Data delivery report

Goshen

Date: 16-12-2022

Author: Ines Lesire

Focus crop(s): Mango & pineapple

Country: Kenya

Introduction

This report elaborates on the primary data collection (PDC) on farmers working with Goshen, cultivating the focus crops mango and pineapple. The mango farmers are located in Makueni and Tana River county and the pineapple farmers in the Kilifi county. Goshen buys these fruits to process them into dry fruit. The goal of the data collection is to assess the current livelihood of these mango and pineapple smallholder farmers as a benchmark for measuring the positive influence of IDH and Goshen 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 Goshen took place between 5-9 December 2022. Two separate surveys were developed, one for the mango farmers and one for the pineapple farmers. This means we have two separate intake forms, separate surveys, separate sampling designs and separate data cleaning files.

Survey design notes

Two surveys were designed: one for the mango farmers and one for the pineapple farmers.
The survey consists of the core questions and some additional case questions:

- Perennial crop questions
- Additional questions mango
- Additional questions pineapple
- Questions on farmer satisfaction and protection

Perennial crop question

Since mango and pineapple are perennial crops, additional questions are asked about the number of mango trees and pineappleplants the farmer has, and in what age category these trees fit.

Additional questions mango

In the survey concerning mango farmers, additional questions about the pruning of mango trees are asked. The frequency of pruning, length of pruning and the age of the tree at moment of pruning are surveyed.

Additional questions pineapple

In the survey for pineapple farmers, an additional question was asked regarding rejuvenating the pineapple farm, since this is a relevant practice for this crop.

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. This is the second case in which these questions are integrated. The plan is to incorporate them in the question library after reviewing their use with the Intelligence team.

Sample characteristics

Mango survey

Goshen works with Mango farmers in Makueni (Mbooni and Kaiti Sub Counties) and Tana river (Hola Sub County). A sample size of 203 farmers was calculated using a population of 1403 farmers that Goshen shared with Akvo through the intake form. 246 farmers were interviewed from both Counties with 165 of them being from Makueni and 81 from Tana River.

Survey question: was the farmer part of the original sample?		
Response	**Nr. of farmers**	**Share of farmers**
No	10	4,06%
No, he/she is an alternative for a sampled farmer that was unavailable	10	4.06%
Yes	226	91.87%

Pineapple survey

A sample size of 83 farmers was calculated using a population of 200 farmers that Goshen shared with Akvo during planning. All the pineapple farmers are located in Magarini Sub County in Kilifi County. A total of 71 farmers were interviewed from Kilifi County.

Survey question: was the farmer part of the original sample?		
Response	**Nr. of farmers**	**Share of farmers**
Yes	71	100%

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, none of the farmers refused to participate so no removals needed to be done.

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 in each survey 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). However, for both mango and pineapple, another measurement unit of "per piece" was introduced. More on this in the section about case specific adjustments.
- A measurement of an area is generally reported by farmers in acres, kilometres squared or hectares. In this case, the farm size was measured only in acres.
- 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 (County and Sub County) are 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 harvest seasons, cal_focus_quant_prod_kg captures the total production during 2 cycles. 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.

However, in the case of mango and pineapple, there is only one season, except for 4 farmers.

Case specific adjustments

During data collection, we monitored incoming data and checked for outliers and inconsistencies using our data monitoring dashboard. Two dashboards were made, for the two different surveys: one for mango and one for pineapple. Variables we check are:

- Land measurement units
- Crop 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

Crop measurement units

When measuring mango and pineapple produced, sold, lost or consumed, many farmers use "per piece" as a measurement unit. It is not feasible to transform this into kgs, since crop sizes vary a lot. The variables indicating quantity produced, sold, lost and consumed are twofold in this case, one "per piece" and one in "kgs". Also for productivity, both productivity_per_acre_kg and productivity_per_acre_per_piece are calculated, to indicate what measurement unit was used.

Quantity produced, sold, lost and/or consumed

Some data entries contained unrealistic numbers on quantity produced, sold, lost and consumed. In the mango survey, one farmer was removed given unrealistic numbers that we could not make sense of. For some farmers, the measurement unit was not known, and we replaced this with “per piece” looking at the quantities reported. These errors were solved in the data cleaning script.

In the pineapple survey, some farmers reported very high production numbers which are not realistic. Contact was made with the responsible enumerator, who corrected the values in conversation with the farmers. One farmer reported a higher number of pineapples sold than produced. Again, the numbers were corrected by the responsible enumerator.

Focus crop price

For the mango survey, some farmers reported a total revenue instead of the unit price for mango. We replaced these high values with the division of this total revenue with the quantity produced. A handful of other farmers indicated a very high unit price that could not make sense in this reasoning. These outliers were put to the median price value of 15 KES/kg.

For the pineapple survey, two extreme outliers in unit price were detected. They were replaced by a value of 50 KES/piec, since this is the standard price for a pineapple per piece.

Calculation of productivity

Mango and pineapple are focus crops often measured per piece instead of in kg when produced and/or sold. This is why productivity was calculated both per kg and per piece, depending on the given measurement unit.

Also for the mango farmers, productivity per tree was measured, both per kg and per piece as well. This leads to 4 productivity variables for the mango data delivery: productivity in kg per acre, productivity per piece per acre, productivity in kg per tree, and productivity per piece per tree. For the pineapple survey, only the first two productivity variables are present.

Notes from data collection

Enumerator Selection and Training

An advert was circulated through our contacts and the company in the two counties. Following receipt of many applications, qualified enumerators were selected. The selected enumerators had

an understanding of the context in the respective counties 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.

General feedback

1. In some counties, mangoes have a main season and 1 'mini season' where farmers harvest a few mangoes. However, since most farmers do not experience the two seasons, all the information was collected and consolidated as one season
2. Approximately 3-4 pieces of mango make up a kg of mango. This is, however, dependent on the mongo variety and the size of the mango. This makes it challenging to change the measurement unit 'per piece' to kgs.
3. The farmers are not registered in groups and they work directly with the company