Colony Registration Analytics & Data Management
A Data-Driven Solution for Church Holiday Camps in Bingerville

Fobah N'gouan Salomon
Student in Data Science – African School of Economics
github.com/fobahsalomon | fobahngouansalomon@gmail.com | My portfolio

1. CONTEXT

The annual holiday camp organized by the network of churches in Bingerville (Côte d'Ivoire) faced recurring logistical and financial challenges: the number of children present on-site frequently exceeded the number of paid registrations. This discrepancy led to budget imbalances, staffing uncertainties, and difficulties in accountability.

As a data science student and member of the organizing team, I proposed a structured data management system to modernize registration, ensure traceability, and support real-time decision-making.

2. OBJECTIVES

The project pursued four key goals:

- Centralize all registrations into a single digital system.

- Ensure complete traceability (child → registering church/guardian).

- Clean and standardize data for analytical reliability.

- Deliver an interactive dashboard for operational monitoring.

3. METHODOLOGY

The workflow was divided into three phases:

3.1. Digital Data Collection

I designed a mobile-friendly registration form using **KoboToolbox**, selected for its **critical offline capability** in areas of low-connectivity. The form captured:

- Child identification (name, age, gender)

- Guardian contact details

- Health and emergency information

- Submitting church or staff member (auto-logged)

The form is publicly accessible at:
https://ee.kobotoolbox.org/x/8fort8cA
Registrations remained open for 2 to 3 weeks, allowing all churches to submit entries reliably even without the Internet.

3.2. Data Cleaning & Standardization

The raw data set (exported as a CSV from KoboToolbox) contained inconsistencies:

- Duplicate registrations

- Inconsistent spellings and capitalization of names

- Unstructured phone numbers

- Missing or misformed age/gender fields

I developed a **Python cleaning script** (available on GitHub) using `pandas` and regex to:

- Deduplicate records

- Normalize names and contact formats

- Validate and standardize age/gender entries

The output was a clean, analysis-ready data set.

3.3. Dashboard Development

The cleaned data was uploaded to **Google Sheets** and connected to **Looker Studio** (formerly Google Data Studio) to build an interactive dashboard displaying:

- Total registrations and daily trends

- Demographic data (age groups, gender distribution)

- Registration sources (by church)

- Health flags and emergency contacts

The live dashboard is available here:
`lookerstudio.google.com/.../LC8IF`
This allowed camp leadership to anticipate needs, allocate resources, and verify attendance on-site against registrations.

4. DELIVERABLES

The final outputs included:

- KoboToolbox registration form (offline-compatible)

- Automated Python data cleaning script

- Clean master dataset (CSV)

- Interactive Looker Studio dashboard

- Technical documentation and user guide

All tools were handed over to the organizing team for autonomous use.

5. CHALLENGES & SOLUTIONS

- **Duplicates**: Solved via fuzzy matching on name/age/church.

- **Inconsistent names**: Addressed with string normalization (lowercase, strip accents, remove extra spaces).

- **Phone numbers**: Standardized using regex to enforce '+225 XX XX XX XX' format.

- **Field errors**: Handled with validation rules and manual review for edge cases.

6. IMPACT

The system delivered immediate operational benefits:

- **Zero registration discrepancies** during camp execution.

- Clear financial accountability (each child linked to a payer).

- Improved medical and logistical preparedness.

- A reusable, scalable framework for future events.

7. CONCLUSION

This project demonstrates how accessible data science tools KoboToolbox, Python, and Looker Studio can solve real-world community challenges with elegance and efficiency. It reflects my ability to:

- Translate operational pain points into data solutions

- Build end-to-end pipelines (collection → cleaning → visualization)

- Deliver actionable insights to non-technical stakeholders

The full code, dashboard, and documentation are available on my GitHub repository.

Report generated on December 1, 2025 — for academic and professional portfolio purposes.