Mini-Grid Innovation Lab:
Overcoming the challenges of big data collection

March 2020

CROSSBOUNDARY

ENERGY4IMPACT
ACCELERATING ACCESS TO ENERGY

Agenda

Executive summary

1. Introduction
   - Introduction to the Mini-Grid Innovation Lab
   - Introduction to Energy 4 Impact and CrossBoundary
2. Objectives of this document
3. Data management
   - Types of data
   - Data collection
   - Data cleaning
   - Data analysis
4. Outcomes
   - Impact on mini-grid developers
   - Socio-economic impact on off-grid communities
5. Observations
   - Challenges of data management
   - In-person vs phone surveys
6. Recommendations
   - Lessons learned on data management
   - Machine learning
7. Next steps and call to action

Executive summary

- Energy4Impact and CrossBoundary collected, cleaned and analysed **550M data points** from January 2017 to December 2019 for the Mini-Grid Innovation Lab using smart meters, field surveys, and financial records coming from **62 sites** of **12 mini-grid developers** (7 in East Africa, 4 in Nigeria and 1 in Zambia).

- We **standardised and outsourced** the data management process to improve quality and impact.

- Some developers had challenges around data management and needed **training and capacity building**, as well as **financial support** to purchase enabling technology (e.g. smart meters).

- We carefully designed **customer surveys** to ensure relevant and unbiased outputs.

- Data outputs suggest the **two prototypes** tested by the Lab (tariff subsidies and appliance financing) have had a **positive impact on developers (increased revenues) and the wider off-grid communities (increased consumption and ability to afford appliances).**

1. Introduction to the Mini-Grid Innovation Lab

The Mini-Grid Innovation Lab is Africa's first R&D Fund exclusively focused on testing new business model innovations for mini-grids

The ROCKEFELLER FOUNDATION

↓

ENERGY4IMPACT ACCELERATING ACCESS TO ENERGY | **Innovation Lab** | CROSSBOUNDARY

↓

The Lab funds developers to test business solutions for mini-grids based in Africa so that they can provide more power, to more people, at less cost

Increase electricity consumption per user	**Reduce costs**	**Increase number of connections**

1. Introduction to Energy 4 Impact and CrossBoundary

Non-profit organisation that has advised and provided on-the-ground support to over 100 mini-grid developers in Africa

Responsibilities in the Lab:

- Leading on data collection, including surveying and cleaning
- Supporting on prototype study design
- Analysing study results
- Disbursing $600K funds to mini-grid developers

Investment and advisory firm with a focus on frontier markets and energy access.

Responsibilities in the Lab:

- Managing the Lab
- Leading on prototype study design
- Analysing study results
- Leading on communication of study results

2. Objectives of this document

Inform public and private players in the energy access sector, including donors, investors, mini-grid developers and NGOs, about the lessons learnt from managing the largest mini-grid dataset in sub-Saharan Africa to date.

Increase understanding around collecting, cleaning and analysing mini-grid datasets from various sources (e.g. household surveys, remote monitoring devices) and how to work with different data providers to ensure consistent data quality.

Gather data to show how mini-grids can increase revenues and lower costs for developers, and create positive social and economic impact for local communities.

3. Types of data

	Remote monitoring data	Socio-economic data	Project economic data
What type of data?	• **Consumption data:** if possible on hourly basis • **Energy payment data:** separate entries for every energy payment made • **Additional data depending on prototype:** loan repayment, wi-fi usage	• Data on **demographic, household, and income**, gathered through in-person or phone surveys in the local language.	• Financial information on past and projected **CAPEX and OPEX** costs. • Information on projected **revenues and consumption**. • Information on **tariff structure**. • Information on **mini-grid sizing**.
Where do we collect it from?	Smart meter and payment systems	Household surveys	Developer's internal systems
Why do we need it?	To show how prototype can increase electricity demand and ability to pay	To determine broader impact of mini-grids and prototypes on rural customers.	To determine the impact of prototypes on the overall mini-grid business model.
When is it needed?	• **Historical data**: last 12 months (min. 6 months) before disbursement of funds • **Monthly data:** until 15th of the following month	• **Baseline survey:** before start of prototype • **Midline survey:** optional, after 6 or 12 months of start of prototype • **Endline survey:** at the end of prototype	• One-off before fund disbursement.
Who is it needed for?	All customers at participating treatment and control sites.	Sample of customers at participating treatment and control sites.	All treatment sites.

3. Data collection

Former manual process

The Lab received smart meter data and other data directly from developers via a shared folder

The Lab followed up with developers if data was incomplete

The Lab cleaned and sense-checked data, and flagged anomalies via a Python script

The Lab reached out to developers to resolve anomalies

The Lab implemented systematic change in their Python script

The Lab analysed data via Python, Excel and Tableau

Odyssey industry platform

New automated process

Odyssey receives smart meter data via API and survey data via uploads from Ipsos

Developers share any monthly uploads in Odyssey

Odyssey cleans and sense-checks data, and flags anomalies

The Lab reviews the data

The Lab reaches out to developers to resolve anomalies

Odyssey implements systematic change on their platform

The Lab analyses data via a dashboard and Tableau

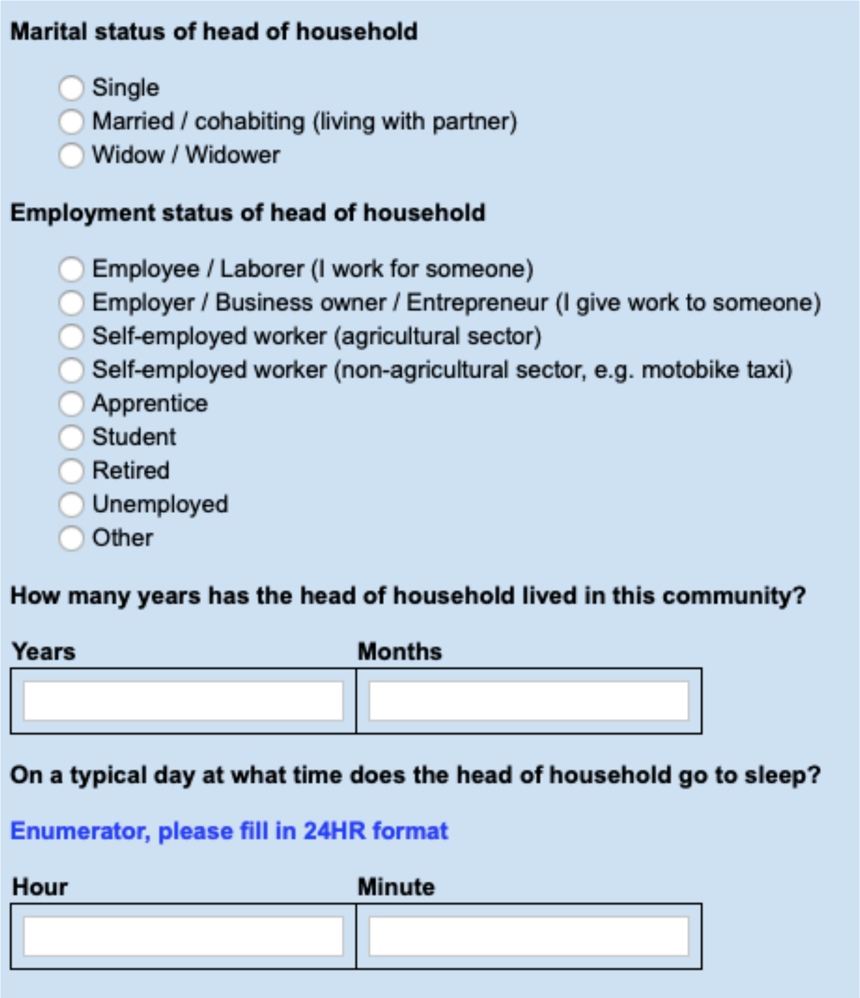
3. Data collection

- Before: Longer survey (30-40 min) with mini-grid customers.
- After: Shorter survey which takes 15-20 min.

- Capture data on change in socio-economic characteristics to explain why a customer may or may not increase consumption after implementation of a prototype.

- Baseline before start of each prototype.
- Midline 6 or 12 months after the implementation of a prototype (depending on the length of the prototype).
- Endline at the end of the prototype (1 year or 5 years after the start).

- Before: Survey was conducted by enumerators who visited the village and collected information from the households using the smartphone app KoboCollect.
- After: Survey conducted via phone by Ipsos enumerators.


Marital status of head of household

- Single
- Married / cohabiting (living with partner)
- Widow / Widower

Employment status of head of household

- Employee / Laborer (I work for someone)
- Employer / Business owner / Entrepreneur (I give work to someone)
- Self-employed worker (agricultural sector)
- Self-employed worker (non-agricultural sector, e.g. motobike taxi)
- Apprentice
- Student
- Retired
- Unemployed
- Other

How many years has the head of household lived in this community?

Years	Months

On a typical day at what time does the head of household go to sleep?

Enumerator, please fill in 24HR format

Hour	Minute

3. Data cleaning

8 Data Cleaning Steps

1. Align format & aggregate
2. Remove duplicates
3. Remove totalisers
4. Remove unclear outliers
5. Remove sales accounts
6. Remove reverse payments
7. Check for missing data and interpolate if sensible
8. Reconcile meters across data sets

Totalisers are meters that act as aggregators at every generation source and are used to detect electricity theft, line losses, or metering errors. Hence, including them would imply higher than actual consumption levels.

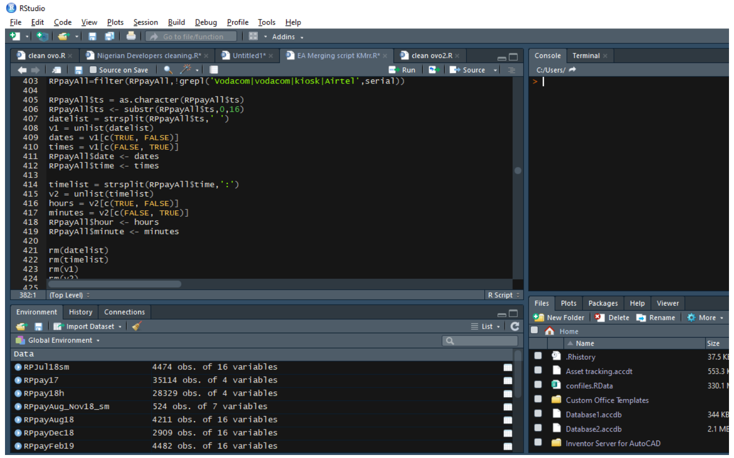
Outliers are values that are too small or large to be realistic (e.g. payments of USD 10,000).

Sales accounts are those in the payment file that do not correspond to a customer account, but to a sales agent or mobile money account.

Reverse payments are those where the customer sends their payment to the wrong account code, e.g. electricity, instead of appliances account. The developer then reverses the payment and re-routes it with the correct account code.

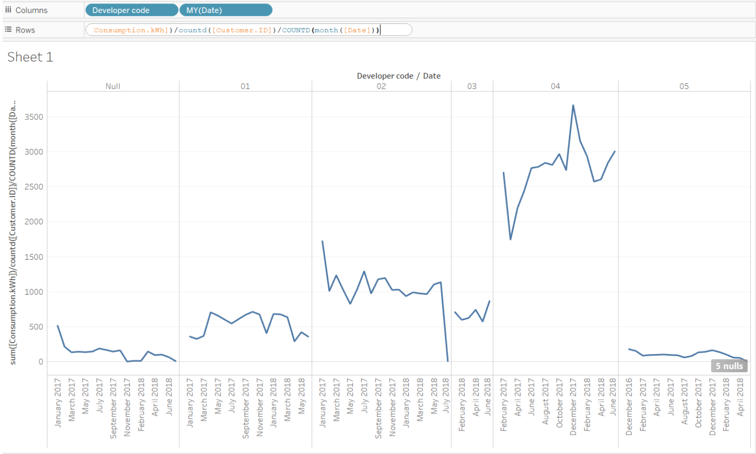
3. Data analysis

Code to clean data



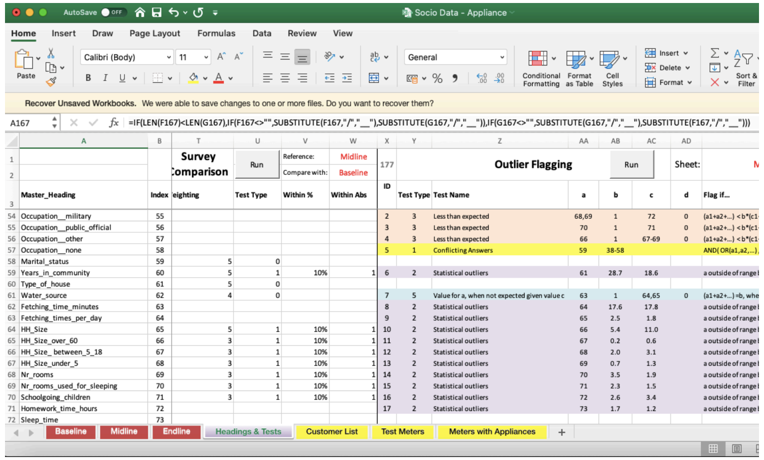
- Standardisation, anonymisation and aggregation of millions of remote monitoring data points.
- Cleaning of data (e.g. removing duplicates, removing sales accounts and totalisers, checking for missing data and outliers).
- Formerly, this step was done via Python and R script. Now it is done on the Odyssey platform.

Data visualization



- Visual analysis of how energy consumption and payments have changed and are correlated to socio-economic factors.
- Creation of intuitive visuals for communication purposes.
- This step is now being done via Odyssey feeding data into Tableau.

Algorithms for regression analysis



- Creation of algorithms for regression and machine learning analysis (e.g. Propensity Score Matching, Random Forrest, LASSO). More details can be found on slide 17.
- This step is still being done manually via Excel, Python or R.

CROSSBOUNDARY ENERGY4IMPACT
ACCELERATING ACCESS TO ENERGY

4. Impact on mini-grid developers

Prototypes

1. **Tariff reduction:** 2 mini-grid developers in Tanzania received a tariff subsidy on two sites, 50% on one and 75% on the other, in May and June of 2018.
2. **Appliance financing:** 7 mini-grid developers got grant funding to sell 663 appliances on credit across 27 sites in Kenya, Tanzania and Nigeria between February and August 2018.

Initial results

- Mini-grid developers may be able to charge lower tariffs and achieve the same or similar revenues because customers increased their consumption almost by the same proportion as the tariff subsidy.
- Mini-grid developers can raise revenues by implementing appliance financing schemes.
- Mini-grid customers can increase electricity consumption if they have access to new electrical appliances. Appliance purchasers consumed nearly twice as much electricity for the first five months following appliance delivery in the regions tested.
- Average revenue per user (ARPU) across all customers (not just those who purchased appliances) increased considerably. In East Africa, revenues were 18% above baseline levels after 11 months and in Nigeria they were 25% above after 5 months.

Challenges

- Procuring and distributing appliances, and tracking loan repayments is operationally complex and requires significant developer resources.
- The dataset for the tariff reduction prototype is relatively small (128 customers across two sites) and the prototype is only nine months into its five-year life. However, the Lab will soon add three more sites, bringing total connections under the prototype to 1,159.

Sources: Innovation Insight: The Price Elasticity of Power for the Rural Poor, April 2019. CrossBoundary
Innovation Insight: Appliance Financing, August 2019. CrossBoundary

4. Socio-economic impact on mini-grid communities

Prototypes

1. **Tariff reduction:** 128 customers in Tanzania received a tariff subsidy on two sites, 50% on one and 75% on another, in May and June of 2018.
2. **Appliance financing:** 3,471 customers were offered appliances and 663 appliances were sold on credit across 27 sites in Kenya, Tanzania and Nigeria between February and August 2018.

Initial results

- Reducing tariffs has an immediate and strong effect on rural customers' use of energy. Rural customers are price-constrained, not demand-constrained.
- Offering appliances on credit has an immediate and strong effect on rural customers' consumption. Rural customers are use- and credit-constrained.
- The most popular of the household and productive use appliances were those used for entertainment. Speakers and TVs made up 393 of the total 663 appliances sold.
- Rural customers principally purchased household appliances. There was little customer appetite for productive use appliances such as carpentry tools.

Challenges

- High electricity tariffs
- Affordability of appliances for rural customers
- Access to credit for rural customers

Sources: Innovation Insight: The Price Elasticity of Power for the Rural Poor, April 2019. CrossBoundary
Innovation Insight: Appliance Financing, August 2019. CrossBoundary

5. Challenges of data management

Challenges on remote monitoring of data

Manual process

- Managing large numbers of data points (550 million) from various sources was time intensive.

Poor data quality

- **Data Gaps:** Missing data for weeks or even months due to system outages, inability to transfer data from smart meters to other systems, connectivity issues.
- **Unique Identifier:** Changes in meter numbers due to replacement or customers adding a second meter number to their account (e.g. business vs household).
- **Manual Processes:** Incomplete data sets and inconsistent formats or units provided by developers as a result of human error.
- **Lack of Smart Meters:** Data collection less frequent and accurate if no smart meters.

Others

- **Repossession of appliances** and cutting off electricity due to non-payments is difficult to implement and has negative consequences on customer relationships.
- **External Factors:** Changes in tariff or promotions by developers that the Lab was not informed about.

Challenges on surveys

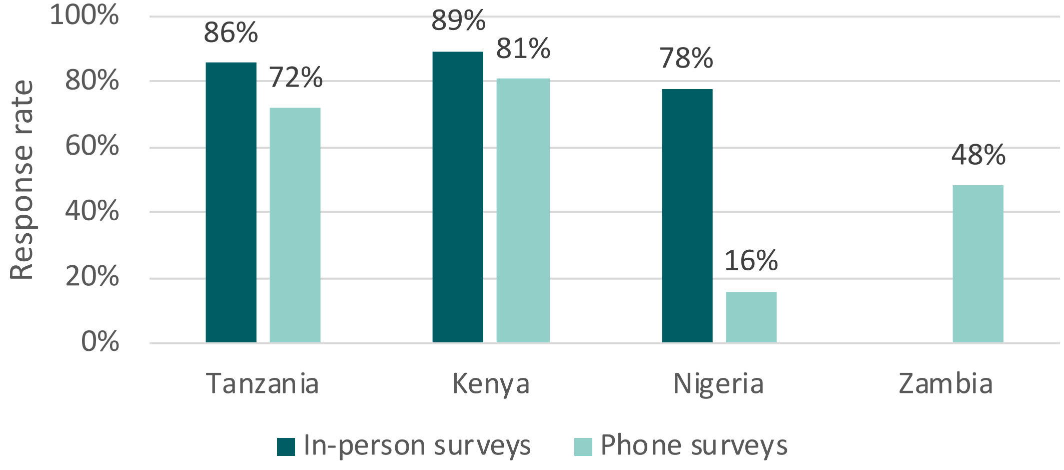
Ensure statistical relevance

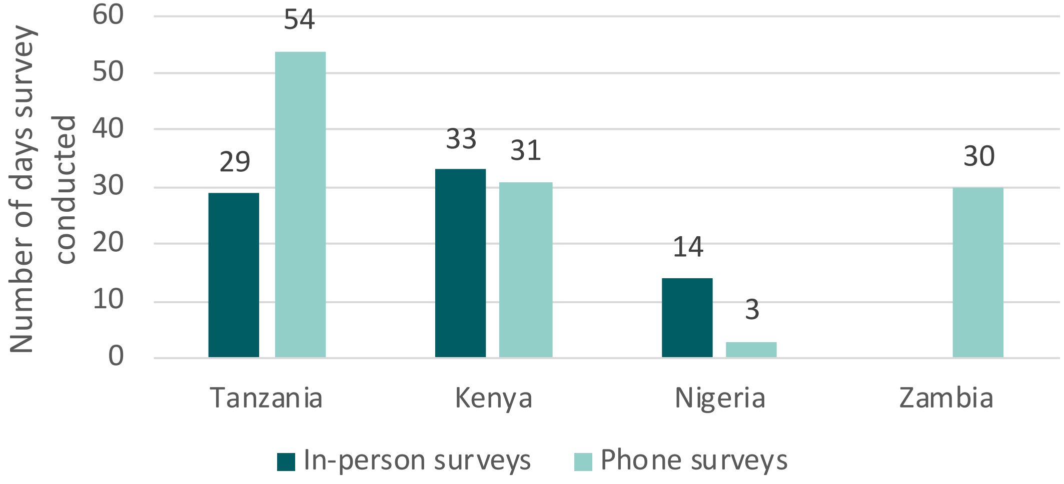
- **Biases:** Order of questions can influence the answers you get (interviewees tend to select the last answer option they hear and the first one they read). Questions about income or other sensitive metrics may be answered in a way to make interviewees look better off or poorer.
- **Survey Fatigue:** If surveys are too long, interviewees tend to answer questions with less rigour.
- **Misinterpretation by enumerators:** If enumerators do not understand a question well, they can misinterpret it.
- **Local language:** Translations into different languages can distort meaning.
- **Attrition between baseline and endline surveys:** People may move or just not respond.
- **In-person vs phone surveys:** In certain geographies phone surveys might not work well due to bad connectivity or low mobile phone penetration.

5. In-person vs phone surveys

Findings: In-person surveys versus phone surveys*

- **Response rate**
  - High response rate for in-person surveys (between 78% and 89%).
  - Lower response rate for phone surveys (between 16% and 81%, with Nigeria being the lowest). Zambia has achieved a response rate of 48% in 30 days at time of writing this report.
  - The baseline surveys in Nigeria were in-person. We then tried using phone surveys for the mid-line. However, due to the weak mobile network in most villages, we reverted to in-person surveys.
- **Response time**
  - To reach a comparable response rate, phone surveys in Tanzania took 54 days to complete compared to 29 days for in-person surveys.
  - The number of days needed for the in-person and phone surveys in Kenya was roughly the same.
  - No in-person surveys were conducted in Zambia, so we cannot make a comparison.
- **Surveying cost**
  - The costs of phone and in-person surveys were about the same, even though we expected the phone surveys to be cheaper. This was mainly due to using local enumerators for the in-person surveys and a single international company for the phone surveys for all countries.





* These results represent six months of phone surveys and one year of in-person surveys. They may be subject to change over time.

CROSSBOUNDARY ENERGY4IMPACT
ACCELERATING ACCESS TO ENERGY

6. Lessons learned on data management

Recommendations on remote monitoring of data

Process

- Receive data directly from smart meter providers and provide templates or APIs to a database to **streamline data sharing** (this enabled us to reduce the time between data collection and cleaning from 8 to 3 weeks).
- Create a **dashboard** for data analysis.
- Set up a rigorous and iterative **data cleaning process**.
- Use **interpolation methods** to account for missing values.
- **Clarify missing or unrealistic values** with developers.

Technology

- Assess in advance what the frequency and type of remote monitoring data is and whether it is necessary to finance **smart meters**.
- Assess convenient **payment systems for customers**.
- Use **machine learning techniques** for predictive modelling and to assess which socio-economic variables can best predict the effect of a prototype on a customer.

Feedback loop

- Appoint **data management staff**, both on programme and developer side.
- Provide **training** on data management for developers to create awareness on importance of collecting high-quality data and avoiding inconsistencies and human errors.
- **Provide regular feedback to developers** to ensure learning.
- Set aside **funds to support developers** with improving on their data management processes.
- Foster **partnerships** with third-party providers to manage loan repayments etc. and ease the burden on developers.

Recommendations on surveys

Design

- **Duration:** Make surveys as short as possible to avoid survey fatigue.
- **Survey method:** Consider whether phone surveys are appropriate for the region / sites based on the following criteria: reliability of mobile network, number and economic status of people owning phones, other aspects (e.g. are customers willing to pick up the phone from an unknown number? Are customers available at specific times of day?)
- **Randomise answer options**: Change order of answer options randomly between customers being surveyed.
- **Ask neutral questions:** For example, ask about expenditures before income to avoid under- or overstatement of income.

Implementation

- Conduct **training** of enumerators so they understand the questions and the rationale for asking these questions.
- Have surveys translated into **local languages** by a trusted and certified translator and train local enumerators.
- **Pilot** surveys among target population and **communicate** to local communities before conducting the surveys (e.g. via SMS).
- Account for **attrition rates** when calculating **sample size**.

6. Machine learning

Machine learning techniques

What are they used for?

1. **Predictive modelling**
   - What will be the effect of the prototype on different types of customers?
2. **Variable importance analysis**
   - **Variables**: socio-economic data about customers, e.g. income, primary mode of transport, education level, etc.
   - Which variables can best predict the effect of a prototype on a customer?
   - Knowing this can help reduce survey length

None of the socio-economic survey variables correlated well with **individual customer electricity consumption** (best median percentage error: 82.5%)*

However, some correlated closely with **the overall community consumption** (percentage error: 9-10%). Based on two different machine learning techniques, we found the following variables had a strong correlation.**

LASSO

- Income/Expenses (e.g. money spent on food and savings)
- Mobile airtime
- Time taken to fetch water

Random Forests

- Income/Expenses (e.g. money spent on food and schooling)
- Mobile airtime
- Lifestyle & other wealth indicators (e.g. sleeping time, transportation methods, number of rooms)
- Appliance ownership & use

* These findings were similar to Blodgett, Courtney, et al. (2017). Accuracy of energy-use surveys in predicting rural mini-grid user consumption. Energy for sustainable development 41: 88-105. https://www.sciencedirect.com/science/article/pii/S0973082617304350?via%3Dihub

** We used a similar methodology and got similar results to Williams, Nathan, Booth, Samuel S, and Baring-Gould, Edward I (2019). NREL. Survey Use in Micro-Grid Load Prediction, Project Development, and Operations: Review and Best Practices. https://www.nrel.gov/docs/fy19osti/72339.pdf

7. Next steps and call to action

- **Data standards.** Build consistent data standards and provide an impartial platform for mini-grid developers to manage their data on, such as the the efforts already underway by the industry-representing organisation AMDA (Africa Minigrid Developers Association).
- **Investment for enabling technology.** Donors should provide more specialised funding for data management systems and equipment (e.g. smart meter, PAYG systems) that allows for better data collection and corresponding training.
- **Specialisation on core processes and automation.** Mini-grid developers should outsource non-core data processes such as loan repayment management or customer surveying to external and specialized providers and should automate internal data processes as far as possible.

Contacts

Energy 4 Impact
Mbaruk Road, Opposite Panorama Court
Off Muchai Drive, Off Ngong Road
P.O. Box 76580 00508
Nairobi, Kenya

www.energy4impact.org
Author: Maria Knodt
maria.knodt@energy4impact.org

CrossBoundary LLC
ABC Place, Waiyaki Way
Nairobi, Kenya

www.crossboundary.com
contact@crossboundary.com