

Analyse et Gestion des Données – Colonie de Vacances

Une solution fondée sur les données pour les camps d'été des églises de Bingerville

Fobah N'gouan Salomon

Étudiant en Data Science – African School of Economics

github.com/fobahsalomon | fobahngouansalomon@gmail.com | [mon portfolio](#)

1. CONTEXTE

Le camp annuel organisé par le réseau d'églises de Bingerville (Côte d'Ivoire) faisait régulièrement face à des défis logistiques et financiers : le nombre d'enfants présents sur site dépassait souvent le nombre d'inscriptions payées. Cet écart entraînait des déséquilibres budgétaires, des incertitudes en matière d'encadrement et des difficultés de traçabilité.

En tant qu'étudiant en data science et membre de l'équipe organisatrice, j'ai proposé un système structuré de gestion des données afin de moderniser le processus d'inscription, assurer la traçabilité et soutenir la prise de décision en temps réel.

2. OBJECTIFS

Le projet visait quatre objectifs principaux :

- Centraliser toutes les inscriptions dans un système numérique unique.
- Garantir une traçabilité complète (enfant → église ou tuteur inscripteur).
- Nettoyer et standardiser les données pour une fiabilité analytique.
- Produire un tableau de bord interactif pour le suivi opérationnel.

3. MÉTHODOLOGIE

Le travail a été structuré en trois phases :

3.1. Collecte numérique des données

J'ai conçu un formulaire d'inscription adapté aux mobiles à l'aide de **KoboToolbox**, choisi pour sa capacité à fonctionner hors ligne essentielle dans les zones à faible connectivité. Le formulaire recueillait :

- L'identification de l'enfant (nom, âge, sexe)
- Les coordonnées des parents ou tuteurs
- Les informations médicales et d'urgence
- L'église ou le membre du staff ayant effectué l'inscription (enregistré automatiquement)

Le formulaire est accessible publiquement à l'adresse suivante :

<https://ee.kobotoolbox.org/x/8fort8cA>

Les inscriptions sont restées ouvertes 2 à 3 semaines, permettant à toutes les églises de soumettre leurs données de façon fiable, même sans connexion internet.

3.2. Nettoyage et standardisation des données

Le jeu de données brut (exporté au format CSV depuis KoboToolbox) présentait plusieurs incohérences :

- Inscriptions en double
- Orthographe et casse des noms irréguliers
- Numéros de téléphone non structurés
- Champs d'âge ou de sexe manquants ou mal formatés

J'ai développé un **script de nettoyage en Python** (disponible sur GitHub) utilisant **pandas** et des expressions régulières pour :

- Supprimer les doublons
- Normaliser les noms et les formats de contact
- Valider et uniformiser les âges et sexes

Le résultat est un jeu de données propre, prêt pour l'analyse.

3.3. Conception du tableau de bord

Les données nettoyées ont été transférées vers **Google Sheets**, puis connectées à **Looker Studio** (anciennement Google Data Studio) pour créer un tableau de bord interactif affichant :

- Le nombre total d'inscrits et les tendances quotidiennes
- La répartition démographique (tranches d'âge, répartition par sexe)
- Les sources d'inscription (par église)
- Les alertes médicales et les contacts d'urgence

Le tableau de bord en ligne est consultable ici :

lookerstudio.google.com/.../LC8IF

Cet outil a permis à l'équipe d'anticiper les besoins, allouer les ressources et vérifier la présence réelle des enfants par rapport aux inscriptions.

4. LIVRABLES

Les produits finaux comprennent :

- Le formulaire KoboToolbox (compatible hors ligne)
- Le script automatisé de nettoyage en Python
- Le jeu de données consolidé (CSV)
- Le tableau de bord interactif Looker Studio
- Une documentation technique et un guide utilisateur

L'ensemble a été remis à l'équipe organisatrice pour une utilisation autonome.

5. DÉFIS RENCONTRÉS ET SOLUTIONS APPORTÉES

- **Doublons** : résolus par une correspondance approximative (fuzzy matching) sur nom/âge/église.

- **Noms irréguliers** : corrigés via une normalisation des chaînes (minuscules, suppression des accents et espaces superflus).
- **Numéros de téléphone** : standardisés avec des expressions régulières au format ‘+225 XX XX XX XX’.
- **Erreurs de saisie** : traitées par des règles de validation et une vérification manuelle pour les cas limites.

6. IMPACT

Le système a généré des bénéfices opérationnels immédiats :

- **Aucun écart** entre inscriptions et présences lors du camp.
- Une traçabilité financière claire (chaque enfant rattaché à un payeur).
- Une meilleure préparation médicale et logistique.
- Un cadre réutilisable et évolutif pour les éditions futures.

7. CONCLUSION

Ce projet démontre comment des outils accessibles de data science KoboToolbox, Python et Looker Studio peuvent résoudre des problèmes concrets au niveau communautaire, avec efficacité et rigueur. Il illustre ma capacité à :

- Transformer des problèmes opérationnels en solutions fondées sur les données
- Construire des pipelines complets (collecte → nettoyage → visualisation)
- Communiquer des insights exploitables à des parties prenantes non techniques

Le code source, le tableau de bord et la documentation sont disponibles sur mon dépôt GitHub.

Rapport généré le 1^{er} décembre 2025 — à usage académique et professionnel.