



Advancing Financial Inclusion

# Data-driven solution for rural agent network management

November 2017

## Requirements:

Call detail record (CDR) data  
Mobile money transactional data  
Agent location data

Established by



Case Study

## Authors

---

Authors: Krista Nordin, Michiel Wolvers,  
Kate Rinehart and Herman Smit

## About insight2impact

---

insight2impact | i2i is a resource centre that aims to catalyse the provision and use of data by private and public-sector actors to improve financial inclusion through evidence-based, data-driven policies and client-centric product design.

i2i is funded by the Bill & Melinda Gates Foundation in partnership with The MasterCard Foundation.

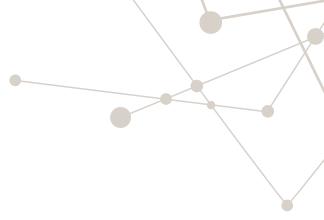
### For more information:

Visit our website at [www.i2ifacility.org](http://www.i2ifacility.org).  
Email Mari-Lise du Preez (Partnership Manager)  
at [mari-lise@i2ifacility.org](mailto:mari-lise@i2ifacility.org).  
Call us on +27 21 913 9510.



# Contents

1.	<b>Introduction</b>	3
2.	<b>Overview of the case study</b>	4
2.1	Business challenge	4
2.2	Data-driven solution	5
3.	<b>Case study approach and results</b>	6
3.1	Data sources	6
3.2	Analytical approach	8
3.3	Results	14
3.4	Applications for results	16
4.	<b>Concluding insights and discussion</b>	18



# Terminology

Throughout this case study, we will reference the following definitions and acronyms:

## Definitions

Gatherings:	An aggregation of active mobile network operator customers in a defined location
Urban/rural:	Urban and rural gathering classifications are based on the categorisation provided by the Urban Extent of Africa map <sup>1</sup>
Weekly:	Gatherings that occur once every seven days
Non-weekly:	Gatherings that do not occur once every seven days
Bi-weekly:	Gatherings that occur twice every week
Bi-monthly:	Gatherings that occur twice every month
Active mobile money customers:	Mobile money users that performed a transaction in the past six months
Far from home:	More than 2 km away from the user's home site
High mobile money penetration:	More than 50% of identified individuals are active mobile money users
Hotspots:	Identified gatherings with high mobile money penetration
Roving agents:	Agents who are not fixed in location, but rather move around to serve multiple locations
Moving average peak detection method:	A method to analyse maximums in a dataset that relies on calculating a moving average indicator which serves as the minimum threshold for detecting the maximums
Discrete Fourier transform:	Analytical method to convert time series data into the frequency domain

## Acronyms

GSM:	Global system for mobile communications
GPS:	Global positioning system
CDR:	Call detail record
MNO:	Mobile network operator
DFT:	Discrete Fourier transform
FSP:	Financial service provider
GIS:	Geographic information system
MMO:	Mobile money operator

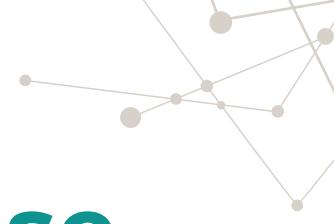
<sup>1</sup> <http://www.ifpri.org/publication/urban-extent-africa-2010>



# 1 Introduction

This document forms part of a series of case studies that explore how new data-driven solutions can overcome financial services delivery challenges. Each case study presents a business challenge faced by a bank, mobile network operator, insurer or fintech in delivering financial services to the financially excluded or underserved. We then present a real-life case study to demonstrate how a specific data source and data application allowed our partners to overcome the business challenge. Our aim with these case studies is to allow for a more granular discussion on how data may contribute to improved financial inclusion.

“ Our aim with these case studies is to allow for a more granular discussion on how data may contribute to improved financial inclusion. ”



# 2 Overview of the case study

Mobile network operators (MNOs) and mobile money operators (MMOs) face several constraints in serving rural customers profitably. One of these constraints relates to establishing and maintaining a sustainable rural agent network for their mobile money business. To do this, MMOs need to select and/or place agents in areas with sufficient economic activity or where mobile money users are likely to require the services of an agent.

## 2.1 Business challenge

Economic activity in rural areas is typically more dispersed than in urban areas where infrastructure such as bus stations, permanent markets or registered commercial entities provide proxies for likely demand for mobile money services. Rural areas lack both formal records of economic activity and the proxy infrastructure that MMOs may rely on to substitute for this data. Agent selection in rural areas typically relies on identifying growth points, e.g. small permanent shopping areas in those rural areas. These growth points are identified using geospatial data, analysing voting patterns or through informant interviews. These data sources are most effective at identifying permanent or fixed points of economic activity. The limitation of this approach relates to the cost of collecting the data, as well as the fact that the data may date quickly and that the data sources do not accurately identify gatherings with high economic activity.

“ MMOs need to select and/or place agents in areas with sufficient economic activity or where mobile money users are likely to require the services of an agent. ”



## 2.2 Data-driven solution

MMOs can use a combination of additional data sources to identify rural gatherings with high economic activity and to place agents in those locations. For the purposes of this case study, we consider call detail record (CDR) data, mobile money transaction data, agent location data and financial access points data. Each of these data sources and their application are described in more detail in the next section. In brief:

- **CDR data** provides us with insights into where rural customers gather throughout the day.
- **Mobile money data** allows us to identify economic activity and mobile money penetration at the identified gatherings (e.g. hotspots).
- **Agent location data** helps us to identify which hotspots are unserved.
- **Financial access point data** allows us to know how far these hotspots are from existing financial access points.

The MMO can use these data sources to inform its rural agent distribution strategy, placing agents at unserved hotspots to ensure agent viability.

### Box 1: About our partners

To produce this case study, we partnered with Airtel Uganda and Masaë Analytics.

**Airtel Uganda** (Airtel UG) is the country's second-largest MNO, with a subscriber base of 7.5 million. Airtel UG is a subsidiary of the Airtel Group, the third-largest mobile service provider in the world, based on number of subscribers (GSMA, 2016). As at December 2016, Airtel Group had over 336 million subscribers across 20 countries.

**Masaë Analytics** is a French start-up at the frontline of innovation in data science, specialised in telco data-mining, predictive models, satellite image processing and enriched social media analytics. Masaë is a sister company of Altai Consulting, a leading consulting and research firm for both public and private sectors in emerging markets.



# 3 Case study approach and results

This section provides details on the data sources and analytical approach used to address our partner's specific business challenge. We aim to provide you, the reader, with sufficient details to assess whether this approach is desirable or feasible within your own organisation or your partner organisation.

## 3.1 Data sources

This sub-section covers the categories of data required and the specific attributes required. It also explains why these data sources are needed to perform the analysis.

The data used to identify unserved hotspots includes CDR data, mobile money transactional data, agent location data, satellite imagery data and existing financial access point data. Descriptions of each data source, why it is needed and the specific attributes that are needed to conduct the analysis are described in detail below.

### CDR data

Call detail records (CDRs) are generated as and when customers interact with their mobile phones. CDRs capture a broad spectrum of relevant information and provide insight into an individual's identity, location, behaviour and movement patterns. As mobile users' locations are recorded throughout the day, it is possible to detect the location and size of customer gatherings.

**“**We aim to provide you with sufficient details to assess whether this approach is desirable or feasible within your own organisation or your partner organisation. **”**



Each CDR contains over 200 attributes; however, only seven are required for this analysis. Descriptions and the purpose for selecting each attribute are indicated below.

**Table 1: Description and purpose of attributes of CDR data used in analysis**

Attribute	Description	Purpose
Record Type	Transaction type; identifying whether it was a voice, SMS or data transaction	To distinguish between types of communication transactions
MSISDN	A number uniquely identifying a mobile phone subscription	To be able to associate each transaction with a unique customer
Cell Tower ID	Unique identifier for the cell tower to which the user is connected while using his/her phone	To identify the location of a mobile phone user
Call Duration	Duration of the connection	
Bytes Up	Number of uploaded bytes (data only)	To identify the volume of the transactions the mobile phone user is engaging in
Bytes Down	Number of downloaded bytes (data only)	
Timestamp	Date and time at which the connection started	To identify the time of the transaction

Source: Authors' own

### Mobile money transactional data

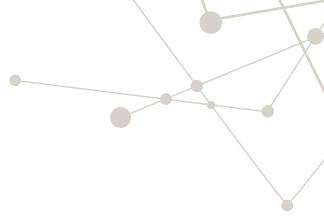
Mobile money transactional data is generated through every monetary transaction between two mobile money users. Mobile money data captures a broad spectrum of relevant information. Pertinent information used in this analysis includes the time and frequency of transactions per user and the originating and receiving phone numbers associated with the transaction. The mobile money transactional data is linked to CDR data to reveal the location of the originating customer and the receiving customer. This information can be used to derive insights into mobile money penetration levels and the level of heightened economic activity at the identified gatherings.

### Agent location data

Agent location data captures information on where current mobile money agents are located. This data is cross-referenced with identified gatherings to determine which gatherings are currently unserved by mobile money agents.

### Satellite imagery data

Satellite imagery data captures the terrain of a country. Google Maps is an open data source that captures this information. Satellite data helps to place identified gatherings within their geographic context relative to the natural landscape (e.g. rivers and mountains) and physical infrastructure (e.g. roads and buildings). It allows for the categorisation of a gathering as a soccer match, a market, religious gathering, and so forth. It is important to note that this data source is not critical to performing the analysis described in the subsequent section.



## Financial access point data

Financial access point data captures the geographic location of financial access points in a country. This data is usually open source and captured by governments, development partners and/or the donor community. In Uganda, this dataset is called OpenStreetMap<sup>2</sup>. It lists the location of 979 bank branches and ATMs in Uganda. Financial access point data provides information on how well identified gatherings are served by already established financial access points. It is important to note that the subsequent analysis can be performed in the absence of this data.

**Combined datasets capture information required to identify where unserved hotspots are.**  
Combining the information from the data sources mentioned above allows one to identify hotspots that are currently unserved or underserved by mobile money agents. Analysing CDR, mobile money transactional and agent data (all internally available to MNOs) provides a basis for understanding where and when gatherings occur, the extent of economic activity at these gatherings and whether mobile money agents are already present. The analysis can be further enriched using external, open-source satellite imagery and financial access point data. Satellite imagery can help qualify the nature of the gathering and financial access point data to understand the proximity of the gathering to an already established financial access point. These combined data sources are the foundation for generating insights into which rural locations represent viable business cases for the placement of agents.

## 3.2 Analytical approach

This sub-section of the report focuses on the analysis required to turn the above-mentioned data sets into insights on where unserved high economic activity customer gatherings are.

The analysis required to convert the data into insights can be grouped into eight steps, two of which are optional (steps 6 and 7). The steps are: (1) data cleaning, (2) developing an indicator for detecting gatherings, (3) populating the indicator for each cell tower, (4) visualising the populated indicator, (5) defining the quantifying factors that dictate which identified locations are gatherings or not, (6) converting the time series data into the frequency domain to check for any further gatherings (optional), (7) qualifying the gatherings against satellite imagery data (optional) and (8) overlaying mobile money transactional and agent location data to the identified gatherings. Once all steps are complete, the location of unserved hotspots can be identified.

---

<sup>2</sup> OpenStreetMap is a free, editable map of the whole world that is being built by volunteers largely from scratch and released with an open-content licence. <https://www.openstreetmap.org>



## How to perform each step:

### Step 1: Data aggregation and cleaning in preparation for analysis.

CDR data is sometimes stored in separate locations dictated by record type (voice, SMS and data). If this is the case for you, it is necessary to first merge all CDR data files. Once all the CDR data is in a single aggregate view, the next step is to prune the data file to only contain attributes needed for this analysis. The seven attributes needed for analysis are detailed in the CDR data section above (see Table 1). At the end of this step, you should have one clean CDR dataset that contains seven attributes per record.

### Step 2: Identify and test proxy indicators for variable of interest – gatherings.

A variety of different indicators were tested<sup>3</sup>. The most predictive indicator of gatherings is the number and distance of customers, within a specific geographical area, from their respective homes.

This indicator can be represented as:

$$\sum_{\substack{\text{customers} \\ \text{connected to} \\ \text{cell tower}}} (| \text{home location} - \text{current location} |)$$

To develop this indicator, two pieces of information need to be known per unique customer: the customer's home location and the customer's location throughout the day. The customer's home location can be calculated by sorting all their CDR records by the *Cell Tower ID* attribute and selecting the cell tower with the highest number of records. The customer's location throughout the day is

already captured in each unique CDR record. The *Cell Tower ID* attribute on the record tells you the customer's location, and the *Timestamp* attribute tells you the time of day.

### Step 3: Create a table that captures the indicator detailed above per cell tower per hour.

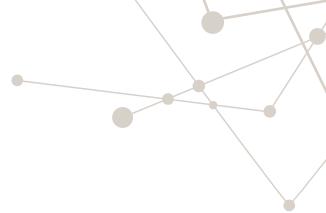
To identify the locations of the gatherings, the indicator developed in Step 2 needs to be calculated per *Cell Tower ID* per hour. At the end of this step, you should have a table structured like Table 2 below.

**Table 2: Example table structure for developing indicator for detecting gatherings**

Cell Tower ID	Day and hour	No. of customers far from home
EA55755	2016070100	42
EA55755	2016070200	30
EA55755	2016070300	14
EA55755	2016070400	9
...	...	...

Source: Authors' own

<sup>3</sup> Other indicators tested included number of connected users, number of calls, number of SMSs, total duration of calls, total data bytes uploaded and downloaded. The most predictive indicator was chosen by testing the indicators against a known sample of gatherings.

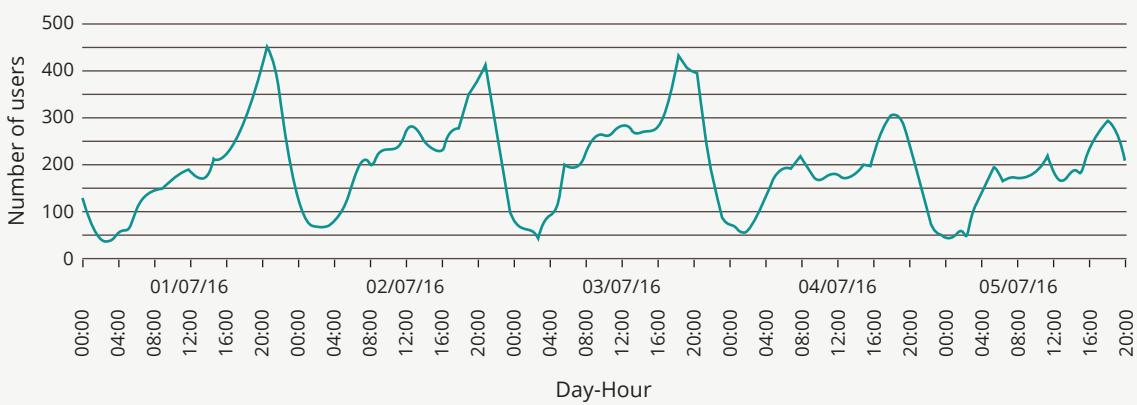


#### **Step 4: Plot the data captured above.**

To further understand the data, it is helpful to generate a line graph with the x-axis capturing the day and hour and the y-axis capturing the number of users far from home who are connected to the cell tower in question (see the example in Box 2 below). Through examining the data by eye, one can start to get a feel for where the peaks are. An example graph of the calculated indicator for one cell tower over a period of four days is in Box 3: Peak detection method visualised, on the next page.

“ Through examining the data by eye, one can start to get a feel for where the peaks are. ”

## **Box 2: Example calculated indicator for one cell tower over four days**



Source: Altai Consulting



### Step 5: Define the minimum threshold that must be met to identify peaks as gatherings.

To assess whether peaks qualify as gatherings, we compared the peak against a seven-day moving average and a seven-day moving standard deviation. These two values combined (the moving average and moving standard deviation) define the minimum threshold that must be met for the peak to be qualified as a gathering. The following equations are used to calculate the seven-day moving average and seven-day moving standard deviation at any given point in time for one cell tower:

Moving average:

$$\frac{\sum_{n=6}^0 \text{customers}_{t_n}}{7}$$

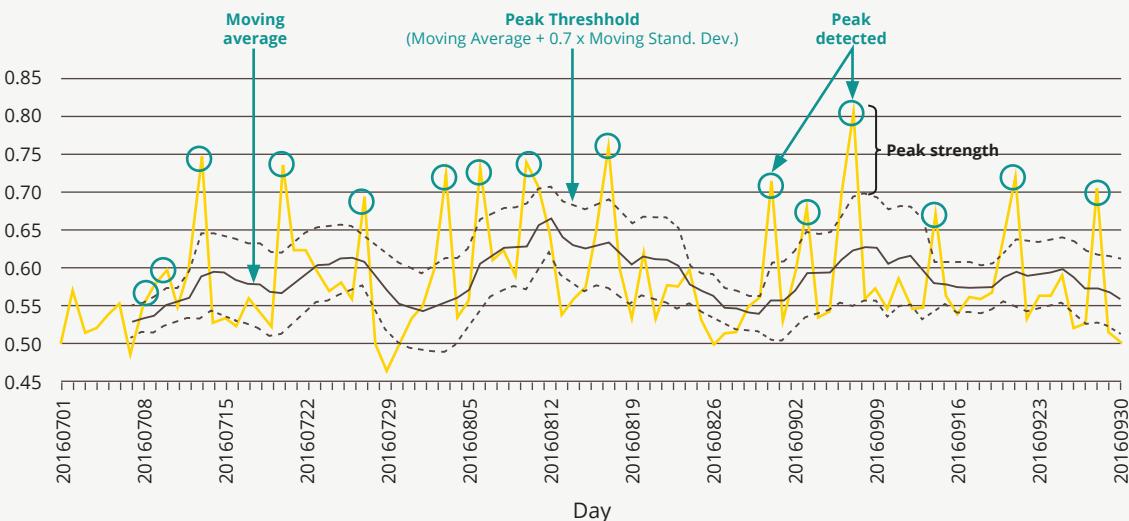
Moving standard deviation:

$$\sigma_{t_0} = \sqrt{\frac{\sum_{n=6}^0 (\text{customers}_{t_n} - \text{movingaverage}_{t_n})^2}{6}}$$

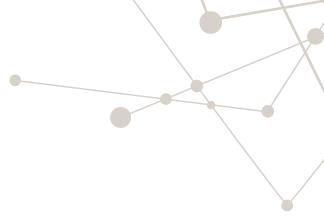
Once the values have been calculated per cell tower per hour, peaks that exceed the defined thresholds can be qualified as gatherings. Box 3, below, depicts this analysis.

### Box 3: Peak detection method visualised

Normalised sum of distances to home site, for a specific site



Source: Masae Consulting



#### **Step 6 (Optional): Check for further gatherings at previously identified gathering locations.**

The moving average peak detection method covered in Step 5 (and Step 6 if completed) is good for detecting weekly gatherings. To check whether there are further gatherings at the identified locations that occur non-weekly (bimonthly, monthly, biweekly, etc.), one can shift the data from the time series domain into the frequency domain. Shifting into the frequency domain simply means changing the table generated in Step 3 and the graph generated in Step 4 to compare signal intensity to various gathering frequencies instead of comparing time to number of customers<sup>4</sup>. One can then look at the new graph and any peaks would signal non-weekly gatherings that occur at the identified location, and the x axis value at these peaks can tell the frequency of that gathering. Box 4 on the next page depicts a juxtaposition of the same data represented in the time domain and frequency domain. Peaks in the frequency domain that correlate with an explainable frequency have been included.

#### **Step 7 (Optional): Gather contextual information on gatherings from satellite imagery.**

It is possible to review the geocoordinates of the identified gatherings on satellite images to further qualify the type of gatherings. If you do not have access to satellite imagery, skip this step and continue to Step 9.

#### **Step 8: Overlay mobile money transactional data and agent location and financial access point data on detected gatherings.**

Once the gathering locations have been identified, the final step is to overlay the mobile money transactional data to understand the level of economic activity and the agent location and financial access point data to distinguish which gatherings are currently served, underserved or not served.

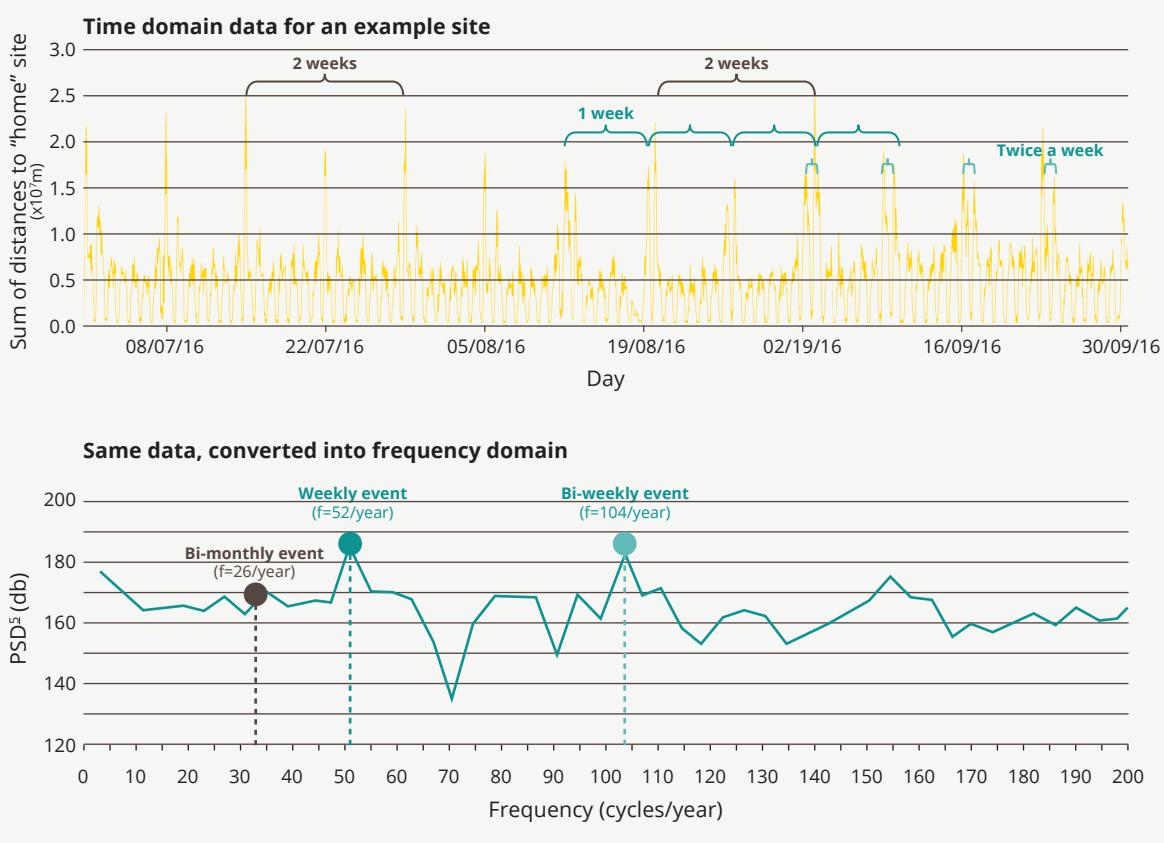
“ Once the gathering locations have been identified, the final step is to overlay the mobile money transactional data to understand the level of economic activity and the agent location data to distinguish which gatherings are currently served, underserved or not served. ”

---

<sup>4</sup> This shift from the time domain to the frequency domain can be done using a discrete Fourier transform.



#### Box 4: Frequency domain peak detection example



Source: Masaë Consulting

5 PSD is the intensity of the signal.

## 3.3 Results

This section covers the insights generated from performing the analysis on Airtel Uganda data. The objectives are to provide readers with a concrete example of the results that can be derived from the analysis above.

**354 unique sites represent 402 customer gatherings identified in Uganda.** The above analysis performed on Airtel Uganda data identified 354 unique sites with 402 dynamic gatherings.

The number of gatherings is higher than the number of sites, as some sites have several gatherings on different days of the week. Of the 402 total gatherings, 195 (49%) were in rural areas and 207 (51%) were in urban areas. The rural gatherings on average attracted 4,700 customers, who travelled on average 36 km one-way from their home site. The urban gatherings on average attracted 15,000 people, who travelled on average 13 km one-way from their home site. The maps in Box 5 on the next page depict the gathering locations, average gathering size and average distance travelled to gatherings.

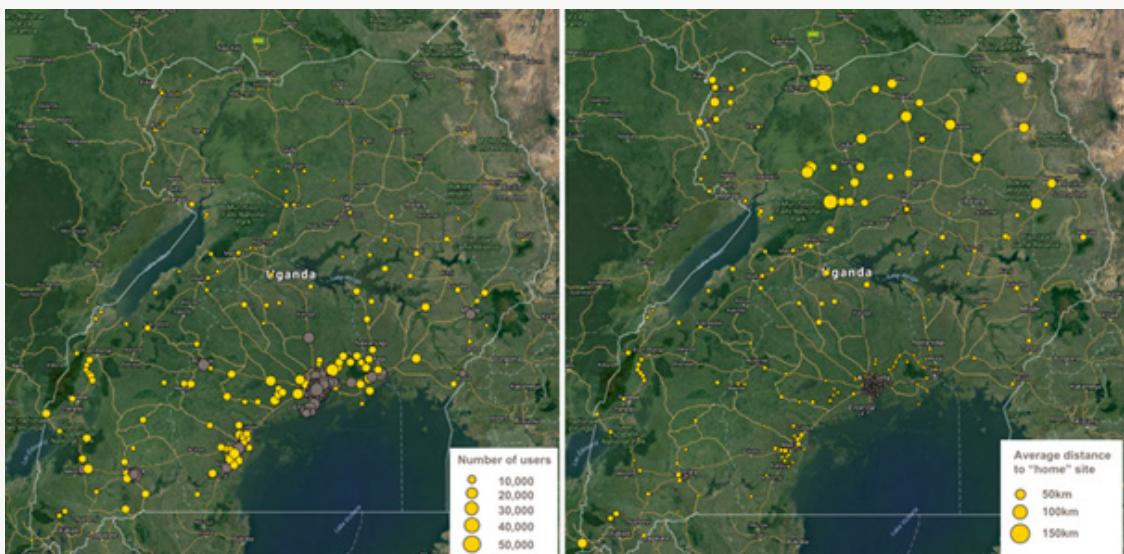
**85% of selected gatherings could be verified through physical field checks.** A sample of gatherings was selected and physically verified. In total, 20 field checks were completed, of which 17 were accurate matches. The remaining three could be false positives or seasonal gatherings that occurred during the months when the CDRs were analysed but not during the time of the field visits.

**68% of the identified gatherings had high mobile money penetration.** From overlaying the Airtel Money data onto the identified gatherings, 272 (68%) of the 402 gatherings were identified to have high mobile money penetration (e.g. hotspots). High mobile money penetration was defined as more than 50% of individuals at the gathering being active mobile money users. 203 of the hotspots were in urban areas, while 69 were in rural areas.

**8% of hotspots don't have detected agents.** From overlaying the agent location data, 23 of the 272 hotspots identified (8%) do not have a mobile money agent. 22 of the unserved hotspots were in urban areas, and one was in a rural area.



**Box 5: Maps of identified markets, depicting gathering size and distance travelled**



Source: Masaë Consulting

“ From overlaying the Airtel Money data onto the identified gatherings, 272 (68%) of the 402 gatherings were identified to have high mobile money penetration (e.g. hotspots). ”





## 3.4 Applications for results

The following section covers how insights generated from the analysis may be applied to business decision-making within MMOs.

The insights generated from this analysis can be applied to inform an MMO rural agent distribution strategy and marketing endeavours. Examples of how and where the insights can be applied for each of the two categories are detailed below.

### Mobile money agent network

The information derived from this analysis could inform the following business decisions related to mobile money agent networks:

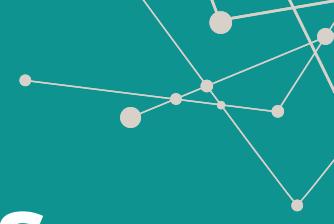
- **Place agents at hotspots without pre-existing agents.** In gatherings with high mobile money penetration but no agent, new agents can be selected or placed.
- **Reduce the number of agents at gatherings with low mobile penetration.** If there is low penetration of mobile money at detected gatherings, that means the volume of cash-in and cash-out transactions is low. The number of agents can be reduced permanently, or agents could be relocated to other hotspots on either a permanent or temporary (roving) basis.

- **Inform the business case for roving mobile money agents.** As the identified gatherings do not occur daily, they do not represent clear economic potential for permanent mobile money agents. The information derived from this analysis can inform the business case for introducing, increasing or decreasing roving agents that serve hotspots.
- **Inform float management systems for agents located in hotspots.** Information on likely consumer transactions allows the MMO to assist mobile money agents to manage increases in mobile money transaction demands.

### Marketing

The analysis provides information on gatherings with low mobile money penetration. The size of the gathering can inform the justifiable resources that the MNO can allocate to customer acquisition.

**“** The analysis provides information on gatherings with low mobile money penetration. The size of the gathering can inform the justifiable resources that the MNO can allocate to customer acquisition. **”**



# 4 Concluding insights and discussion

This case study aims to demonstrate how existing internal MNO data can be leveraged to aid decision makers in improving the viability of mobile money agent networks in rural areas. The analysis allows MMOs to select or place agents with more certainty into the viability of the placement. While the data analysis model presented here already provides increased accuracy for MMO agent placement, there are further improvements that can be made to the model, including:

- I. **Learn from agent placing.** To operationalise the results of this data analysis, the next step for the MMO is to place agents at the identified unserved or underserved hotspots. Implementation learnings can then be used to further improve the model.
- II. **Include additional data sources.** Four data sources have been used for this model. CDR data provides the basis for the analysis, and the other three are used to increase the model's accuracy. Including additional data sources will further increase the accuracy of the model. For example, geolocation mapped population data could be used to estimate the total potential market around the identified hotspots.
- III. **Use more historical CDR data.** Three months' historical CDR data was used for this analysis, which proved useful to identify hotspots. However, given the influence of seasonality on gatherings in rural areas, the model could be further improved through incorporating one annual cycle of historical CDR data.
- IV. **Use real-time data.** While historical data provides insights for future agent placement, hotspot locations are likely to change over time. Having access to the real-time data will further improve the model accuracy by identifying new hotspots and changes in current hotspots in real time.



### Box 7: Key takeaways from the case study

#### For executives at MMOs:

- Existing internal data can be leveraged to improve rural agent network sustainability.
- Required analysis can be performed in-house or completed by data solution providers in approximately four weeks.

#### For analysts at MMOs:

- CDR data can be used to identify where rural customers gather throughout the day:
  - The most predictive indicator for gatherings is the number and distance of customers at a location, from their respective homes.
  - Plotting the indicator against a moving threshold is appropriate for identifying gathering locations.
  - Further gatherings at identified locations can be determined through analysing the indicator in the frequency domain.
- Mobile money transactional data can be used to understand mobile money penetration and economic activity at the identified gatherings to determine whether the gatherings should qualify as hotspots.
- Agent location data can be used to understand which hotspots are currently not served.

#### For data solutions providers:

- MMOs may need assistance in replicating this data analysis.

Source: Authors' own

# How to find us:

Get involved. Contact us.

**Krista Nordin**

[krista@i2ifacility.org](mailto:krista@i2ifacility.org)

**Michiel Wolvers**

[info@i2ifacility.org](mailto:info@i2ifacility.org)

**Kate Rinehart**

[kate@i2ifacility.org](mailto:kate@i2ifacility.org)

**Herman Smit**

[herman@i2ifacility.org](mailto:herman@i2ifacility.org)

+27 21 913 9510

[i2ifacility.org](http://i2ifacility.org)

 @i2ifacility

 /insight2impact

 /insight2impact

 /i2ifacility

Established by



Sponsored by

