

Sex disaggregated data analysis guide

Best practices for data analysis

This sex disaggregated data analysis guide provides details and guides on how once sex disaggregated data is collected.

it can be used to understand gender issues and eventually support in tailoring products and services that better serve the needs of women smallholders.

Directions to use the sex disaggregated data analysis guide:

1. It is recommended that this guide is used after or along with the sex disaggregated data collection guide to derive the most value. The sex disaggregated data collection guide is available [here](#).
2. Go through the tips and examples on data analysis to understand the best practices on how to perform gender analysis.
3. This tool is best used by members of an organization who have prior experience or training in data analysis and have sufficient proficiency in MS Excel or statistical software packages.
4. At the end of the tool we have provided a glossary of some commonly used terms which will aid in better understanding of some of the indicators.



Primary audience
Gender unintentional and intentional organisations



Tool type
Guide



Personnel
Analytics teams, strategy teams



Duration
1-3 hours

GOALS

(1) Understand some of the best practices in gender data analysis.

REQUIREMENTS

(1) Fully integrating this tool will require financial investments in data analysis or statistical packages to analyze the data.

(2) Trained data analysts will be key, or there will be a need to invest in such training of existing employees.

(3) For gender analysis of sex disaggregated data it is useful to have an internal gender expert/advisor to lead the analysis and hypothesis creation.

(4) If there are terms which are unclear we suggest referring to [Gender concepts glossary](#) tool or consult your Gender Advisor if you have one.

Sex disaggregated data analysis guide

Best practices for data analysis

BEST PRACTICES ON SEX DISAGGREGATED DATA ANALYSIS

While collecting sex disaggregated data is a crucial first step, the next step of analysing the same to generate gender insights is important to understand underlying gender norms, barriers, and gender based constraints. These analyses and insights can in turn help (re)design better products and services which serve women smallholders and increase their agency.

Below are some of the best practices to keep in mind while performing sex disaggregated data analysis:

(1) Sex disaggregated data, wherever possible, should be cross tabulated with other relevant variables

This will allow for the analysis to highlight differences between males and females in agriculture and in sub-groups which would otherwise be invisible. Other variables that can be used include age, education levels, labour force participation, income, asset ownership, etc.

Example: A digital financial services provider can collect sex disaggregated data on loan redemption values and cross tabulate it with data on collateral ownership or asset ownership by sex. This will enable the provider to develop hypotheses on varying loan redemption rates by women and men smallholders. It is possible, that even when women and men smallholders are approved loans at the same rate, lack of or lower asset ownership among some groups of women could lead them to hesitate from actually redeeming the loans and hence running the risk of losing their assets in case they are unable to pay back. The digital financial service provider would not be able to develop such hypotheses if they only analysed single data points on loan redemption rates by sex.

(2) Sex disaggregation of data and cross tabulation should be done with specific hypotheses in mind

This will allow for the analysis to test meaningful correlations and relationships between variables. This will also help avoid generating excess quantum of data which in the best case can be unwieldy and difficult to analyse, but in the worst case can lead to analysis of false correlations between data points and even confuse them for causation.

Example: Continuing with the previous example, the digital financial service provider can cross-tabulate data to test the hypothesis that eventual loan redemption rates by women and men smallholders is correlated to the collateral or asset ownership of the farmers. If the digital financial service provider did not have this hypothesis, they would end up cross tabulating loan redemption rates data with all other types of variables (such as farm machine ownership, livestock ownership, number of family members, education levels of farmers, time spent in doing household activities, etc.) and would have a difficult time analysing all this information. In worst case, they may end up looking at spurious correlations derive insights based on that, rather trying to prove or disprove a hypothesis using the sex disaggregated data collected.

Sex disaggregated data analysis guide

Best practices for data analysis

(3) Use descriptive analysis of data for gender-focussed comparisons, but they are better suited for large datasets

Descriptive analysis of data includes computation of simple measures of distribution that can facilitate straightforward gender-focussed comparisons. Such measures include proportions and percentages, ratios, medians, quintiles, means, and standard deviations. It is key however to limit the usage of these measures to large datasets, where the insights or conclusions can be safely extrapolated to a large population. For smaller datasets, actual numbers should preferably be used instead of descriptive measures.

Example: Continuing with the previous example, if the digital financial service provider pilots a new loan product with 50 farmers then they should report and analyse the uptake and usage of the product using actual numbers. This would mean that if 35 men farmers took the loan product and 5 of them defaulted and 15 women farmers took the loan product and 2 of them defaulted the numbers should be used as is, instead of using proportions and percentages in the gender analysis. In this small dataset, it would not be appropriate to use percentages and say that a similar percentage of women and men smallholders (~14% and ~13% respectively) defaulted on the loan product as it could be misconstrued and mis-extrapolated to a larger population or taken as a generic insight.

(4) Combine various types of datasets with sex disaggregated data collected, wherever possible

This will allow to highlight important nuances. However, care must be taken to ensure that the combined datasets are comparable in terms of coverage and completeness, time period, alignment of key terms and concepts.

Example: Continuing with the previous example, the digital financial service provider can combine the sex disaggregated data on loan redemption rates with other data relating to weather patterns in the relevant regions and sexes of the village agents and field extension officers. By combining these varied datasets, they might be able to draw hypotheses and insights on whether loan redemption rates varies by gender and if weather plays a role (bad weather conditions might prohibit some some farmers to redeem loans as they know they will not be able to repay). The digital financial service provider might also test if the loan redemption rates vary between genders and is correlated to the gender of the village agent or the field extension officer, i.e., test if women smallholders do not feel very comfortable dealing with male agents and thus may not redeem loans where they have to primarily deal with men or send their husbands or other male members instead.

(5) Combine sex disaggregated data with qualitative data and information to derive more nuanced gender analysis

This will allow to uncover nuances around gender and social norms, gender based constraints, and gender biases - all of which may not be readily captured or understood through numerical data. Such qualitative information can be collected through interviews (in-person or telephonic),

Sex disaggregated data analysis guide

Best practices for data analysis

(5) Combine sex disaggregated data with qualitative data and information to derive more nuanced gender analysis

focus group discussions, informal conversations with women and men farmers and other key stakeholders and their families and through secondary research.

Example: Continuing with the previous example, if the digital financial service conducts informal conversations with women and men smallholder farmers they might uncover that the loan redemption rates by women smallholders is highly dependent on spousal approval. This is a gender and social norm that indicates towards women to obtain permission from male members of their households before engaging in any financial decision. By combining sex disaggregated data with qualitative information, the digital financial service provider is able to derive gender insights on the key norms and barriers that prevent women smallholder farmers from fully engaging with financial service products. To address this challenge, the provider can now modify their customer acquisition process and ensure that male members are engaged from the very beginning and can allow women farmers to take the financial decision of redeeming the loan a few weeks down the line.

Sex disaggregated data analysis guide

Best practices for data analysis

RELEVANT GLOSSARY

As you work through this tool, you may have come across terms that are unfamiliar or confusing. Below is a list of some terms, knowing which would aid in better understanding of the tool. If there still are terms that you do not understand, it is recommended to refer to the [Gender Glossary](#) and other linked sources or consult with your Gender Advisor (if your organization has one).

Gender analysis refers to the critical examination of the differences between men and women, boys and girls, in their access to and control of resources, their roles and responsibilities, and the constraints they face in a given situation or context relative to each other. For instance, an assessment of the differences in access to and control of digital agricultural solutions between women and men smallholder farmers.

Sex disaggregated data refers to quantitative and qualitative data collected and presented by sex that allows for measurement of socio-economic and socio-cultural differences between men and women. For example, collection of data on the number of women and men smallholder farmers that use digital tools for agriculture separately in order to measure the gendered differences.

Gender bias refers to the conscious or unconscious, explicit or implicit prejudiced actions or thoughts that could lead to unfair treatment of people based on their perceived gender. For instance, male farm laborers hired by women farm owners are likely to be less hardworking and reliable due to the lower perceived strictness of women.

Gender gap refers to the disparity in men and women's social, political, intellectual cultural and/or economic condition or position in society, often based on underlying socio-cultural norms. It reflects the unequal distribution of resources, opportunities and outcomes across genders. For instance, women smallholder farmers have less access to digital financial services indicating a gender gap in access to resources.

Gender/ social norms collectively held standards to which people are expected to conform at different stages of their lives based on their sex or gender identity. These are norms that seek to represent beliefs and values of what it means to male or female in different societal contexts. For instance gendered expectations for women to engage in unpaid caregiving work limits their time to engage in productive agricultural activities.

Gender based constraints refer to formal laws, attitudes, perceptions, values, or practices that limit people's access to resources and/or opportunities based on their sex or gender identity. For instance, cultural norms that limit women's ownership of agricultural land are gender-based constraints.