



Advancing Financial Inclusion

Network analysis to **identify and segment farmers**

April 2018

Requirements:

Customer relationship management (CRM) data
Customer transactional data

Established by



Case Study

Authors

Krista Nordin, Renée Hunter, Herman Smit and Louise de Villiers, Client Insights

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.

Email Mari-Lise du Preez (Partnership Manager) at mari-lise@i2ifacility.org.

Call us on **+27 21 913 9510**.





Contents

1.	Introduction	3
2.	Overview of the case study	4
	2.1 Business challenge	4
	2.2 Data-driven solution	5
3.	Case study approach and results	6
	3.1 Data sources	6
	3.2 Analytical approach	8
	3.3 Results	11
	3.4 Applications for results	13
4.	Concluding insights and discussion	14



Terminology

The following definitions and acronyms have been used in this case study:

Definitions

Agribusiness:	A business involved in agricultural production
Agriculture value chain:	The various actors involved in the production, processing and sale of agricultural products (For this case study, the main categories of value chain actors are agribusinesses and farmers.)
Customer:	An individual or institution who has an account with the bank
Customer relationship management (CRM) data:	Information the financial institution captures on who its customers are and their subsequent interactions. This information is typically captured at the time of account opening and is, in some cases, updated at predefined customer interaction points with the bank (e.g. service calls, emails and account updates).
Customer segments:	Customers of the bank who engage in similar economic activity (in this instance, farming) and who share similar financial needs (e.g. agricultural finance)
Network analysis:	This broadly relates to the breakdown and analysis of a complex interconnected network. In the financial sector, network analysis has been used primarily to identify fraudulent activities and individuals. In this case study, network analysis is used to infer characteristics of a customer segment (farmers) based on their position within a network (agricultural value chain).
National ID:	A national identification number assigned by a country's government to confirm the identity of its citizens, permanent residents and temporary residents
Regular expression:	A sequence of symbols and characters expressing a string or pattern to be searched for within a larger piece of text
Transactional data:	Data generated each time a customer performs a financial transaction using their bank account or makes a payment facilitated by a branch or agent.
Transactional-data-driven network analysis:	Network analysis on the sender and receiver of payments as captured by transactional data
Unique ID:	A unique identification number assigned by a financial service provider to allow them to uniquely identify each customer

Acronyms

CRM:	Customer relationship management
ID:	Identification
KCB:	Kenya Commercial Bank
SACCOs:	Savings and credit cooperative organisations



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 examples, or 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

Traditional banks serve a diverse set of customers that may include entities such as schools, churches, agricultural cooperatives, small enterprises, SACCOs, farmers, households and individuals (to name a few). Each of these customer segments engage in similar economic activity, have specific financial needs and require different products and services. The information that banks have on these customers segments, or that they can infer, often differs. This case study considers how banks can identify and segment customers that are engaged in farming activity by analysing their transaction patterns and relationships (in a defined network) with other customer segments (such as agricultural cooperatives or dealers) that the banks may have more information on.

2.1 Business challenge

Existing customer identification and segmentation approaches rely heavily on information captured on the customer at the time the account was opened. However, this data can be inaccurate (for a myriad of reasons¹), can simply be left blank or may no longer represent a useful customer segment, which undermines the role that CRM data can fulfil in improving operational efficiency.

The absence of reliable customer data inhibits the key business functions related to:

- 1 Understanding the nature and prevalence of specific customer needs
- 2 Monitoring the effectiveness of financial services in meeting the needs of specific customer segments
- 3 Cost-effectively reaching customer segments with new or more appropriate financial offerings

The data-driven solution described below focuses specifically on overcoming these challenges in the service of farmers.

¹ For example: due to capturing errors as the information is captured manually; the account owner satisfying several categories at the time of opening and, therefore, the agent picking one; or the customer's core category changes with time and the attribute is not updated.



2.2 Data-driven solution

Traditional banks have a wealth of internal data on their customers' transactional behaviour. This data is available for both underserved and well-served customers. Transactional data further provides insights into the relationship between customers on whom the banks may have more information (e.g. commercial entities such as agro-dealers) and those on whom the banks may have less information (e.g. individual farmers). Traditional banks can leverage this data – through network analysis – to supplement or update their existing CRM data. The additional customer attributes derived from transactional data allows banks to (i) identify customers engaged in farming activity and respond by (ii) offering farmers more appropriate product and service offerings.

This transactional-data-driven network analysis (as elaborated on in Section 3) uses transactional and CRM data to infer characteristics about existing customers. This will allow financial service providers to use this data more effectively to reduce customer acquisition cost and increase the value they provide to customers. The applications of the results are further discussed in Section 3.4.

“ This transactional-data-driven network analysis uses transactional and CRM data to infer characteristics about existing customers. This will allow financial service providers to use this data more effectively to reduce customer acquisition cost and increase the value they provide to customers. ”

Box 1: About our partners

To produce this case study, we partnered with KCB Kenya and Pula Advisors.

KCB Bank Kenya Limited is a commercial bank headquartered in Nairobi, Kenya. As of 2015, KCB was the largest commercial bank in Kenya. KCB Kenya is a subsidiary of KCB Group, which has over 10 million customers.

Pula Advisors is a boutique data analytics consultancy based in Nairobi, Kenya.



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 sufficient details to assess whether this approach is desirable or feasible within one's own organisation or one's partner organisation.

3.1 Data sources

The data required to perform the network analysis and subsequent customer segmentation includes CRM data and customer transactional 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.

CRM data

CRM databases contain information on a firm's customers, and the data captured is aggregated from various points of contact between the customer and the company. These databases typically contain demographic information on customers and the products that they use. Organisations that use CRM systems effectively often capture all engagements between the customer and the firm (across different channels, such as website, telephone and email).

Sophisticated CRM data can contain hundreds of attributes; however, for this analysis only five basic attributes are required. Descriptions and the purpose for selecting each attribute are indicated on the next page.

“Sophisticated CRM data can contain hundreds of attributes; however, for this analysis only five basic attributes are required.”



Variable	Description	Purpose
Unique ID of account holder	A unique ID is generated internally for each account holder	Used to obtain a single-customer view across datasets internally
Name of account holder	The name of the account holder	Contains key words that can be used to identify which sector an account holder operates in
Industry of account holder	The industry ² that the account holder self-identifies with at the time of account opening	Used to identify the industry the account holder is in
Type of account	The type of the account (e.g. personal or business account)	Used to delineate between small-scale farmers and commercial farmers/ agribusinesses
Telephone number	Telephone number of the account holder	Used to cross-reference customers across external data sets

Transactional data

Transactional data comprises a record of each transaction made by a customer, including transactions facilitated by bank agents and branches. An individual data entry for each transaction includes a number of variables from which customer attributes can be derived. Pertinent information required to undertake this analysis and segmentation includes: the unique ID of both sender and receiver, the timestamp, the transaction amount and the free-text field. Descriptions and the purpose for selecting each attribute are indicated below.

Variable	Description	Purpose
Unique ID of sender Unique ID of receiver	A unique ID of the sender and receiver for the transaction	To map transactions between larger and smaller players in the agriculture value chain (e.g. between agribusinesses and farmers) identified in the CRM database
Category of sender	An identification of the category of the entity that the transaction is originating from (e.g. teller, agent or customer)	To determine where the transaction originated from (e.g. teller, agent or customer)
Free-text field	A field where the originator can choose to include additional information; Often used to record the purpose of a transaction or – in the case of an agent or teller transacting on behalf of an individual – the identity (usually their name, national ID number or phone number) of the sending individual	To identify the sender or receiver when the transaction is completed through an agent or branch teller

The nine variables listed above contain the required information to undertake a network analysis of customers to identify customers engaged in farming activity. The analysis of these variables is discussed in Section 3.2 below.

² Example industries include agriculture, education, consumer goods, telecommunications, etc.



3.2 Analytical approach

The network analysis approach followed can be broken down into two broad steps:

- 1 Identify bank customers who are known agricultural value-chain actors (e.g. agribusinesses and agri-based government institutions) that would typically transact with farmers.
- 2 Analyse transaction patterns to and from the identified players in Step 1 to identify farmers.

Once one has completed the steps, one will have a list of identified farmers within one's customer base.

Throughout this analysis, it is important to bear in mind that queries can take a long time if one runs them on large amounts of data. We recommend that queries be tested on small subsets of data before running them on the entire CRM or transactional database.

How to perform each step:

Step 1: Identify bank customers who are known agricultural value chain actors (e.g. agribusinesses and agri-based government institutions) that would typically transact with farmers.

1.1 Run two queries³ in the CRM system:

- Filter on "industry of account holder" for the identified target industry (in this case, agriculture).
- Search in the "name of account holder" column using regular expressions (regexes).
 - Create regexes that will match for relevant agricultural key words.⁴

See Box 2 for a description of regular expressions and example regexes one could use.

“ Queries should be tested on small subsets of data before running them on the entire CRM or transactional database. ”

³ A query is a request for data or information from a database table or combination of tables.

⁴ If you are unfamiliar with the key players in the industry, search online for the names of a few of the large agri-based government agencies and agribusinesses to inform your key word search. Examples in this case might be maize, dairy, tobacco and tea. Alternatively, consult business directories.

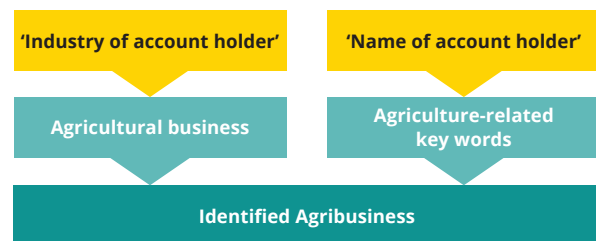
Box 2: Description and examples of regular expressions

A regular expression is a sequence of symbols and characters that expresses a string or pattern to be searched for within a longer piece of text.

An example regex that one can use to identify agri-based government institutions and agri-businesses is `\agri\`. This will match to any string⁵ that contains the substring “agri”. For example, if the name of account holder field was “ABC Agri-production” it would return this account holders name as a match.

“Agri”, as used above, can be replaced with any agriculture-related keyword to identify agribusinesses. For example, maize, dairy, tea or ag.

A visualisation of this step is below.



At the end of this step, one should have a list of agribusinesses.

1.2 This step is optional and can be used to identify additional agribusinesses not identified in Step 1.1.

- Pull all transactional data associated with the identified agribusinesses found in Step 1.1.
- For each identified agribusiness:
 - Sort the transactions by number of transactions per “unique ID of sender” and number of transactions per “unique ID of receiver”. Select the top few for each category.
 - Run the top unique IDs of senders and receivers through the CRM system. If the “type of account” associated with these IDs is “business”, add it to the list of agribusinesses.⁶

5 A string is a sequence of characters.

6 The names of these additional agribusinesses found in Step 1.2 can be used to update the key word list used in Step 1.1.



Step 2: Analyse transaction patterns to and from the identified players in Step 1 to identify farmers.

- 2.1 For all the agribusinesses generated from Step 1, pull their transactional data from the transactional database.
- 2.2 Filter on transactions in which these businesses are the receiving party.
- 2.3 Filter on “category of sender”
 - Transactions from customers:
 - If the “type of account” of the customer is personal, add this account holder to the list of farmers.
 - Transactions from third-party accounts (i.e. agents or tellers):
 - Analyse the “free-text field” to determine on behalf of whom the agent or teller was transacting.
 - Create a regex that will identify potential phone numbers that have been entered in this field. See Box 3 below for example regular expressions for this.
- 2.4 For returned potential phone numbers:
 - Run the phone number through the CRM system to see whether there is a match
 - If so, and if the “type of account” is personal, add the individual to the list.

Box 3: Regular expressions for identifying phone numbers

Two sample regular expressions that can be used to identify phone numbers in a text field are:

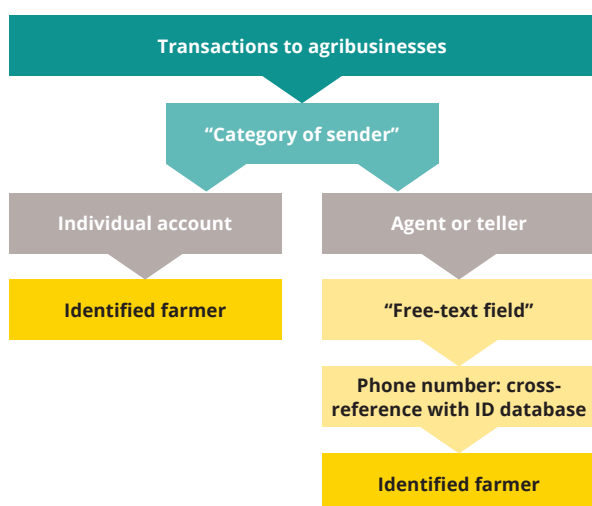
- `\d{10}` (This will match to any substring that contains 10 digits back to back. For example, if the open text field was “Mwangi 0705558989 450kg”, it would return “0705558989” as a match.)
- `\d{3}-\d{3}-\d{4}` (This will match to any series of digits that fits the format XXX-XXX-XXXX, where Xs are digits 0–9. For example, if the open text field was “2018-01-23 Fertilizer 070-123-4567”, it would return “070-123-4567” as a match.)

More complicated regexes can be used if one would like to include potential parentheses around area codes, spaces between numbers, etc.

“ At the end of this step, one should have a list of farmers that transact with the large agribusinesses and agri-based government institutions. ”



See a visualisation of this step below.



At the end of this step, one should have a list of farmers that transact with the large agribusinesses and agri-based government institutions.

3.3 Results

This section covers the insights generated from performing the analysis on KCB Kenya data. The objective is to provide a concrete example of the type of results the analysis may provide.

5,671 agribusinesses and government agricultural agencies identified. The above analysis performed on KCB Kenya data identified 5,563 agribusinesses and 108 government agricultural agencies. The segmentation on “industry of account holder” revealed 187 agribusinesses. An additional 5,376 agribusinesses were identified through analysing the “name of account holder” field and the network of agribusinesses.

2,560 farmers found. From analysing the transactions going to these identified agribusinesses, 2,560 farmers were identified.

“ An additional 5,376 agribusinesses were identified through analysing the ‘name of account holder’ field and the network of agribusinesses. ”





3.4 Applications for results

The insights generated from this analysis can be applied to strengthen a traditional bank's CRM database or create a table of a bank's customers that share common attributes. This table, or strengthened CRM data, can in turn advance the bank's marketing efforts, cross-selling strategies, and development of products and services that better meet the customers' needs.

Examples of how and where the insights can be applied for each of the categories are detailed below.

Product development

The analysis provides three valuable sources of information for product development. The first is the main source of income or economic-sector classification. This allows financial services providers to group together customers with similar financial needs (e.g. farmers). The second is the size of the customer segment. This information can inform the prioritisation and design of appropriate products. The third is transaction sizes and frequencies for the customer segment. This information can inform product features, such as loan sizes.

Cross-selling

The information derived from the analysis provides two valuable information sources for cross-selling. The first is the likelihood of uptake for a specific product offering (e.g. agricultural finance). The second is information on how to reach the customer or customer segments (e.g. mobile phone number).

Marketing

Marketing efforts can be advanced by economic-sector classification and the timing of transactions. Economic-sector classification can improve the positioning of the product to ensure that it relates to the economic reality of the target segment. Knowledge of the timing of financial transactions by customer segment allows for improved timing of marketing and sales activities.

“ The insights generated from this analysis can be applied to strengthen a traditional bank's CRM database or create a table of a bank's customers that share common attributes. This can in turn advance the bank's marketing efforts, cross-selling strategies, and development of products and services that better meet the customers' needs. ”

4 Concluding insights and discussion

This case study demonstrates how existing CRM data can be strengthened or supplemented through transactional-data-based network analysis. In this approach, a simple methodology was used to leverage customers on whom one has more information (e.g. agribusinesses and agri-based government agencies) to identify customers on whom one has less information (e.g. farmers) based on the transaction patterns between these customers.

While the data analysis model presented here already identifies the customer segment in question, there are additional steps that could be applied to the model for further impact, including:

- Use transaction sizes and frequency for credit scoring.
- Identify potential customers who currently interact with the bank (through bank branches and agents) but who do not have an account with the bank.
- Automate the analytical approach to continue to update the list of identified farmers as new data is captured.
- Add in a location element to identify further potential farmers based on the locations in which they are transacting.
- Apply this network analysis approach to other value chains outside of the agricultural sector.



Box 4: Key takeaways

For executives at traditional banks:

- Customer transactional data can be used to strengthen one's CRM database.
- Required analysis can be performed in-house or completed by data solution providers in approximately one week.

For analysts at traditional banks:

- CRM data can be used to identify key agribusinesses and agri-based government agencies.
- Customer transactional data can be used to identify individual farmers that transact with these larger commercial entities.

For data solutions providers:

- Network analysis can improve the data quality of existing CRM data.

Source: Authors' own

How to find us:

Get involved. Contact us.

Krista Nordin
krista@i2ifacility.org

Renée Hunter
renee@i2ifacility.org

Herman Smit
herman@i2ifacility.org

Louise de Villiers
louise@cenfri.org

+27 21 913 9510
i2ifacility.org

 @i2ifacility

 /insight2impact

 /insight2impact

 /i2ifacility

Established by



Sponsored by

BILL & MELINDA
GATES *foundation*

