

Anis Guechtouli

guechtoulianiss7@gmail.com | +213 555648000 | Algiers, Algeria | in/anis-guechtouli/ | /GuechtouliAnis

PROFILE

Motivated Data Scientist with one year of experience. Completed a 5-month Data Analyst internship at CAAR, focusing on data cleaning, visualization, and reporting. Recently finished a 7-month Data Scientist internship at Djezzy, where I worked on an NLP project to develop a conversational AI solution. Passionate about leveraging data and machine learning to drive innovation and solve real-world problems.

EDUCATION

Master's Degree in Data Science and Analytics 09/2023 – 06/2025
University Of Algiers 1 - Benyoucef Benkhedda

Bachelor's Degree in Computer Science 11/2020 – 07/2023
University Of Algiers 1 - Benyoucef Benkhedda

EXPERIENCE

Data Scientist Intern – Djezzy, Algiers, Algeria 11/2024 – 06/2025

- Developed an **NLP-based** conversational AI solution using **Large Language Models** to automate repetitive information delivery tasks.
- Processed** and **structured** Q&A data to support model training and evaluation.
- Applied **prompt engineering** techniques to improve the consistency and relevance of generated answers.

Data Analyst Intern – CAAR, Algiers, Algeria 02/2023 – 07/2023

- Developed a user-friendly data analysis dashboard to visualize and analyze several metrics in the automobile industry for an insurance agency.
- Worked closely with the Vice President of the Automobile branch and the IT team to deliver accurate insights into the agency's performance.
- Successfully identified significant changes in key metrics over specific periods by visualizing and analyzing the data.

SKILLS

- Languages** : Arabic (Native speaker), English (Advanced proficiency), French (Intermediate proficiency)
- Programming** : SQL, Git, Python (key libraries: PySpark, Pandas, Numpy, Matplotlib, Plotly, Scikit-learn, LangChain, Streamlit)
- Data Managment** : Spark, SQLite, Oracle, PostgreSQL
- Dashboarding & BI Tools** : Excel, PowerBI, Tableau
- Communication & Presentation** : Canva, PowerPoint, Markdown

PROJECTS

Expected Goals (xG) Model 01/2025 – present

- Collected free StatsBomb data, building a dataset of approximately **12 million rows** from multiple competitions.
- Used **PySpark** to preprocess and clean the large dataset, ensuring quality and consistency.
- Engineered **7 features**, including **shot distance**, **angle**, dummy variables, and player attributes like **preferred foot**.
- Logistic Regression** achieved ~**90%** accuracy, **98.7%** matching rate with StatsBomb xG, and strong regression performance (MSE **0.0037**, R^2 **0.83**) in predicting shot outcomes.

Exploratory Data Analysis in Retail 07/2024 – 07/2024

- Conducted **exploratory analysis** on a retail dataset of **9,994 sales**, identifying key issues such as **unprofitable products** and **excessive discounting** in specific states.
- Developed **visualizations**, a **dashboard**, and a **detailed report** to clearly communicate insights and guide **strategic decision-making**.

Football Web Scraping 02/2024 – 06/2024

- Scraped data from FBref using **BeautifulSoup**, **Requests**, and **Pandas**.
- Prepared the data for further analysis of various player metrics.