Data delivery report

Landmark Millers

Date: 27-05-2022
Author: Jildemarie Brouwer
Focus crop(s): Maize & Sorghum
Country: Uganda

Introduction

This document contains details of the Landmark Millers primary data collection and data cleaning steps. Please read the PDC process guide before consulting this document.

The data collection for Landmark Millers took place between 15-21 May 2022 in the regions Soroti and Amuria, Uganda. Initially, 383 farmers were interviewed during this data collection, all of them participated. 3 farmers were deleted from the datasets because of reporting inconsistent answers. This will be explained in more detail in the chapter on Data Cleaning.

Survey design notes

The survey consisted of the usual core and some optional questions:

- Questions about the farmer agent model
- Questions tagged as “income diversification” in the question library

The survey differs from other surveys with regards to questions about the focus crop, because we had 2 focus crops: maize and sorghum. The questions about labour and input use were unaffected.

The following information is of help to understand specifics of the Landmark Millers survey. It was received during conversation with the company.

Survey questions on agent model

Please check the questions with variable names starting with “*cs_sdm_company_agent..” f*or the questions on agents that were added upon request by Landmark Millers.

Survey questions for income diversification information

Several survey questions can be of help to get knowledge on income diversification, as was requested by Landmark Millers. These were the questions tagged as income diversification in the question library, and some additional. Topics/survey questions to check are listed below.

- Size of land used to produce maize/sorghum
- Crops produced - we extended the questions asking for the type of crop that generates the highest income, second highest income, third highest income, .. etc. See section “Farm information”.
- Household labour used

- f_integrated_farming
- Interest in loans - what would they do with it
- Future outlook section

Two focus crops in one survey

Crop seasons

Farmers produced maize and/or sorghum throughout the year. They can produce maize in season 1 and season 2, maize in season 1 and sorghum in season 2 (or other way around), or sorghum in season 1 and season 2.

There are 2 seasons: Feb-June and Sep-Dec.

The same type of labour and inputs is used for the 2 crops, therefore we did not change those questions.

Certification

No certification schemes exist, hence the questions related to this topic were removed.

Maize & Sorghum varieties

Maize

1. Longer 10h
2. Longer 5

Sorghum

1. SC Sila
2. SESO 1, 2 & 3
3. chromatin
4. Gadam
5. Narosorghum

Landmark millers indication on production costs, including labour

Landmark millers provided an indication on production costs, including cost of labour. This can be used as a reference.

S/N	ACTIVITIES/INPUTS	QUANITITY	AMOUNT
1	Seeds (Maize)	10kgs	60,000
2	Hiring garden	1 acre	100,000
3	Land clearing	2 men	60,000
4	1st ploughing	tractor	120,000
5	2nd ploughing	tractor	120,000
6	Planting	8 pple	40,000
7	First weeding	12 pple	80,000

8	Second weeding	12 pple	80,000
9	Fertilizer Application	100kgs	100,000
10	Spraying	4 pple	25,000
11	Harvesting	8 pple	40,000
12	Threshing	3 women	30,000
13	Winnowing	3 women	30,000
14	Transportation		50,000
15	Storage		60,000
16	Bags	15bags	15,000
			1,010,000

Landmark millers existing data

The following data can be of help, please check with Landmark Millers

- Agents have paper-based records of farm transactions
- The mills have records of payments to agents

Sample characteristics

To get an accurate picture of the farmers we take a random sample of the farmer group. However, we often run into issues with the farmer list supplied by the company. Farmers for example might not have an address and we are dependent on a local contact person to take us to the right farmer. We often find the farmer is not the farmer that is part of the sample. As the random component of sampling is very important for the reliability of our findings we try to register whether the farmers we speak to were part of the original sample.

The amount of farmers part of the original sample in case of Landmark Millers can be found below:

Survey question: Was the farmer part of the original sample?		
Response	**Nr. of farmers**	**Share of farmers**
No	80	21%
No, he/she is an alternative for a sampled farmer that was unavailable	51	13%
Yes	252	66%

Data cleaning steps

Introduction

This section contains an overview of the different steps that are taken to clean the data. These steps have been drawn up in cooperation with IDH-FarmFit analysts and will be discussed in the following order:

- Removing Farmers from the Set
- Text cleaning
- Determining and handling outliers
- Looking at missing values
- Anonymizing
- Repeated question groups
- Case specific adjustments

Removing Farmers from the Set

Farmers are only removed from the set if they refused to participate in the survey. The only data we have from these farmers is the name, location and sometimes a phone number. None of the farmers refused to participate in the survey

Text Cleaning

In order to make the FarmFit data more accessible a few general steps are taken to clean the data.

- All columns and text values are set to lowercase
- Flow sets spaces to points; we set them to '_'.
- Dummy variables get the prefix 'X..OPTION…' by Flow, these are removed from the cleaned data set.
- A few free text options that have been found often in the data are set to similar text in order to make them comparable. An example is: 'don't know', 'doesn't know', 'I am not sure' are all changed to: 'I don't know'.
- In case the measurement of crop is supplied by farmers in a measurement unit other than Kilogram (e.g bags, boxes, crates, etc.), we have identified the value of the alternative measurement units in KG. The variable cal_focus_measurement_prod_kg captures a numeric value, indicating the number of kg that is in the measurement unit that is used (similar for measurement units use d to report quantities sold, lost, or own consumption)

- A measurement of an area is generally reported by farmers in acres, kilometres squared or hectares. In the cleaned data the measurements are set to acres, which can be seen in the column heading (_acre).
- Some redundant columns are removed, for example columns with Flow details unimportant for the FarmFit analyses.

Determining and Handling Outliers

To determine outliers for the numerical questions of the survey, a cut off of three standard deviations from the corresponding mean is set. All values are compared to this cut off. When the value is either higher than three standard deviations above the mean or lower than three standard deviations below the mean, it is set to '9997'.

Addressing Missing Values

The structure of the FarmFit survey prevents having actual missing values. All multiple-choice questions have the options 'I don't know' and 'I prefer not to say' and are mandatory. The numerical questions are also mandatory. Enumerators are instructed to answer them with '9999' in case a farmer doesn't know the answer, and '9998' when the farmer doesn't want to give the answer. This way all missing values are defined. In case of numerical questions, these values are not usable in aggregations and will give incorrect descriptive values. Therefore, all values containing '9999', '9998' and '9997', including those resulting from outlier handling, are set to 'NA'.

Anonymizing

In order to anonymize the data, farmer names, phone numbers, geolocation (longitude and latitude) and location except the highest administrative level (e.g. region or district) is removed from the set.

Repeated question groups

When recording the amount of crop produced, sold, lost or used for own consumption, we use 'repeated question groups'. This means farmers can provide input per season or for the whole year. In the cleaned data we only present one row of calculated values for each farmer. So if farmers reported production for 2 seasons, cal_focus_quant_prod_kg captures the total production during 2 seasons. For the amount produced, sold, lost and used for own consumption, we add the values of every season to get an idea of what happens throughout the year.

This process is applied for farmers that reported quantities produced, sold, lost, or consumed for multiple seasons; farmers that reported labour for multiple seasons; and farmers that reported input use and costs for multiple seasons.

Case specific adjustments

During data collection, we monitored incoming data and checked for outliers and inconsistencies using our data monitoring dashboard. Variables we check are:

- Measurement units
- Quantities produced, sold and lost (and consumed) of the focus crop
- Price received for the focus crop
- Farm size, including farm size of focus crop

Outliers were detected and adjusted in collaboration with the enumerators. In the majority of cases, the outliers were caused by typos.

In addition to this and the above described standard data cleaning steps, we have made case-specific adjustments to the data to ensure that data is reported in the correct way. We only assess the need for such adjustments for the above list variables, considering we have only limited time for data cleaning:

Measurement units

We added a land measurement unit "Decimals" in the survey because this is one of the main land measurement units in Uganda. During data cleaning, we calculated the farm size in acre, as usual.

- Quarter an Acre=25 decimals
- Half an Acre=50 decimals
- One acre=100 decimals

Quantity produced, sold, lost and/or consumed

Measurement units

Our contact person at Landmark Millers, indicated that farmers sell in bags of 100 KG (both Maize and Sorghum) and generally know how much KG they produced and sold in total. However ,some farmers forgot the number of Kgs sold, so they interpret these in basins, cups and sacks measurements. We asked Landmark Millers to give us actual measurements for this.

- Basin maize=15kgs
- Basin Sorghum -25kgs,
- 1 cup of maize=half kg
- 1 cup of sorghum=half kg
- sack of maize=100
- sack of sorghum=120-125kg fresh, 95kg dried and overstayed

We produced the values for cal_*focus*_measurement_prod_kg, cal_*focus*_measurement_sold_kg, and cal_*focus*_measurement_lost_kg accordingly.

Quantities reported

As will be described in the "Notes from data collection" chapter, there was a drought that caused many farmers to have no or very little production (in both seasons, and for both crops). We checked with the enumerators whether production levels were really so low (e.g. 4kg of maize in a season), and each of them confirmed that farmers reported these low numbers. Hence, the productivity ranges indicated on the intake form are not applicable to the past seasons, this was also confirmed by Landmark Millers. Note that this also affected the overall income from sorghum/maize of the farmers. Some made investments but had no or lower revenues from the crops.

Next, we only made adjustments to the quantities reported ("*cal_focus_quant_…"* variables) for which there was inconsistency in the measurement units reported and the related quantities. This was done with care, based on the knowledge we have of the measurement units.

As an example, a farmer reported to produce 3 KG, sold 200 KG and consumed 100 KG - we interpret this as that the farmer produced 3 bags of 100 kg, sold 2 bags and consumed 1 bag.

3 observations were deleted from the dataset, this was for farmers where there was inconsistency in the quantities reported for produced, sold and lost, and the accompanying measurement units. We were not able to identify the correct numbers in collaboration with the enumerators. Given the large sample size, we decided to leave out those observations.

Price/kg for sorghum and maize

- Some farmers reported the total revenues from when asking for the price/kg. For those, we calculated the price per kg (cal_maize_price or cal_sorghum_price) by dividing the reported value for price by the quantity sold.
- Some farmers reported a very low or very high price/kg. We did not want to exclude them from the data. To ensure we could still make the calculations, we replace their reported price by the median price:
    - Sorghum: entries that reported a price/kg lower than 150 UGX/g and higher than 2000 UGX/kg, received the value of the median price reported: 1000 UGX/kg
    - Maize: entries that reported a price/kg lower than 100 UGX/g and higher than 1800 UGX/kg, received the value of the median price reported: 800 UGX/kg
    - Note that the boundaries are not related to the boundaries reported at the intake form, instead they reflect specific entries in the dataset.

Farm size, including farm size of focus crop

No major adjustments were made, except for manual changes during data collection after checking odd numbers with the enumerators.

Other notes

- We were for now unable to add the descriptive results of the farmer agent questions to the sheets: “Cat. desc. multi all farmers” and “Cat. desc. multi by gender” in the data delivery file. We will aim to fix this problem but we still went ahead with delivering the data.
- Since we have 2 focus crops, calculated variables can have different names. Each variable that starts usually with “cal_focus_..” now has 2 versions: “cal_maize_…” or “cal_sorghum_..”
- There is 1 variable that captures net-income from focus crop: cal_maize_sorghum_income
- For more information on the calculation, please check the Excel where we explain the calculation.

Notes from data collection

Introduction

Our data collection lead took notes during data collection. These notes related to qualitative findings in the field, context-specific events the SDM team needs to be aware off, and feedback.

Enumerator Selection

Enumerators were selected from the 2 Districts on recommendation by a member of the ACODE, an organisation that carries out different data collection in the country.

Lesson Learnt: Using enumerators that have an experience in data collection in the field and using different applications is a bonus because they are easy to train and also they quickly interpret the questions and make sense out of them. During the training it was evident that some of them have done data collection with farmers before because they could mention things like unit measure and calculations were easier for them.

Agents

Landmark Millers requested us to collect information about the work of agents that work for them, in addition to the core survey questions. However, we only interviewed 12 agents. We had aimed for interviewing more. Some of the Agents that were given to us to work with were not available for various reasons.

- They had family emergencies
- Some agents left Landmark

- Some of the agents had to attend to saving groups
- Market days where the agents also had to sell their produce

Solutions: We used some Chairpersons to lead us to the farmers in the sample list and also to other farmers that were not in the sample list.

Lessons Learned: We should consider in preparation to request the companies to also share contacts of farmer group chairpersons.

Finding sampled farmers

We had a list of sampled farmers and the target was to reach 370 farmers in 5 days. In both Districts some of the sampled farmers said they do not know Landmark Millers. Some of the farmers sampled had left the groups and were independent farmers. Below are example reasons why they left the groups:

Tubur Subcounty

Landmark millers demanded all members to open accounts in equity bank so they could receive funds for farming finance and had expected to receive money by Dec which they didn't till to date. Group members paid 13,500@ for group registration, constitution signing and membership which they are now demanding Chairpersons to refund.

Many members are not even aware that they are members of a given a group, some were shocked to see their names hence many of the sampled have never grown any of those focus crops

Gweri Subcounty

- Delay supply of seed associated with procurement issues. Agent had to make a refund of the seed money collected.
- Crop failure due to climate. Some farmers failed in season one while others failed in season 2.
- Youthful farmers especially those that are still in school rely on their proxies to do the farming. Their parents are practically the group members. Even savings in VSLA is done by these.
- Group dynamics. Some people didn't even know they were in groups. Others thought the Landmark arrangement had not seen the light of day.

Lessons Learnt: the non-sampled farmers were willing to participate and gave more information as opposed to the sampled farmers. This may be because they are not keeping the interests of Landmark Millers or have not been asked to give specific information.

General Feedback

- In some sub counties like Achuna-Tubur, farmers said while they had prepared farms ready to plant maize and sorghum, the seeds were not available to them in time by Landmark Millers hence they missed the planting season. This also made them missing out on any other crop
- In Tubur some groups did not plant anything in season one even having prepared the gardens. This was because they continued waiting for the seeds they had been assured of by Landmark Millers which did not come. By the time they realized it was already too late to plant anything. Unfortunately, they did not receive an explanation as to why their group missed out.
- Most parents enrolled their children in the group so they could be able to access Landmark services as Landmark Millers was specifically interested in working with youthful farmer groups.
- In Amuria District , Asamuk subcounty in Alengo village names of members sampled were not available on ground
- Members paid 4 seeds expensively with hope of getting more but after harvest selling price drops which affects their profits and hence do not find it as an advantage to be part of the Landmark groups created
- Some members sampled were not aware of the groups and did not join with consent. This was evident because they did not even know the group names that they were being told they belonged to.
- Some groups never received seedlings since formation and are still hopeful of getting support from Landmark Millers
- Most groups are under the leadership of chairpersons as opposed to agents as directed by Landmark millers
- In groups in town boards (Asamuk and Orungo) some members do not know Landmark Millers agents and they have never bought seeds from the agents instead its other non-registered farmers who buy seeds from the agents.
- Some of the group members sampled were unaware of being members of the stated groups that they are purported to be belonging to.
- The lists mixed up some members who were nonresidents of the area and are not known to residents.

Lessons Learnt:

Companies sometime give information that they have not updated and hence working with snowball sampling is also important