Oussama Essalehi

Data Analyst

+33 7 52 07 31 00 | oussama.essalehi@gmail.com | linkedin.com/oussama-essalehi | github.com/Oussama-essalehi | oussama-essalehi.dev | Toulouse, France

Formation

Université Toulouse III – Paul Sabatier Sept. 2024 – Août 2026
Master MIAGE – Ingénierie des Données et Protection (IDP) *Toulouse, France*
Cours : *Data Science, Data Engineering, Machine Learning, Big Data, Gen AI, Cloud & DevOps, Gestion de projet agile, Gouvernance des données*

Expérience

Data Analyst Sept. 2024 – Août 2026
Wewise *Toulouse, France*

- Conception et automatisation de tableaux de bord Power BI (DAX) à partir des données de l'ERP, pour le suivi de l'activité, des ventes, des achats et des stocks.
- Création de flux Power Query et automatisation de traitements récurrents via Excel VBA.
- Développement de requêtes SQL Server pour l'extraction, le nettoyage et la consolidation des données de gestion.
- Développement de scripts Python (pandas, numpy) de contrôle qualité : détection des doublons, des valeurs manquantes et des incohérences entre sources.
- Documentation des traitements et des indicateurs, formation ou accompagnement des utilisateurs.

Projets

Plateforme de données e-commerce batch & streaming | Demo | Code
Python, Kafka, Spark, Airflow, dbt, PostgreSQL, Great Expectations, Docker

- Conception de pipelines d'ingestion batch (API, CSV) et streaming (commandes via Kafka), orchestrés par Airflow avec reprises et alertes.
- Structuration d'un Data Lake en architecture médaillon (bronze, silver, gold) sur Delta Lake, puis modélisation d'un Data Warehouse en schéma en étoile avec dbt.
- Mise en place de contrôles qualité automatisés (Great Expectations, tests dbt) : complétude, unicité, fraîcheur des données, avec suivi des anomalies.
- Mise à disposition de data marts pour les usages analytiques (Power BI) et data science (features pour la prévision de la demande).
- Documentation des pipelines, du lignage et des schémas via dbt docs, et conteneurisation de l'ensemble avec Docker Compose.

VéloScope Toulouse – analyse et prévision des vélos en libre-service | Demo | Code
Databricks, PySpark, Delta Lake, MLflow, scikit-learn, Streamlit, Power BI

- Collecte continue de l'état des 450 stations VélôToulouse (flux GBFS) et de la météo, via des jobs Databricks planifiés.
- Lakehouse Delta Lake en architecture médaillon (bronze, silver, gold) en PySpark et Spark SQL, avec mise en quarantaine des lignes incohérentes et alertes sur le taux de rejet.
- Reconstruction des départs et arrivées à partir des variations de stock entre deux relevés, les API ne publiant que l'état des stations.
- Segmentation des stations par clustering et prévision à 24 h par gradient boosting, comparée à une référence simple et suivie dans MLflow.
- Application Streamlit (carte en direct, prévisions, rééquilibrage recommandé) et tableau de bord Power BI en DirectQuery.

Immoscope France – marché immobilier 2021-2025 | Demo | Code
Python (pandas, SciPy), PostgreSQL, SQL, Docker, Power BI (DAX), React, Plotly

- Pipeline Python de téléchargement et de nettoyage de 385 fichiers publics DVF (18 M de lignes) ramenés à 3,8 M de ventes fiables, avec tests automatisés.
- Modélisation en schéma en étoile sous PostgreSQL (Docker) et requêtes SQL d'analyse : fonctions de fenêtre, médianes, GROUPING SETS.
- Analyse statistique des écarts entre villes (tests de Mann-Whitney, correction de Bonferroni, tailles d'effet), avec contrôle croisé des médianes SQL et Python.
- Restitution par un tableau de bord Power BI de 4 pages (mesures DAX, simulateur de budget) et un site React bilingue avec carte interactive.

Compétences techniques

Langages : Python (pandas, numpy), SQL, Java, JavaScript
Data Engineering : Spark / PySpark, Kafka, Airflow, dbt, Delta Lake, Databricks, ETL / ELT, Great Expectations
Cloud & DevOps : AWS, Azure, Docker, Git, FastAPI
IA & LLM : scikit-learn, MLflow, Hugging Face, LangChain, RAG, LLM APIs, Streamlit
Bases de données : PostgreSQL, SQL Server, MySQL, MongoDB, Oracle
Visualisation & BI : Power BI (DAX), Power Query, Excel (VBA), Tableau
Langues : Français, Anglais

Centres d'intérêt

Échecs, poker et jeux de stratégie · Course à pied et fitness.