



SOUHAYLA ELMEFTAHI

Fourth-year data engineering student

Looking for an PFA internship

☎ (+212) 668546041

🌐 souhaylael.github.io

🌐 souhayla Elmeftahi

✉ souhaylaelmeftahi@gmail.com

📍 MARROC

🌐 <https://github.com/souhayla18>

TECHNICAL SKILLS

Programmation: Python,java, R,C

Cloud: azure ,cloud, gcp (bigquery)

Big data tools: hadoop ecosystem

Databases : SQL,NewSQL ,noSQL

Scripting : shell bash scripting

Data Engineering : Selenium,mage , PowerBi, LookerStudio dbt,NLP ,machine learning,deep learning

Data Modeling : Entity-relationship Merise, UML

Web development : HTML/CSS, typescript

OUTILS: docker, git/github

Operating system: Linux , windows

CERTIFICATS

Machine Learning - Python

supervisor & unsupervised learning

Certificate by DeepLearning.AI, Stanford University - coursera

Big Data

Hadoop ecosystem masterclass

Certificate by Udmey

Date Engineer

+12 data (engineering/science/analytics)

certificates Certificate by DataCamp

LANGUAGES

- ARABIC (NATIVE)
- FRENCH (PROFICIENT)
- ENGLISH (PROFICIENT)
- GERMAN (BEGINNER)

SOFT SKILLS

- Teamwork •Technology Monitoring
- Problem solving • Communication
- Adaptability • Active Listening

INTERESTS

- Avid reader • Blogging
- Sport: kickboxing,trekking

EXTRACURRICULAR

- dataClub : member
- 01 club : member

EDUCATION

📍 **National School of Applied Sciences of AI- Hoceima**

data engineering field

2022 - to be completed in 2025

📍 **National School of Applied Sciences of AI- Hoceima**

Preparatory cycle: Sciences & Techniques for the Engineer
2020 - 2022

📍 **Faculty of Science and Technics of AI Hoceima**

Mathematics, Computer Science, Physics

2019 - 2020

PROJECTS

End-to-End azure Cloud Migration and CI/CD pipeline Implementation

PROJET DATA ENGINEERING [GITHUB](#) [DEMO](#)

- Connect an on-premise postgres database to Azure Cloud.
- Develop a batch pipeline using Azure Data Factory to import data, transform it with Databricks, and load it into a Synapse data warehouse,create report with PowerBI.
- Implemented a robust CI/CD pipeline on Azure DevOps with Git workflows and branch protection for code reliability and stay up-to-date with changes.

Tools used: SQL Server, Azure Data Factory, Data Lake, Synapse, Spark, Databricks, PowerBI,Azure DevOps (CI/CD).

Web Scraping & Unsupervised Learning on ResearchGate Articles

PROJET DATA science & ENGINEERING [GITHUB](#) [article](#)

- Investigate scientific articles on ResearchGate through web scraping (Selenium).
- Use neural networks for dimensionality reduction.
- Apply unsupervised learning algorithms (DBSCAN, k-means, k-prototype) for clustering patterns.
- Leverage Pandas and NumPy for data preprocessing Utilize Matplotlib for visualization

Tools Used:Selenium,Neural ,Networks ,Pandas ,NumPy ,Matplotlib ,Unsupervised Learning Algorithms (DBSCAN, k-means, k-prototype ...)

Airbnb Company Data Exploration and Reporting

PROJET DATA ENGINEERING [GITHUB](#) [DEMO](#)

- Develop an ETL pipeline using mage for data extraction, transformation, and loading.
- Ingesting data from various sources, including Azure Blob storage, local storage, HTTP API, and AWS storage.
- Consolidate data in Snowflake, clean using dbt, and generate Power BI reports.

Tools Used:Azure , Blob Storage ,Snowflake, dbt ,Power BI, mage ,Docker ,Postgres

EXPERIENCE

Data Engineering & web development

2023 June-July

📍 **IT department at ENSAH** [GITHUB](#) [DEMO](#)

- Explore and visualize E-commerce data.
- Utilize SQL for querying , BigQuery as the database, Looker Studio for dashboarding.

Tools Used: Sal , Bigquery , Looker Studio

• Develop a Contact Manager and Group Management Java Desktop Application.

Tools Used: Java , Spring,JDBC,H2

Data Analyst

2024 Jun-febr

📍 **virtual internship training with Meriskill**

[GITHUB](#)

- HR Analytics
- Healthcare Analytics

Tools Used: Looker Studio, Power BI