

# Anis Guechtouli

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

## PROFIL

Data Scientist motivé avec un an d'expérience en entreprise. Stage de 5 mois en tant que Data Analyst chez CAAR, axé sur le nettoyage de données, la visualisation et le reporting. Dernièrement, stage de 7 mois chez Djezzy en Data Science, où j'ai travaillé sur un projet de NLP pour développer une solution d'IA conversationnelle. Passionné par l'exploitation des données et du machine learning pour innover et résoudre des problèmes concrets.

## ÉDUCATION

**Master Analyse et Sciences de Données** 09/2023 – 06/2025  
Université d'Alger 1 - Benyoucef Benkhedda

**Licence Informatique** 11/2020 – 07/2023  
Université d'Alger 1 - Benyoucef Benkhedda

## EXPÉRIENCE

**Stagiaire Data Scientist – Djezzy, Alger, Algerie** 11/2024 – 06/2025

- Développé une solution d'IA conversationnelle basée sur le **NLP** et les **Large Language Models** pour automatiser des tâches répétitives de transmission d'informations.
- Traitement et structuration** des données au format Q/R pour l'entraînement et l'évaluation du modèle.
- Appliqué des techniques de **prompt engineering** pour améliorer la cohérence et la pertinence des réponses générées.

**Stagiaire Analyste de Données – CAAR, Alger, Algerie** 02/2023 – 07/2023

- Développement d'un tableau de bord intuitif pour analyser plusieurs indicateurs du secteur automobile au sein d'une agence d'assurance.
- Collaboration avec le vice-président de la branche automobile et l'équipe IT pour fournir des analyses précises sur la performance de l'agence.
- Identification de variations significatives sur des indicateurs clés grâce à la visualisation des données.

## COMPÉTENCES

- Langues** : Arabic (Native speaker), English (Advanced proficiency), French (Intermediate proficiency)
- Programmation** : SQL, Git, Python (bibliothèques principales : PySpark, Pandas, Numpy, Matplotlib, Plotly, Scikit-learn, LangChain, Streamlit)
- Gestion des données** : Spark, SQLite, Oracle, PostgreSQL
- Visualisation des données** : Excel, PowerBI, Tableau
- Communication et présentation** : Canva, PowerPoint, Markdown

## PROJETS

**Modèle Expected Goals (xG)** 01/2025 – Actuel

- Collecte de données gratuites StatsBomb, création d'un dataset de **12 millions de lignes** sur plusieurs compétitions.
- Prétraitement et nettoyage avec **PySpark** pour garantir qualité et cohérence.
- Ingénierie de **7 caractéristiques** clés : distance, angle, variables binaires, pied préféré.
- Modèle de **régression logistique** atteignant environ **90 % de accuracy**, **98,7 % de taux de correspondance** avec le modèle xG de StatsBomb, avec de solides performances en régression (**MSE 0,0037**, **R<sup>2</sup> 0,83**).

**Analyse Exploratoire des Données dans le Commerce de Détail** 07/2024 – 07/2024

- Réalisé une **analyse exploratoire** sur un dataset retail de **9 994 ventes**, identifiant des problèmes clés comme des **produits non rentables** et des **remises excessives** dans certains états.
- Créé des **visualisations**, un **tableau de bord** et un **rapport détaillé** pour communiquer clairement les résultats et orienter les **décisions stratégiques**.

**Extraction de données du football** 02/2024 – 06/2024

- Extraction des données du site FBref en utilisant **BeautifulSoup**, **Requests** et **Pandas**.
- Préparation des données pour une analyse future des différents aspects des joueurs.