

ELGHAZI SOUFIANE

Data Engineer

@ soufiane.el-ghazi@outlook.com

soufiane-el-ghazi

+212 6533-61122

SoufianeElghazi

Casablanca, Morocco

elghazisoufiane.vercel.app/

M elghazisoufiane02

</> Soufiane-EL



SUMMARY

”Data is a story waiting to be told—you just need to know how to listen.”

Data Engineer with over a year of experience, I specialize in turning raw data into structured, actionable insights. Proficient in **Python**, **SQL**, and **Spark**, I design and optimize **ETL pipelines**, build high-performance **data architectures**, and tackle complex challenges with a sharp analytical mindset. Always seeking innovation, I adapt quickly to **emerging technologies** to deliver reliable data solutions.

Combining engineering expertise with advanced analytics, I implement **machine learning models** while maintaining robust data infrastructure. My experience spans from developing **real-time analytics pipelines** to building **interactive dashboards**, ensuring scalable solutions.

EDUCATION

Data and Knowledge Engineering

School of Information Sciences [ESI]

October 2021 – June 2024

Rabat

Preparatory Classes

Ibn Tahir High School

September 2019 – June 2021

Errachidia

Baccalaureate in Mathematics B

Ibn Al Haitam Technical High School

September 2018 – June 2019

Ouarzazate

WORK EXPERIENCE

Data Consultant

ENOVA RT • Full-time • On-site

September 2024 – Present • 7 Month

Rabat

- Developed ETL processes using Pandas integrated with PostgreSQL
- Implemented a centralized data warehouse for BI operations with ETL processes, dimensional modeling, and comprehensive documentation
- Suggested and developed a CV matching application leveraging NLP to improve recruitment process
- Proposed and implemented a local chatbot for documentation using RAG and Ollama to streamline knowledge sharing
- Developed and maintained backend APIs using Flask and FastAPI for internal data services
- Developed SQL and PL/SQL scripts for data injection, extraction, and updates
- Designed and deployed dynamic dashboards for hospital, billing, and payments metrics, enhancing decision-making speed

Python

PowerBI

SQL

Talend

DBT

RAG

Java

Fastapi

Computer Vision & AI Engineer

FIVERR • Freelance • Remote

May 2024 – September 2024 • 4 Months

Paris

Card Detection and Extraction

- Developed a verification feature for a mobile application using Flutter, OCR for text extraction, and OpenCV for image enhancement, focused on national ID card (CIN) validation during user login
- Developed custom web scraping for client-specific data collection tasks
- Built backend APIs with Flask for data retrieval
- Collaborated with the client to define project requirements and deliver tailored validation solutions

Python

Beautifulsoup

OpenCV

Flask

React

OCR

Yolo

Data Science PFE Intern

 ENOVA RT • Internship • On-site

 February 2024 – August 2024 • 6 Months

 Rabat

Predictive Modeling of Nosocomial Infections

- Developed a 92%-accurate model for predicting nosocomial infections, deployed it in a web app, and automated the ETL/ML processes
- Implemented a real-time prediction pipeline and Engineered an 88%-accurate model for classifying 50 cancer types

Python Kafka Spark Airflow Django Celery Redis

PROJECTS

Real-Time Churn Prediction:



Developed a pipeline using Kafka, PySpark, and Snowflake to process real-time churn signals and built dashboard in Power BI for stakeholder insights

- **Challenge:** Managing high-velocity data streams while maintaining prediction accuracy

Spark Kafka Power BI Snowflake

Financial Statement Extraction Tool :



Developed an automated tool to extract and display financial statements from PDF financial reports

- **Challenge :** Extract key financial data from complex PDF documents

OCR OpenCV PyTesseract Flask React

SKILLS

Programming Languages

Python SQL Java M Language DAX

Machine Learning / Deep Learning

Computer Vision NLP Transformers LLMs RAG Ollama

Tools / Software

Docker Linux Kafka Talend Airflow Power BI Git

Databases

PostgreSQL SQL Server Mongodb Cassandra Snowflake Chroma db

Big Data & Cloud Computing

Spark Hadoop Azure

Soft Skills

Problem-Solving Communication Adaptability Time Management

CERTIFICATIONS

-  Microsoft Azure Fundamentals (AZ-900)
-  Microsoft Power BI Data Analyst Associate (PL-300)
-  Microsoft Fabric Data Engineer Associate (DP-700)
-  Microsoft Azure Data Fundamentals (DP-900)
-  Google Cloud Certified Professional Data Engineer
-  AWS Certified Data Engineer - Associate
-  Databricks Certified Associate Developer for Apache Spark

LANGUAGES

- **Arabic:** Native
- **English:** Fluent (Speak, Read, Write)
- **French:** Fluent (Speak, Read, Write)

MY ENDEAVORS

-  My Microsoft Power BI Udemmy Course
-  Understanding Power BI's Engines
-  Data Engineering Concepts Article
-  My Learn Git Udemmy Course
-  Understanding DAX Contexts