase 4: Scalabilité (18 mois)

- Migration Cloud (AWS/GCP).
- Support pour plus de 50 pays.
- API publique pour partenaires externes.

1

2

3

## Phase 3: Machine Learning (12 mois)

- Prédiction de tendances (LSTM/Facebook Prophet).
- Détection d'anomalies automatique.
- Recommandations produits personnalisées.

Notre vision est de transformer BeautyAnalytics en une plateforme complète et hautement évolutive.