Data delivery report

Farmworks

Date: 18-11-2022

Author: Ines Lesire, Mercy Makena

Focus crop(s): Sweetcorn & French beans

Country: Kenya

Introduction

This report elaborates on the primary data collection (PDC) on farmers working with Farmworks, cultivating sweetcorn and French beans in Kirinyaga County in Kenya. The goal of the data collection is to assess the current livelihood of these smallholder sweet corn and French beans farmers as a benchmark for measuring the positive influence of IDH and Farmworks activities on their livelihoods and improve on the service delivery model for the company which in turns is expected to improve the smallholder farmer income.

Further information and context will be provided on the survey design, the actual sample characteristics, the data cleaning steps, and qualitative observations in the field during data collection. The purpose of this document is to handover contextual knowledge so that analysis can be done in the most optimal and efficient way.

The data collection for Farmworks took place between 7-11 November 2022 in the regions of East, Central and West Kirinyaga in Kenya.

Survey design notes

The survey consists of the core questions and some new questions:

- Questions about farmer satisfaction and protection
- Questions on off-taker of focus crops

Further, optional questions on future outlook, household roster, and the poverty probability index were added.

The survey differs from other PDC surveys because the crops of focus were sweet corn and french beans which are horticulture crops and they are not grown in seasons. The questions on labour and income were set using the harvest cycles in place of seasons.

Survey questions on farmer satisfaction and protection

Please check the questions with variable names starting with "*fsp_*" for the questions on satisfaction and protection of farmers that were added upon request by the SDM team, since this is a returning topic. The section contains a total of 28 questions that address the farmers' perception on the company's - service

delivery, their responsible pricing, loans taken from the company, their transparency, fair and respectful treatment, privacy of the client data, and complaint resolution. The plan is to incorporate these survey questions in more cases in the future after reviewing their use with the Intelligence team.

Survey questions on off-taker

In the section on revenues from focus crop, additional questions -on whether or not all produced focus crop was sold to the off-taker (Farmworks), and what happens with the produce that is not sold to them were added.

Sample characteristics

Theoretical sample size calculation

We selected a subset of farmers that are representative and from which conclusions can be drawn about the whole population. For this, the population-based sampling technique was used to calculate the sample size. The different CBO's also referred to as the collection centres were taken into account for disaggregation. The calculation was based on a farmer list we received from Farmworks.

A response rate of 80% was taken into account, presuming that 20% of the farmers in the sample will not be able to answer the survey. This led to a calculated sample size of 299 farmers, divided over the five different CBO's (collection centres)

Sample size allocation in the field

In the field, finding each farmer part of the theoretical sample calculation was a big challenge. Akvo used farmer guide from the PU groups to identify farmers from the sample. To ensure that we get data on the income, labour costs and inputs, we targeted farmers who had at least one cycle of either french beans and/or sweet corn. The farmers might have sold the crop to farmworks or any other offtaker however, it had to be within the last 12 months.

The table below shows to what extent the sampled farmers were part of the original sample calculation.

Survey question: was the farmer part of the original sample?		
Response	**Nr. of farmers**	**Share of farmers**
No	23	6.6%

No, he/she is an alternative for a sampled farmer that was unavailable	33	9.4%
Yes	301	84%

Figure 1 illustrates that the percentages of farmers in each CBO in the actual sample, corresponding to the farmer list received from Farmworks.

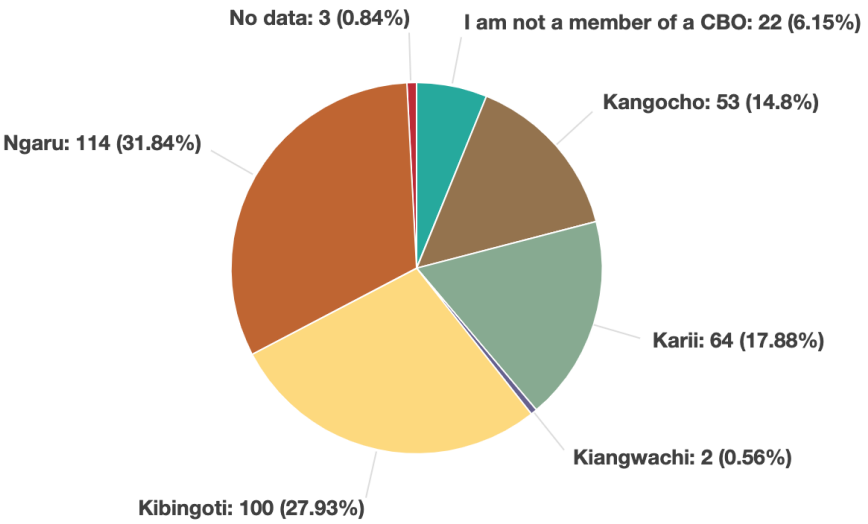

Figure 1: sample farmers in each CBO

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. For this case, only **1 farmer** refused to participate and was deleted from the dataset.

Text cleaning

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

- Set the submission date variable to date format
- 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 used 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 this case, the farm size was measured in acres, squared meters, units and plots. In the cleaned data the measurements are set to acres, which can be seen in the column heading (_acre). This is explained in more detail in section "Case specific adjustments".
- Some redundant columns with Flow details which are unimportant for the FarmFit analyses, are removed from the data.

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', which means that the value is missing (see next section).

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’.

Anonymising

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) are removed from the set. The farmer can still be identified by the unique number in the “identifier” column.

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 cycles; farmers that reported labour for multiple cycles; and farmers that reported input use and costs for multiple cycles.

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

Measurement units

We added a land measurement unit “Squared meters” and “plot” in the survey because they are one of the main land measurement units in Kirinyaga County in Kenya. During data cleaning, we calculated the farm size in acre, as usual.

- 1 squared meters = 0,000247105 acres
- 1 plot = 500 squared meters = 0.125 acres

Quantity produced, sold, lost and/or consumed

Some farmers reported a much higher number for the quantity sold than the quantity produced of sweetcorn or french beans. It turned out that these farmers reported the total revenue of the quantity sold instead of the amount (in kg). We assume this because the division by the quantity produced gives us the number 25, which is the unit price (in KES) reported by the majority for sweetcorn. These numbers were thus replaced by dividing the total revenue by a unit price of 25 KES. This was done for 4 farmers.

For one farmer, numbers for produced and sold focus crops were odd, and sense-making was not possible. This farmer was removed from the dataset.

Furthermore, some farmers reported unrealistic production numbers compared to a small focus crop farm size. These were probably mistakes from errors of either the enumerators or the farmer himself. We decided to put these production numbers to NA.

Focus crop price

Some reported focus crop prices (both sweetcorn and french beans) were unrealistically high. When analyzing into more detail, it seemed that these farmers reported the total revenue from the focus crop instead of the price per unit (in KES). We solve this by dividing the reported value for price by the quantity sold, and impute this value in cal_sweetcorn_price or cal_frenchbeans_price.

Farm size

Lastly, some farmers reported a focus crop size quantity produced, but a focus crop farm size of 0. This is not feasible and gives errors in the productivity calculation. This is why these values were put to NA, for a total of 10 farmers.

Calculation of productivity

As already mentioned, the respective focus crops of sweetcorn and french beans are planted continuously throughout the year in rotation. Above that, most farmers cultivate only one of the two focus crops (figure 2).

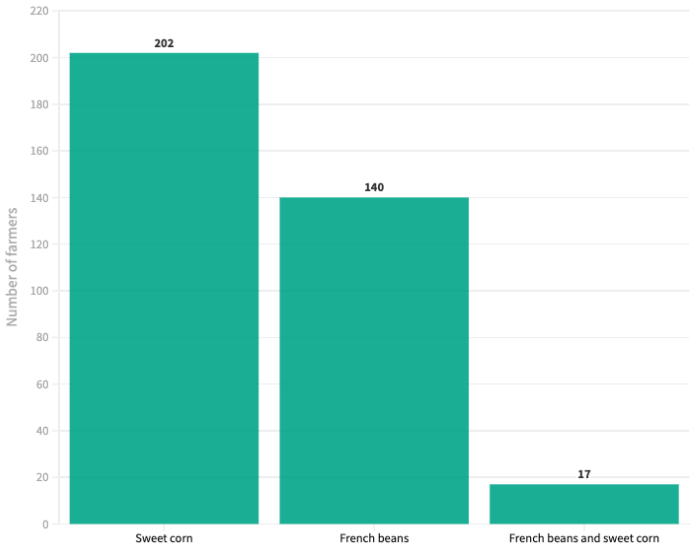

Figure 2: number of farmers cultivating each focus crop

Because of these reasons, the productivity for each focus crop is calculated separately, through the following formulas:

$$cal\_sweetcorn\_productivity\_acre = \frac{cal\_sweetcorn\_quant\_prod\_kg}{f\_focus\_crop\_size\_acre}$$

$$cal\_frenchbeans\_productivity\_acre = \frac{cal\_frenchbeans\_quant\_prod\_kg}{f\_focus\_crop\_size\_acre}$$

Notes from data collection

Enumerator Selection and Training

An advert was circulated through our contacts in the region. Following receipt of many applications, qualified enumerators were selected. The selected enumerators had an understanding of the context in Kirinyaga county and they could speak the local language. A one day training was conducted where the enumerators were taken through the use of Akvo flow for data collection and the survey.

Finding Samples farmers in the field

Farmers who had harvested at least one cycle of either sweet corn and/or french beans were identified as part of the sample. To identify these farmers on the ground, Akvo worked with farmer leads who are

usually the production unit (PU) leads. Some of the farmers who were interviewed had not sold any crop to farmworks, or they had sold a different crop to farmworks, for instance, tomatoes, capscum, chilli, sweet potatoes in the past but they were just in the initial stages of growing the crops of focus. Most of these farmers, however had grown and harvested the crop before and sold to different off takers in the past 12 months hence were also included in the sample.

Farmers from one of the CBO, Kiangwachi, were not included in the sample because they had not harvested any cycle of the focus crops.

General feedback

- Most of the farmers are interested and happy to grow sweet corn and french beans. The main reason for this is because it is not rainfed and they can have income all year round. Farmers felt that they start getting income from french beans only 2 months from the time of planting
- Farmers depend on irrigation for the production of sweet corn and french beans
- Farmers dedicate only ⅓ of their farms to the production of sweet corn and french beans. In addition, farmers work with plots for the crop production where each farmer works with either 1 or 2 plots (0.125 to 0.25 acres) of land at the moment for both french beans and sweet corn
- The company is still in the process of registering the farmers that they work with
- In case the crop is not acceptable for the market, and Farmworks or the offtaker does not buy from the farmer, the farmer feed the produce to livestock
- Farmers working with Farmworks are organised into production units. Several production units form a CBO. The collection centres are located in places where farmworks have a CBO and each CBO is equivalent to a collection centre..