ELGHAZI SOUFIANE

Junior Data Engineer

@ soufiane.el-ghazi@outlook.com   in soufiane-el-ghazi   📞 +212 6533-61122   📍 Casablanca, Morocco   🌐 elghazisoufiane.vercel.app/

EDUCATION

Data and Knowledge Engineering	🗓 Oct. 2021 – June 2024
🎓 **School of Information Sciences [ESI]**	📍 Rabat
Preparatory Classes	🗓 Sept. 2019 – June 2021
🎓 **Ibn Tahir High School**	📍 Errachidia
Baccalaureate in Mathematics B	🗓 Sept. 2018 – June 2019
🎓 **Ibn Al Haitam Technical High School**	📍 Ouarzazate

PROFESSIONAL EXPERIENCE

Data Consultant
💼 **ENOVA RT • Full-time • On-site**

🗓 Sept. 2024 – Present • 1 year   📍 Rabat

- Implemented a centralized data warehouse for BI operations with ETL processes, dimensional modeling, and complete documentation
- Suggested and developed a CV matching application using NLP to improve recruitment efficiency
- Proposed and deployed a local chatbot for documentation using RAG and Ollama to optimize knowledge sharing
- Developed SQL and PL/SQL scripts for data ingestion, extraction, and updates
- Designed and deployed dynamic dashboards for hospital, billing, and payment metrics

`Python` `PowerBI` `PostgreSQL` `SQL Server` `Talend` `RAG` `Fastapi`

Data Science Intern (PFE)
💼 **ENOVA RT • Internship • On-site**

🗓 Feb. 2024 – Aug. 2024 • 6 months   📍 Rabat

Predictive Modeling of Nosocomial Infections
- Developed a predictive model for nosocomial infections with 92% accuracy, deployed in a web application, with automated ETL/ML pipelines
- Built a classification model for 50 cancer types achieving 88% accuracy

`Python` `Kafka` `Spark` `Airflow` `Django` `Celery` `Redis`

PROJECTS

Financial Statement Extraction Tool:
Developed an automated tool to extract and display financial statements from PDF financial reports
`OCR` `OpenCV` `PyTesseract` `Flask` `React`

Real-Time Customer Churn Prediction:
Real-time data processing for predictive churn analysis
`Spark` `Kafka` `Power BI` `Snowflake`

CERTIFICATIONS

- ☁ **Microsoft Azure Fundamentals (AZ-900)**
- ☁ **Microsoft Azure Data Fundamentals (DP-900)**
- 📊 **Microsoft Power BI Data Analyst Associate (PL-300)**
- ☁ **Microsoft Fabric Data Engineer Associate (DP-700)**

LANGUAGES

- **Arabic**: Native
- **English**: Fluent (Speaking, Reading, Writing)
- **French**: Fluent (Speaking, Reading, Writing)

ACHIEVEMENTS

- 🎓 **My Microsoft Power BI Udemy Course**
- 🎓 **My Learn Git Udemy Course**
- Ⓜ **Data Engineering Concepts Article**
- Ⓜ **Understanding DAX Contexts**

SKILLS

Programming Languages
`Python` `SQL` `DAX`

Databases
`PostgreSQL` `Elasticsearch` `MongoDB` `Snowflake` `Cassandra`

Big Data & Cloud Computing
`Spark` `PySpark` `Hadoop` `Kafka` `Git` `Docker` `Azure`

ETL & Data Modeling
`Talend` `DBT` `Dimensional Modeling` `Airflow` `Power BI`

Soft Skills
`Problem Solving` `Communication` `Adaptability` `Time Management`