MCC Evaluation Microdata
Data Package

Section 1: Cover Sheet

Overview of Data Package

The package contains a dataset comprising the responses from a household-level survey conducted on 125 farmers and traders between 08-28-2018 and 09-27-2018. Although similar, the package includes two datasets; one for farmer survey and one for trader survey. The farmer survey dataset reflects responses that can be broadly divided into eight modules. The first module contains basic information on the respondent including age, gender, marital status, education, membership at farmers' groups etc. The second module focuses on information related to farm characteristics, production, revenue and cost. The third module contains data related to sales and marketing such as amount of sales, exports, transportation to market etc. The fourth module includes data on drip irrigation management, usage, satisfaction, challenges, etc. The fifth module contains information related to the training component under the WMAS Project. The sixth module includes information on post harvesting practices. The seventh module contains information on various aspects of credit services and the last module includes data on costs and benefits of agricultural production to inform a value chain basic information, quality standards, market information, training, post-harvest practice, credit, and cost and benefits related data.

The trader survey dataset reflects responses that can be broadly divided into seven modules. The first module includes basic information on the respondents. The second module includes information on how the traders maintain quality standards. The third module contains various market related information such as where the produces are sold or to whom the produces are sold, transportation to the market, purchase price, sales price, etc. The fourth module was designed to collect information on the training component. The fifth module focuses on post-harvest practice followed by the sixth module which includes data on credit services usage by the traders. The seventh module includes various cost and price information to inform a value chain analysis.

This data package includes the following components:
Proposed for public release
- Data Package Component 1: Farmer survey responses
Not proposed for release, even under restricted access (see Publication Notes below)
- Data Package Component 2: Trader survey responses

Publication Notes

Both components used a consent form that promised, "*Any information you provide that can identify you will be kept strictly confidential*". The consultant was able to de-identify the Farmer data to prevent re-identification of the respondents, although this included masking many analysis variables to prevent them from being used for re-identification.

However, given the Trader data's small sample size of 11 and the large number of identifying data points, the consultant concluded that it is impossible to de-identify the Trader data and still have a usable set of data.

Complementary Data
N/A

Data Package Folder Contents

Data Package				
Component	**Worksheet**	**Informed Consent**	**Questionnaire**	**Other Documents**
FOR PUBLICATION 1: Farmer data` • De_identified_farmer_survey_responses_WMAS_Cape_Verde.dta NOT TO BE PUBLISHED IN ANY FORM 2: Trader data • De_identified_trader_survey_responses_WMAS_Cape_Verde.dta	37 - Cape Verde - Watershed Mangement - De-Identification.doc	• 37 - Cape Verde - Watershed Management - MCC CV Consent Form.pdf	1: Farmer Survey Questionnaire_WMAS_Cape Verde.pdf 2: Trader survey questionnaire_WMAS_Cape_Verde.pdf	• 37 - Joe Green - Cape Verde - Watershet Management and Agriculture Support Activity.do • 37 - Cape Verde - Watershet Management and Agriculture Support Activity - data info.xlsx

Section 2: <mark>Data Component</mark> Preparation Overview

	Response		Discussion/Explanation
Data + Code Completeness	Complete	1: Farmer data The data is incomplete because many important treatment and analysis variables were either dropped or had their values grouped to prevent re-identification of the respondents. 2: Trader data (not included in files to be published)	*To be considered Complete: The available data must allow new users to replicate evaluator analysis to the extent allowable by providing the full data set + analysis code. The constructed variables may also be included in a dataset, but if the dataset+code produces those variables, it is not necessary.* *To be considered Incomplete: The available data only provides a sub-section of data as produced by the survey and/or the constructed variables only. Incomplete data files are limited in terms of full verification of analysis and/or broad usability of data and must be justified.*
	Incomplete		
Data Round(s):	Baseline only	Endline only	*MCC is willing to trade-off broad use of individual rounds for more consistent de-identification protocols across rounds of data. Therefore, unless there is specific demand for the baseline/interim only data, or contractual requirements, MCC prefers contractors to prepare all data rounds in one package.* *If one stage only – please (i) confirm demand and/or contractual justification and (ii) discuss how preparation and release of this data as presented to the DRB may affect future data round releases.* *If combination, please discuss if this file replaces any previously published datasets.*
	Interim only		
	Endline only		
	Combination of rounds		
Informed Consent and IRB	High restriction	Medium restriction	*MCC assumes DIRECT identifiers are always removed from any public-use file. With this assumption: Please refer to the informed consent statement – does it require: High restriction: access to data that includes indirect identifiers is limited to the contractor only; Medium restriction: access to data that includes indirect identifiers is limited to the contractor and qualified researchers, including MCC; Low restriction: data with indirect identifiers may be made public.* *Please discuss how the promises of confidentiality in the informed consent informed de-identification efforts. Please include any additional guidance provided by the IRB as applicable.*
	Medium restriction		
	Low restriction		

Geographic Identifiers	Highest	Island	Removed. Removed due to the low sample size of <100 farmers per island, and the large number of indirect identifiers.	*Please provide justification on the identification/de-identification/complete removal of specific geographic regions. De-identifying at a higher geographic level may support privacy protection, but it may also reduce data usability. Please provide justification for recommendation.*
	--(i.e. District)	N/A	N/A	
	--(i.e. State)	N/A	N/A	
	Lowest	Municipality	Removed. Removed due to the low sample size of <100 farmers per island, and the large number of indirect identifiers.	
Knowledge of Treatment	High risk	Low risk The treatment area as well as the sample size is small. However, the evaluator has removed all possible identifiers including name, address, GIS and plot location. Additionally, the consultant removed two treatment variables that identified groups of less than 5 farmers. Even though it is not difficult to tell who the treated farmers or traders are, the risk of identifying individual information from the de-identified data is low.	*In some cases, general knowledge of treatment areas and/or inclusion of a treatment variable can significantly increase re-identification risk depending on the population affected. Please provide assessment of this re-identification risk and recommendation if considered high/medium risk.*
	Medium risk			
	Low risk			
Publication Type	Public-use only	1: Farmer data Public-use only. The de-identified dataset can be used by anyone to replicate some of the evaluation results. 2: Trader data Neither. The consultant concluded this data should not be release in any form (see Publication Notes above).	*Please state for this data package: will there be public-use data only, restricted-use data only, or both and provide justification as this relates to enabling verification of evaluation results and/or broad usability of the data.*
	Restricted-use only			
	Both			

Section 3: Data Component Preparation Details

Specific Issues	Risk Analysis		Risk Mitigation	
	Instructions	*Response*	*Instructions*	*Response*
1. Who has significant financial, legal, cultural, or other incentives to re-identify survey respondents?	*List all potential threats[1]*	Government and financial institutions		
2. What is the potential value to these intruders?	*List all uses (for example: capture delinquent tax payments, or stigmatize the respondent)*	Knowledge of the farmers' delinquent water or loan payments.		
3. What is the expected cost to these intruders to re-identify the data?	*Describe degree of difficulty for re-identification*	Given the extensive de-identification of direct and indirect identifiers, it would be nearly impossible to re-identify the data.		
4. Assess availability of 'linkage' data that can be used to re-identify respondents. This includes other datasets or archives with information that can be used to re-identify individuals in the dataset.	*List all potential existing data*	Evaluator's response: None available.	*Describe how to mitigate link to existing data that enables re-identification*	N/A

---

[1] As stated in NIST 2016, de-identification practitioners should assume that de-identified US government datasets will be subjected to sustained, world-wide re-identification attempts, and they should gauge their de-identification requirements accordingly. Although a specific dataset may not be seen as sensitive, de-identifying that dataset may be an important step in de-identifying another dataset that is sensitive. Alternatively, the adversary may merely wish to embarrass the US government agency or its partners. Thus, adversaries may have a strong incentive to re-identify datasets that are seemingly innocuous.

Specific Issues		Risk Analysis		Risk Mitigation	
		Instructions	*Response*	*Instructions*	*Response*
5.	**Identity Disclosures:** What are the DIRECT identifiers in the raw data?	*List the DIRECT identifiers (names, addresses, geographic information, government-issued ID numbers, etc.)*	Response Code, IP Address, Interviewer Name, Interviewer ID, Household ID, Respondent's name, Respondent's Phone Number, Address Area, GPS/GIS information, and Farm Location.	*List all DIRECT identifiers removed from the dataset.*	Removed all direct identifiers.
6.	**Attribute Disclosures:** For GIS/GPS data, this distance data can be a direct identifier that is VERY useful analytically. Therefore, please describe how GIS/GPS data VALUE/USABILITY can be retained.	*List all GPS and/or GIS data.*	This data has been completely removed by the evaluator in the de-identified dataset received by the consultant.	*Describe process for de-identification. For example: introduce random errors into geographic data (GPS, GIS, etc.). Displace urban points 0-2 km, rural points 0-5 km, and additional 1% of rural points 0-10 km².*	N/A
7.	**Attribute Disclosures:** What variables have OUTLIERS that create INDIRECT identifiers are in the raw data?	*List the identifying items/variables*	See Addendum A: Additional removed or grouped data	*Describe top/bottom coding: set upper & lower bounds to remove outliers for continuous. Specify: are values set to the median, or other? For large categories/datasets, the OMB suggests top coding at least the highest .5%; for smaller categories/datasets, top code the highest 3-5%. The same principles apply to bottom coding.[3]*	See Addendum A: Additional removed or grouped data

---

[2] ICF International, Demographic & Health Surveys
[3] Office of Management and Budget, Checklist on Disclosure Potential of Proposed Data Releases (current link)

Specific Issues	Risk Analysis		Risk Mitigation	
	Instructions	*Response*	*Instructions*	*Response*
			Describe any variables that require collapse and describe construction of new variable	N/A
			Describe any global re-coding to group observations into categories (e.g., age 0-5, 5-10, 65+, etc.). Ensure that the categories are neither too broad nor too narrow.	N/A
8. **Attribute Disclosures:** What variable combinations produce UNIQUE observations that create INDIRECT IDENTIFIERS (for example: individuals with high incomes, ages, or unique combinations, such as 17-year old widowers or contextually unusual racial/ethnic backgrounds)	*List the identifying items/variables:*	See Addendum A: Additional removed or grouped data	*For each identified rare data, describe the local suppression techniques employed to remove unique and rare data. Specify: are values set to missing, the median, or other?*	See Addendum A: Additional removed or grouped data

Addendum A: Additional removed or grouped data

Variable Group 1) Converted specific date to month and year
* DraftedOn = Drafted On
* SubmittedOn = Submitted On
* Dateofinterview = Date of interview
* DateofinterviewTimeAnswered = Date of interview (Time Answered)

Variable Group 2) Removed variables that are too identifying or sensitive
* Island = Island:
* AddressMunicipalityundefi = Address: Municipality: - undefined
* FarmLocationMunicipality = Farm Location: Municipality: - undefined
* age = Age of respondent:
* gender = Gender of respondent
* MaritalStatus = Marital Status
* Marital_status_other = Marital Status (Other (please specify)) - specify
* household_head_other = Are you the head of the household? (No (please specify who is the head of the h
* primary_operator_other = primary_operator_other
* cooperation = Do you belong to any cooperatives?
* coop_years = ¬†How many years have you been in the cooperative?*
* coop_decision = To what extent do you participate in the cooperative's decision making?
* farm_years = How many years have you been operating your farming business?
* Q28_other = What other activities do you receive income from, and how much income do you rec
* Q37_7 = What are the main reasons for your income change? Others specified.
* Q45_other = What types of crops did you grow before the Infrastructure Development & Trainin
* Q46_other = What types of crops did you grow, and please provide information on plot size, h
* Q48_comments = What other expenses did you make in the last agricultural year and before the In
* Q49_comments = What amount (in kg) of your production did you and your family sell, store, cons
* Q50 = Do you sell your agricultural products?
* Q51_5 = Who do you sell the majority of the products to? - Processors

* Q51__6 = Who do you sell the majority of the products to? - Farmer Organization
* Q52_4 = Where do you sell? - Another market
* Q52_5 = Where do you sell?- Another Island
* Q52_6 = Where do you sell? - Other
* Q53_2 = What is the destination market?- Export
* Q53_3 = What is the destination market?- Both In-ciuntry & Export
* Q54_1_comments = Why dont you export? (Other (please specify)) - specify
* Q55_* = Please mention the island you sell? - *
* Q57_* = How do you go there (means of transportation)? - *
* Q58 = On average, how much do you spend on transportation every time?
* Q59_* = How often do you export to other islands? - *
* Q61 = What certifications are required for exporting commodities?
* Q62 = Are certification requirements mandatory?
* Q63 = What are the rules and regulations you need to comply with to get certified?
* Q65_6_comment = Greatest challenges to agricultural production last agricultural year- Other spe
* Q66 = Relative to a typical year, during the past agricultural year, how much rainfall
* Q67 = Is there any water association in your area?
* Q68 = Are you or a family member a member of any water association?
* Q70_other_comment = What equipment do you have? (Other (please specify)) - specify
* Q72_a = Did you pay for the irrigation water you used? (CVE)
* Q72_b = Did you pay for the irrigation water you used? (CVE) (Don't Know)
* Q72_1_a = How much did you pay, on average, per cubic meter of irrigation water? (CVE)
* Q72_2_a = Why didn't you pay for the irrigation water you used? - No money
* Q72_2_b = Why didn't you pay for the irrigation water you used? - Costly
* Q72_2_comments = Why didn't you pay for the irrigation water you used? (Other (please specify)) -
* Q73_3 = What irrigation source did you use to irrigate your land? - Private provider
* Q73_5_comments = What irrigation source did you use to irrigate your land? Other specified.
* Q74_3_a = How satisfied are you with the water volume? : Private Provider - Fully Sat
* Q74_3_b = How satisfied are you with the water volume? : Private Provider - Somewhat
* Q74_3_c = How satisfied are you with the water volume? : Private Provider - Neither S

* Q74_3_d = How satisfied are you with the water volume? : Private Provider - Somewhat
* Q74_3_e = How satisfied are you with the water volume? : Private Provider - Fully Uns
* Q74_3_f = How satisfied are you with the water volume? : Private Provider - N/A
* Q75_3_a = How satisfied are you with the water quality? : Private provider - Fully Sa
* Q75_3_b = How satisfied are you with the water quality? : Private provider - Somewhat
* Q75_3_c = How satisfied are you with the water quality? : Private provider - Neither
* Q75_3_d = How satisfied are you with the water quality? : Private provider - Somewhat
* Q75_3_e = How satisfied are you with the water quality? : Private provider - Fully Un
* Q75_3_f = How satisfied are you with the water quality? : Private provider - N/A
* Q76_3_a = How satisfied are you with the timely water delivery? : Private Provider -
* Q76_3_b = How satisfied are you with the timely water delivery? : Private Provider -
* Q76_3_c = How satisfied are you with the timely water delivery? : Private Provider -
* Q76_3_d = How satisfied are you with the timely water delivery? : Private Provider -
* Q76_3_e = How satisfied are you with the timely water delivery? : Private Provider -
* Q76_3_f = How satisfied are you with the timely water delivery? : Private Provider -
* Q77_1_a = How satisfied are you with the cost of the water you received? : Private Pr
* Q77_1_b = How satisfied are you with the cost of the water you received? : Private Pr
* Q77_1_c = How satisfied are you with the cost of the water you received? : Private Pr
* Q77_1_d = How satisfied are you with the cost of the water you received? : Private Pr
* Q77_1_e = How satisfied are you with the cost of the water you received? : Private Pr
* Q77_1_f = How satisfied are you with the cost of the water you received? : Private Pr
* Q78_1_a = How satisfied are you with the services of the providers (billing, ordering, etc
* Q78_1_b = How satisfied are you with the services of the providers (billing, ordering, etc
* Q78_1_c = How satisfied are you with the services of the providers (billing, ordering, etc
* Q78_1_d = How satisfied are you with the services of the providers (billing, ordering, etc
* Q78_1_e = How satisfied are you with the services of the providers (billing, ordering, etc
* Q78_1_f = How satisfied are you with the services of the providers (billing, ordering, etc
* Q81_1 = Q81_1
* Q81_2 = Q81_2
* Q81_3 = Q81_3

* Q81_4 = Q81_4
* Q81_comments = Q81_comments
* UW = How did you finance the irrigation for your land in the last agricultural year?
* UX = How did you finance the irrigation for your land in the last agricultural year?
* VA = How did you finance the irrigation for your land in the last agricultural year?
* Haveyoupaidyourirrigationfe = Have you paid your irrigation fees for the current period?
* Q92_other = What were some of the topics covered in the training? SELECT ALL THAT APPLY
* Q99 = Do you still receive any extension/training services from the Ministry of Agricu
* WR = Which of the following practices do you follow or use equipment for? SELECT ALL
* WV = Which of the following practices do you follow or use equipment for? SELECT ALL
* XB = From whom or what did you learn about these practices or tools? SELECT ALL THAT
* XD = From whom or what did you learn about these practices or tools? SELECT ALL THAT
* Doyoustoreaportionoftheha = Do you store a portion of the harvest?
* Q107_1 = Where did you store the portion of crops that you consumed/ sold? SELECT ALL THA
* Q107_2 = Where did you store the portion of crops that you consumed/ sold? SELECT ALL THA
* Q107_3 = Where did you store the portion of crops that you consumed/ sold? SELECT ALL THA
* Q107_5 = Where did you store the portion of crops that you consumed/ sold? SELECT ALL THA
* Q107_6 = Where did you store the portion of crops that you consumed/ sold? SELECT ALL THA
* Didyoudrythiscommodityadequ = Did you dry this commodity adequately to reduce spoilage during storage?
* Didyoutreatthecommoditywith = Did you treat the commodity with chemicals during storage to control insect pest
* Q110 = Q110
* Whatwasthemaincauseofloss = What was the main cause of loss during storage, if any? (1)
* XS = What was the main cause of loss during storage, if any? (2)
* XT = What was the main cause of loss during storage, if any? (3)
* XU = What was the main cause of loss during storage, if any? (4)
* XV = What was the main cause of loss during storage, if any? (5)
* XW = What was the main cause of loss during storage, if any? (6)
* XX = What was the main cause of loss during storage, if any? (7)
* Q113_3 = Why didn‚Äôt you store in a post-harvest
* Q113_4 = Why didn‚Äôt you store in a post-harvest center?

* Q113_comments = Why didn‚Äôt you store in a post-harvest center?
* Didyousellaportionofyours = Did you sell a portion of your stored commodities?
* Ofthestoredcommoditywhatpe = Of the stored commodity, what percentage did you sell at a later period
* Didyougetahigherpricefort = Did you get a higher price for the stored products?
* Didthecontentofthetraining = Did the content of the training positively affect your export business, i.e. hel
* Q129 = Do you still apply marketing practices in your export business were taught duri
* ZH = If yes, how long have you been a customer of this institution? (2)
* ZI = If yes, how long have you been a customer of this institution? (3)
* ZK = If yes, how long have you been a customer of this institution? (5)
* Q134 = Have you applied for a loan, but got rejected?
* Q135 = What do you think was the main reason for rejection? [One answer]
* Q135_comment = What do you think was the main reason for rejection? [One answer] (Other (pl
* Q136_5 = Why didn‚Äôt you apply for a loan?
* Q136_6 = Why didn‚Äôt you apply for a loan?
* Q143 = What was the main purpose of the loans you applied for?¬†
* Q143_comments = What was the main purpose of the loans you applied for?
* Q144_comments = Did you encounter any difficulties in applying for the loan? If yes, what were s
* Q146 = What was the outcome of your application for MFI Loan?
* Q147_1 = For what reason was your loan rejected? No collateral
* Q147_2 = lack of business plan
* Q147_3 = For what reason was your loan rejected? - Low performance
* Q147_4 = For what reason was your loan rejected? - Requested loan too high
* Q147_5 = For what reason was your loan rejected? - Refused to answer
* Q147_6 = For what reason was your loan rejected? SELECT ALL THAT APPLY (Other (please
* Q147_comments = For what reason was your loan rejected? SELECT ALL THAT APPLY (Other (please
* ABK = For what reason was your loan rejected? SELECT ALL THAT APPLY (Don't Know)
* Q148_4 = What are the terms and conditions of the loan? : Repayment schedule (weekly
* Q148_5 = What are the terms and conditions of the loan? : What was the value of guar
* ABQ = What are the terms and conditions of the loan? : amount paid per schedule -
* Whenyoureceivedtheloanwhat = When you received the loan, what purpose did you end up using it for? SELECT ALL

* ABV = When you received the loan, what purpose did you end up using it for? SELECT ALL
* ABW = When you received the loan, what purpose did you end up using it for? SELECT ALL
* ABX = When you received the loan, what purpose did you end up using it for? SELECT ALL
* ABY = When you received the loan, what purpose did you end up using it for? SELECT ALL
* ACB = When you received the loan, what purpose did you end up using it for? SELECT ALL
* ACL = If answered (b) NO above, why not? (Other (please specify)) - specify
* Q156_1_comments = What were the reasons of your satisfaction? (Other (please specify)) - specify
* Q156_2_h = Why are you dissatisfied? (Other (please specify)) - specify
* Q157 = Do you have any recommendation for this loan program (e.g. what could have been
* ADJ = What do you think was the main reason for rejection?
* ADK = What do you think was the main reason for rejection? (Other (please specify))
* Q159_2 = How did you finance your business before you received loan from this MFI under t
* Q159_4 = How did you finance your business before you received loan from this MFI under t
* Q159_7 = How did you finance your business before you received loan from this MFI under t
* Whatwerethetermsandconditio = What were the terms and conditions of the loan? : Amount (CVE) - Amount
* ADW = What were the terms and conditions of the loan? : Interest rate (%) - Amoun
* ADX = What were the terms and conditions of the loan? : Term (months) - Amount
* ADY = What were the terms and conditions of the loan? : Repayment schedule (weekl
* ADZ = What were the terms and conditions of the loan? : What was the value of gua
* AEA = What were the terms and conditions of the loan? : Amount paid per schedule
* Whatwastheoverallexperience = What was the overall experience of beneficiaries with this financial products?
* AEC = What was the overall experience of beneficiaries with this financial products?
* AED = What was the overall experience of beneficiaries with this financial products?
* AEE = What was the overall experience of beneficiaries with this financial products?
* AEF = What was the overall experience of beneficiaries with this financial products?
* Whatisthetotalvalueofthese = What is the total value of these other outstanding loans?_____ (CVE)
* AEL = For what purposes? SELECT ALL THAT APPLY (2)
* AEM = For what purposes? SELECT ALL THAT APPLY (3)
* AEN = For what purposes? SELECT ALL THAT APPLY (4)
* AEO = For what purposes? SELECT ALL THAT APPLY (5)

* AEP = For what purposes? SELECT ALL THAT APPLY (6)
* AES = For what purposes? SELECT ALL THAT APPLY (9)
* AEU = For what purposes? SELECT ALL THAT APPLY (Other (please specify)) - specif
* Cropnamemostproducedbythe = Crop name (most produced by the farmer)
* AFB = How much does it cost on the following heads to produce in one agricultural year
* AFC = How much does it cost on the following heads to produce in one agricultural year
* AFE = How much does it cost on the following heads to produce in one agricultural year
* AFG = How much does it cost on the following heads to produce in one agricultural year
* AFJ = How much does it cost on the following heads to produce in one agricultural year
* AFK = How much does it cost on the following heads to produce in one agricultural year
* own_land_change =
* rent_cult_change =
* yam_price_pre_avg =
* yam_price_post_avg =
* tomato_price_pre_avg =
* tomato_price_post_avg =
* beet_price_pre_avg =
* beet_price_post_avg =
* onion_price_pre_avg =
* onion_price_post_avg =
* pepper_price_pre_avg =
* pepper_price_post_avg =
* cabbage_price_pre_avg =
* cabbage_price_post_avg =
* papaya_price_pre_avg =
* papaya_price_post_avg =
* carrot_price_pre_avg =
* carrot_price_post_avg =
* lettuce_price_pre_avg =
* lettuce_price_post_avg =

* cassava_price_pre_avg =
* cassava_price_post_avg =
* beans_price_pre_avg =
* beans_price_post_avg =
* maize_price_pre_avg =
* maize_price_post_avg =
* yam_harvest_pre_val =
* tomato_harvest_pre_val =
* beet_harvest_pre_val =
* onion_harvest_pre_val =
* pepper_harvest_pre_val =
* cabbage_harvest_pre_val =
* papaya_harvest_pre_val =
* carrot_harvest_pre_val =
* lettuce_harvest_pre_val =
* cassava_harvest_pre_val =
* beans_harvest_pre_val =
* maize_harvest_pre_val =
* yam_harvest_post_val =
* tomato_harvest_post_val =
* beet_harvest_post_val =
* onion_harvest_post_val =
* pepper_harvest_post_val =
* cabbage_harvest_post_val =
* papaya_harvest_post_val =
* carrot_harvest_post_val =
* lettuce_harvest_post_val =
* cassava_harvest_post_val =
* beans_harvest_post_val =
* maize_harvest_post_val =

Variable Group 3) Aggregate values into larger groups
* Education = Education

Variable Group 4) Aggregate values into equal-numbered groups
2 equal-numbered groups
* household_members = How many people live in your household (including yourself)?
* group_years = How long have you been part of the organization? (years)
* land_owned = If own, how much land you own?
* own_cult = How much of the owned land you cultivated in the last year?
* land_rented = If rented or crop sharing, how much land do you rent and how much did you cultiv
* rent_cult = If rented or crop sharing, how much land do you rent and how much did you cultiv
* land_owned_pre = How much land you owned before the project?
* own_cult_pre = If owned, how much of your land was cultivated before the Infrastructure Develo
* rent_cult_pre = If rented, how much land could you have cultivated before the Infrastructure Dev
* sales_pre = Agricultural sales before the project
* sales_post = Agricultural sales after the project
* inc_pre = Agricultural income before the project
* inc_post = Agricultural income after the project
* health_expense = How much does your household spend per year on: : Health - Amount (CVE)
* food_expense = How much does your household spend per year on: : Food - Amount (CVE)
* utilities_expense = How much does your household spend per year on: : Utilities - Amount (CVE
* Whatwastheamountoftheloan = What was the amount of the loan you applied for?_____(CVE)
* Q148_1 = What are the terms and conditions of the loan?   : Amount (CVE) - Amount
* Q148_2 = What are the terms and conditions of the loan?   : Interest rate (%) - Amount
* Q148_3 = What are the terms and conditions of the loan?   : Term (months) – Amount
3 equal-numbered groups
* land_rented_pre = If rented, how much land could you have cultivated before the Infrastructure Dev
* Q47_1 = Inputs Use: How much did you spend on agricultural inputs last agricultural year
* Q47_2 = Inputs Use: How much did you spend on agricultural inputs before project

* Q71 = How much water did you use to irrigate your land per agricultural year? (m3)
* Q83 = Q83
* Q96 = How long did the training take (duration)?_____(days)
* Amountoflandinm2 = Amount of land (in m2)
* productionamountinoneagricul = production amount in one agricultural year (in KG)
* Averagepriceinmarketperkg = Average price in market (per kg)
6 equal-numbered groups
* employ_change
9 equal-numbered groups
* expense_post
* expense_pre
* total_production_val_pre
* total_production_val_post
* profit_pre
* profit_post
* profit_change
* total_harvest_pre
* total_harvest_post
* paid_worker_pre
* paid_worker_post
* total_worker_pre
* total_worker_post
* total_sold
* total_harvest
* total_sold_percent
* land_cult_post
* productivity_post
* land_cult_pre
* productivity_pre

Variable Group 5) Aggregate values into 2 groups: 0, and 1+
* bus_inc = What other activities do you receive income from- Business
* other_job_inc = Alternative income from business
* spouse_inc = What other activities do you receive income from- Spouse salary
* family_contr = What other activities do you receive income from- Other family contributions
* pension_inc = What other activities do you receive income from,- Pension
* remit_inc = What other activities do you receive income from- Remittance Income
* other_inc = What other activities do you receive income from?- Other sources
* edu_expense = How much does your household spend per year on: : Education - Amount (CVE)
* other_expense = How much does your household spend per year on: : Others - Amount (CVE)
* Q171 (variables Howmuchdoesitcostonthefol AFA AFD AFF AFH AFI) = How much does it cost on the following heads to produce in one agricultural year

Variable Group 6) Aggregate male + female workers to total workers, and aggregate values into 9 equal-numbered groups
* Q42_* = male and female workers

Variable Group 7) Group responses into ranges
* crop_cycle_post = How many times a year do you grow your crops (crop cycle)?
* crop_cycle_pre = How many times a year did you grow your crops (crop cycle) before the Infrastruc

Variable Group 8) Make measurement missing if fewer than 10 farmers were growing, or fewer than 10 farmers were not growing. Otherwise replace positive values with median of positive values
* Q45 crops: aggregate land area, irrigated land area, harvested kgs, and price into groups

Variable Group 9) If all values are 0: leave as-is. If <10 values are 1+: remove all values. If 10+ values are 1+: group values into 0 and 1+.
* Q48 expenses
* Q49 crop production

Variable Group 10) Group together -1 and 0 values

* own_cult_change
* land_rented_change
* sales_change