



THIERNO SADOU DIALLO

Stage Data Analyst / Business Analyst

SQL · R · Python · Power BI · Modélisation statistique · Data Quality

Caen (14), mobile toute la France | thiernosadoudiallo25@gmail.com

linkedin.com/in/thierno-sadou-diallo-a30179208

Stage de fin d'études · 4 à 6 mois · Dès janvier 2027

PROFIL

Étudiant en **Master 2 Statistiques appliquées et analyse décisionnelle (SAAD)** à l'Université de Caen Normandie, après **5 ans d'enseignement des mathématiques**. Je transforme des données brutes en **indicateurs fiables et analyses utiles à la décision** : nettoyage et contrôle qualité, modélisation statistique, entrepôt de données et tableaux de bord. Mon expérience d'enseignant m'a appris à rendre des résultats techniques clairs pour tous.

PROJETS DATA

Parcoursup : analyse de la tension des formations (R, Shiny) 2025–2026

- Nettoyage des données nationales 2024 (**14 079 formations**, 118 variables) : pipeline R automatisé, contrôle qualité, traitement des valeurs extrêmes.
- Focus Normandie 2020–2024 (**près de 3 000 observations**) : KPI de pression, remplissage, taux de proposition et d'acceptation, suivi des évolutions.
- Typologie des formations (ACP/AFDM, CAH, k-means) et modélisation de la tension (régressions, Random Forest, XGBoost).
- Restitution des résultats dans des **dashboards interactifs R Shiny**.

Data Warehouse « Fil Rouge » : ETL décisionnel (Apache Hop, SQL) 2025

- Intégration de **plusieurs sources** (base opérationnelle, référentiel des communes, catalogue de prestations) dans un **schéma en étoile** (1 table de faits, 4 dimensions).
- 5 pipelines ETL** sous Apache Hop : nettoyage, lookups, chargement de **15 000 ventes** et 10 000 clients ; historique clients en **SCD Type 2**.
- Vues SQL analytiques et requêtes OLAP : CA mensuel par ville et catégorie, ventes hors région, durées d'intervention.

Campus sans tabac : enquête statistique (R, RMarkdown) 2024–2025

- Nettoyage et analyse d'une enquête de **713 répondants** (87 variables) auprès des étudiants et personnels de l'Université de Caen.
- 14 tests d'indépendance** (χ^2 , Fisher) pour profiler fumeurs et vapoteurs.
- Résultat clé : **78 % d'adhésion** au projet de campus sans tabac ; restitution dans un rapport RMarkdown.

Accidents de la route : modèle de données (Excel) 2025

- Import et transformation de 8 tables avec **Power Query** ; modèle en étoile sous **Power Pivot** et analyses par tableaux croisés dynamiques.

EXPÉRIENCE PROFESSIONNELLE

Professeur de mathématiques, Conakry (Guinée) 2018–2023

- Enseignement des mathématiques en collège et lycée ; préparation des cours et des évaluations, suivi des résultats des élèves.
- Explication de notions abstraites à un public non spécialiste.

FORMATION

Master 2 Maths appliquées, statistique (SAAD) 2025–2027

Université de Caen Normandie

Master 1 Maths appliquées, statistique (SAAD) 2024–2025

Université de Caen Normandie

Master 1 MEEF Mathématiques – Université de Bordeaux 2023–2024

Licence Maths appliquées – Université de Conakry 2015–2018

SQL & DONNÉES

SQL : jointures, agrégations, vues
SQLite
Modélisation en étoile
ETL (Apache Hop)
SCD Type 2
Nettoyage et contrôle qualité

BI & VISUALISATION

Power BI
Tableau
R Shiny, ggplot2
Excel avancé (TCD)
Power Query, Power Pivot
Construction de KPI

STATISTIQUE & ML

Tests statistiques (χ^2 , Fisher)
Régression linéaire
ACP / AFDM
CAH, k-means
Random Forest, XGBoost

LANGAGES & OUTILS

R (tidyverse)
Python (Pandas, NumPy, scikit-learn)
SAS
RMarkdown, LaTeX
Git

SAVOIR-ÊTRE

Pédagogie et vulgarisation
Rigueur
Esprit de synthèse
Autonomie

LANGUES

Français : courant
Anglais : opérationnel

INFORMATIONS

Permis B, véhicule personnel
Mobilité : toute la France
Disponible dès janvier 2027